

míst, která by ženám umožnila plynulý návrat do pracovního života a zbavila je tak i tlaku ze strany zaměstnavatelů, který je na ně v současnosti vyvíjen.

Nejde si však nepovšimnout, že obě dvě zmíněné problematiky jsou nerozlučně propojené. Aby se matky vracely do práce, musí jim být poskytnuta jistota, že je zde zázemí, které jim zajistí dostatečnou péči o jejich děti a vice versa. Toto zázemí by pak mělo vycházet jak ze státní podpory (jesle, dětské skupiny atd.), tak z rodiny – tedy například od otců formou otcovské dovolené.

Dle mého názoru jde tak v konečném důsledku o nacházení rovnováhy mezi životem profesním a životem matky/rodiče obecně. To se myslím Francouzům daří více než nám, a proto je naším úkolem se z těchto flexibilnějších modelů péče o dítě učít a převzít či přizpůsobit danou úpravu našim podmínkám.

### 13.

## UMĚLÁ INTELIGENCE JAKO PŘEDMĚT PRÁVNÍ REGULACE: BENEFITY A RIZIKA

Vratislav Slavík

Na obecné úrovni je tématem této kapitoly vztah práva a jedné z technologií, která se vedle jiných technologií (např. nanotechnologie, biotechnologie, informační technologie) zcela zásadně podílí na hluboké přeměně nastavení, fungování a podoby stávajícího světa, a to umělé inteligence (dále jen „UI“). Podle některých autorů<sup>659</sup> byl však systém práva po určitou dobu k dopadům, stejně jako k rizikům přinášěných integrací UI do společností, zcela netečný.

Diskuse o tom, nakolik je právo jako systém schopné udržovat tempo s technologickým vývojem, se odehrává především v rámci etablovaného akademického pole, které se označuje jako právo a technologie. V této diskusi je vztah práva a současných technologií (včetně UI) často pojímán jako závod. Právo je v něm vykreslováno jako závodník, který je vždy pomalejší a který zaostává za technologiemi. V této roli se právo ocitá zejména proto, že ve vztahu k technologiím selhává v naplňování svého primárního účelu. Ten je spatřován v poskytování včasných a dostatečně efektivních opatření pro usměrňování vývoje technologií takovým směrem, který nebude představovat riziko a hrozbu pro stabilitu a existenci společnosti.<sup>660</sup>

Ve spojitosti s touto diskusí se pak debatuje o načasování regulačních zásahů a hledá se nejvhodnější moment pro zavedení nových zákonů, které zakazují, usměrňují nebo podporují nově se objevující

659 BLACK, J. / MURRAY, A. Regulating AI and machine learning: setting the regulatory agenda. *European Journal of Law and Technology*, 2019, sv. 10, č. 3, s. 9; CHESTERMAN, S. *We, the Robots?: Regulating Artificial Intelligence and the Limits of the Law*. Cambridge: Cambridge University Press, 2021, s. 11.

660 ALLENBY, B. R. Governance and Technology Systems: The Challenge of Emerging Technologies. In: MARCHANT, G. E. / ALLENBY, B. R. / HERKERT J. R. (eds.) *The Growing Gap Between Emerging Technologies and Legal-Ethical Oversight: The Pacing Problem*. Springer Dordrecht, 2011, s. 10.

technologie. V tomto kontextu se odkazuje ke konceptu Collingridgova dilematu, v němž je problém kontroly nad technologiemi a regulace technologií zachycen.<sup>661</sup>

Z celkového pohledu na problematiku regulace nových technologií, a to včetně regulace UI, lze říci, že představuje komplexní téma, které vyžaduje zvážení potenciálních rizik a přínosu konkrétní technologie, jakož i různorodých perspektiv a zájmů, zastávaných zúčastněnými stranami v procesu vyjednávání o způsobech a podobách regulace. Označení zúčastněné strany zahrnuje především tři hlavní skupiny: (1) regulátory (vláda, zákonodárci), (2) soukromé korporace a (3) jednotlivci.<sup>662</sup> V rámci diskuse o možných podobách regulace UI se stal dominantním názor, který prosazovaly a systematicky šířily technologické nadnárodní korporace stojící za rozvojem UI, tedy že regulace je vždy negativní silou, která brání inovacím a pokroku, a území, které nespouští regulace, je pro technologické inovace určitým způsobem nezbytnou podmínkou.

S obdobně konstruovanými názory na regulaci technologií a zásahy státu do procesu technologické změny se setkáme v období, které bylo podobně zlomové jako naše současnost, a to v počátcích modernity a průmyslové revoluce, která se z pohledu technologií točila v první fázi okolo parního stroje a železnice, následně spalovacího motoru a využívání ropy.<sup>663</sup>

Hlavní postavy průmyslu (*robber barons*) této doby, vybaveny myšlenkami Herberta Spencera o funkci státu (kterou měla být převážně ochrana práv jednotlivců a zajištění kolektivní obrany), doktrínou o přežití nejschopnějších, která měla být jedinou regulací jejich obchodních aktivit (tzn. samoregulace), a představou, že země je jeden velký a nekonečný zdroj surovin, které je potřeba vytěžit, jsou jedním z kořenů pro pochopení současnosti. S problémy, které sebou přinesly technologie spojené s počátkem modernity, se společnost vyrovnává dodnes. Tyto problémy ohrožují samotnou možnost existence lidstva na této planetě a zároveň zřetelně usvědčují systémy tuto katastrofu způsobující, tedy především logiku systému kapitalismu založeného na nekonečném růstu, fungování soukromých korporací spočívajícím

661 MOSES, L. B. / ZALNIERIUTE, M. Law and Technology in the Dimension of Time. In: RANCHORDAS, S. / ROZNAI, Y. (eds.) *Time, Law, and Change: An Interdisciplinary Study*. Oxford: Hart Publishing, 2020, s. 309.

662 BLACK / MURRAY, *Regulating AI and machine learning*, s. 4.

663 ALLENBY, *Governance and Technology Systems*, s. 7.

v neustálém zvyšování zisku, a na individuální úrovni pak prostou touhu po moci a obranu práv kapitálu.<sup>664</sup>

Druhým bodem v rámci historie technologií, s nimiž nás názory na jejich regulaci uvedené výše propojují, je období poloviny 90. let dvacátého století, konkrétně pak diskuse o regulaci internetu, která v tomto období probíhala. Toto téma bylo rámováno obecnou debatou o regulaci odehrávající se uvnitř akademického právního pole. Ta byla primárně zaměřena na kritiku podoby regulace a její reformu. Hlavní proud akademického právního myšlení té doby asimiloval ekonomické myšlenky Friedricha Hayeka a Milтона Friedmana, a neoliberalní ideologie se tak stala optikou, kterou bylo nahlíženo na regulaci a regulační úlohu státu. Jak ukazuje Short<sup>665</sup> ve studii analyzující 1 400 článků z právních časopisů na téma regulace, které vyšly mezi lety 1980 do 2005, byl diskurz vystavěn na hlubokých obavách z donucovací moci státu a na spojení tématu regulace s pojmy, jako je tyranie a autoritářství. Současně s tím převládl názor, že neúčinnějším nástrojem proti centralizaci moci a nátlakovému charakteru státní regulace je samoregulace. Tato myšlenka je postavena na přístupu zdola nahoru a současně absenci donucení. Na rozdíl od tradičního centralizovaného přístupu k nastavování pravidel a jejich vynucování se tato myšlenka stala na dlouhou dobu určující v debatě o regulaci internetu, a setkáváme se s ní do až současnosti, kdy plynule přešla do debaty o regulaci UI a stojí za skokovým nárůstem produkce *soft law* dokumentů obsahujících etické pokyny pro její vývoj.

Tyto vlivy se otiskly do veřejného mínění a také do nastavování vládních politik vztahujících se k regulaci internetu a přijetí postoje k tomuto tématu ze strany států, který je v literatuře označován jako digitální liberalismus.<sup>666</sup> Způsob, jakým státy přistoupily k řešení problému regulace internetu, spočíval v tom, že se zaměřily na řešení dílčích problémů, například porušování autorských práv či zneužívání osobních údajů, avšak koherentní strategie regulace tohoto prostředí a rizik s ním spojených se nepodařilo dosáhnout. Takto nastavený

664 ZUBOFF, S. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. Profile Books, 2019, s. 112; CRAWFORD, K. *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. Yale University Press, 2021, s. 225.

665 SHORT, J. L. The paranoid style in regulatory reform. *Hastings Law Journal*, 2012, sv. 63, č. 3, s. 654.

666 DE GREGORIO, G. The rise of digital constitutionalism in the European Union. *I•CON*, 2021, sv. 19, č. 1, s. 44.

system regulace a kontroly digitálního prostředí vyprodukoval mimo jiné pět hlavních globálních korporací (Google, Facebook, Amazon, Microsoft a Apple), které ovládají to, jak v současnosti vypadá podoba našeho života v tomto prostředí a jak je toto prostředí zažíváno. Dále se stal jedním z klíčových faktorů, který ovlivnil rychlost rozvoje strojového učení a UI.<sup>667</sup>

V roce 2016 proběhly tři události, které měly zásadní podíl na tom, že UI vstoupila do širšího povědomí, a s ní spojené otázky se následně staly tématem veřejné debaty (v době psaní této kapitoly by bylo možné přidat ještě další, a to ChatGPT). První z nich je uvedení chatbota s názvem Tay prostřednictvím Twitteru společností Microsoft, který byl po 16 hodinách působení na této sociální síti odstaven, a to z toho důvodu, že v interakci s ostatními uživateli začal zveřejňovat zprávy (tweets) které byly rasistické, sexuálně urážlivé a popíraly holocaust.

Druhou událostí bylo vítězství AlphaGo, UI vyvinuté společností Google DeepMind ve hře Go, která byla svojí složitostí považována za hru, kterou mohou hrát pouze lidé, protože vyžaduje využití intuice a strategického myšlení.

Třetí a poslední událostí z roku 2016 spojovanou s UI, která je v literatuře zmiňována, je uvedení gynoida Sophie na veřejnost zakladatelem společnosti Hanson Robotics David Hansonem, které bylo počátkem globálního spektaklu odehrávajícího se po tři roky, a který vyvrcholil tím, že Sophii bylo uděleno občanství Saudské Arábie.

Tyto tři uvedené události směřují k základnímu okruhu problémů a otázek, které se staly stálou součástí uvažování o technologiích jako je UI a robotika, a to problému kontroly a obavy z její ztráty, a problému konkurence mezi stroji a lidstvem, která povede k ovládnutí lidstva stroji. Takto vytyčený základní okruh problémů nachází svůj zdroj ve dvou klasických literárních narativích zachycujících představy o vztahu mezi člověkem a jeho výtvořem, vykazujícím známky vlastního myšlení. Tyto motivy nacházíme například v románu *Frankenstein* Marry Shelley, v legendě o Golemovi, v povídkách Issaca Asimova<sup>668</sup> a samozřejmě v dramatu *R.U.R.* Karla Čapka.

V současnosti posunuly myšlenku o tom, že UI bude posledním výnosem lidstva, který povede k jeho konci, do popředí veřejné debaty o umělé inteligenci překvapivě samy vlivné osoby stojící za vývojem této technologie jako jsou Elon Musk, Ray Kurzweil, Stuart Russell

či Nick Bostrom. Do debaty navíc vnesly nové pojmy jako superinteligence, technologická singularita či transhumanismus. Koncept superinteligence předpokládá, že se vědcům podaří vyvinout tzv. obecnou UI, neboli inteligenci, která se vyrovná inteligenci člověka, a na tomto základě se vytvoří její další úroveň, která už lidskou převyší. Tato UI přinese dramatickou změnu, kdy už sami nedokážeme pochopit, co se vlastně děje, a život tak jak ho známe, skončí.

### 13.1 Vývoj oboru UI a témata právního výzkumu z oblasti UI

V rámci teoretické právní literatury je téma UI přítomné od doby konsolidace UI jako samostatné oblasti výzkumu a kopíruje vývoj této oblasti. Událostí, která je obecně považována za zrod vědeckého pole se zaměřením na výzkum a vývoj UI, je konference v Dartmouthu, která proběhla v roce 1956. V následujících letech, a to až do roku 1974, je výzkum UI spojen s velkým optimismem, ustaluje se koncept UI a dochází k průlomovým objevům v oblasti zpracování přirozeného jazyka a strojového vidění. Problémy, jako byl nedostatek dat a výpočetního výkonu, však začaly brzdit počáteční nadšené snahy o vytvoření UI podobné té lidské. K těmto problémům se připojila i kritika a obavy ohledně cílů projektu vývoje UI, což vedlo k šest let trvajícímu období nazývanému první zimou UI, které se projevilo omezením finanční podpory výzkumných projektů umělé inteligence a celkovou ztrátou zájmu o toto téma.

Z pohledu právního výzkumu v oblasti UI v tomto období (1956–1974) byla hlavním tématem charakteristika procesu řešení právního problému a právní argumentace, a možné způsoby zapojení UI v těchto procesech (jinými slovy možnost nahrazování člověka v právnických profesích UI). To, co však stojí mimo zájem právní vědy v tomto období, je snaha o konstruování právních pojmů a jejich obsahu.

Na počátku 80. let dochází po kritice z let předešlých k obnově financování výzkumu UI. Vědci se odklonili od původních cílů a velkých abstraktních konceptů a nově se výzkumy zaměřili na vývoj znalostních (expertních) systémů. Na konci 80. let však zájem investorů o oblast UI znovu ochabl, protože se zcela nevydařil přechod od vývoje technologie k širokému přijetí veřejností a naplňování ekonomických cílů. V tomto období tak nastává druhá zima UI.<sup>669</sup>

667 ZUBOFF, *The Age of Surveillance Capitalism*, s. III.

668 Srov. například *Konflikt nikoli nevyhnutelný*.

669 BECCERA, S. D. *The Rise of Artificial Intelligence in the Legal Field: Where*

V hranicích „právního pole“ (*juridical field*),<sup>670</sup> zejména jeho akademické části, patří mezi témata diskutovaná autory, kteří se věnují UI v těchto letech (1980–1987), na jedné straně znovu úvahy o možnostech zapojení UI v právních profesích, možnosti zpracování informací a vyhledávání textů, na druhé straně i otázky nové, týkající se samotného využívání UI a případné odezvy ze strany práva.

Třetí vlna vzestupu výzkumu UI začíná počátkem 90. let dvacátého století a trvá do současnosti. Toto období je charakterizováno výzkumem zaměřeným na strojové učení a odklonem od znalostních systémů. Nejrozšířenějšími metodami strojového učení se stávají učení s učitelem, bez učitele a posilované učení. Po roce 2010 se do popředí výzkumu a aplikace UI dostávají zejména neuronové sítě a hluboké učení. Z pohledu právního pole a jeho propojení s tématem UI dochází v této poslední vlně, zvláště pak od roku 2016, k významnému nárůstu literatury věnované tomuto tématu. Výzkum, který byl na počátku úzce zaměřený na popsání a pochopení mentálních procesů používaných při výkonu právních profesí, jejich formalizaci a aplikaci UI při řešení praktických právních problémů, se mění na méně specializovaný a pokrývá více obecných témat.<sup>671</sup>

V rámci témat právního výzkumu lze identifikovat tři základní okruhy, kterým je věnována pozornost, a to: (1) dopady UI na základní koncepty, na nichž je právo postaveno, jako jsou například právní subjektivita, právní odpovědnost, princip nediskriminace, základní lidská práva; (2) okruh věnující se oblastem využití UI a identifikaci potenciálních rizik a příležitostí, a (3) třetí okruh, který se soustředí okolo témat navrhování modelů regulace a správy UI a návrhů možných podob právní regulace.<sup>672</sup>

### 13.2 Možné varianty definice umělé inteligence pro účely regulace

Všichni odborníci (vědečtí, političtí), jejichž předmětem zájmu je UI, se shodnou na skutečnosti, že neexistuje jednoznačná a široce přijímaná definice UI – ať už na poli výzkumu a vývoje UI, nebo taková, která je obecně použitelná pro účely regulace.<sup>673</sup>

K UI lze přistoupit jako k sociálnímu konstrukt, tedy tak, že UI je to, na čem se určitá skupina shodne, a to v určitém kontextu a pro určitý účel, což současně znamená, že pojem UI je používán odlišnými způsoby různými skupinami, a může označovat mnoho různých věcí. Při bližším pohledu je však možné zjistit, že to, co tento pojem zachycuje, závisí především na tom, zda se hovoří o UI jako o vědním oboru, technice/metodě výstavby systému UI, nebo konkrétní aplikaci systémů UI. Dvěma hlavními aktéry z pohledu podílu na utváření obsahu pojmu UI jsou skupina vědeckých pracovníků z oboru UI, a skupina politických aktérů vytvářejících podobu právní regulace, přičemž nesoulad mezi těmito skupinami v pojmání UI může bránit adekvátní odezvě ze strany práva. Z hlediska regulace a obsahu politiky je proto důležité zahrnout do pojmání UI jak ty aspekty technologie, které vyžadují právní zásahy, tak i současné a budoucí technické možnosti UI.<sup>674</sup>

### 13.3 Umělá inteligence jako vědní obor

Jako označení pro vědní obor se UI pojí s letní školou v Dartmouth organizovanou Johnem McCarthym v roce 1956. V souvislosti s touto událostí je zveřejněna i jedna z prvních definic UI, v současnosti označovaná za klasickou, která je zkonstruována na základě uceleného popisu této technologie.<sup>675</sup> Výzkumný projekt letní školy vycházel pře-

We Are and Where We Are Going. *The Journal of Business, Entrepreneurship & the Law*, 2018, sv. 11, č. 1, s. 35.

670 Ke konceptu pole srov. například text Pierre Bourdieu'a *The Force of Law: Toward a Sociology of the Juridical Field* z roku 1987.

671 BECCERA, The Rise of Artificial Intelligence in the Legal Field, s. 42.

672 GOANTA, C. / VAN DIJCK, G. / SPANAKIS, G. Back to the Future: Waves of Legal Scholarship on Artificial Intelligence. In: RANCHORDAS, S. / ROZNAI, Y. (eds.) *Time, Law, and Change: An Interdisciplinary Study*. Oxford: Hart Publishing, 2020, s. 341.

673 MARTINEZ, R. Artificial Intelligence: Distinguishing Between Types & Definitions. *Nevada Law Journal*, 2019, sv. 19, č. 3, s. 1016; SCHERER, M. U. Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies. *Harvard Journal of Law & Technology*, 2016, sv. 29, č. 2, s. 359.

674 KRAFFT, P. M. / YOUNG, M. / KATELL, M. / HUANG K. / BUGINGO, G. Defining AI in Policy versus Practice. In: *Proceedings of the 2020 AAAI/ACM Conference on AI, Ethics, and Society (AIES '20)*. New York: Association for Computing Machinery, New York, 2020, s. 6.

675 COECKELBERGH, M. *AI ethics*. Cambridge, MA: The MIT Press, 2020, s. 66.

devším z předpokladu, že každý aspekt učení nebo jakýkoliv jiný rys lidské inteligence lze v zásadě popsat tak přesně, že je možné vytvořit stroj, který jej bude schopen simulovat. Úsilí mělo být směřováno k řešení problému UI, za který se považuje přímět stroj, aby se choval způsobem, který by byl označen za inteligentní, pokud by se tak choval člověk. Taková definice je však problematická z mnoha důvodů, například není zodpovězena otázka, zda je opravdu možné redukovat inteligenci a mysl člověka na pouhý sled výpočetních procesů.

V následujících letech dochází k mnoha dalším pokusům o definování pojmu UI. Najít jednoznačně přijímanou definici se nepodařilo, což však podle některých autorů paradoxně pomohlo oboru růst a široce se rozvinout. Jak popisují autoři Russell a Norvig<sup>676</sup> ve své klasické učebnici *Artificial Intelligence: A Modern Approach*, která je považována za základní a vstupní text do studia oboru UI, lze i přesto definice rozčlenit na základě čtyř obecných přístupů, a to podle toho, jakou charakteristiku používají při hodnocení systému jako inteligentního.

Definice jsou rozděleny podle dvou dimenzí, a to lidské proti racionálnímu myšlení, a myšlení proti jednání. Výsledné kombinace definic jsou: (a) definice zaměřené na lidské myšlení, podle nichž je UI systém, který myslí jako člověk; důraz je předně kladen na zpracování informací způsoby inspirovanými lidským poznáváním a imitací těchto způsobů, (b) definice zdůrazňující lidské chování popisují UI jako systém, který napodobuje způsob chování člověka při plnění konkrétního úkolu nebo v konkrétní situaci, (c) třetí možnou podobou definic UI jsou definice postavené na znaku racionálního myšlení; vidí UI jako systém zpracovávající informace a uvažující na základě matematických pravidel (jako například deduktivní logiky a pravidel odvozování), (d) poslední forma definic UI se opírá o charakteristiku, kterou je racionální chování. Z tohoto pohledu je UI takový systém, který optimálním způsobem řeší specifikované úkoly nebo dosahuje naprogramovaných cílů. Samotný způsob chování systému může být neobvyklý nebo případně člověku neznámý, avšak efektivní z pohledu plnění určitého úkolu.

<sup>676</sup> RUSSELL, S. / NORVIG, P. *Artificial Intelligence: A Modern Approach*. 3rd edition. PEARSON Education Limited, 2016, s. 3.

### 13.4 Definice použitelná v rámci právní regulace, charakteristiky a rizika

Obor UI již od počátku svého vzniku usiloval o vývoj a naplnění formy obecné (silné) umělé inteligence. Tato forma staví na tom, že je srovnatelná s lidskou inteligencí, což znamená, že má schopnost řešit širokou škálu úkolů a přizpůsobovat se novým problémům prostřednictvím osvojování si nových postupů. Takto nastavený cíl však zůstává jen malou částí výzkumu a aplikací UI. Většina úsilí je spíše zaměřena na vývoj systémů spadajících pod úzkou (slabou) UI, která slouží k řešení úzce vymezených úkolů prostřednictvím dosažení určitého cíle nebo souboru cílů.<sup>677</sup>

Ačkoli je vývoj úzké UI velice vzdálený od složitého a komplexního úkolu vytvoření obecné UI, zaměření se na její vývoj vedlo k významným pokrokům v rámci celého vědního oboru UI. Právě použití této formy má zásadní dopad na každodenní svět a je tak i hlavní současnou výzvou pro právo a regulaci. Některé z těchto systémů UI dosahují cíle způsobem, který je silně ovlivňován lidskými vstupy, zatímco jiné vyžadují minimální vstupy člověka, protože se učí a aktualizují svůj rozhodovací model. Co přesně se učí a jakého cíle dosahují, závisí předně na konkrétním prostředí.<sup>678</sup>

Různé definice UI používané v technické literatuře, která je spojena s výzkumem UI, mohou být užitečné pro pochopení jejího fungování, ale jak zmiňuje Schuett,<sup>679</sup> nejsou vhodné jako základ pro její regulaci. A i když určitá nejistota může být novým technologiím vlastní, je problematické soustředit nové zákony a regulaci obecně okolo nejasných konceptů bez pevně vymezených hranic, jako je UI.

Z pohledu práva a regulace souvisí téma definice a vymezení UI s úsilím, které se soustředí na hledání základních charakteristik systémů UI a na problém identifikace hlavních zdrojů rizik, které jsou s těmito systémy spojeny. Tyto zdroje rizik lze rozčlenit do tří základních kategorií, přičemž názvy těchto kategorií se pojí s konkrétní od-

<sup>677</sup> KÖNIG, P. D. / KRAFFT, T. D. / SCHULZ, W. / ZWEIG, K. A. Essence of AI: What Is AI? In: DIMATTEO, L. A./ PONCIBO, C./ CANNARSA, M. (eds.) *The Cambridge Handbook of Artificial Intelligence: Global Perspectives on Law and Ethics*. Cambridge University Press, 2022, s. 19; SCHERER, Regulating Artificial Intelligence Systems, s. 361.

<sup>678</sup> KÖNIG / KRAFFT / SCHULZ / ZWEIG, Essence of AI: What Is AI?, s. 24.

<sup>679</sup> SCHUETT, J. Defining the scope of AI regulations. *Law, Innovation and Technology*, 2023, sv. 15, č.1, s. 62.

povědí na obecnou otázku „co je zdrojem rizika systému založených na UI“. Rizika lze rozčlenit na rizika plynoucí z (1) technického přístupu k tvorbě UI („jak je systém UI navržen“), (2) aplikace („k čemu se technologie používá“), což značí rizika plynoucí z použití, účelu a kontextu, a (3) schopnosti UI („co technologie dokáže“). Tyto tři kategorie následně tvoří základní prvky pro konstrukci právních definic a představují možný klíč k regulaci UI.<sup>680</sup>

### 13.5 Strojové učení jako potenciální zdroj rizika

Kategorie označená jako technický přístup spočívá v tom, že určité techniky a metody používané k výstavbě systémů UI považuje ze své podstaty rizikové. Touto rizikovou technikou je myšleno především strojové učení. Z pohledu současných systémů UI představuje jejich klíčovou složku algoritmus strojového učení a od něj odvozená schopnost automatického učení se na základě dat. Základním úkolem, pro který je strojové učení využíváno, je rozpoznávání vzorců a pravidel v dostupném souboru dat, na jejichž základě jsou předpovídána data budoucí. Schopnost učení se tak umožňuje systémům upravovat a přizpůsobovat pravidla pro generování výstupů (pro jejich chování, případně rozhodování).

Hlavní problém, spojený s používáním této techniky (např. hloubkové učení využívající neuronové sítě), spočívá v tom, že při jejím zapojení se stává charakteristickou vlastností rozhodovacího procesu neprůhlednost a nesrozumitelnost rozhodnutí, a to v tom smyslu, že jednotlivce, kterého se rozhodnutí týká, může často jen velmi stěží pochopit, jak nebo proč byl určitý vstupní údaj v dané situaci kategorizován, a v jakém rozsahu měl vliv na vytvoření určitého výstupu.

Důvodem je velmi zjednodušeně řečeno to, že rozhodnutí se neopírají o úplný řetězec kauzálních vztahů, ale pouze o korelace mezi proměnnými. Pokud je proces takto nastaven, je dotčeným jednotlivcům znemožněno smysluplně reagovat na tato rozhodnutí, a to například v podobě změny svého chování, zahájením debaty o tom, zda kritéria, která systém využívá k třídění a kterým byl jednotlivce podroben, jsou přijatelná v demokratické společnosti, či posouzením, zda naplňuje všechny zákonné podmínky pro jeho učinění.<sup>681</sup>

680 Tamtéž, s. 77.

681 BUITEN, M. C. Towards Intelligent Regulation of Artificial Intelligence. *Euro-pean Journal of Risk Regulation*, 2019, sv. 10, č. 1, s. 49.

V literatuře<sup>682</sup> je pak možné rozlišit dva postoje k řešení problému neprůhlednosti systémů UI. Pro první přístup je charakteristické, že cestu spatřuje v pochopení fungování vnitřních mechanismů systému, a to prostřednictvím zveřejnění zdrojového kódu a současně vytvořením obecného modelu, který se přibližuje tomu, jak se systém UI chová. Tento přístup sice může sloužit k nalezení pravidelností v chování systému UI, ale neznamená skutečné vysvětlení toho, jak dospívá ke svým výstupům. Informace, kterou tento přístup poskytuje, je informace o tom, že určitý výstup byl vytvořen na základě ovlivnění konkrétním neuronem v rámci neuronové sítě. Tato informace však neodpovídá požadavkům práva na transparentní rozhodovací proces, ani neposkytuje z hlediska práva vhodný základ pro účely vysvětlení rozhodnutí.

Na rozdíl od prvního přístupu, který klade důraz na technickou stránku problému a popis vnitřního fungování systému za účelem vytvoření obecného rozhodovacího modelu, se druhý přístup obrací k pochopení faktorů, které ovlivňují konkrétní rozhodnutí, a k vysvětlení toho, proč bylo dané rozhodnutí učiněno tak, jak bylo, a zda by změna jedné ze vstupních proměnných vedla k jinému výsledku.<sup>683</sup>

Takto pojatá forma transparentnosti a vysvětlitelnosti se tak zaměřuje pouze na ty skutečnosti, které měly pro konkrétní rozhodnutí význam, a jde jí o zdůvodnění konkrétního výsledku. Cílem této formy transparentnosti a vysvětlitelnosti je vybavit osobu, které se rozhodnutí založené na systému UI týká, především takovou informací, na jejímž základě se bude schopna právně relevantním způsobem bránit.<sup>684</sup>

Přítomnost součástí, které umožňují systémům UI učit se, a které mohou měnit chování systému na základě zpracovávaných vstupních dat, může těmto systémům propůjčovat určitou míru nezávislosti a také nepředvídatelnosti. Systém založený na UI tak může dospět k chování, které bylo od počátku nezamýšlené a zcela nepředvídatelné tím, kdo systém vyvinul. Jednou z charakteristických vlastností současných systémů UI, která je v literatuře popisována,<sup>685</sup> je schopnost fungovat bez lidského zásahu. Fungování těchto systémů je tak běžně

682 CHESTERMAN, *We, the Robots?*, s. 63; KÖNIG / KRAFFT / SCHULZ / ZWEIG, *Essence of AI: What Is AI?*, s. 30.

683 SURDEN, H. Artificial Intelligence and Law: An Overview. *Georgia State University Law Review*, 2019, sv. 35, č. 4, s. 1317.

684 BUITEN, *Towards Intelligent Regulation of Artificial Intelligence*, s. 54.

685 CHESTERMAN, *We, the Robots?*, s. 31.

označované za autonomní. To však neznamená, že by tyto systémy měly svobodnou vůli či rozum a byly schopny sebeuvědomění. Je to právě tato vlastnost, která činí tuto technologii výjimečnou ve srovnání s předešlými, a která s sebou přináší řadu výhod. Především v podobě toho, že stroje založené na UI jsou na sebe schopny brát stále větší množství úkolů. Vzhledem k tomu, že UI nemusí být vázána určitými způsoby myšlení, které jsou lidem vlastní, stejně jako není omežována schopnostmi mozku, přichází často s neobvyklými způsoby řešení problémů. Atraktivita těchto systémů spočívá nejen v tom, že jsou v současnosti schopny nacházet řešení problémů tam, kde se člověku nedařilo, ale že do budoucna budou schopny přicházet s řešením problémů, které si člověk ani neuvědomuje.

Současně s tím, jak se mění rozložení práce a poměr rozhodnutí, která jsou vykonávána UI a která člověkem, dostávají se do středu pozornosti otázky týkající se systému a koncepce pravidel odpovědnosti za inteligentní systémy. Pro tuto debatu jsou charakteristické otázky, které se ptají na to, kdo ponese odpovědnost v případě újmy způsobené systémem UI a kdy je tento systém schopen přijímat rozhodnutí, která nelze zpětně přímo připsat vývojáři ani provozovateli systému. Druhým charakteristickým tématem provázaným s problematikou autonomie systémů UI je téma udělení právní subjektivitě a přiznání práv a povinností těmto technickým artefaktům. Toto zavedení by mohlo představovat jedno z možných řešení problému, v literatuře označovaného jako mezera v odpovědnosti, kdy není jasné, jak a komu připsat odpovědnost. Odpovědným by tak byl samotný systém UI.<sup>686</sup>

### 13.6 Schopnosti umělé inteligence, rozhodovací proces a problém zkreslení

Dalším souvisejícím tématem se strojovým učením je kvalita rozhodnutí a problém zkreslení, kterým může být rozhodovací proces ovlivněn. Z pohledu rizik lze toto riziko zařadit do kategorie vyplývající ze schopností UI. V tomto případě se jedná o schopnost těchto systémů kategorizovat, klasifikovat a hierarchizovat, přičemž tyto schopnosti a postupy se vztahují k člověku. Základem, na kterém jsou postaveny současné systémy UI, jsou tréninková data. Tyto vstupní

686 KÖNIG / KRAFFT / SCHULZ / ZWEIG, *Essence of AI: What Is AI?*, s. 33.

soubory dat utvářejí hranice, jimiž se řídí fungování UI, a také hranice toho, jak může UI vidět prostředí, v němž je nasazena.<sup>687</sup>

Z hlediska kvality jakéhokoliv rozhodnutí systému UI to znamená, že jeho kvalita se primárně odvíjí od kvality souboru dat, která byla použita k jeho trénování. Zkreslený soubor dat může vést k rozhodnutím UI, která diskriminují určité jednotlivce a skupiny lidí. Ke zkreslení výstupů dochází například v situaci, kdy tréninková data nejsou reprezentativní pro reálné prostředí, ve kterém má systém fungovat. Systém UI tak může produkovat neočekávané a nežádoucí výsledky, pokud se setká se situací, která se výrazně liší od dat, na kterých se učil. Další možnost zkreslení výsledku vychází z toho, že v datech se odráží existující sociální nerovnost a její projevy v podobě předsudků a stereotypizací, které jsou následně vtěleny systémem do pravidel pro řešení konkrétního úkolu. Stručně řečeno, tímto způsobem může UI reprodukovat ve svých rozhodnutích existující vzorce diskriminace tak, jak se vyskytují ve společnosti.<sup>688</sup>

### 13.7 Oblast použití a účel použití umělé inteligence jako možný zdroj rizika

Poslední kategorie se věnuje rizikům spojeným s konkrétním používáním systémů založených na UI. Opírá o to, že dopady daného systému UI vždy vyplývají z širšího kontextu, v němž je systém implementován. Technická specifikace systému UI v tomto případě sehrává pouze podružnou roli a nepředurčuje ani dopad, který systém bude mít, ani rizika, která s sebou jeho implementace přinese.<sup>689</sup>

Systémy UI jsou konkrétním technologickým řešením daného problému, které může být v jedněch podmínkách zcela neškodné, případně schopné přinášet značný užitek. V případě přenesení z tohoto prostředí do jiného se však může stát, že v novém prostředí bude mít radikálně odlišné sociální důsledky a bude tak způsobovat z hlediska regulace hluboké problémy. Například systém pro rozpoznávání tváře,

687 BUITEN, Towards Intelligent Regulation of Artificial Intelligence, s. 51; DEVILLÉ, R. / SERGEYSSELS, N. / MIDDAG, C. Basic Concepts of AI for Legal Scholars. In: DE BRUYNE, J. / VANLEENHOVE, C. (eds.) *Artificial Intelligence and the Law*. Cambridge / Antwerp / Chicago: Intersentia (Centrum voor Verbintenissen-en Goederenrecht), 2021, s. 4.

688 KÖNIG / KRAFFT / SCHULZ / ZWEIG, *Essence of AI: What Is AI?*, s. 29; CHESTERMAN, *We, the Robots?*, s. 70.

689 SCHUETT, *Defining the scope of AI regulations*, s. 73.