

PRÁVNICKÁ FAKULTA MASARYKOVÉJ UNIVERZITY

MUNI
PRÁVNICKÁ
FAKULTA

**Užitie automaticky
zozbieraných dát k strojovému
učeniu ako autorskoprávne
užitie diela**

ŠTUDENTSKÁ VEDECKÁ ODBORNÁ ČINNOSŤ

Andrej Krištofík

Brno, jar 2019

Vyhlásenie

Vyhlasujem, že som predloženú prácu vypracoval samostatne len s použitím uvedenej literatúry a prameňov.

Andrej Krištofík

Abstrakt

Táto práca sa zameriava na posúdenie otázky, či v súčasnej autorsko-právnej úprave existuje možnosť podradiť automatický zber autorských diel pre potreby strojového učenia pod niektoré z existujúcich zákonných užití, a ako áno, či k takému existuje zákonná licencia. Práca sa tiež ďalej zaoberá otázkou budúceho vývoja tejto problematiky, najmä s ohľadom na Smernicu o autorskom práve na digitálnom jednotnom trhu.

Abstract

This paper focuses on the question whether web scraping of copyrighted works with intent to use them as training dataset for machine learning, can be currently viewed as a copyrighted usage and if so, whether there is a space for legale license for such. Furthermore this work tries to look at future developments of said question, namely in the light of Directive on copyright in the Digital single market.

Klíčové slová

Autorské právo, zákonné licencie, užitie diela, web scraping, strojové učenie, umelá inteligencia

Keywords

Copyright law, legal licences, use of copyrighted works, web scraping, machine learning, artificial intelligence

Podakovanie

Týmto chcem poďakovať JUDr. Janovi Zibnerovi za usmernenie pri výbere témy, poskytnutie podkladov a nápadov ako aj za vedenie a konzultácie práce. Ďalej chcem poďakovať BSc. Jakubovi Kmecovi, za čiastkové konzultácie v technologických otázkach.

Obsah

1	Úvod	11
1.1	<i>Pojem umelej inteligencie</i>	11
1.2	<i>K autorskému právu</i>	13
2	Užitie diela	15
2.1	<i>Rozmnožovanie diela</i>	16
2.2	<i>Vzťah medzi modelom a pôvodným dielom</i>	18
3	Vývoj pro futuro	23
3.1	<i>Nové užitie a Smernica o autorskom práve na jednotnom digitálnom trhu</i>	24
3.1.1	<i>Možné východiská</i>	25
4	Záver	27

1 Úvod

Profesor *Yoshua Bengio*, kanadský vedec venujúci sa výzkumu umelej inteligencie, povedal že (ľudská) kreativita je výsledkom výpočtov, ktorých procesom rozumieme, a adaptácia tohoto procesu pre stroje, je len otázkou výpočtovej techniky.¹ Súčasný stav technológie už pomaly speje k takejto simulácii kreatívneho procesu, a je schopný vytvárať autorské diela na základe predchádzajúcich diel ľudí, ktoré sú mu dané za predlohu. Práve takémuto využívaniu autorských diel by som sa chcel venovať v tejto práci.

V prvej časti predstavím niektoré základné znalosti a definície, z ktorých budú vychádzať nasledujúce časti. Následne sa pozriem na súčasnú úpravu autorského práva a ako do nej zapadá súčasný stav tejto technológie. V tretej časti sa budem venovať otázke Smernice o autorskom práve na jednotnom digitálnom trhu, ktorá operuje aj s dataminingom ako časťou procesu opisovaného aj v tejto práci. Na záver predstavím niekoľko možných východísk pre budúcu úpravu tejto problematiky.

1.1 Pojem umelej inteligencie

Termín umelá inteligencia funguje ako "dáždnikový termín" zahŕňajúci mnohé prístupy k vytvoreniu "inteligentných" strojov.² Medzi tie v súčasnosti najrozšírenejšie patrí strojové učenie (tzv. Machine learning)³, ktorému sa budem venovať v tejto práci. Algoritmy strojového učenia sa odlišujú od bežných počítačových programov v tom ohľade, že nie sú programované priamo k vykonávaniu určitej úlohy, no skôr k tomu, aby sa vykonávanie tejto úlohy takpovediac "naučili". Po naprogramovaní je do tohto systému nahrané obrovské množstvo dát, tzv. dataset, ktorý obsahuje "ukážky" výstupov, aké má systém produkovať. Daný systém by následne z tohoto datasetu mal vyvodíť, akým

1. REESE, Byron. Episode 1: A Conversation with Yoshua Bengio. *Voices in AI* [online]. GIGAOM, [cit. 03.03.2019].

2. FETZER, James H. *Artificial Intelligence: Its scope and limits*. Dordrecht: Kluwer Academic Publishers, s. 3 - 5.

3. K bližšiemu opisu tejto techniky viď. MAŘÍK, Vladimír a kol. *Umelá inteligencia (1)*. Praha: Academia, 2001, s. 168 - 183.

1. ÚVOD

spôsobom dospeje k produkovaniu takýchto výsledkov.⁴ Nie len že sa jedná o výpočtovo veľmi náročnú úlohu, k jej správne fungovaniu je nutné poskytnúť systému čo najväčší dataset. Čím komplexnejšia úloha má byť daným systémom umelej inteligencie vykonaná, tým obsiahlejší dataset je nutné jej poskytnúť.⁵ S postupom času sa preto stáva vytváranie datasetov úlohov, ktorej rozsah presahuje medze mechanickej ľudskej činnosti, a preto sú čím ďalej využívané rôzne automatizované systémy pre zber dát, ako napríklad web crawleri pre zber dát z internetu, pomocou techniky tzv. web scrapingu.⁶ Tento automatizovaný web crawler najprv stiahne webovú stránku (tzv. fetching), v ktorej zdrojovom kóde následne hľadá pred-definované informácie, ako kontaktné údaje, či napríklad nadpisy článkov ako je tomu často pri vytváraní datasetov. Takto zozbierané informácie ukladá lokálne u programátora vo forme, ktorá umožňuje ich následné využitie (napr. vo formáte .csv pre datasety).⁷

4. Jedným z nedávno predstavených príkladov použitia takéhoto systému môže byť napríklad funkcia "smart reply" emailu od spoločnosti Google. Ako ukážky použil Google spárované emaily s odpoveďami, z ktorých mal systém Smart Reply zistiť vzťah medzi obsahom prichádzajúceho emailu a obsahom odpovede. Na základe toho sa systém teraz snaží predpovedať a ponúkať užívateľom niektorú časť odpovede, ako napríklad rôzne zdvorilostné frázy, či odkazy na prílohy v predchádzajúcich správach.

5. Napríklad webová stránka StatMT poskytuje študentom dataset pre vytvorenie jednoduchého prekladača, podobného službe Google Translate, ktorý má pre prekladanie z Angličtiny do 5 ďalších jazykov veľkosť 26 gb, teda približne obsahuje 11 555 555 556 normostrán textu, pričom sa má jednať len o jednoduchšiu verziu reálnych prekladačov. Zdanie s datasetom je dostupné na: <http://www.statmt.org/wmt14/translation-task.html>.

6. Na takomto princípe stoja v súčasnosti aj všetky internetové vyhľadávače a preto je zrejme tým najpoužívanejším v súčasnosti webcrawler spoločnosti google - GoogleBot. Spoločnosť Google takmer neustále necháva svoje webcrawleri prehľadávať internet. Tie konštantne sťahujú obsah webových stránok a posielajú ho na servery spoločnosti Google (k tejto časti sa referuje ako k web scrapingu, keď program zbiera obsah stránky), ktorá ich následne analyzuje a radí na tzv. Google Index, podľa toho ako obsah vyhodnotí. Tento vnútorný proces je verejnosti neznámy a stojí za úspechom tohoto vyhľadávača. Keď používateľ zadá termín, ktorý chce nájsť, Google skontroluje práve tento Index a podľa toho ako vyhodnotil obsah stránky, používateľovi ponúkne odkazy na daný obsah.

7. LAKSHMI, Prasanna Kaspá a kol. *Towards Extended Data Mining: An Examination of Technical Aspects*. In: *Procedia Computer Science*, 2018, vol. 139, s. 49 - 55.

1.2 K autorskému právu

Autorské právo, upravené v autorskom zákone⁸, je právom absolútnym, pôsobiacim *erga omnes*, teda postavenie osoby autora v základnom nastavení vylučuje kohokoľvek iného z dispozície s autorským dielom, či k nemu prislúchajúcimi právami.⁹ Medzi jedno z najpodstatnejších (majetkových) práv, ktorými autor disponuje, je právo dielo užiť,¹⁰ ako vyjadrenie absolútneho majetkového panstva nad dielom.¹¹ Iné osoby môžu práva náležiacie autorovi vykonávať iba v prípade, že im k tomu bola autorom udelená licencia, ako výraz dispozície s právami, prípadne že túto licenciu nadobudli zo zákona, na základe stanovených dôvodov.

Autorský zákon obsahuje výpočet výlučne pozitívnych užití diela, ktoré sú zákonom vnímané ako spôsoby autorskoprávneho užitia, teda také, ktoré prislúchajú autorovi alebo subjektu, na ktorý autor licenciou tieto oprávnenia preniesol.¹² Užitia, ktoré autorský zákon za autorskoprávne neoznačuje, sú v teórii užitiami voľnými,¹³ i keď je spornou otázkou, či je zákonný výpočet užití výpočtom taxatívnym, či demonštratívnym.¹⁴ Vo svojom komentári *Telec* poukazuje na fakt, že skrz nie úplne šťastnú legislatívnu techniku došlo novelou autorského zákona z roku 2006 k zavedeniu demonštratívneho výpočtu užití diela, naproti stávajúcemu výpočtu taxatívnemu.¹⁵ Ďalej potom však túto otázku *Telec* prenecháva rozhodovacej praxi a právnej náuke, nakoľko odstránením taxatívneho výpočtu došlo aj k odstráneniu akejkoľvek definície pojmu užitie diela.¹⁶

Subsumpcia užitia autorského diela ako súčasť datasetu potrebného k učeniu systému umelej inteligencie pod niektoré zo zákonom

8. Zákon č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon), v znení neskorších predpisov. Ďalej len "autorský zákon".

9. TELEC, Ivo; TŮMA, Pavel. *Autorský zákon: Komentář*. Praha: C. H. Beck, 2007, s. 4.

10. § 12 autorského zákona.

11. TELEC, op. cit., s. 165.

12. § 12 - § 13 autorského zákona.

13. § 30 autorského zákona.

14. TELEC, op. cit., s. 166.

15. Ibid.

16. Ibid., s. 165.

1. ÚVOD

vymedzených autorskoprávných užití, případně možnost jejich využití v tomto režimu iným způsobem, který neobsahuje zákonný demonstrativní výčet, je otázkou, kterou se budem zabírat v první části této práce. Takáto subsumpcia je podstatná pro vyřešení otázek právního vztahu třetích osob k dílům autora. Ak by sme prišli k záveru, že sa nejedná o autorské užitie diela, tretie osoby nemusia k takému mať licenciu.

Ak by sa o autorské užitie diela jednalo, tretie osoby túto licenciu nadobudnúť k užitiu musia, a to buď priamo od autora, ako výraz jeho dispozície s právami, preto nie je nutné sa takýmto nadobúdaním ďalej zaoberať,¹⁷ alebo skrz zákonnú licenciu, čo je otázka, ktorou sa budem eventuálne zaoberať v druhej časti tejto práce.

17. V prípade kedy by autor nadobúdateľovi k takejto činnosti udelil licenciu, jednalo by sa o jeho prejav vôle, a súčasná právna úprava, či teoretické východiská civilného práva pre takýto postup nepredstavujú žiadne problémy.

2 Užitie diela

Táto kapitola si berie za cieľ zaoberať sa otázkou, autorskoprávneho využitia diel pri učení systémov umelej inteligencie, teda pred ich vznikom. K užitiu bude teda dochádzať prostredníctvom programátora, nie systému umelej inteligencie, čo by predstavovalo inú otázku. V prvom rade bude pre potreby stanovenia spôsobu užívania autorských diel do datasetov potrebné zistiť, či aj takéto užitie možno považovať za užitie autorskoprávne, teda za také majetkové právo, ktorého výkon prináleží autorovi. Autorský zákon v § 12 a násl. obsahuje výčet práv, ktoré je za takéto možno považovať. Predtým však, než sa budem venovať jednotlivým využitiam a možnosti ich aplikácie na vyťažovanie diel pre datasety, je nutné zohľadniť, či je súčasný zákonný výpočet charakteru taxatívneho alebo demonštratívneho. Ak by sme totižto v tejto práci dospeli k záveru, že v súčasnosti takéto využívanie autorských diel nie je uvedené ako niektoré z autorských užití a tento výčet je taxatívny, nie je nutné sa ďalej zaoberať otázkou legitímnosti užívania diela tretími osobami, respektíve možnosťou nadobudnutia zákonnej licencie takýmito osobami.¹⁸ Ak by však výpočet bol len demonštratívny, bolo by potrebné sa ďalej zaoberať možnosťou vnímania takéhoto užívania ako užívania autorskoprávneho, a doposiaľ neupraveného, a teda aj možnosťami zákonných licencií.

S ohľadom na súčasné využitie tejto technológie, rovnako tak ako aj na jej budúci potenciál majú oba závery širšie dopady. Na jednej strane existuje záujem chrániť majetkové práva autora, ako jedna z hlavných pohnútok k tvorbe (nových) diel,¹⁹ a to s ohľadom na už spomínaný veľký ekonomický potenciál technológií, pre ktoré sa dáta využívajú. Naproti tomu však stojí záujem užívateľov takto vyťažovaných dát na ich užití, ako aj záujem spoločnosti na inováciach, ktoré by tieto technológie mohli priniesť. V neposlednej rade je tu taktiež otázka právnej istoty prevádzkovateľov systémov, ktorý už v súčasnosti takúto

18. V takýchto prípadoch je samozrejme zachovaná ochrana osobnostných práv autora, ako napríklad dehonestujúce užitie a podobne. Týmito aspektami sa však v tejto práci zaoberať nebudem, nakoľko užívanie diel pre strojové učenie nie je viditeľné, či akýmkoľvek spôsobom vnímané verejnosťou, a preto autorova osoba, ktorá má byť v tomto prípade chránená, s takým užitím nie je spájaná.

19. SRSTKA, Jiří a kol. *Autorské právo a práva související*. Praha: Leges, 2017, s. 43 - 45.

technológiu využívajú a doteraz sa tým pohybovali tak povediac v "šedej zóne".

Zákonný demonštratívny výpočet užití diel je možné rozdeliť na dve kategórie; mechanické, resp. práva k nepriamemu zdelovaniu, ktoré sa viažu na hmotné nosiče a nakladanie s nimi, a na práva zdelovania diela verejnosti - priame zdelovanie; táto druhá forma sa už na nosič neviaže a spôsob zdelovania diela verejnosti prebieha v nehmotnej "efemérnej podobe".²⁰ Zo všetkých možností ponúkaných (demonštratívnym) výpočtom autorského zákona sa s ohľadom na vyššie popísaný technologický proces môže javiť ako najvhodnejšie rozmnožovanie diela, ktorému sa budem venovať v nasledujúcej časti.

2.1 Rozmnožovanie diela

Prvým zákonným užitím diela je jeho rozmnožovanie, ako jedno z autorskoprávných užití, viazaných na hmotný nosič, i keď druhý odstavec § 13 autorského zákona rozširuje rozmnožovanie aj na digitálne vyjadrené elektronické rozmnoženiny.

Užitie spôsobom rozmnožovania je voči všetkým ostatným užitiam vymenovaným v § 13 až § 23 autorského zákona rozdielne v tom, že rozmnožovanie zo zásady autorský zákon nechápe ako sprístupňovanie diela verejnosti v niektorej z jeho foriem.²¹ Teda verejnosť nie je jednou z podmienok klasifikácie užitia ako užitia rozmnožovaním. Čo však má pre užitie rozhodný charakter, je či k nemu dochádza za účelom osobného využitia rozmnoženiny fyzickou osobou, nakoľko také sa doktrinálne za autorskoprávne užitie nepovažuje.²²

Samotný akt rozmnožovania je neoddeliteľne spätý s hmotným nosičom, so zachytením či prevedením diela na takýto nosič. V našom prípade ho možno za také zrejme považovať, i keď literatúra sa vyslovuje o takýchto prípadoch nezmieňuje, zozbierané dáta neexistujú "v priestore" a sú zapisované na dátový nosič, či už na serveri zbierateľa alebo na "cloud", kde však tiež v konečnom dôsledku k zápisu na hmotný nosič dochádza taktiež.

20. TELEEC, Op. cit. s. 175.

21. KŘÍŽ, Jan a kol. *Autorský zákon: Komentář a předpisy souvisící*. Praha: Linde Praha, a. s., 2005, s. 95.

22. *Ibid.*, s. 96.

Pri takomto druhu užívania dát je zaujímavé poukázať práve na požiadavku verejnosti užitia, ktorá zrejme bude častou prekážkou pre užitia iného charakteru definovaného autorským zákonom. I keď je distribúcia nosičov diela verejnosti samostatne upravená, je spätá s predchádzajúcou výrobou kópií. Pri zbere dát pre potreby strojového učenia však k distribúcii, či akejkoľvek inej forme zverejňovania diela nedochádza, nakoľko jediným kto diela "číta", či s nimi obdobnou formou zachádza sú algoritmy, podobne ako tomu bolo v prípade *Authors Guild v. Google Books*.²³ V súčasnosti už však zverejňovanie vyhotovených rozmnoženín nie je mandatorným znakom rozmonožovania diela.

S odkazom na túto skutočnosť je vhodné sa pozrieť na iné delenie rozmnoženín, a to na delbu z hľadiska ich hospodárskeho významu. Ten môže byť samostatného charakteru, teda rozmnoženina je "samoučelná" a hospodársky prospech pramení priamo z nej. Druhou možnosťou je rozmnoženina hospodársky nesamostatná, ktorá má (hospodársky) význam iba ako súčasť iného celku, či procesu.²⁴ V takomto prípade však súčasná úprava požaduje, aby bol tento technologický proces využitý k sprístupňovaniu diela. Od hospodársky nesamostatného diela je potrebné tiež odlíšiť rozmnoženiny zámerné a náhodné, ktoré rovnako tak vznikajú spoločne s iným autorským užitím, avšak nie sú jeho zámerom.

Pri vytváraní zámernej rozmnoženiny je nutná prítomnosť vôle rozmnoženinu vytvoriť, bez ohľadu na účel, či vôľu alebo vedomie o porušovaní autorských práv, na také sa potom nevzťahuje zákonná bezúplatná licencia, ako je tomu v prípade náhodnej rozmnoženiny, ktorá vzniká bez zámeru ako vedľajší produkt.²⁵

I keď vytváranie rozmnoženín nie je hlavným zámerom ich automatizovaného zberu pre naše potreby strojového učenia, len ťažko ich možno považovať za rozmnoženiny náhodného charakteru, nakoľko samotný zber je zámerný a činený vyslovene za cieľom nadobudnutia (rozmnoženiny) diela.

23. Rozsudok Odvolacieho Súdu (Druhý obvod) vo veci The AUTHORS GUILD v. GOOGLE, INC.) Dostupné z https://scholar.google.cz/scholar_case?case=11449470169883565960hl=en&dt=6as&is=1oi=scholar

24. TELECOM, Op. cit., s. 182.

25. TELECOM, Op. cit., s. 181.

Na základe vyššie uvedeného by sa mohlo javiť, že web scraping je možné považovať za rozmnožovanie diela, pred tým než však budeme môcť urobiť tento záver, je vhodné sa pozrieť na následné využitie takto rozmnožených diel. I keď je pravdou, že vo všeobecnosti nie je potrebné skúmať následné využívanie rozmnoženín, nakoľko ich autorskoprávnym užitím je už samotné vyhotovenie rozmnoženiny bez ohľadu na to, akým spôsobom sa následne využije, v našej situácii - teda v momente keď sú rozmnoženiny zhotovované ako dataset pre potreby strojového učenia, ja na mieste aj táto úvaha, nakoľko by v konečnom dôsledku mohla odstrániť niektoré z podmienok aplikovateľnosti § 13 autorského zákona.

Pre strojové učenie, a pre narábanie s datasetmi pre jeho potreby, je typický práve samotný proces "učenia". Ten v takýchto prípadoch prebieha tak, že algoritmus si vezme nahrané súbory, tvoriace jeho učiacu vzorku, teda akúsi učebnicu, a ich postupov analýzov sa z nej snaží "naučiť", akým spôsobom sú jemu prezentované vzorky vytvárané. Čiže namiesto kopírovania alebo relatívne náhodilostného preskladávania prvkov v učiacej vzorke, sa snaží prísť na vzťah medzi jednotlivým prvkami, a rôzne matematicky vyjadriteľné vzťahy, ktoré sú esenciou daného diela, respektíve súboru diel - datasetu. S určitou mierou nadsádzky sa dá povedať, že v momente, kedy prebieha učenie, sa počítač snaží kvantifikovať prezentovanú myšlienku, tak aby následne pomocou svojich "vedomostí" takto nadobudnutých vedel vytvoriť dielo úplne nové, na základe myšlienky, či spôsobu tvorby, diela predchádzajúceho.

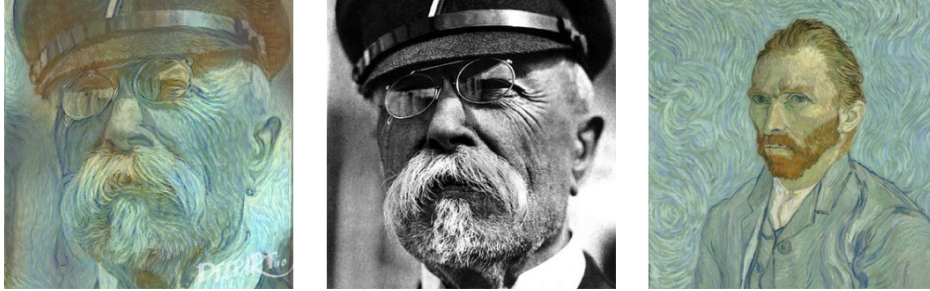
2.2 Vzťah medzi modelom a pôvodným dielom

Na lepšiu demonštráciu vyššie opísaného by som tento proces chcel priblížiť na príklade platformy *DeepArt*²⁶, pomocou ktorej môže používateľ pretvárať ním nahrané obrazové materiály (fotografie) v štýle rôznych známych umelcov.

To síce operuje s určitou formou zachytenia (resp. vyjadrenia) diela, avšak ako vyplýva z použitého technologického procesu, a ako to bude

26. <http://www.deepart.io>

opísané nižšie, celý tento proces sa snaží o akúsi syntézu myšlienky²⁷ diela. Na konci tohoto procesu bude teda stáť forma a obsah diela ako dve zložky.



Obr. 1.1: Fotka T. G. Masaryka upravená v štýle V. Van Gogha skrz *DeepArt*^a.

a. Fotografia T. G. Masaryka prebratá z: <https://pospolitost.files.wordpress.com/2009/10/masaryk.jpg>, Portrét V. Van Gogha poskytnutý platformou *deepart.io*

Celý tento proces využíva systému neurónových sietí²⁸, ktorý je jedným z v súčasnosti najperspektívnejších prístupov k strojovému učeniu.²⁹ Neurónová sieť sa skladá z množstva jednoduchých procesorov, tvoriacich jednotlivé "neuróny", ktoré sú vzájomne medzi sebou poprepájané rôznymi spôsobmi. V konečnom dôsledku pripomína takáto sieť neurónovú sieť centrálnej nervovej sústavy človeka, či iných živočíchov, odkiaľ pochádza aj jej názov.³⁰ Jednotlivé poprepájania bodov (neurónov) v sieti vytvárajú tzv. vrstvy. Každá vrstva obsahuje niekoľko takýchto procesných bodov (neurónov), ktorých spoločnou vlastnosťou je ich vzdialenosť (počet predchádzajúcich neurónov) od vstupného bodu siete. Počet vrstiev sa niekedy označuje aj ako hĺbka systému a čím "hlbší" daný systém je, tým hodnovernejšie vie napodobniť fungovanie CNS, teda jeho výsledky sú často bližšie k

27. V tomto prípade bude za myšlienku považovaný aj proces tvorby diela, či iné kreatívne procesy autora spejúce k špecifickému vyjadreniu.

28. Respektíve konvolučných neurónových sietí ako užšie vymedzenie tohoto prístupu, no pre potreby tejto práce budem používať oba termíny zameniteľne.

29. STERGIOŮ, Christos; SÍGANOS, Dimitrios. *Neural Networks*. Londýn: Imperial College of Science [online], [cit. 03.03.2019].

30. MAŘÍK, 2001, op. cit., s. 217.

zamýšlaným výstupom. Pridávanie vrstiev do systému samozrejme so sebou nesie požiadavky na zvyšovanie výpočtovej sily potrebnej k jej prevádzkovaniu.³¹

Takejto vrstvenej architektúry využíva aj proces na pozadí platformy *DeepArt* a v podstate slúži ako také "sito" oddelujúce obsah (vyobrazovaný objekt) od formy (umelecký štýl stvárnenia). Nahraný obrazový vstup, ktorý má v tomto prípade slúžiť pre učenie systému, prechádza u každej vrstvy neurónov určitou "grafickou" úpravou, ktorá má za úlohu pomôcť systému rozlíšiť obsah od jeho zachytenia. Tieto "grafické" úpravy sú tiež označované ako filtre a každý bod siete, ich môže mať niekoľko, v závislosti od komplexnosti úlohy. Tento filter funguje na základe určitých predprogramovaných kritérií a všíma si rozdiely v hodnotách jednotlivých pixelov na obraze a vyhodnotený rozdiel zaznamenáva do novej pixelovej siete. Teda v podstate, vníma rozdiely v zafarbení na jednotlivých častiach obrazu a na základe toho sa snaží determinovať jeho vlastnosti ako hrany a podobne. Takýto "výsledok" následne putuje do ďalšej vrstvy siete, ktorá obsahuje iný filter s inými vlastnosťami. Na konci takéhoto postupu sieťou môžeme teda nájsť dáta, na ktorých na jednej strane stojí "obsah" (to čo je predmetom obrazu, čo je v prípade *DeepArt* vedľajšia informácia) a "forma", teda spôsob akým bol daným umelcom tento predmet zachytený na plátne, či iné médium.³²

Následne je proces v podstate otočený. Systém prijme vstupný obraz (fotografiu) a nechá ju prejsť podobnou sieťou neurónov, ktorej filtre detekujú obsah podobne ako pri vstupnom obraze. Následne na takto zmapovaný obraz aplikuje filter, ktorý využíva dáta z výstupu "formy" pôvodného obrazu a aplikuje ho na obsah nového obrazu, tak aby pretvorený v podobnom štýle.

Akokoľvek teda môže operovať celý systém s autorským dielom, v tomto prípade s obrazom, potrebným údajom, resp. relevantnými dátami je forma zachytenia obsahu. Špecifickosť konvolučnej neurónovej siete, ktorá je využitá v prípade *DeepArt*, je práve jej "konvolučný" charakter, pre ktorý je možné do určitej miery pôvodné vyjadrenie diela "prehádzať", či pozmeniť, ako je opísané vyššie (nakolko však

31. MAŘÍK, 2001, op. cit., s. 219 - 222.

32. GATYS, Leon A. a kol. *A Neural Algorithm of Artistic Style*. arXiv.org [online]. Cornell University, publikované 26.08.2015 [cit. 02.03.2019].

system stále potrebuje odlišovať “formu” a “obsah”, takéto rozhádzanie je možné robiť v častiach, ktoré z pohľadu diela stále dávajú zmysel). Konvolučná vlastnosť siete spočíva v podstate v cyklickom spracovávaní drobných výsekov vstupných údajov (často o veľkosti 3x3 pixle), v ktorých postupne prechádza celý obraz a aplikuje na ne nastavený filter.³³

33. SIMARD, Patrice Y. *Best Practices for Convolutional Neural Networks Applied to Visual Document Analysis*. In: *Proceedings of the Seventh International Conference on Document Analysis and Recognition*, vol. 2, s. 958–962, 2016.

3 Vývoj pro futuro

I keď koncept umelej inteligencie nie je naprosto novou otázkou; či konceptom, ktorý sa objavil len pred nedávnom, v súčasnosti je možné sledovať jeho veľký rozmach a jeho stále obširnejšiu inkorporáciu do bežného života, a to najmä s ohľadom na neustály vývoj v dostupnej výpočtovej kapacite, ktorá výskum umelej inteligencie robí stále dostupnejším, ako aj výzkum nových prístupov implementujúcich koncept umelej inteligencie a vytváranie frameworkov³⁴ uľahčujúcich a urýchľujúcich vývoj, ktorý je takto prístupný dokonca "laickej" verejnosti. Preto je možné predpokladať, že úprava rôznych aspektov tekejto novej technológie bude stále potrebnejšia, nakoľko sa jedná o skutočne nové technológie, ktoré zákonodarca nemal šancu reflektovať v súčasnej právnej úprave.

Jasná úprava takýchto právnych otázok je dôležitá aj s ohľadom na ekonomický potenciál, ktorý so sebou tieto technológie prinášajú. Jedným z účelov autorského zákona, resp. autorskoprávnej ochrany, je ochrana investície (či už v podobe finančnej alebo v podobe vynaloženého času a tvorivého úsilia), ktorá má okrem samotnej ochrany práva autora k výsledku tvorivej činnosti, motivovať ďalšiu tvorbu (umeleckých) diel skrz záruku návratnosti autorovej investície, ako spoločensky žiadaný jav.³⁵ Je teda evidentné, že úpravu tejto problematiky nemôže zákonodarca do budúcnosti odbyť ako niečo nepodstatné, či len ako akýsi prechodný problém.³⁶

Otázka, s ktorou sa treba do budúcnosti teda vysporiadať, je otázka, či by mala byť, s ohľadom na vývoj technológií a ekonomický záujem autora, pôsobnosť autorského zákona rozšírená aj na prípady, kedy je syntetizovaná forma spracovania diela a aplikovaná v systémoch strojového učenia, alebo či je súčasná úprava dostačujúca a rozšírenie

34. Spoločnosť Google vyvíja a poskytuje verejnosti framework, alebo knižnicu kódu, pod názvom *TensorFlow*, ktorá obsahuje bloky kódu, ktoré môže programátor použiť vo svojom kóde iba skrz jednoduchý odkaz. Vhodnou právnickou metaforou by mohlo byť pripojenie externých obchodných podmienok k zmluve jednoduchým odkazom v texte na prílohu, či ich umiestnenie na webe. Pri využívaní takejto knižnice (frameworku) sa teda programátor nemusí zdržiavať tým, že bude dookola písať ten istý kód, či implementovať rovnakú myšlienku.

35. SRSTKA, Jiří. Op. cit. s. 43 - 45.

36. Ibid.

ponímania autorského diela, kedy nie je rozhodné jeho zachytenie je neprimeraným rozšírením, ktoré by zabránilo inováciám na poli technológií, ktoré sú rovnako potrebné z pohľadu spoločnosti. Prípadne, či existuje nejaká stredná cesta, ktorá by sa zamerala na samotný zber dát, ako na ich zachytávanie na hmotný nosič (a prípadnú následnú distribúciu) bez ohľadu na ich využitie pri učení systémov umelej inteligencie.

3.1 Nové užitie a Smernica o autorskom práve na jednotnom digitálnom trhu

V súčasnosti je prejednávaná Smernica o autorskom práve na jednotnom digitálnom trhu,³⁷ ktorá už v prípravnej fáze vzbudila mnoho ohlasov.³⁸ Jeden z jej článkov, článok 3, sa venuje aj "hlbkovej analýze dát".³⁹ Na základe tohoto článku by mala vo vnútroštátnych úpravách vzniknúť výnimka z ochrany, resp. zákonná licencia, pre hlbkovú analýzu údajov, ak ide o extrakciu a rozmnožovanie diela výskumnými organizáciami za účelom výzkumu.

Extrakcia dát za účelom výzkumu je v podstate webscraping dát za účelom nadobudnutia datasetu, na ktorom bude následne počítačovo činená hlbková analýza. Jedná sa teda o identický proces, ako v prvom kroku web scrapingu za účelom získania datasetu pre účely strojového učenia.

Znamená to teda, že pre túto činnosť by mala existovať zákonná výnimka? Nie nutne. V prvom rade, navrhovaná výnimka sa vzťahuje len na vedecký výzkum, ktorý už svoju kategóriu v autorskom zákone má, a omnoho páľčivejšou otázkou v súčasnosti je situácia komerčne využívaných systémov. Dalšia otázka, ktorú si treba položiť, je či teda požiadavka existencie zákonnej licencie implikuje aj autorskoprávne užitie.

37. Návrh Smernice Európskeho Parlamentu a Rady o autorskom práve na digitálnom jednotnom trhu 2016/0280. Ďalej len "Smernica".

38. Návrh priniesol četné diskusie o obmedzovaní slobody prejavu na internete, či cenzúre; viď. napr.: SULÍK, Richard. *Smernica EÚ o autorskom práve obmedzí slobodu prejav* [online]. sulík.sk, publikované 13.06.2018 [cit. 04.03.2019]. Dostupné z: <https://sulik.sk/smernica-eu-o-autorskom-prave-obmedzi-slobodu-prejavu/>.

39. Smernica, op. cit.

I keď technologický prístup k zberu dát je viac menej identický, tento faktor pre takýto záver nie je dostačujúci. Ako bolo opísané v častiach 2.1 a 2.2 vyššie, i keď sa štandardne neskúma následné využívanie zozbieraných údajov, v tomto prípade to potrebné je, nakoľko následné užitie môže odstrániť niektorý z predpokladov uplatnenia predmetných autorskoprávných užití, či dokonca celej autorskoprávnej ochrany. Navyše stanovením výnimky pre (následný) vedecký výzkum, nás k takémuto zamysleniu vyzýva aj Smernica.

Ako bolo opísané vyššie, pre využívanie dát k strojovému učniu nie je potrebné zachovanie ich vyjadrenia (teda toho, čo autorský zákon chráni) a taktiež primárnym zámerom je určitým spôsobom kvantifikovať tvorivý proces, ktorý k vytvoreniu diela (a jeho zachytenia) vedie. Takéto užívanie teda nemožno k využívaniu pre hĺbkovú analýzu pre výzkumné účely prirovnávať.

I keď sa Smernica zaoberá podobným prístupom a technologickým konceptom, nemožno ich stavať na roveň, či vyvodzovať zákonné usmernenie pre strojové učenie z úpravy dátovej analýzy. Ak niečo, tak možno vnímať ako dôležitý fakt to, že na základe tejto Smernice možno vytváranie datasetov pomocou webscrapingu možno považovať za rozmnožovanie, ktoré by následne bolo autorskoprávne chránené. Nemožno však z nej vyvodzovať takúto ochranu aj pre situáciu, kedy je takýto dataset bezprostredne využitý pre potreby strojového učenia.

3.1.1 Možné východiská

Na základe súčasnej úpravy ako aj snahy Európskeho parlamentu je možné dospieť k záveru, že samotné vytváranie datasetu, je skutočne autorskoprávnym užitím, v konečnom dôsledku sa jedná o vytváranie súborného diela. Jeho bezprostredné využitie pre potreby strojového učnia však vytvára problematickú situáciu, nakoľko už autorskoprávna ochrana prestáva byť jednoznačná, a presne to je otázka, ktorú treba do budúca riešiť.

Z môjho pohľadu sa v súčasnosti núka niekoľko východisiek, pričom každé nesie svoje pre a proti. Prvá z možností je zachovať súčasný stav, pričom by sa nemalo jednať len o nereflektovanie technologického vývoja. Mal by to byť vedomý krok, ktorého cieľom je technologické napredovanie a inovácie, ktoré sú potrebné pre zachovanie trhovej konkurencieschopnosti, ako aj pre potencionálny prínos, ktorý by mohli

mať pre spoločnosť. Na druhej strane je tu ochrana investícií autorov a ich práva na "plody svojej práce", ktoré má okrem ich ochrany slúžiť aj ako motivácia pre vytváranie nových diel, čo je rovnako v záujem spoločnosti. Takýto prístup možno rovnako tiež vnímať ako aj najjednoduchšie implementovateľný, nakoľko sa podľa neho do veľkej miery operovalo od samotného vzniku takýchto technológií.

Druhá možnosť je prirodzene tou opačnou k predchádzajúcej. Strojové učenie, či iné formy komerčného data miningu by boli vnímané ako autorskoprávne užitie, s odkazom na súčasnú taxatívnu ochranu (ktorá bola zavedená ako taká najmä s odkazom na nemožnosť predpokladania stavu rýchlo sa vyvíjajúcich technológií) ako aj na vyššie spomínanú ochranu investícií. Prirodzene by sa situácia otočila a plusom by bola ochrana práv autorov, mínusom potencionálny negatívny dopad na technologický rozvoj. Rovnako tak by to do veľkej miery vytvorilo neistotu pre už existujúce platformy a projekty, ktoré doteraz operovali v tak trochu "šedej" zóne na základe modelu číslo jedna.

Poslednou možnosťou, ktorá by bola implementačne najkomplexnejšou, je vytvorenie nových, či rozšírenie pôsobnosti súčasných autorských ochranných zväzov, ktoré by reflektovali aj užívanie chránených diel pre potreby strojového učenia⁴⁰ (pričom by sa v tomto prípade však muselo jednať o autorské užitie, a teda tento model je rozšírením modelu číslo 2). Tento model by teda hľadal určitý kompromis medzi autormi, ktorí by stále profitovali zo svojho diela, a subjektami využívajúcimi tieto diela, nakoľko by sa im uľahčil (legálny) prístup k nim. Negatívom je však komplexnosť takejto implementácie rovnako tak ako aj ani zďaleka nie ideálne fungovanie autorských zväzov v ich súčasnom rozsahu pôsobenia.

40. Tak ako v súčasnosti napríklad OSA zastupuje súbor autorov a poskytuje užívateľom licenciu napríklad k reprodukcii ich hudby, by takéto existujúce zväzy mohli poskytnúť licenciu užívateľom, k využívaniu diel pre potreby strojového učenia. Ideálne by reflektovali skutočnosť, že nebude dielo zverejňované a bola by nimi k takému udeľovanej iná licencia, či by vznikol nový zväz k udeľovaniu výhradne licencií pre strojové učenie.

4 Záver

Táto práca mala za cieľ konfrontovať súčasnú autorskoprávnu úpravu s ohľadom na rozvíjajúce sa technológie, najmä s ohľadom na umelú inteligenciu. Práca mala zväziť autorskoprávne užitie diel pre vytváranie datasetov za účelom učenia systémov umelej inteligencie.

K záveru tejto práce bolo jasné, že s ohľadom na technologické vlastnosti, rozpísané v tejto práci, nie je možné za súčasného nastavenia autorskoprávnej ochrany možnú takúto činnosť považovať za autorskoprávne užitie. Tento záver vychádza najmä z dichotómie medzi formou a obsahom, respektíve vyjadrením a myšlienkou. Pre autorský zákon je v súčasnosti rozhodné zachytenie diela, pre systém umelej inteligencie zase myšlienka, a vo svojom procese spracovávajúce dáta tieto dve náležitosti oddeľuje.

Otázkou v závere práce ostáva budúci vývoj v tejto problematike. Je možné v súčasnosti pozorovať určitú vôľu na európskej úrovni, s ohľadom na rozoberanú Smernicu, avšak ešte stále nezohľadňuje všetky aspekty, ktoré prináša so sebou rozvoj umelej inteligencie - z pohľadu tejto práce je to následné využívanie, nie len zbieranie dát.

V závere sa táto práca snažila predstaviť niekoľko potenciálnych riešení. V prvom rade by sa mohlo jednať o zachovanie súčasného stavu, kedy takéto užitie nie je klasifikované ako autorskoprávne, riešenie ideálne z pohľadu inovácií a technológií, ktoré autorské diela v súčasnosti takto už užívajú. Druhou možnosťou bolo vytvorenie nového autorskoprávneho užitia, ktoré by zohľadňovalo aj následné použitie zozbieraných autorských diel, teda riešenie vhodné pre autorov. Poslednou možnosťou by bolo vytvorenie určitej formy kolektívnej správy autorských práv, možnosť v teórii vhodná pre obe skupiny, v praxi však kolektívna správa nefunguje úplne ideálne.

Do budúcnosti bude teda rozhodne nutné sa vysporiadať s režimom takéhoto užitia autorských diel, ideálne v čo najkratšom časovom horizonte, tak aby neboli užívatelia ponechaný v neistote, tak ako je tomu teraz. S ohľadom na zavádzanie takejto ochrany bude tiež nutné sa vysporiadať s otázkou dichotómie formy a obsahu, ktorá v súčasnosti zabraňuje autorskoprávnemu riešeniu takéhoto problému.

Použitá literatúra

1. DeepArt UG (haftungsbeschränkt). <https://deepart.io>.
2. FETZER, James H. *Artificial Intelligence: Its scope and limits*. Dordrecht: Kluwer Academic Publishers, s. 337. ISBN 978-0-7923-0548-4.
3. GATYS, Leon A. a kol. *A Neural Algorithm of Artistic Style*. arXiv.org [online]. Cornell University, publikované 26.08.2015 [cit. 02.03.2019]. Dostupné z: <https://arxiv.org/abs/1508.06576>.
4. REESE, Byron. Episode 1: A Conversation with Yoshua Bengio. *Voices in AI* [online]. GIGAOM, [citovné 03.03.2019]. Dostupné z: <https://voicesinai.com/episode/1-a-conversation-with-yoshua-bengio/>.
5. KŘÍŽ, Jan a kol. *Autorský zákon: Komentář a předpisy souvisící*. Praha: Linde Praha, a. s., 2005, 792 s. ISBN 80-7201-546-X.
6. LAKSHMI, Prasanna Kaspá a kol. *Towards Extended Data Mining: An Examination of Technical Aspects*. In: *Procedia Computer Science*, 2018, vol. 139, s. 49 - 55. ISSN 1877-0509. Dostupné z: <https://www.sciencedirect.com/science/article/pii/S1877050918318854>.
7. MAŘÍK, Vladimír a kol. *Umělá inteligence (1)*. Praha: Academia, 2001, 264 s. ISBN 80-200-0496-3.
8. Návrh Smernice Európskeho Parlamentu a Rady o autorskom práve na digitálnom jednotnom trhu 2016/0280. In: EUR-Lex [právny informačný systém]. Úrad pre publikácie Európskej únie [cit. 03. 03. 2019]. Dostupné z: <https://eur-lex.europa.eu/legal-content/SK/TXT/?uri=CELEX%3A52016PC0593>.
9. Rozsudok Odvolacieho Súdu (Druhý obvod) vo veci The AUTHORS GUILD v. GOOGLE, INC. Dostupné z: <http://madisonian.net/downloads/copyright/authorsguild.pdf>.

4. ZÁVER

10. Shared Task: Machine Translation. USA: Baltimore, publikované 26.06.2014 [cit. 01.03.2019]. Dostupné z: <http://www.statmt.org/wmt14/translation-task.html>.
11. SIGANOS, Dimitrios; STERGIUO Christos. *Neural Networks*. Londýn: Imperial College of Science [online], [cit. 03.03.2019]. Dostupné z: <https://bit.ly/236gERF>
12. SIMARD, Patrice Y. *Best Practices for Convolutional Neural Networks Applied to Visual Document Analysis*. In: Proceedings of the Seventh International Conference on Document Analysis and Recognition, vol. 2, s. 958–962, 2016. Dostupné z: <http://www.cs.cmu.edu/~bhiksha/courses/deeplearning/Fall.2016/pdfs/Simard.pdf>.
13. SRSTKA, Jiří a kol. *Autorské právo a práva související*. Praha: Leges, 2017, 414 s. ISBN 978-80-7502-240-0.
14. TELEČ, Ivo; TŮMA, Pavel. *Autorský zákon: Komentář*. Praha: C. H. Beck, 2007, 971 s. ISBN 978-80-7179-608-4.
15. Zákon č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon). In: *Beck online* [právní informační systém]. C. H. Beck [cit. 03.03.2019].