

Autoreferát disertační práce

Disertační práce: Právo, společnost a nové technologie 21. století
Studijní program: Teoretické právní vědy – Teorie, filozofie a sociologie práva
Autor: JUDr. Jan Hořeňovský
Školitel: Mgr. Petr Agha, LL.M., Ph.D.

1 Shrnutí disertační práce v českém jazyce

1.1 Cíle a obsah disertační práce

Disertační práce se věnuje široce vymezenému tématu vztahu práva, společnosti a nových technologií 21. století. Text je rozdělen na část obecnou, která se zaměřuje na oblasti tématu společné, a část zvláštní, která je úzce zaměřena na určité konkrétní problémy spojené s partikulárními technologiemi, jako jsou umělá inteligence nebo sociální sítě. Každá z těchto částí je dále rozdělena do tří kapitol. Texty vznikaly původně samostatně, avšak myslím, že se mi je podařilo seřadit a upravit způsobem, který poskytuje koherentní celek. I tak ale vnímám, že jde o text velmi široce rozkročený, který spoustu témat nevyhnutelně pouze lehce načrtne. Jde o daň za zpracovávání širokého tématu, kterou však v dnešní době stále rostoucí partikularizace dle mého mínění stálo za to zaplatit.

První, úvodní, kapitola vychází ze sociologie práva a jejím účelem je vytyčit základní terén, který je později hlouběji rozveden v dalších kapitolách ve vztahu ke konkrétním socio-technologickým změnám. Tato část vychází ze systémové teorie, která má podle mého názoru nejlepší předpoklady popsat fungování evoluce/adaptability práva v kontextu měnícího se socio-technologického prostředí. Jde o část, která, byť nedokonale, vytváří důležité uvedení tématu a rámec pro popsání základních jevů. Teorie zde představená přitom funguje obecně na studium jakýchkoli technologií v kontextu práva a společnosti.

Musím na upozornit na jednu důležitou limitaci. Aplikace tohoto rámce je ze své podstaty omezena v tom smyslu, že předpokládá, že právní systém je relativně stálá praxe, která omezeným způsobem reaguje na dynamický socio-technologický kontext. Stále však musíme mít na paměti i možnost, že socio-technologické změny, k nimž dochází v okolí práva, mohou mít více než jen dopad na účinnost práva, slouží-li regulačním politikám. Tyto změny mohou nakonec zcela proměnit i samotnou povahu práva, například směrem ke stavu, který někteří autoři nazývají *právní singularitou*. Takový stav je však již mimo záběr této práce.

Druhá kapitola se věnuje netradičnímu tématu technologické stagnace. Pojem „technologie“ totiž konceptuálně nezahrnuje pouze tzv. „informační technologie“, ale i ty ostatní. Varující je skutečnost, že i samotné předporozumění slova „technologie“ se v posledních letech proměňuje. Například „centrum práva a technologií“ by ještě před několika desítkami let znamenalo pracoviště, které řeší jaderné reaktory a jejich právní úpravu. Dnes bychom si takřka výlučně představili pracoviště práva informačních technologií. Jde o posun v jazyce, který nám říká něco důležitého o společnosti a prostředí, ve kterých žijeme.

Technologie totiž stále znamenají i oblast energetiky, stavebnictví, léků, dopravy atd., přičemž v těchto oblastech stále více autorů dochází k názoru, že zažíváme technologickou stagnaci. Originalita mého přínosu v tomto podtématu spočívá v pokusu o hlubší propojení těchto hlasů s otázkami fungování a problémů současného práva.

Třetí kapitola, poslední v obecné části, je silně experimentální. Snaží se totiž v kontextu systémového pohledu na právo vysvětlit roli technologií ve vztahu k entropii a komplexitě jak těchto systémů, tak jejich prostředí. Věnuji se zde otázce, jak technologie ovlivňují praxi práva a co mohou být benefity a omezení s tímto jevem spojená. Jde však pouze o prvotní náčrt tohoto dosud velmi málo probádaného tématu.

Čtvrtou kapitolu považuji za nejvýznamnější, a to jak z pohledu originality a inovativnosti, tak svojí důležitostí pro společnost a právo. Věnuje se totiž otázce dlouhodobých aspektů umělé inteligence, přičemž velmi silně akcentuje otázku bezpečnosti těchto systémů. Propojuji zde oblast teorie práva a teorie lidských práv s technickým pohledem, jak by se dala taková bezpečnost zajistit. Konkrétně je zde představen návrh, jak by se dalo inkorporovat dodržování negativních lidských práv do samotné technologie, včetně doprovodných regulativních mechanismů mezinárodního práva. Dále se zde věnuji důležitému tématu zapojování umělé inteligence do forem vládnutí a otázce právní subjektivity těchto strojů.

Pátá kapitola se věnuje aktuální otázce svobody projevu na veřejných internetových fórech. Snažím se zde hájit liberální pozici, která akcentuje důležitost této svobody a vyvrací argumenty pro její omezování spojené s morální panikou například ve vztahu k otázce šíření tzv. „dezinformací“. V této kapitole se snažím shrnout významné poznatky k tomuto tématu a představit pohled, který nabízí konstruktivní společenská a právní řešení.

Poslední kapitola reflektuje dobu pandemie Covid-19 a řeší některé otázky spojené s využíváním moderních technologií v boji s pandemií.

Tematicky jde o text, který by měl sloužit všem zájemcům o pochopení role práva v regulaci nových technologií. Je určen pro ty, kteří věří, že technologicky vyspělejší a lepší budoucnost je dosažitelná a že musíme aktivně hledat cesty, jak jí dosáhnout. Věřím, že partikulárně bude tato práce užitečná především všem právním akademikům, kteří se podobným tématům budou chtít věnovat. Pro ty by myšlenky zde obsažené mohly fungovat jako dobrý odrazový můstek pro vlastní výzkumné projekty. Přál bych si však, aby se tato práce propsala i do právní praxe, a to jak v cestě legislativní, tak judiciální, neboť jedině tak bude moci reálně ovlivnit společnost. V neposlední řadě si myslím, že je můj text psaný dostatečně srozumitelně,

aby jej mohla číst i širší veřejnost. Věřím proto a doufám, že na moji práci navážou další autoři a že jim tento text může být jak inspirací, tak praktickým pomocníkem.

Cílem této práce bylo obrazně rozepsat část knihy, kterou v mnoha důležitých detailech budou muset dokončit jiní. Toto reflektuje skutečnost, že téma vztahu dynamiky technologického vývoje a společenských a právních mechanismů je ohromným a dosud nedostatečně probádaným polem. Právní systém jistě může sloužit jako důležitý ochranný mechanismus společnosti před technologickou disrupcí, avšak jako každý lidský výtvar, má svoje přirozené limity. V této práci jsem se snažil jak možnosti práva, tak jeho limity, v základech představit.

S tématem technologií jsou navíc silně spojeny i otázky etické a politické, související s tím, zda vůbec jako společnost máme funkční veřejné prostory, kde by se například o potenciálně nebezpečných technologiích (a o podobách regulace) dalo smysluplně diskutovat. Absence smysluplných společenských debat je v tomto kontextu zásadním problémem. Technologie nevyhnutelně vyvolávají nepříjemné otázky, neboť vedou k hlubokým změnám ve společnosti. Otázky, zda právní normy zůstanou relevantní, a pokud ne, jak by se měly změnit, jsou zcela fundamentální a v základu se dotýkají každého z nás. Pokud nebude, kde je smysluplně adresovat, obávám se, že se s nimi jako společnost dostatečně nevy pořááme.

Jak bylo také již upozorněno, samotný pojem „technologie“ se v posledních letech významově zúžil v podstatě jen na otázku informačních technologií. Osobně bych si velmi přál, abychom během našich životů zažili výrazný technologický progres v nejširším – dnes již skoro zapomenutém – slova smyslu, zahrnujícím i v současnosti stagnující oblasti jako stavebnictví, energetika či doprava. Zda se technologický progres přenesení do těchto oblastí, bude však do značné míře záležet na nás samotných. Nové technologie nutně potřebujeme a existují cesty, jak strukturovat (nejen) právní regulaci způsobem, který k takovému progresu povede. Tato práce některé cesty, jak toho dosáhnout, nastínila. Jde o důležitý první krok.

Toto století také hrozí novými ohroženími základních liberálních svobod. Technologie mohou v tomto kontextu být velmi nebezpečným hráčem, ale i dobrým pomocníkem. Existují přitom elegantní cesty, jak základní práva a svobody v technologickém kontextu chránit. Některé prvotní myšlenky, jak toho dosáhnout, zde byly představeny. Za nejvýznamnější v této oblasti považuji návrh inkorporace negativních lidských práv do samotných AI systémů. Jsem si však vědom, že bez další – právní i technické – práce, a hlavně bez praktické implementace, tento návrh nebude mít takový efekt, jaký by byl potřeba.

V neposlední řadě žijeme v období společenských změn, kdy se demokratické režimy dostávají do konfrontace jak s autoritářskými a technokratickými způsoby vládnutí, tak s vnitřní tendencí k sebedestrukci liberálních principů svobodné otevřené společnosti. Proč musíme těmto vnitřním tendencím, i externím nepřátelům čelit, bylo posledním výrazným tématem této práce.

Domnívám se, že téma vztahu společnosti k technologiím je v současnosti nezdravě narušené. Za to lze vinit skutečnost, že technologický progres v poslední dekádě přinesl obří zisky několika technologickým korporacím v malé části USA, avšak zbytek společnosti – včetně demokratického zřízení – z něj buď neprofitoval vůbec, nebo jen velmi málo. Myslím si však a doufám, že toto nebude příběh této dekády a že pokrok v umělé inteligenci se skutečně může pozitivně propsat do životní úrovně většiny lidí. Aby se tak stalo, potřebujeme zvrátit negativní pohled na budoucnost a technologie. Budoucnost může být lepší a již v současnosti existují zcela konkrétní technologické cesty, jak takové budoucnosti dosáhnout.

Pokud tato práce alespoň trochu přispěje tomu, abychom takové budoucnosti dosáhli, budu považovat její cíl za splněný.

1.2 Výsledky disertační práce a charakteristika původního řešení

KAPITOLA 1: Adaptabilita právních systémů v éře socio-technologických změn

Ve svém konceptuálním rámci vycházím ze systémové teorie, která považuje právní systém za *operativně uzavřený autopoietický systém*.¹ Pro právní systém je z tohoto pohledu klíčové jeho oddělení od prostředí.² Buď něco je součástí právního systému, nebo je to prostředím, ve kterém právní systém působí. Uzavřenost v tomto kontextu však nelze považovat za úplnou separaci (odloučenost), pouze se tím chce říct, že se právní systém reprodukuje skrze „*vytváření vlastních operací*“.³ Autopoietická uzavřenost žádným způsobem nevylučuje kauzální působení systému se svým prostředím ani možnou závislost systému na prostředí (a naopak závislost prostředí na systému).⁴

Operativní oddělenost mezi právním systémem a jeho socio-technologickým prostředím přináší nutnost *relačního pohledu*.⁵ Pochopení vztahu (relace) přitom vyžaduje zachycení vzájemné dynamiky. Analýza každého prvku zvláště samozřejmě může přinést cenné partikulární poznatky, nepomůže nám však s pochopením vysvětlujícího rámce relačních jevů. Jinými slovy, potřebujeme nejdříve pochopit dynamiku vztahu, abychom zjistili, jak a proč záleží na vlastnostech jednotlivých prvků relace.⁶

Vzhledem k relační povaze adaptace práva nelze vytvořit stabilní produktovou teorii, jak by se měl právní systém ve vztahu k prostředí přizpůsobovat nebo dokonce jak by měl optimálně trvale vypadat. Ze své podstaty se totiž vůbec nejedná o jakkoli stabilní a tudíž předvídatelnou kategorii. To, co dobře funguje dnes, nemusí být vhodné zítra, a co bude vhodné zítra, může být zcela nepoužitelné v delším časovém horizontu, kdy se prostředí výrazně změní. Jde o stejný problém, jako bychom chtěli vymyslet optimálního biologického predátora. Žralok operuje výborně v moři, ale pokud toto prostředí vyschne a on se ocitne na poušti, jeho dříve vynikající vlastnosti budou zcela nepoužitelné.

Z těchto důvodů platí, že je možné přijít pouze s procedurálními přístupy vysvětlujícími, jak může být adaptabilita/evoluce právního systému ve vztahu k měnícímu se socio-

¹ LUHMANN, N. *Introduction to Systems Theory*. Cambridge: Polity Press, 2013, s. 63–83.

² Tamtéž, s. 41.

³ Tamtéž, s. 41.

⁴ BAXTER, H. Niklas Luhmann's Theory of Autopoietic Legal Systems. *Annual Review of Law and Social Science*, 2013, vol. 9, no. 1, s. 170.

⁵ K relačnímu přístupu v sociálních vědách obecně viz SELG, P.; VENTSEL, A. *Introducing Relational Political Analysis. Political Semiotics as a Theory and Method*. Palgrave Macmillan, 2020.

⁶ Jde o myšlenku Adama Růžičky vyjádřenou ve vztahu ke krizím současných demokracií.

technologickému prostředí kontextuálně realizována.⁷ Takový proces spočívá v rozbití vnitřní struktury systému a jeho opětovném začlenění do nové struktury. Jedná se o vývojový proces sebeorganizace systému.

Jak vyplývá z výše uvedeného, právní systémy jsou vztahově provázané se svým socio-technologickým prostředím a vstupují do komplikovaných relací s různými dalšími společenskými sub-systémy (např. s ekonomikou či politikou). Může se přitom stát, že se socio-technologické prostředí, v němž právní systém působí, výrazně změní – toto nazýváme *socio-technologickou změnou* – přičemž právní systém tuto změnu po dlouhou dobu nemusí kognitivně vnímat a vnitřně reflektovat. V takovém případě právní systém, právě vzhledem ke své operativní uzavřenosti, pokračuje v reprodukci závěrů (např. rozhodnutí), jako kdyby tato změna vůbec nenastala, což může narušovat dříve dobře fungující vztah. Toto může mít nepříznivý vliv jak na ostatní společenské sub-systémy, tak i na legitimitu samotného právního systému.

Jakmile dokážeme identifikovat relevantní socio-technologické změny a jejich škodlivé společenské důsledky, musíme si položit následující otázku: Jak se může právní systém vnitřně přizpůsobit změněné realitě? Toto reflektuje skutečnost, že evoluce práva může být popsána jako „*souhra mezi endogenními evolučními mechanismy právního systému a exogenní evolucí společnosti jako celku*“.⁸ A že pohledem Luhmanna: „*evoluce probíhá pouze tehdy, pokud jsou ve vztahu mezi systémem a prostředím zachovány jak odlišnosti, tak adaptace*“.⁹

Představená východiska posléze propojují se třemi základními modely regulace nových technologií, které lze současně považovat za způsoby vnitřní změny právního systému, a ještě o krok dál za metodologii v oblasti doktrinální právní vědy. Pro další výzkum ponechávám otevřený sociologický pohled na tyto modely, který by nad rámec zde představeného musel vysvětlit také podmínky, jež způsobují, že se právní systém rozhoduje pro jeden z těchto přístupů. Takovou ambici nemám, tato část proto spíše zůstává na úrovni teorie regulace nových technologií.

První možností, jak může právní systém reagovat na měnící se socio-technologické prostředí, je *nedělat nic a tvářit se, jako by ke změně nedošlo*. Jedná se o pasivní model, jehož podstatou je zachování nezměněného vnitřního fungování a reprodukce právního systému.

⁷ Pozn. podobně jako u biologické teorie evoluce nebo teorie realizace relevance v kognitivní vědě.

⁸ TEUBNER, G. Substantive and reflexive elements in modern law. *Law & Society Review*, 1983, vol. 17, no. 2, s. 263.

⁹ LUHMANN, N. *Law as a Social System*. Oxford: OUP, 2004, s. 231.

Tento přístup je z pohledu tvůrců práva, správních orgánů či soudců přímočarý a lákavý, ale v mnoha případech také velmi problematický vzhledem k výše diskutované možnosti technologicky zapříčiněných policy driftů.

Druhým možným přístupem je tzv. *law-centric model*. V tomto modelu se na socio-technologickou změnu nahlíží optikou stávajícího právního rámce, případně jeho možných drobných modifikací. V rámci tohoto přístupu by právníci měli reinterpretovat stávající právo nebo provést skrze legislativu drobné změny ve stávající právní infrastruktuře tak, aby byl zachován účel právní úpravy, který je v důsledku socio-technologických změn a technologiemi způsobených policy driftů v nové situaci nenaplněván.

Třetím možným přístupem je tzv. *emergentní model*. Podstatou tohoto modelu je zavedení nové komplexní regulace, popřípadě dokonce nového odvětví práva. Do tohoto modelu bych také zahrnul vysoce inovativní přístupy, které jsou nyní nepředstavitelné, ale k nimž může socio-technologický a právní vývoj vést.

Tato kapitola načrtla některá základní východiska vztahu práva, společnosti a nových technologií, z nichž některá jsou podrobněji rozvedena v následujících kapitolách. Zároveň ukázala, proč přílišná konzervativnost práva může značně uškodit nejen společnosti, ale i samotné legitimitě právního systému. Věřím, že se tato kapitola může stát dobrým výchozím bodem pro uvažování o regulaci nových technologií obecně.

KAPITOLA 2: Právní a kulturní aspekty teze o technologické stagnaci

Dominantním narativem, převažujícím v dnešní západní společnosti, je přesvědčení, že žijeme v době nebývalého technologického pokroku, jenž historicky nemá obdoby. Máme sociální sítě, chytré telefony, internet a směřujeme k vyspělé umělé inteligenci. Myslíme si, že se nacházíme v éře bezprecedentního technologického progresu, že se věci mění stále rychleji a že technologie mohou být velmi destruktivní záležitostmi, proto je lepší vůči těm novým zachovávat obezřetnost.

Tento narativ je však v posledních letech napadán řadou významných autorů jako Vijg nebo Thiel, jejichž perspektiva v českém a evropském veřejném prostoru schází.¹⁰ Tito autoři poukazují na skutečnost, že progres byl v posledních zhruba 50 letech zaměřen především na informační technologie (a částečně biotechnologie), zatímco v ostatních oblastech, jako je

¹⁰ Viz VIJG, J. *The American Technological Challenge: Stagnation and Decline in the 21st Century*. New York: Algora Publishing, 2011; THIEL, P.; MASTERS, B. *Zero to One. Notes on Start Ups, or How to Build the Future*. London: Ebury Publishing, 2014.

energetika, doprava či stavebnictví, zažíváme výraznou stagnaci.¹¹ Jejich argument je komparativní a je založen na tvrzení, že se ve srovnání s první polovinou 20. století tempo technologických změn ve fyzickém světě výrazně snížilo.

Tato kapitola z tohoto paradigmatu vychází a klade si dva cíle. Prvním je seznámit čtenáře s tezí o technologické stagnaci. Druhým je analyzovat, jaké mohou být příčiny tohoto jevu a jakou roli v něm může sehrávat právní systém. Mluví-li se zde přitom o právním systému, moje pojetí má nejbližší, jak již bylo zmíněno dříve, k systémové teorii práva.¹² Toto pojetí dobře ilustruje, že mění-li se socio-technologické prostředí, ve kterém právo působí, mohou se společenské dopady vnitřně statického právního systému znatelně lišit od původně zamýšlených. Do pojmu právního systému přitom zahrnuji také faktickou institucionální aktivitu, nikoli pouze textaci právních norem a principů.

Ambicí této části není tento problém vyřešit ani čtenáře přesvědčit, že všechny argumenty v něm obsažené jsou nutně správné, nýbrž spíše na tuto problematiku upozornit českou a slovenskou odbornou veřejnost. Zároveň věřím, že tento text může otevřít dveře dalšímu výzkumu na poli jednotlivých oborů práva, včetně potenciální empirické analýzy. V závěru nabízím diskusi, ve které některé své hypotézy zpochybňuji.

KAPITOLA 3: Právní systém a zvyšující se komplexita

Třetí kapitola obecné části se věnuje analýze, co pro právní systém znamená zvyšující se komplexita jak vnějšího světa, tak samotného právního systému a jakou roli v tomto jevu hrají moderní technologie. Jde o téma, které již bylo v předchozích kapitolách naznačeno, avšak až zde se mu dostává hlubšího rozpracování. Tato kapitola ukazuje, že existují hranice entropie a komplexity právního systému a že tyto limity mohou nové technologie posouvat. Existuje také fenomén zvyšování entropie a komplexity ve vnějším světě, na čemž mají technologie významný podíl. Podoba vývoje právního systému do budoucna je velmi nejasná a těžko předvídatelná. Domnívám se však, že zde pojmenované fenomény mohou pomoci při úvahách o proměně právního systému v současnosti a budoucnosti a že mohou ukázat na některé obecné jevy, které v evoluci práva nastávají.

KAPITOLA 4: Umělá inteligence a právo v dlouhodobé perspektivě

Tato kapitola se zabývá oblastmi, které se tématu práva, společnosti a umělé inteligence týkají v horizontu nejbližších desetiletí. Na úvod proto diskutuji vůbec opodstatněnost takového

¹¹ BETZ, U. Is the force awakening? *Technological Forecasting & Social Change*, 2018, vol. 128, s. 299–300.

¹² LUHMANN, N. *Law as a Social System*. Oxford: Oxford University Press, 2004, s. 76–142.

přístupu v kontextu právní vědy. V první části se posléze věnuji otázce právní subjektivity umělé inteligence. V druhé části se zaměřuji na některé existenční hrozby spojené s umělou inteligencí a na způsoby, jak by jim právo mohlo pomoci zabránit. V poslední části se věnuji tématům AI totalitarismu a AI technokracie, což je diskuse týkající se nově se objevujících forem vlády v kontextu využití umělé inteligence ve správě věcí veřejných.

Právní subjektivita a AI

V první části této kapitoly je probírána umělá inteligence a možné reakce práva *de lege ferenda* ve vztahu k otázkám, které může vyvolávat potřeba nechat se zastupovat umělou inteligencí v právních vztazích. Jedním ze zde prezentovaných řešení je učinit z umělé inteligence plnohodnotný subjekt práva, který by se zařadil vedle osob fyzických a právnických. Druhou probíranou variantou je užití právní konstrukce otroctví z římského práva, která se ve vztahu k umělé inteligenci jeví, při jiném pojmenování, vhodně. Třetí variantou je ponechat umělé inteligenci status věci; tato možnost ale může velmi brzo začít narážet na potřeby společnosti.

Právo a dlouhodobá bezpečnost AI

Toby Ord ve své knize *Nad propastí: Existenční riziko a budoucnost lidstva* uvádí pravděpodobnost jedna ku deseti, že lidstvo do konce tohoto století vymře kvůli pokročilé AI.¹³ Existuje-li zde takto velké riziko, že naše jednání bude mít v horizontu několika málo generací zcela fatální následky, pravděpodobně jde o téma určité pozornosti hodné i v současnosti. I kdyby tato pravděpodobnost totiž ve skutečnosti byla o řád nižší, například na úrovni jedna ku stu, pořád by šlo o nezanedbatelný problém. A to především kvůli možným dopadům v podobě úmrtí miliard lidí žijících v současnosti a zabránění životu desítek až stovek miliard lidí žijících v budoucnosti.

Podobně varovný pohled nabízí například Bostrom, který situaci, ke které pomalu směřujeme, analogicky popisuje, jako by myši vyvíjely umělého člověka za účelem, aby jim poskytoval sýr.¹⁴ Rozdíl v inteligenci mezi námi a pokročilou AI může být podobně významný jako mezi myší a člověkem. Je realistické předpokládat, že by myši byly v takové situaci schopné člověka plně ovládat a že by člověk dělal pouze to, co myši původně zamýšlely, aby

¹³ ORD, T. *Nad propastí: Existenční riziko a budoucnost lidstva*. Praha: Argo, 2022, s. 166 a násl.

¹⁴ BOSTROM, N. *Superintelligence: až budou stroje chytřejší než lidé*. Praha: Prostor, 2017.

dělal? Spíše nikoliv. V kontextu dlouhodobé bezpečnosti AI systémů jsou proto klíčové dva koncepty:

- 1) problém kontroly;¹⁵
- 2) sladění AI systémů s lidskými zájmy a hodnotami (tzv. AI alignment).¹⁶

V prvním případě jde o stále více naléhavou otázku, jak udržet kontrolu nad něčím, co bude každého z lidí převyšovat ve schopnostech i intelektu. V druhém případě jde o problém, jak zajistit, aby AI systémy jednaly v souladu s našimi nejlepšími zájmy a hodnotami, a to nejen úmyslně, ale i skrze nezamýšlené důsledky jejich jednání. Podíváme-li se na tento problém z pohledu práva, bude nás zajímat, jak může právní věda či regulace přispět v dlouhodobém horizontu k tomu, aby vývoj a vznik pokročilé umělé inteligence nevedl ke katastrofickým následkům. Obecně půjde o následující čtyři oblasti výzkumu:

- 1) snižování existenčních rizik spojených s umělou inteligencí;
- 2) snižování rizik spojených s utrpením v důsledku umělé inteligence;
- 3) společenské sdílení přínosů umělé inteligence;
- 4) meta-výzkum v oblasti umělé inteligence.¹⁷

V kontextu právní vědy zde přitom dochází k zajímavému větvení. Prvním směrem jdou otázky týkající se regulace výrobců, vývojářů, uživatelů atd. Toto je oblast, kterou jako právníci známe velmi dobře a umíme s ní pracovat. Druhý směr je komplikovanější, neboť se dotýká regulace samotné pokročilé umělé inteligence. Zahrnuje složité a interdisciplinární otázky typu, jak přenést dodržování právních pravidel a hodnotových pozic do samotné technologie nebo jak využít poznatky obsažené v právní vědě k ovlivnění jednání technologie, která ale v ničem nebude připomínat lidskou mysl a lidskou inteligenci. V této části se věnuji právě tomuto druhému směru.

Jak se systémy vytvořené člověkem stávají stále více autonomními – tzn. schopnými jednat bez lidského vedení ve stále širším spektru situací – budeme je muset vybavit obecnými principy, které jim umožní rozhodnout, které činnosti nebo výsledky jsou pro člověka žádoucí a kterým je třeba se naopak vyhnout. Při použití úzce vymezených pokynů totiž hrozí, že umělá inteligence bude směřovat k řešením, která tyto úzce specifikované pokyny sice splní, ale zároveň budou v rozporu s širšími lidskými hodnotami a zájmy.

¹⁵ RUSSELL, S. *Human Compatible. AI and the Problem of Control*. London: Allen Lane, 2019, s. 10–15.

¹⁶ GABRIEL, I. Artificial Intelligence, Values, and Alignment. *Minds & Machines*, 2020, vol. 30, s. 411–437.

¹⁷ WINTER, Ch. (eds.). *Legal Priorities Research: A Research Agenda*. Legal Priorities Project, 2021, s. 36.

Například jsme se již setkali s tvrzeními, že doporučovací systémy na sociálních sítích mohou uživatele tlačit k extrémnějším názorovým pozicím. To sice algoritmům pomáhá naplnit jejich úzký cíl maximalizace času, který uživatel stráví na webové stránce tím, že se uživatelé stanou předvídatelnějšími, ale zároveň jde o postup, který je v rozporu s širšími lidskými preferencemi ohledně toho, co by měl algoritmus dělat.^{18 19}

Vzhledem k tomu, že AI systémy budou do budoucna stále schopnějšími, pravděpodobně v čase významně porostou náklady na takzvaná *selhání specifikace* (specification failures). Z tohoto důvodu se již v současnosti objevují výzvy, aby se pokročilé AI systémy učily lidským hodnotám nebo preferencím, a to s cílem vyřešit tzv. *AI alignment problem*.²⁰ O jaké hodnoty by se však mělo jednat? Pojem *lidské hodnoty* vyvolává široké neshody mezi lidmi ohledně jejich vlastních hodnot a preferencí. To se stává vážným problémem, pokud uvažujeme o potřebě závazné regulace, která bude nezbytná pro globální řešení negativních externalit vyplývajících z umělé inteligence – včetně existenčních rizik pro lidstvo.

Pokud má regulace vyžadovat dodržování souboru zásad nebo hodnot, musí alespoň nepřímým způsobem specifikovat, o jaké hodnoty se jedná. A má-li být takový regulační rámec přijat a prosazován na mezinárodní úrovni, potřebujeme soubor hodnot, které jsou samy o sobě široce uznávány mezinárodním společenstvím. Tedy zde existuje dvojitý požadavek na hodnoty – ty by měly být zároveň obecné a široce přijímané.

Návrh, který byl publikován v podobě článku *Negative Human Rights as a Basis for Long-term AI Safety and Regulation*,²¹ ze kterého tato část vychází, spočívá v aplikaci konceptu *negativních lidských práv* jako minimálního souboru principů, které by měly respektovat všechny AI systémy a které jsou už dnes široce přijímány a chráněny. Návrh omezuje na *negativní práva*, jelikož se nám požadavek, aby všechny systémy umělé inteligence napomáhaly plnění pozitivních lidských práv, nezdá smysluplný (přirozeným nositelem povinností je v tomto případě stát). Mezi negativními lidskými právy se zpočátku navrhuje

¹⁸ RUSSELL, S. *Human Compatible: Artificial Intelligence and the Problem of Control*. London: Allen Lane, 2019.

¹⁹ Pro další příklady viz KRAKOVNA, V. et al. Specification gaming: the flip side of AI ingenuity. In: *Google Deep Mind* [online]. 21. 4. 2020. Dostupné z: <https://deepmind.google/discover/blog/specification-gaming-the-flip-side-of-ai-ingenuity/>.

²⁰ YUDKOWSKY, E. Complex Value Systems in Friendly AI. In: SCHMIDHUBER, J.; THORISSON, K. R.; LOOKS, M. (eds.). *Artificial General Intelligence. AGI 2011. Lecture Notes in Computer Science, vol 6830*. Berlin, Heidelberg: Springer Publishing, 2011.

²¹ HORENOVSKY, J.; BAJGAR, O. Negative Human Rights as a Basis for Long-term AI Safety and Regulation. *Journal of Artificial Intelligence Research*, 2023, vol. 76, s. 1043–1075.

soustředit především na ty, na nichž již panuje elementární shoda – například právo na život nebo ochranu vlastnictví.

Abychom tomuto zdánlivě abstraktnímu konceptu dali konkrétní význam, inspirujeme se soudním systémem, který je navržen tak, aby dával autoritativní odpověď, zda je určité jednání dovoleno. Tímto způsobem jsou již dnes lidská práva chráněna, a to prostřednictvím mezinárodního i vnitrostátního práva a soudních systémů. V tomto duchu náš návrh chápe určité jednání umělé inteligence jako souladné s lidskými právy pouze tehdy, kdy by je konkrétní soudní systém posoudil jako souladné na základě úmluvy vymezující, která konkrétní práva jsou chráněna.

V praxi by se rozhodovací systém přímo uplatňoval pouze v malé množině případů – cílem je především navrhnout budoucí systémy umělé inteligence tak, aby samy předvíдалy, které chování může porušovat negativní lidská práva, a vyloučily takové strategie ještě před jejich provedením. Náš návrh také řeší související technický úkol klasifikovat jakékoli konkrétní (hypotetické) jednání jako souladné či nesouladné s negativními lidskými právy, přičemž rozhodovací systém by sloužil jako zdroj referenční pravdy. Toto je druh úkolu vhodný pro paradigma strojového učení. Náš návrh tedy řeší jak institucionální struktury, které mají dlouhodobě zajistit bezpečnost umělé inteligence, tak zarámování souvisejícího technického problému.

Náš návrh přispívá ke stavu poznání následujícími body:

- Poskytujeme nové rámování *AI alignment* problému, přičemž zdůrazňujeme potřebu oddělit pozitivní cíle a bezpečnostní omezení. V souladu s tím se zaměřujeme na to, jak by se systém neměl chovat, nikoli na to, jak by se chovat měl.
- Na základě poznatků z právní teorie navrhujeme koncept negativních lidských práv jakožto hodnotový základ bezpečnostních omezení a vysvětlujeme, proč jsou pro tento účel vhodná.
- Vysvětlujeme, jak by se dodržování negativních lidských práv mohlo využít při vytváření technických řešení pro bezpečnou umělou inteligenci, a nastiňujeme některé související výzvy jako inspiraci pro budoucí technickou práci. Rovněž uvádíme některé stávající oblasti, na nichž by taková práce mohla stavět.
- Oproti tomu, co považujeme za standardní způsob regulace – tedy sérii úzce zaměřených pravidel cílených na konkrétní problémy, které se již vyskytly – nabízí náš

návrh obecnější pokyny pro systémy umělé inteligence s větší šancí uplatnění v nepředvídaných či nepředvídatelných budoucích situacích.

- Pro konkrétnější ilustraci předchozích bodů předkládáme návrh, jak by mohl fungovat mezinárodní regulační systém založený na našem návrhu.

Upozorňujeme také na následující tři omezení:

- Poskytujeme pouze částečný příspěvek k řešení problému specifikace při návrhu bezpečné umělé inteligence, který bude třeba doplnit vhodnými pozitivními cíli (ve směru *AI alignmentu*), jakož i dalšími bezpečnostními opatřeními, včetně právních.
- Neposkytujeme konkrétní technická řešení, ale spíše rámec, který vymezuje dodržování negativních lidských práv jako úlohu strojového učení, která je přístupná budoucí technické práci.
- V některých částech předpokládáme pokročilé schopnosti budoucích AI systémů, zejména pokud jde o předvídaní dopadů jejich činnosti. Zatímco schopnosti AI systémů se rychle rozvíjejí, jejich konkrétní podoba a míra zůstávají do určité míry předmětem spekulací, což znamená, že některé naše argumenty zůstávají odpovídajícím způsobem nejisté.

Text je koncipován následovně: Po přehledu některých existujících studií nejprve stručně vysvětlujeme, na jaké systémy umělé inteligence se zaměřujeme. Poté nastiňujeme naše pojetí lidských práv pro tento účel a paralelu s jejich využitím v právu. Následně pokračujeme technickými úvahami o bezpečnosti umělé inteligence založené na negativních lidských právech. Na závěr popisujeme návrh, jak by bylo možné tuto bezpečnost začlenit do systému mezinárodních organizací.

AI a hrozba nových forem vlády

V poslední části se na pokročilou AI zaměřuji v souvislosti s tím, jak lidstvo spravuje věci veřejné, což skýtá další potenciální výzvy týkající se budoucnosti. Ty přitom máme možnost začít ovlivňovat již v současnosti. Umělá inteligence se totiž již dnes využívá v mnoha systémech používaných výkonnou mocí, kde významně zlepšuje jejich fungování. Tyto systémy pomáhají v mnoha světových armádách, vesmírných programech, ve veřejném zdravotnictví. Je důležité zdůraznit, že ve většině případů se jedná o využití systémů, které zlepšují efektivitu a účinnost veřejného sektoru, a můj argument by neměl být chápán jako kritika těchto systémů (za předpokladu, že splňují základní bezpečnostní principy). Přesto však jde o trend, který v dlouhodobém horizontu vyvolává důležité a znepokojivé otázky. Zde

bychom přitom měli rozlišit dvě možné formy vlády, které z toho vyplývají. Jednou je **AI totalitarismus**, druhou **AI technokracie**.

KAPITOLA 5: Sociální síť, svoboda projevu a informační veřejný prostor

Britský právní teoretik Raz v roce 1991 napsal, že svoboda projevu je klíčová především pro lidi, kteří mluví a projevují názory veřejně. Ostatním, kteří se veřejně neprojevují, se nemusí jevit natolik důležitá, proto mohou považovat za podstatnější jiná práva, která se jich naopak bezprostředně dotýkají.²² V jeho úvaze hrálo roli, že pouze málo lidí mělo v minulosti přístup do veřejného prostoru. Sociální síť a internet tuto realitu v posledních letech zcela zásadně změnily, neboť zpretrhaly rozdíl mezi veřejným a soukromým projevem. Každý projev se tak potenciálně stal projevem veřejným a svoboda projevu se začala dotýkat mnohem většího množství lidí. Internet a sociální síť tudíž představují důležitý socio-technologický milník, na který je třeba reagovat. A o to se tato kapitola snaží.

Nejdůležitější závěry této kapitoly: Jakákoli budoucí regulace by měla především zohledňovat zachování svobody projevu v co nejšířší možné podobě a přistoupit k omezování pouze tehdy, je-li to absolutně nezbytné. Regulačním cílem ve vztahu k sociálním sítím by přitom mělo být zajištění stavu, aby svoboda projevu byla chráněna i na nich. Vhodným prostředkem je propojení regulace sociálních sítí se současným, ať již americkým či evropským, lidskoprávním standardem, který představuje existující souhrn pravidel určující, co se ve veřejném prostoru smí a nesmí říkat. Sociální síť by naopak neměly být činěny odpovědnými za obsah, který se na nich nachází. Je také třeba se, pokud možno, vyvarovat restriktivního a kriminalizujícího přístupu v boji s dezinformacemi. Místo toho je vhodné zvolit alternativní prostředky čelení těmto negativním společenským jevům.

KAPITOLA 6: Technologie, právo a covid-19

Tato poslední kapitola se věnuje užití technologií při boji s pandemií onemocnění Covid-19 a jejich právním implikacím, přičemž je ve svém základu rozdělena na tři části. První část se věnuje diskusi ohledně elektronického sledování osob, které vyvolává otázky především ohledně ochrany soukromí a osobních údajů. Tématem druhé části je tzv. covid pas a důsledky z něj vyplývající pro otázku diskriminace. Třetí část se věnuje možnostem užití umělé inteligence v boji s epidemií, kterou český stát v podstatě nevyužil, i když by to v mnoha situacích velmi pomohlo.

²² RAZ, J. Free Expression and Personal Identification. *Oxford Journal of Legal Studies*, 1991, vol. 11, no. 3, s. 303.

Obecně jde v této kapitole o otázku, kde jsou (*de lege lata*) a kde mají být (*de lege ferenda*) limity veřejné moci při užití technologií v boji s epidemií a jakou roli zde sehrává problém diskriminace a požadavek na zachování lidských práv při mimořádných situacích. Současně je nezbytné upozornit na míru emotivnosti, kterou některé tyto otázky, spolu například s návrhem povinného očkování, po celou dobu pandemie vyvolávaly. Tato kapitola se pokouší od těchto emotivních přístupů oprostít, neboť jejím účelem není ovlivňovat přístup k nedávné epidemii, nýbrž především zprostředkovat zkušenost užitečnou pro zvládnutí epidemií budoucích. V tomto kontextu se proto kapitola pokouší podat spíše obecný přehled, v jehož rámci zdůrazňuje nejrelevantnější podněty.

Nejdůležitější závěry této kapitoly: V boji s epidemií pomocí technologických nástrojů je klíčovým prvkem v demokratické společnosti důvěra občanů ve stát, který je transparentní a který je schopen zajistit efektivní ochranu soukromých údajů. Pokud jsou tyto podmínky naplněny, jsou lidé více ochotní užívat dobrovolné trasovací aplikace a být více aktivní v zabraňování šíření viru ze své vlastní iniciativy. Pokud tyto podmínky naplněny nejsou, lidé odmítají dobrovolné trasovací aplikace využívat i pokud jsou fakticky bezpečné. Používá-li navíc stát oba druhy trasování (dobrovolné i nedobrovolné), podkopává tím ochotu lidí užívat dobrovolné aplikace.

1.3 Použité zdroje

V práci je především čerpáno ze zahraniční a domácí odborné literatury (252 záznamů). Přestože většina zdrojů je anglických, kde to šlo, pokoušel jsem se využívat i české prameny. Užití judikatury je spíše okrajové (20 záznamů). Stejně platí o použitých právních předpisech (20 záznamů). Poměrně extenzivně využívám tzv. ostatní zdroje, kam spadají různé odborné blogové příspěvky, akademické přednášky nebo záznamy konferencí (56 záznamů).

2 Stručné shrnutí disertační práce v anglickém jazyce

This work focuses on the relationship between legal systems and society's rapidly evolving socio-technological environment. In the first part, it explores the question of the evolution/adaptability of legal systems in response to dynamic socio-technological developments, using the theoretical framework of systems theory. The first part also analyses different models for regulating new technologies, placing them within the logic of systems theory and illustrating them with concrete examples such as Uber and Airbnb and the challenges associated with artificial intelligence and big data.

The second chapter explores the legal and cultural aspects of the technological stagnation thesis, the increasing complexity of legal systems and their socio-technological environments, and questions about how these systems respond to social resistance to new technologies and risk aversion. The trade-off between innovation and security is presented. The third chapter develops the theme of legal system complexity and its relationship to entropy and modern technology.

In the second part, the thesis addresses specific issues, such as the position of AI in law and the possibility of AI becoming a legal entity, including ethical and legal implications. It also discusses the integration of human rights into the development of AI and presents a proposal for an international regulatory system. Finally, it discusses the potential threats of new forms of AI-mediated government, such as AI totalitarianism and AI technocracy.

The fifth chapter focuses on social networks and freedom of expression, reviews digital platforms' role in public discourse and examines different approaches of regulation.

The final chapter focuses on the impact of technology and law in the context of the COVID-19 pandemic, exploring issues of electronic surveillance and COVID-19 passports, and how big data and AI can contribute to the fight against the pandemic.