

Roboti ve světle teorií subjektivních práv

TRISTAN FLORIAN*

Robots in the Context of Theories of Rights

Summary: The paper applies widely accepted theories of rights to the topic of robots. It first explains the different approaches to the definition of artificial intelligence and then deals with the definition of a robot. Next, it discusses the two fundamental theories of rights – will theory and interest theory. The will theory emphasizes rights as the protection of choices and is often criticized as non-inclusive of those who cannot make choices – animals, newborns, or people in a vegetative state. However, the author discusses that this may not be the case for robots, as robots can make more qualified decisions than humans in some areas. If rights are to serve the function of protecting choices, then robots that can make choices should have them. However, this would imply that an artificial entity would have a higher status than a living one, a conclusion that the text presents as a flaw of the will theory. The author discusses that one can argue in favour of robot rights through the interest theory, which emphasizes rights as the protection of interests of ultimate value – we can accept stronger legal protections for robots because such protections serve humans. However, the article argues against calling such protections robot rights when in fact, due to their artificial nature, they do not even have the potential to have any interests, and we can only use the interests of other actors to justify such protections. According to the author, the fact that we can justify robot rights through the widely accepted theories should make us challenge these theories rather than make us accept robot rights.

Keywords: robot rights, philosophy of artificial intelligence, theories of rights, will theory, interest theory

Nacházíme se v době rozmachu umělé inteligence, kdy počítače zvládají více věcí, než si lidé ještě před deseti lety dokázali představit. V právní filozofii se tento technologický pokrok projevuje silnými debatami o právech robotů. Diskutuje se o reálných schopnostech robotů, ale také hůře uchopitelných problémech, jako například jejich vůli.

V tomto textu aplikuji teorie subjektivních práv na problematiku robotů. Abych tak mohl učinit, nejprve musím vysvětlit pojem umělé inteligence, který tvoří základ, od něhož se odvíjí veškeré další úvahy. Představím jak filozofické dělení, tak praktické pojetí, které je oproštěno od metafyzických otázek o lidské mysli a dá se vědecky lépe uchopit. Poté se budu věnovat samotnému pojmu robot, včetně praktických příkladů, jak robot může navenek vypadat.

Text následně přejde k problematice subjektivních práv. Budu se soustředit na hohfeldovské incidenty jakožto jejich hlavní složky. Dále se budu zabývat volní a zájmovou teorií, což jsou hlavní teorie subjektivních práv, které stojí v centru pozornosti současné právně filozofické literatury. Tyto teorie pak

využiji k tomu, abych se zabýval otázkou práv robotů. Na konci také nastíním základní otázky, které vyvstávají ohledně právní odpovědnosti robotů.

Cílem textu není přesvědčit čtenáře o tom, že by roboti měli mít práva, ale spíše problematizovat dominantní teorie subjektivních práv. Aplikací teorií subjektivních práv na problematiku robotů chci ilustrovat, k jak pochybným závěrům můžeme dojít, když hlavní teorie subjektivních práv přijmeme.

Rád bych na tomto místě zdůraznil, že text nespadá do oboru civilistiky, nýbrž primárně do právní filozofie, a reflektuje především debaty, které můžeme najít v zahraniční literatuře. Od toho se odvíjí také používaná terminologie, která se od té, jež používá česká civilistika, může lišit.

Umělá inteligence

Pojem umělé inteligence poprvé použil americký kognitivní vědec John McCarthy v roce 1956, když navrhoval letní workshop podobně zaměřených vědců na Dartmouth College.¹ McCarthy se snažil vyhnout pojmu „počítač“

* Celým jménem Oldřich Tristan Florian působí jako interní doktorand na Katedře právní teorie PrF MU. E-mail: florian@otflorian.com. <https://orcid.org/0000-0003-1058-2289>.

¹ RUSSELL, S. – NORVIG, P. *Artificial Intelligence: A Modern Approach*. Essex 2016, str. 17.

nebo „výpočetní“, a to kvůli americkému matematikovi Norbertu Wienerovi, který tyto pojmy spojoval spíše s analogovými zařízeními než s digitálními technologiemi.² Tento pojem pak začal žít vlastním životem a začal se používat i pro interpretaci filozofických úvah, které se objevily před tím, než pojem umělé inteligence vůbec vznikl.

V souvislosti s pojmem umělé inteligence se dnes rozlišuje řada přístupů. Na více než tisíce univerzit po celém světě se například používá pojetí počítačových vědců Stuarta J. Russella a Petera Norviga.³ Ti rozlišují přístupy „jednat lidsky“ (*acting humanly*), „myslet lidsky“ (*thinking humanly*), „myslet racionálně“ (*thinking rationally*) a „jednat racionálně“ (*acting rationally*).⁴ Každý z těchto přístupů upřednostňuje jiné charakteristiky jako směrdatné pro to, abychom o něčem řekli, že se jedná o umělou inteligenci. Sami autoři se pak přiklání k poslednímu jmenovanému přístupu, přičemž racionální jednání definují jako jednání, kterým dosáhneme buďto nejlepšího výsledku, nebo alespoň nejlepšího z očekávaných výsledků.⁵

Text Russella a Norviga týkající se definice umělé inteligence je však velice stručný. Je spíš úvodem do tématu pro technicky zaměřené čtenáře a sám o sobě nemá moc velkou filozofickou hodnotu. Pojetí Russella a Norviga nereflektuje ani to, jaké dělení pojmu umělé inteligence se používá ve filozofii mysli. Pro filozofické zkoumání tohoto pojmu je tedy daleko důležitější pojetí amerického filozofa Johna Searla.

Searle se sice v poslední době zabývá především sociální ontologií, ale do historie filozofie se zapsal svými publikacemi v oblasti filozofie mysli. Právě tento autor přišel se základním dělením pojmu umělé inteligence, které se ve filozofii mysli ujalo.

Podle Searla musíme rozlišovat *weak AI* (slabá umělá inteligence) a *strong AI* (silná umělá inteligence). *Weak AI* je v jeho pojetí označení pro filozofický přístup, který uznává umělou inteligenci jako užitečnou simulaci lidské mysli.⁶ *Strong AI* je pak označení pro filozofický přístup, podle kterého je mysl pro mozek to, co software pro hardware

– mysl tedy Searle chápe jako počítačový program.⁷

Oba tyto přístupy mají jiné filozofické implikace. Přístup *weak AI*, ke kterému se hlásí i Searle, uznává počítačovou simulaci lidské mysli jako prostředek, kterým můžeme lidskou mysl zkoumat.⁸ Tedy pouze prostředek, nic více. Z přístupu *strong AI* naopak plyne, že správně naprogramovaný počítač, který dokáže napodobovat lidskou mysl, je myslí sám o sobě, nikoliv pouhým prostředkem k jejímu zkoumání.⁹ Jinak řečeno, podle *weak AI* je umělá inteligence nástrojem k vysvětlení, zatímco podle *strong AI* je umělá inteligence samotným vysvětlením.¹⁰ Tedy pokud budeme schopní vytvořit umělou inteligenci na úrovni člověka, pak bude mít, stejně jako člověk, svoje vědomí.

Searle se vůči teorii o *strong AI* vymezuje ve svém známém myšlenkovém experimentu, pro který se vžilo pojmenování *Chinese room argument* (argument čínského pokoje). Searle chce, abychom si představili, že jsme zamčení v místnosti, dostaneme velké množství čínských znaků a návod, jak tyto znaky spojit, aby dohromady tvořily smysluplné věty a odpovědi na ně. Tvrdí, že jakmile se naučíme dobře manipulovat s čínskými znaky, budou naše odpovědi nerozeznatelné od odpovědí rodilých mluvčích. Nikdo, kdo se k našim odpovědím dostane, nepozná, že vlastně neumíme čínsky.¹¹

Searle tak poukázal na to, že ačkoliv navenek to může působit, že čínštinu ovládneme, ve skutečnosti jenom následujeme sadu instrukcí, kterou jsme dostali. Ačkoliv se to navenek jeví jinak, čínštině jsme neporozuměli. A podobným způsobem funguje i umělá inteligence. Zpracovává nějaké informace a na základě nich se snaží dopátrat výstupu. Podle přístupu *weak AI* tak však nečiní vědomě – jedná se jenom o simulaci.

Řada výše nastíněných otázek má přesah do dlouhodobého filozofického sporu o tom, co je vlastně vědomí. Případně jak spolu vůbec mohou komunikovat tělo a vědomí a zda existují ve stejném světě.¹² Jedná se o metafyzické otázky, které nejspíš nemají jednoznačné řešení. V praktických problémech nám nepomohou, a proto nejsou pro počítačové

² Ibidem.

³ Viz 1440 Schools Worldwide That Have Adopted AIMA. In: *Artificial Intelligence: A Modern Approach* [online]. Dostupné na <<http://aima.cs.berkeley.edu/adoptions.html>>.

⁴ RUSSELL – NORVIG, *Artificial Intelligence: A Modern Approach*, str. 2–6.

⁵ Ibidem, str. 4.

⁶ SEARLE, J. R. *The Mystery of Consciousness*. New York 1997, str. 9.

⁷ Ibidem.

⁸ SEARLE, J. R. Minds, brains, and programs. *Behavioral and Brain Sciences*, 1980, roč. 3, č. 3, str. 417.

⁹ Ibidem.

¹⁰ Ibidem.

¹¹ Viz ibidem, str. 417–418.

¹² Pro tyto debaty se vžilo označení mind-body problem. Viz HEIL, J. *Philosophy of Mind: A Contemporary Introduction*. New York 2013, str. 18.

výzkumníky tolik zajímavé. Pro použití v oblasti vývoje technologií se tedy ujalo jiné dělení umělé inteligence – úzká umělá inteligence (*narrow artificial intelligence*), obecná umělá inteligence (*artificial general intelligence*) a umělá superinteligence (*artificial superintelligence*).¹³

V případě úzké umělé inteligence (označované mnohdy také jako *weak AI*, v nesouladu s terminologií Johna Searla) se jedná o umělou inteligenci, která je schopná efektivně plnit úlohy, které jí zadal člověk.¹⁴ Obecná umělá inteligence (označovaná jako *strong AI*, opět v nesouladu s terminologií Johna Searla) je pak umělá inteligence, která se dokáže chovat nerozeznatelně od člověka.¹⁵ Tuto formu umělé inteligence někteří autoři označují jako velký sen v oblasti výzkumu umělé inteligence.¹⁶ Konečně, umělá superinteligence je umělou inteligencí, která dokáže člověka v některých aspektech předstihnout.¹⁷

Definice umělé inteligence existuje nesmírné množství. Jak říká publicista Paul McCloskey, je jich „zhruba tolik, kolik výzkumníků tuto technologii vyvíjí“.¹⁸ Obecně se ale dá říct, že pojem *umělá inteligence* tak, jak se používá ve vědeckých studiích, vždy naplňuje dvě klíčové charakteristiky. První charakteristikou je řešení velmi složitých úkolů, druhou přízpusobením se prostředí.¹⁹ Takové pojetí umělé inteligence tedy již nemá přesah do metafyzických otázek o lidské mysli. Určující je v něm pouze to, co umělá inteligence dokáže, tedy jak se nám jeví navenek, nikoliv co probíhá uvnitř a jestli to zakládá její vědomí.

Robot

Už tedy máme představu o tom, jak můžeme přistupovat k pojmu umělé inteligence. Co ale pojem *robot*?

Stejně jako je tomu u umělé inteligence, ani pro tento pojem neexistuje jednoznačně přijímaná definice. Slovo *robot* pochází z českého slova *robota* a poprvé bylo použito v divadelní hře R.U.R. spisovatele Karla Čapka.²⁰

Jeho počátky tedy sahají k literatuře science-fiction. Právě umělecký žánr sci-fi má dodnes velký vliv na to, co si pod slovem *robot* lidé představují.²¹

V různých příbězích mají přitom roboti různou podobu. Například v R.U.R. jsou na první pohled nerozeznatelní od lidí, jelikož tak vypadají a zároveň se tak chovají.²² Také uvnitř jsou stejní jako lidé, ale odlišuje je, že nejsou naučení některým lidským konceptům, jako například smrti či lítosti.²³ Obdobné je to u robotů v podání například Phillipa K. Dicka, jehož nejslavnější dílo se dočkalo kultovního filmového zpracování z roku 1982 pod názvem *Blade Runner*. Roboti (které Dick v příběhu nazývá replikanty) vypadají jako lidé, přičemž někteří si třeba ani neuvědomují, že lidmi nejsou.²⁴

Odlíšným typem je Terminátor ze stejnojmenné filmové série. Jedná se totiž o kyborga – tedy entitu, která kombinuje stroj a člověka.²⁵ Uvnitř je to obrněný stroj poháněný mikroprocesorem, ale na povrchu má svaly, krev, kůži či vlasy, čímž se podobá člověku.²⁶ Robotům, kteří chováním i vzhledem připomínají lidi (jako právě zmínění R.U.R., replikanti či Terminátor), se obecně říká android. Ne všichni roboti ale vypadají jako lidé. Typickou ukázkou může být Johnny 5 z filmu *Číslo 5 žije (Short Circuit)*. Je čistě jen stroj z oceli, bez lidské kůže na povrchu. Má hlavu ve tvaru kvádru, ruce se třemi prsty a místo nohou kola s pásy.

Z filmů a literárních děl tedy máme nějakou představu, jak roboti myslí a vypadají. Tyto představy jsou pak určující i pro to, jak budou vypadat roboti, kteří jsou konstruováni v reálném světě. Jak ale můžeme robota obecně definovat?

Příhodnou definici podle mého názoru poskytují Neil M. Richards a William D. Smart. Ti definují robota jako zkonstruovaný systém, který se po fyzické i mentální stránce jeví jako aktér, avšak v biologickém smyslu není živý.²⁷ Tedy že robot je něco vytvořeného, co se pohybuje po světě, tváří se, jako by činilo

¹³ SARACCO, R. Artificial Super Intelligence. In: *IEEE Future Directions* [online]. 2018. Dostupné na <<https://cmte.ieee.org/futuredirections/2018/09/10/artificial-super-intelligence/>>.

¹⁴ Ibidem.

¹⁵ Ibidem.

¹⁶ BAUM, S. A Survey of Artificial General Intelligence Projects for Ethics, Risk, and Policy. *SSRN Electronic Journal*, 2017, str. 7.

¹⁷ SARACCO, R. *Artificial Super Intelligence*.

¹⁸ PAUL, M. What's AI, and what's not. In: *GCN* [online]. 10. 5. 2017. Dostupné na <<https://gcn.com/articles/2017/03/10/defining-ai.aspx>>.

¹⁹ LÜCK, N. *Machine Learning-Powered Artificial Intelligence* [online]. 2019, str. 2. Dostupné na <<https://www.jstor.org/stable/resrep26193.5>>.

²⁰ HORÁKOVÁ, J. 85 let Čapkových Robotů a jejich proměn. *Musicologica Brunensia*, 2006, roč. 55, č. H41, str. 63–66.

²¹ GUNKEL, D. J. *Robot Rights*. Cambridge, Massachusetts 2018, str. 16.

²² Viz ČAPEK, K. *R.U.R. Rossum's Universal Robots: kolektivní drama o vstupní komedii a třech dějstvích*. Praha 1940, str. 23.

²³ Viz ibidem, str. 23–25.

²⁴ Viz DICK, P. K. *Do Androids Dream of Electric Sheep?* London 1999.

²⁵ Viz CAMERON, J. *Terminator: The Script, Fourth Draft*. 1983, str. 117–118.

²⁶ Ibidem.

²⁷ CALO, M. R. – FROMKIN, M. – KERR, I. (eds). *Robot Law*. Cheltenham, UK 2016, str. 6.

racionální volby, ale zároveň je strojem.²⁸ V této definici je obsažena jak umělá inteligence, tak nějaký fyzický projev, který nemusí být přitom nutnou součástí systému umělé inteligence. Pod umělou inteligencí si totiž můžeme představit například i aplikaci, která nám podle naší historie prohlášení doporučuje videa, jež by se nám mohla líbit. Jelikož se však nijak nepohybuje po světě, bylo by neintuitivní nazývat to robotem.

Výhodou definice Richardse a Smarta je její všeobecnost. Mnohým filozofům činí problém uznat, že by měl umělý výtvar, který je schopný se navenek projevat podobně jako člověk, svoje vlastní vědomí. Tedy že může prožívat stavy vědomí, kterých jsou schopní lidé. Tato definice však hledí spíše na vnější charakteristiky a není směrodatné, co se skutečně děje uvnitř. Důležité tedy je, zda se daný stroj tváří jako entita, která je schopná racionálně jednat, ale už ne to, jestli tak skutečně jedná.

Teorie subjektivních práv

Hohfeldovské incidenty

Nyní už tedy máme představu, co je to robot. Jak je to ale s jeho subjektivními právy? Pro zkoumání pojmu subjektivních práv je potřeba nejprve uvést jejich čtyři základní komponenty, se kterými přišel Wesley N. Hohfeld a vžilo se pro ně označení hohfeldovské incidenty (*Hohfeldian incidents*). Právě tyto incidenty nám totiž umožní analyticky uchopit pojem subjektivních práv tak, jak jej používá mimo jiné také česká a slovenská právní teorie.²⁹

Mezi hohfeldovské incidenty patří privilegium (*privilege*, někdy také svoboda – *liberty*), nárok (*claim*),³⁰ pravomoc (*power*) a imunita (*immunity*).³¹ Lze je definovat následovně (A/B jsou aktéři a ϕ je sloveso):³²

1. Privilegium (svoboda): A má *privilegium* k ϕ jen a pouze tehdy, když A nemá

povinnost k non- ϕ . Podle Hohfelda je opakem povinnost (*duty*). Korelativem (jak Hohfeld označuje to, co má B, když A má nárok) je podle něj ne-nárok (*no-claim*), tedy opak nároku.³³

2. Nárok: A má *nárok*, aby B ϕ , jen a pouze tehdy, když B má *povinnost* vůči A k ϕ . Opakem je podle Hohfelda ne-nárok (*no-claim*), korelativem pak povinnost (*duty*).³⁴
3. Pravomoc: A má *pravomoc* jen a pouze tehdy, když A má *možnost* pozměnit právní situaci sebe nebo ostatních. Za opak Hohfeld považuje nemožnost (*disability*), za korelativ pak odpovědnost (*liability*).³⁵
4. Imunita: A má *imunitu* jen a pouze tehdy, když B nemá *možnost* pozměnit právní situaci aktéra A. Jako opak Hohfeld uvádí odpovědnost (*liability*), jako korelativ pak nemožnost (*disability*).³⁶

Teorie subjektivních práv se pak snaží najít nějaký jednotící prvek těchto incidentů, tedy jaké jsou mezi nimi vztahy a co mají společného.³⁷ Tyto teorie pak tradičně rozdělujeme na volní teorii (*will theory*)³⁸ a zájmovou teorii (*interest theory*).

Práva jako ochrana volby

Jako nejstarší zastávce volní teorie je mnohdy zmiňovaný Immanuel Kant,³⁹ ačkoliv podle některých autorů se současné pojetí volní teorie od Kantovy teorie subjektivních práv liší.⁴⁰ Volní teorie totiž obvykle považují práva za prostředek k tomu, aby aktér mohl činit rozhodnutí ohledně jejich vynucení,⁴¹ zatímco pro Kanta mají práva naopak neinstrumentální povahu.⁴² V současnosti nejčastěji zmiňovaným proponentem volní teorie je pak H. L. A. Hart.

Hart vysvětluje práva na možnosti kontroly povinnosti ostatních, přičemž zmiňuje tři prvky takové kontroly. Prvním je, že se držitel práva může vzdát výkonu s ním korespondující

²⁸ Ibidem.

²⁹ ŠEJVL, M. Subjektivní práva. In: SOBEK, T. – HAPLA, M. (eds). *Filosofie práva*. Brno 2020, str. 313.

³⁰ Pojem *nárok* má zde jiný význam než v české civilistice. V originálním textu Hohfeld psal o *right*, což používal střídavě s *claim*. V textech, které na něj navazovaly, se pak pro lepší přesnost ustálilo používání *claim*, tedy *nárok*, viz např. WENAR, L. Rights. In: ZALTA, E. N. (ed.). *The Stanford Encyclopedia of Philosophy* [online]. 2021. Dostupné na <<https://plato.stanford.edu/archives/spr2021/entries/rights/>>.

³¹ DUFEK, P. – BAROŠ, J. Teorie lidských práv. In: HOLZER, J. – MOLEK, P. (eds). *Demokratizace a lidská práva: středoevropské pohledy*. Praha 2013, str. 84.

³² GUNKEL, *Robot Rights*, str. 28–30.

³³ HOHFELD, W. N. Some Fundamental Legal Conceptions as Applied in Judicial Reasoning. *The Yale Law Journal*, 1913, roč. 23, č. 1, str. 30.

³⁴ Ibidem.

³⁵ Ibidem.

³⁶ Ibidem.

³⁷ GRAHAM, P. The Will Theory of Rights: A Defence. *Law and Philosophy*, 1996, roč. 15, č. 3, str. 259.

³⁸ Český překlad používám stejný jako DUFEK – BAROŠ, *Teorie lidských práv*, str. 85; dá se také přeložit jako teorie vůle, viz HAPLA, M. *Lidská práva bez metafyziky: Legitimita v (post)moderní době*. Brno 2016, str. 35. Jednotný překlad v české literatuře není.

³⁹ Viz např. ONDŘEJEK, P. *Lidská práva v soukromém právu*. 2009, Univerzita Karlova, Právnická fakulta, str. 37.

⁴⁰ Viz ZYLBERMAN, A. Kant's Juridical Idea of Human Rights. In: FOLLESDAL, A. – MALIKS, R. (eds). *Kantian Theory and Human Rights*. New York 2013, str. 27.

⁴¹ FRYDRYCH, D. What Is the Will Theory of Rights? *Ratio Juris*, 2019, roč. 32, č. 4, str. 462.

⁴² ZYLBERMAN, *Kant's Juridical Idea of Human Rights*, str. 27.

povinnosti, nebo tuto povinnost ponechat v platnosti.⁴³ Stejně jako Kant tedy vnímá právo jako nárok, který logicky koresponduje s povinností někoho jiného.⁴⁴ Druhým prvkem je, že porušenou povinnost může držitel práva nechat vynutit, případně ponechat nevy nucenou.⁴⁵ Třetím je, že má držitel práva možnost nechat zaniknout nutnost náhrady škody za porušení povinnosti.⁴⁶

Každé právo tedy obsahuje *pravomoc* nad určitým *nárokem* a umožňuje nositeli práva získat kontrolu nad povinnostmi druhých.⁴⁷ Aktéra, který má kontrolu nad povinnostmi ostatních, pak Hart nazývá malým suverénem (*small-scale sovereign*).⁴⁸ Pokud si například koupím lístek na vlak, pak jsem uzavřel smlouvu s přepravcem, aby mi poskytl místo v konkrétním vagónu, a přepravce má tedy povinnost mi to zajistit; avšak abych mohl získat tento *nárok*, musím mít *pravomoc* si koupit lístek.⁴⁹ Práva podle volní teorie tedy spočívají v tom, že prostřednictvím *pravomocí* vytváříme a rušíme *nároky*.⁵⁰ Spojují tedy práva s autoritou aktéra nad určitými oblastmi jednání.⁵¹

Problém nám tato teorie však může přinést u otázky, kdo má být nositelem práv. Když v souladu s volní teorií přijmeme, že práva chrání volby, pak nemohou být nositeli práv ti, kteří volby nejsou schopni činit – například novorozenci.⁵² To je však v rozporu s intuicemi většiny z nás, a proto se jedná o zásadní slabinu této teorie.⁵³ Běžně totiž uznáváme, že i novorozenci či lidé ve vegetativním stavu mají mít práva a není to (na rozdíl třeba od práv nenarozených dětí) předmětem zásadních morálních a politických sporů.

Z volní teorie rovněž vyplývá, že aktéři nemají žádná práva vztahující se k trestnímu právu – nemají totiž volbu ohledně povinností z tohoto právního odvětví vyplývajících.⁵⁴ To je však také zvláštní, protože bychom

přece intuitivně řekli, že máme právo nebýt pobodaný na ulici.⁵⁵ To stejné platí například u sportu, kde sportovci, kteří jsou faulovaní (tedy jim bylo narušeno právo, které vyplývá z pravidel hry), stejně nerozhodují o vymáhání takových pravidel.⁵⁶

Volní teorie je obecně často kritizovaná právě kvůli své neinkluzivnosti a není kompatibilní ani s ideou práv zvířat, která se v posledních dekáдах dostává čím dál víc do popředí.⁵⁷ Jako řešení problémů s neinkluzivností se tedy nabízí zájmová teorie.

Práva jako ochrana zájmů

Počátky zájmové teorie sahají k Jeremymu Benthamovi, nejpobulárnější pojetí zájmové teorie však pochází od Josepha Raze. Podle Raze má aktér právo jen a pouze tehdy, když má způsobilost mít práva a zároveň má nějaký aspekt zájmů⁵⁸ tohoto aktéra dostatečný důvod k tomu, aby byla jiná osoba podřízena povinnosti.⁵⁹ Způsobilost mít práva pak definuje tak, že aktér je právně způsobilý jen a pouze tehdy, když mají jeho zájmy silnou hodnotu nebo je právníčkou osobou (např. korporace).⁶⁰ Matthew H. Kramer uvádí, že pojmy *právo* a hohfeldovský incident *nárok* jsou v rámci zájmové teorie to samé (ačkoliv by tedy *nárok* ještě měl v praxi obsahovat *imunitu*, která zabrání jeho zrušení).⁶¹

Razově teorii se však problémy nevyhnu. Co když totiž nejsou zájmy držitele práva dostatečně silné, aby odůvodnily povinnosti ostatních?⁶² Raz sám zmiňuje například právo novinářů chránit zdroje svých informací. Tato ochrana však nemůže být plně zdůvodněna jenom samotným zájmem novinářů získávat informace – ke zdůvodnění se musí přidat fakt, že je tento zájem hodnotný kvůli tomu, jaký užitek přináší veřejnosti.⁶³ Zájem aktéra zde není dostatečným důvodem a tím

⁴³ HART, H. L. A. *Essays on Bentham: Studies in Jurisprudence and Political Theory*. Oxford 1982, str. 184.

⁴⁴ ZYLBERMAN, *Kant's Juridical Idea of Human Rights*, str. 27.

⁴⁵ HART, *Essays on Bentham: Studies in Jurisprudence and Political Theory*, str. 184.

⁴⁶ *Ibidem*.

⁴⁷ DUFEK – BAROŠ, *Teorie lidských práv*, str. 85.

⁴⁸ HART, *Essays on Bentham: Studies in Jurisprudence and Political Theory*, str. 183.

⁴⁹ Příklad je obdobou GRAHAM, *The Will Theory of Rights: A Defence*, str. 261.

⁵⁰ *Ibidem*.

⁵¹ DUFEK – BAROŠ, *Teorie lidských práv*, str. 85.

⁵² MEYERSON, D. Persons and Their Rights in Law and Morality. *Australian Journal of legal philosophy*, 2010, roč. 35, str. 123.

⁵³ DUFEK – BAROŠ, *Teorie lidských práv*, str. 85.

⁵⁴ WENAR, L. The Nature of Claim-Rights. *Ethics*, 2013, roč. 123, č. 2, str. 203.

⁵⁵ *Ibidem*.

⁵⁶ *Ibidem*.

⁵⁷ STUCKI, S. Towards a Theory of Legal Animal Rights: Simple and Fundamental Rights. *Oxford Journal of Legal Studies*, 2020, roč. 40, č. 3, str. 541.

⁵⁸ Raz pojem *zájem* používá zaměnitelně s pojmem *well-being*, který podle něj značí dobrý život, resp. snahu usilovat o hodnotné vztahy a cíle, viz RAZ, J. The Role of Well-Being. *Philosophical Perspectives*, 2004, roč. 18, str. 269.

⁵⁹ RAZ, J. *The Morality of Freedom*. Oxford 2009, str. 166.

⁶⁰ *Ibidem*.

⁶¹ KRAMER, M. H. Some Doubts about Alternatives to the Interest Theory of Rights. *Ethics*, 2013, roč. 123, č. 2, str. 246–247.

⁶² WENAR, *The Nature of Claim-Rights*, str. 204.

⁶³ RAZ, *The Morality of Freedom*, str. 179; WENAR, *The Nature of Claim-Rights*, str. 204.

se tento příklad vymyká definici. Podobným příkladem jsou přídatky na děti – ačkoliv se jedná o právo rodiče, je zdůvodněno zájmy dítěte.⁶⁴

Zájmová teorie tedy občas bývá vnímána jako nedostatečná, protože se jí vymykají situace, kdy lidé mají práva, i když nemají dostatečně silné zájmy.⁶⁵ Domnívám se však, že jde pouze o vnitřní problém Razovy definice. Kramerova definice je totiž přesnější, zmíněný problém řeší a zní následovně:⁶⁶

1. Nutným, avšak nedostačujícím předpokladem toho, aby byl aktér držitelem práva, je, že povinnost, která s právem koresponduje, normativně chrání některé aspekty situace aktéra, které jsou typicky v zájmu lidské bytosti, kolektivu nebo zvířete.
2. Ani nutným, ani dostačujícím předpokladem toho, aby byl aktér držitelem práva, je to, že je aktér způsobilý a oprávněný požadovat nebo se vzdát vynutitelnosti povinnosti, která s právem koresponduje. (Tj. negace volní teorie.)

Zájmová teorie tedy počítá s tím, že právní subjektivita nemusí být nutně omezená na zdravé dospělé lidi. Podle Kramera je totiž rozšíření na novorozence, duševně nemocné, mrtvé lidi a některá zvířata logickým vyústěním základních premis, ze kterých zájmová teorie vychází.⁶⁷ A to nejen v případech, kdy budeme chtít pouze zohledňovat třeba schopnost zvířat cítit bolest, abychom jim přiznali právo nebýt týraná. Můžeme totiž stanovit i jejich právní povinnosti.⁶⁸

Volby robotů

Volní teorií můžeme obhájit práva robotů. I když je totiž volní teorie neinkluzivní vůči zvířatům, novorozencům, lidem ve vegetativním stavu a dalším entitám, které by mohly být vnímány jako morálně relevantní osoby, v případě práv robotů tomu může být jinak. Co když totiž vznikne umělá entita, která bude z faktického hlediska schopná

vykonávat *pravomoc* nad určitým *nárokem*? Tedy jinými slovy, co když bude schopná jednání, které má být podle volní teorie subjektivními právy chráněno?

Počítačové programy před nástupem umělé inteligence spoléhaly na zásah programátora, protože to byl právě on, kdo musel napsat sérii podmíněných příkazů, kterými se program řídil.⁶⁹ Pro ilustraci, představme si například proměnnou x , která může nabývat libovolné celočíselné hodnoty. Dále si představme jednoduchý program, který uživatele vyzve, aby nastavil hodnotu x . Program také obsahuje podmíněný příkaz s logickým výrazem $x \geq 0$. Pokud se tento logický výraz vyhodnotí jako pravda, program je nastavený tak, aby zobrazil fotografii koťátka. Pokud jako nepravda, zobrazí se fotografie slona. Využívalo se tedy výhradně deduktivní usuzování. Někdo musel vytvořit obecná pravidla, která určovala chování programu v konkrétních případech.

Součástí umělé inteligence je však vlastnost zvaná strojové učení (*machine learning*), která do věci vnáší induktivní usuzování.⁷⁰ Jako trénovací data jsou zadána ta, která jsou k dispozici, a algoritmus strojového učení si na jejich základě vytváří jejich představu.⁷¹ Po dokončení tohoto trénování mohou být algoritmu předložena data, která předtím neviděl, a s přiměřenou přesností a jistotou by měl být schopen na základě předchozích dat tato data klasifikovat.⁷²

Umělá inteligence se tak učí hledáním opakujících se vzorců ve velkém množství dat.⁷³ Měla by být schopná na základě dílčích pozorování sama tvořit obecné závěry a na jejich základě se pak chovat. Strojové učení vede k tomu, že je umělá inteligence schopná se přizpůsobit prostředí, a následně díky tomu také řešit složité úlohy. Ve výsledku se tak můžeme setkat s tím, že je počítač schopný sám vytvořit nahrávku nějaké písně podle těch, které si načte předtím,⁷⁴ přeložit odborný text z cizího jazyka tak, že je čím dál obtížnější jej rozeznat od profesionálního

⁶⁴ WENAR, *The Nature of Claim-Rights*, str. 204.

⁶⁵ DUFEK – BAROŠ, *Teorie lidských práv*, str. 86.

⁶⁶ KRAMER, M. H. Refining the Interest Theory of Rights. *The American Journal of Jurisprudence*, 2010, roč. 55, č. 1, str. 32.

⁶⁷ KRAMER, M. H. Do Animals and Dead People Have Legal Rights? *Canadian Journal of Law & Jurisprudence*, 2001, roč. 14, č. 1, str. 54.

⁶⁸ Viz *ibidem*, str. 41.

⁶⁹ LÚCK, *Machine Learning-Powered Artificial Intelligence*, str. 2.

⁷⁰ THAGARD, P. Philosophy and Machine Learning. *Canadian Journal of Philosophy*, 1990, roč. 20, č. 2, str. 261; pojem strojové učení slouží také jako označení pro vědní obor, který se na strojové učení zaměřuje, viz KOHAVI, R. – PROVOST, F. Glossary of Terms. *Machine Learning*, 1998, roč. 30, č. 2/3, str. 273.

⁷¹ NATH, V. *Autonomous robotics and deep learning*. New York 2014, str. 2.

⁷² *Ibidem*.

⁷³ BISHOP, C. M. *Pattern recognition and machine learning*. New York 2006, str. 1.

⁷⁴ CHOW, A. R. Musicians Are Using AI to Create Otherwise Impossible New Songs. In: *Time* [online]. Dostupné na <<https://time.com/5774723/ai-music/>>.

překladač,⁷⁵ nahradit obličej ve videu za jiný,⁷⁶ napodobit cizí hlas,⁷⁷ případně vést hluboké konverzace s člověkem.⁷⁸ Samozřejmě s určitými omezeními, ale postupem času se jednotlivé překážky překonávají.⁷⁹

Řada autorů, kteří by byli ochotní v budoucnu uznat práva robotů, tvrdí, že jakmile roboti dosáhnou určitých vlastností, pak by jim měla být přiznána práva.⁸⁰ Mezi tyto vlastnosti se pak může počítat i schopnost se o práva přihlásit.⁸¹ V nedaleké budoucnosti pak může vzniknout umělá entita, která bude z faktického hlediska schopná vykonávat *pravomoc* nad určitým nárokem – což by znamenalo, že její případná práva budou moct plnit funkci, kterou práva podle volní teorie plnit mají.⁸²

Právě hlášení se o svá práva a faktické vykonávání svých pravomocí není podle mého názoru něco, co by vzhledem k současnému pokroku v oblasti umělé inteligence bylo nepředstavitelné. Právě již zmíněné automatické generování hudebních nahrávek, nahrazování obličejů ve videích, věrohodný překlad, napodobení hlasu a schopnost vést hluboké konverzace je daleko komplexnější. A to všechno již umělá inteligence umí a neustále se v tom zlepšuje.

Schopnost se přihlásit o svá práva vyznívá podle mého názoru triviálně, vzhledem k tomu, co již dnes umělá inteligence dokáže. Kauza, kdy uživatelé Twitteru v roce 2016 zvládli z umělé inteligence od Microsoftu udělat během jednoho dne krajně pravicovou aktivistku,⁸³ ukazuje, že se umělá inteligence může naučit prezentovat morální postoje. Objevují se také úvahy o tom, že v dohledné budoucnosti bude moct manažerská rozhodnutí čím dál efektivněji dělat právě umělá inteligence, proto pokrok v této oblasti může vést i k rušení manažerských pozic.⁸⁴ Domnívám se, že pokud budeme posuzovat nějaké

entity čistě podle jejich dovedností, můžeme se dostat dokonce do stavu, kdy z toho ve výsledku jako lidé vyjdeme hůře.

V rámci volní teorie plní práva funkci ochrany volby. Dává ale smysl tato práva odepírat entitám, které jsou schopné volby dělat, byť by takové entity vznikly uměle? Podle mého soudu nikoliv. Nemusíme se uchýlovat k zájmové teorii, pokud bychom chtěli vztáhnout práva i na roboty. Ačkoliv volní teorie nepasuje na koncept práv zvířat a kvůli této neinkluzivitě je kritizovaná, v případě práv robotů tomu může být jinak. A to navzdory tomu, že zvířata jsou živé bytosti, zatímco roboti pouhé umělé výtvořky.

Prostřednictvím volní teorie tak můžeme dojít k tomu, že umělé entitě přisoudíme vyšší status než živé. Domnívám se, že jde o ukázkou, jak může být volní teorie problematická, když jejím prostřednictvím můžeme dojít k takto pochybnému závěru.

Zájmy robotů

Volní teorie není jedinou, v jejíž rámci by se dalo argumentovat pro práva robotů. Také zájmová teorie je právům robotů otevřená. Umožňuje totiž aktérovi mít práva nejen kvůli jeho vlastním zájmům, ale i kvůli relevantním zájmům jiných osob, které se jeho situace účastní. Výhodou je, že taková argumentace je nezávislá na schopnostech robotů. Nemusíme tedy čekat na okamžik, kdy roboti získají vlastnosti, které bychom mohli vnímat jako morálně relevantní.⁸⁵

V duchu zájmové teorie argumentuje Kate Darling, která tvrdí, že máme silný zájem na tom, aby roboti získali naše uznání a ochranu.⁸⁶ Náš vztah k robotům se totiž postupem času proměňuje, například v případě sociálních robotů, kteří napodobují projevy, jež máme asociované s lidskými pocity.⁸⁷ Jelikož

⁷⁵ VÁCLAVÍK, L. Google Translate našel přemožitele. Vyzkoušeli jsme překladač DeepL. In: *Živě.cz* [online]. 28. 4. 2021. Dostupné na <<https://www.zive.cz/clanky/google-translate-nasel-premozitele-vyzkouseli-jme-prekladac-deepL/sc-3-a-209820/default.aspx>>.

⁷⁶ TOEWS, R. Deepfakes Are Going To Wreak Havoc On Society. We Are Not Prepared. In: *Forbes* [online]. Dostupné na <<https://www.forbes.com/sites/rotoeows/2020/05/25/deepfakes-are-going-to-wreak-havoc-on-society-we-are-not-prepared/>>.

⁷⁷ NAST, C. The Ethics of a Deepfake Anthony Bourdain Voice. In: *The New Yorker* [online]. 17. 7. 2021. Dostupné na <<https://www.newyorker.com/culture/annals-of-gastronomy/the-ethics-of-a-deepfake-anthony-bourdain-voice>>.

⁷⁸ Czech Technical University team wins Alexa Prize SocialBot Grand Challenge 4. In: *Amazon Science* [online]. 16. 8. 2021. Dostupné na <<https://www.amazon.science/academic-engagements/czech-technical-university-team-wins-alexa-prize-socialbot-grand-challenge-4>>.

⁷⁹ KOVAL, V. – DUČKOVÁ, A. Kamarád, který dokáže dobře poradit, popisují vývojáři svoji představu budoucnosti konverzačních chatbotů. In: *Radiožurnál* [online]. 13. 8. 2020. Dostupné na <<https://radiozurnal.rozhlas.cz/kamarad-ktery-dokaze-dobre-poradit-popisuji-vyvojari-svoji-predstavu-budoucnosti-8271828>>.

⁸⁰ GUNKEL, *Robot Rights*, str. 79.

⁸¹ ASARO, P. M. What Should We Want From a Robot Ethic. *International Review of Information Ethics*, 2006, roč. 6, č. 12, str. 12.

⁸² Viz DUFEK – BAROŠ, *Teorie lidských práv*, str. 85.

⁸³ HUNT, E. Tay, Microsoft's AI chatbot, gets a crash course in racism from Twitter. In: *The Guardian* [online]. 24. 3. 2016. Dostupné na <<http://www.theguardian.com/technology/2016/mar/24/tay-microsofts-ai-chatbot-gets-a-crash-course-in-racism-from-twitter>>.

⁸⁴ Viz CITRIX – QUARTZ CREATIVE. What if AI could manage better than your manager? In: *Quartz* [online]. Dostupné na <<https://qz.com/1917008/what-if-ai-could-manage-better-than-your-manager/>>; *Work 2035: How people and technology will pioneer new ways of working* [online]. Fort Lauderdale, Florida 2020, str. 65. Dostupné na <https://www.citrix.com/content/dam/citrix/en_us/documents/analyst-report/work-2035.pdf>.

⁸⁵ Viz GUNKEL, *Robot Rights*, str. 133–134.

⁸⁶ *Ibidem*, str. 134.

⁸⁷ *Ibidem*, str. 133.

mohou roboti sloužit jako nástroj lidské sociality, měli bychom se k nim podle tohoto přístupu chovat dobře pro naše dobro.⁸⁸

Tento přístup přenáší pozornost z robotů na nás – práva jim totiž nechceme přiznat kvůli jim samotným, nýbrž kvůli nám a našim zájmům. Přidávají se tam zájmy jiných entit. Podobně, jako je tomu v případech již zmíněných práv novinářů na ochranu zdrojů informací, u kterých jde ve výsledku o zájem nikoliv novináře, nýbrž celé společnosti.

Podobně můžeme obhajovat práva lidí ve zvláštním postavení. Například podle českého práva je vražda dítěte mladšího patnácti let trestána přísněji než vražda dospělé osoby.⁸⁹ Zabití novorozence je trestáno přísněji, než když je zabitý osmnáctiletý člověk. Je tento rozdíl však ospravedlnitelný výhradně s ohledem na novorozence, aniž bychom zvažili dopad daných činů na společnost? Novorozence je kognitivně méně způsobilý než dospělý člověk. Dokonce je i kognitivně méně způsobilý než prase,⁹⁰ ačkoliv usmrcení prasete není z hlediska práva ani většinové morálky nijak kontroverzní. Zde opět hraje roli spíše ochrana zájmů společnosti než dítěte samotného.

Paralelu lze najít s právy zvířat, respektive s nerovným právním postavením různých druhů. V celé Evropské unii je například zakázáno uvádět na trh, dovážet do něj či z něj vyvážet kůže koček a psů a výrobky obsahující tyto kůže.⁹¹ Preambule Nařízení č. 1523/2007, které tento obchod zakazuje, hovoří o znepokojení spotřebitelů kvůli výrobkům z kůže koček a psů. Dále také o tom, že občané EU považují kočky a psy za zvířata v zájmovém chovu, a proto je nepřijatelné používat jejich kůže nebo výrobky obsahující tyto kůže.

Taková úprava v konečném důsledku vede k tomu, že kočky a psi jsou chráněni víc než ostatní zvířata. Lidé je totiž nemůžou chovat za účelem, který by vyžadoval jejich zabití. Toto výhodnější právní postavení oproti jiným zvířatům se přitom nezakládá na tom, že by byli kočky a psi kognitivně způsobilejší než ostatní obratlovci, ale pouze na tom, jak jejich chov většina lidí v našem kulturním prostředí vnímá.⁹²

Nemusíme ale zacházet až k odlišnému zacházení s různými druhy. Například zákony

proti týrání zvířat nechrání jenom zvířata, ale i společnost. Jak by asi lidé reagovali, kdyby najednou bylo možné beztrestně týrat zvířata? Kdyby hromadné sdělovací prostředky byly plné záběrů, jak lidé špatně zacházejí se zvířaty, a čtenáři by věděli, že se takovým lidem nic nemůže stát? Je taky možné, že by se s rozšířením krutosti vůči zvířatům zvedla míra krutosti vůči lidem. Tyrani by totiž mohli ztratit zábrany a svoji zlost by si pak mohli snadněji vybit na lidech.⁹³ Ve výsledku tedy zvířata nechráníme jen kvůli nim samotným, ale i kvůli zachování společenského řádu.

Blahobyt zvířat tedy může sloužit také jako zástupný důvod a ve výsledku nám jde především o ochranu lidí. Stejným způsobem lze přistoupit i k právům robotů. Nejen že můžeme argumentovat, že společnost bude pobouřena, když lidé budou ničit roboty, kteří působí jako lidé. Můžeme podle mého názoru argumentovat i tím, že když budou roboti vypadat jako lidé, společnost bude desenzibilizovaná k násilí na něčem, co vypadá jako člověk, a proto útoky na „skutečné“ lidi budou psychicky jednodušší. Proto by bylo žádoucí takovým útokům zabránit donucovací silou státní moci.

Mám za to, že pokud bychom silnější právní ochranu robotů chtěli zdůvodnit zájmy lidí, roboti navzájem by se nenacházeli v rovném postavení. Silnější ochrana by se totiž týkala především androidů, tedy robotů, kteří vypadají jako lidé. Ostatní typy robotů by z větší ochrany byly buďto úplně vyřazeny, nebo by nebyla tak vysoká, protože by na ní lidé měli menší zájem. Silnější právní ochranu robotů totiž přijmeme zejména kvůli tomu, že taková ochrana poslouží lidem, aby se jim ve společnosti lépe fungovalo. Pokud by nějaký robot nedokázal vzbudit soucit či natrénovat lidi k tomu, aby byli většími násilníky, pak by byl zájem lidí na ochraně takového robota menší.

Je pak ale vhodné takovou koncepci nazývat *práva robotů*, když ve skutečnosti nejde o jejich zájmy, ale o zájmy jiných osob? Dá se namítnout, že to by se pak nedalo mluvit ani o právech zvířat či některých lidí. Jenže roboti na rozdíl od lidí a zvířat nejsou živými entitami, nýbrž umělými výtvoři. Samozřejmě i právnické osoby, které zájmová teorie zahrnuje, jsou

⁸⁸ Ibidem, str. 158.

⁸⁹ Ustanovení § 140 zákona č. 40/2009 Sb., trestního zákoníku.

⁹⁰ Například nedávná studie nasvědčuje, že jsou prasata schopná se naučit ovládat joystick, viz CRONEY, C. C. – BOYSEN, S. T. Acquisition of a Joystick-Operated Video Task by Pigs (*Sus scrofa*). *Frontiers in Psychology*, 2021, roč. 12.

⁹¹ Čl. 3 Nařízení Evropského Parlamentu a Rady (ES) č. 1523/2007 ze dne 11. prosince 2007.

⁹² O této problematice jsem psal již v jednom ze svých předešlých článků, FLORIAN, O. Proč pouze lidská? Potíž s univerzalitou lidských práv. *Právník*, 2018, roč. 157, č. 11, str. 960–961.

⁹³ Viz FLYNN, C. Acknowledging the „Zoological Connection“: A Sociological Analysis of Animal Cruelty. *Society & Animals*, 2001, roč. 9, č. 1, str. 74.

umělé. V případě právnické osoby však nemluvíme o zájmech ostatních osob, ale o zájmech právnické osoby jako takové. Tyto zájmy jsou pak konstitutivní pro to, abychom právnickou osobu uznali (nehledě na zájmy individuálních osob, které právnickou osobu tvoří) jako subjekt práv.⁹⁴ Roboti však ani nejsou živí, ani nemají žádné zájmy a ve výsledku o ně vlastně vůbec nejde. Jde spíše o to, abychom my, lidé, byli méně pohoršení a případně se nevystavovali takovému nebezpečí.

Podle mého soudu může pojem práv robotů sloužit nanejvýš jako užitečná zkratka. Neměl by však vyvolávat dojem, že jsou roboti osobami v morálně relevantním smyslu, případně že jsou dokonce rovni lidem. Nejsou totiž živí, natož aby například dokázali cítit bolest. Nemají žádné zájmy a jejich případná právní ochrana bude zaváděna zejména kvůli ochraně zájmů lidí. Zůstanou stále pouze prostředkem pro zlepšení našich životů. To, že můžeme prostřednictvím některých teorií zdůvodnit jejich práva, by nás tedy mělo přimět k tomu, abychom takové teorie spíše zpochybnili, než je brali jako prostředek k přijetí práv robotů.

Odpovědnost robotů

V předchozích kapitolách jsem se zabýval zejména morálním statusem robotů a tím, zda by se pohledem volní či zájmové teorie dali roboti považovat za držitele práv. Účelem bylo poukázat na slabiny volní a zájmové teorie. V praxi nám ale kromě práv robotů vystane také otázka jejich odpovědnosti. Byť tato kapitola neproblematizuje úvahy z předešlých a je spíše samostatným tématem, pro úplnost a případný navazující výzkum v této oblasti považuji za vhodné na závěr alespoň nastínit otázky, které se mohou v debatách o odpovědnosti robotů objevit.

Jedním ze zásadních autorů, kteří se problematikou práv robotů v poslední době zabývali, je finský právní filozof Visa A. J. Kurki. Ve svých publikacích kritizuje „ortodoxní pohled“ (*Orthodox View*) na subjektivní práva, který říká, že narozené lidské bytosti

mají/mohou mít práva, zatímco nenarozené, že korporace mají/mohou mít práva, že zvířata nemají práva a že otroci neměli práva.⁹⁵ Kurki si totiž všímá, že zájmová ani volní teorie pro zdůvodnění tohoto ortodoxního pohledu neobstojí,⁹⁶ proto navrhuje alternativní způsoby, jak se dívat na právní osobnost, držení práv a jejich vzájemný vztah. Svoje pojetí pak využívá pro lepší analytické uchopení některých problémů týkajících se například práv zvířat, novorozenců či otroků. Zároveň se však věnuje také právům robotů, přičemž za důležitou pokládá také problematiku jejich právní odpovědnosti. V této kapitole tedy představím Kurkiho pojetí právní osobnosti a následně přejdu k otázkám spojeným s právní odpovědností robotů (respektive umělé inteligence obecně, bez ohledu na její fyzickou schránku).

Kurki zastává pojetí právní osobnosti, které nazývá *Bundle Theory* (možné přeložit jako „balíková teorie“). Spočívá v tom, že se právní osobnost dá dělit na různé druhy, podle toho, jakým „balíkem“ incidentů právní osobnosti daná osoba disponuje. Rozděluje přitom pasivní a aktivní incidenty právní osobnosti.

Za aktivní incidenty považuje způsobilost spravovat ostatní incidenty bez zastoupení (například způsobilost vstupovat do smluv).⁹⁷ Dále také to, co nazývá jako součást „zatěžující právní osobnosti“ (*onerous legal personhood*) – tedy zejména trestní a deliktní odpovědnost.⁹⁸

Pasivní incidenty pak rozděluje na dvě skupiny. První z nich jsou „substantivní pasivní incidenty“. Mezi ty řadí fundamentální ochranu (ochranu života, svobody a tělesné integrity), způsobilost být beneficentem zvláštních práv (zvláštními právy Hart označuje práva, která vyplývají ze závazků),⁹⁹ způsobilost vlastnit majetek a nepřípustnost být vlastněn.¹⁰⁰ Druhou skupinou pasivních incidentů jsou pak „incidenty nápravy“ (*remedy incidents*). Mezi ty řadí pravomoc vystupovat před soudy a jinými úředními orgány, možnost být vnímaný jako oběť v rámci trestního práva a způsobilost být vystavený právním sankcím.¹⁰¹

⁹⁴ PASTERNAK, A. From Corporate Moral Agency to Corporate Moral Rights. *The Law & Ethics of Human Rights*, 2017, roč. 11, č. 1, str. 145.

⁹⁵ KURKI, V. Legal Personhood and Animal Rights. *Journal of Animal Ethics*, 2021, roč. 11, č. 1, str. 51.

⁹⁶ Kurki uvádí, že podle zájmové teorie měli otroci ve Spojených státech v období před občanskou válkou ve skutečnosti určitá omezená zákonná práva, protože byli chráněni před určitými druhy fyzického násilí. Zároveň také byli nositeli i právních povinností, protože mohli být stíháni za trestné činy. Stejně tak i zvířata jsou dnes nositeli práv ve smyslu zájmové teorie, jelikož jsou chráněna zákony o ochraně zvířat. To je přitom v rozporu s ortodoxním pohledem. Podobné problémy má podle Kurkiho volní teorie. Podle některých verzí totiž nemohou mít práva malé děti, což by však znamenalo, že na děti nemůžeme nahlížet jako na osoby (v právním smyslu), zatímco na otroky ano (otroci by se podle Kurkiho například mohli odvolat proti svému odsouzení za trestný čin, což je právo ve smyslu volní teorie). To je pak taky v rozporu s ortodoxním pohledem – na děti totiž jako na osoby běžně nahlížíme. KURKI, V. A. J. *A Theory of Legal Personhood*. Oxford 2019, str. 17; KURKI, *Legal Personhood and Animal Rights*, str. 51–52.

⁹⁷ KURKI, *Legal Personhood and Animal Rights*, str. 53.

⁹⁸ Ibidem.

⁹⁹ Viz KURKI, *A Theory of Legal Personhood*, str. 102.

¹⁰⁰ KURKI, *Legal Personhood and Animal Rights*, str. 53.

¹⁰¹ Ibidem.

Podle toho, jakými incidenty osoba disponuje, Kurki vyjmenovává následující druhy právní osobnosti, které vnímá jako typické:¹⁰²

1. Čistě pasivní právní osobnost (*purely passive legal personhood*). Osoba požívá nároků (v Hohfeldově smyslu), s tím, že jiné osoby mají vůči této osobě korespondující povinnosti.¹⁰³ Zároveň může požívat imunity.¹⁰⁴ Nemůže však vůbec spravovat svou vlastní právní platformu a nemůže být odpovědná. (Pojmem „právní platforma“ Kurki označuje souhrn práv a povinností, které se nějakým způsobem navzájem ovlivňují.)¹⁰⁵ Typicky jí disponují novorozenci, případně zvířata.
2. Závislá právní osobnost (*dependent legal personhood*). Osoba má určité kompetence, které se týkají vlastní právní platformy. Tato způsobilost je ale omezená a podléhá dohledu. Typicky jí disponují děti.
3. Nezávislá právní osobnost (*independent legal personhood*). Osoba je autonomní (*sui iuris*) a má rozhodující slovo při správě své právní platformy. Typicky jí disponují dospělé osoby, které nejsou omezeny na svéprávnosti.
4. Čistě zatěžující právní osobnost (*purely onerous legal personhood*). Osoba je odpovědná, nepožívá však prakticky žádných výhod právní osobnosti. Typicky jí disponují otroci.

Odpověď na otázku, zda je nějaká entita osobou v právním smyslu, tedy z pohledu balíkové teorie není pouze ano/ne. Právní osobnost nám totiž rozděluje do několika různých druhů, takže například novorozenci disponují jiným druhem právní osobnosti než dospělí, plně svéprávné osoby.

Pro téma tohoto článku je ale důležité zejména to, jakým způsobem Kurki přistupuje k otázce právní osobnosti umělé inteligence. Kurki u debat o této otázce rozlišuje tři kontexty. V prvním kontextu, který Kurki nazývá jako „kontext nejvyšší hodnoty“ (*ultimate*

-value context), se ptáme, jestli má umělá inteligence nejvyšší hodnotu, a zda je tedy hodna toho, aby se jí dostalo některé z ochran, které obvykle požívají například děti či další entity, které vnímáme jako osoby v právním smyslu.¹⁰⁶ Tedy zda na ni můžeme vztáhnout alespoň třeba fundamentální ochranu. Druhý kontext, nazývaný jako „kontext odpovědnosti“ (*responsibility context*), se zabývá tím, jestli může být umělá inteligence právně odpovědná za své jednání.¹⁰⁷ Zde přichází do hry aktivní incidenty, tedy konkrétně ty, které se týkají zatěžující právní osobnosti. Třetí kontext, označovaný jako „komerční kontext“ (*commercial context*), souvisí s fungováním umělé inteligence jako aktéra v komerčních vztazích.¹⁰⁸ Úvaha z předešlých kapitol tohoto článku zapadá primárně do prvního kontextu; na tomto místě však nastíním otázky spojené s kontextem odpovědnosti.

Kurki uvádí „andělský ideál“, kdy mají tvůrci umělé inteligence povinnost ji naprogramovat tak, aby dodržovala určitá předem stanovená pravidla, nebo případně tak, aby vždy dodržovala právní povinnosti, které jsou jí ukládány.¹⁰⁹ V takovém případě bychom problematiku její odpovědnosti nemuseli řešit. Uvědomuje si ale, že snaha o tento ideál je komplikovaná.

Zmiňuje situaci, kdy tvůrce odmítne tento ideál dodržovat a sice vytvoří umělou inteligenci, která rozumí právním požadavkům, ale spíše je zahrne do celkového kalkulu.¹¹⁰ Potenciální právní důsledky tak nebude brát jako nadřazené, ale spíš jako jedno z mnoha vodítek, zda provést nějakou akci.¹¹¹ Kurki dále uvádí, že když vytvoříme umělou inteligenci tak, aby byla schopná porušovat své právní povinnosti, tak občas můžeme dosáhnout společensky prospěšných důsledků.¹¹² Je tomu tak například v situaci, kdy vyhodnotí, že jednání v rozporu se smlouvou bude ekonomicky nejefektivnější pro všechny strany.¹¹³ Zároveň se umělá inteligence může občas

¹⁰² Pokud není uvedeno jinak, hlavním zdrojem v následujícím výčtu je KURKI, *A Theory of Legal Personhood*, str. 151.

¹⁰³ KURKI, *Legal Personhood and Animal Rights*, str. 53.

¹⁰⁴ *Ibidem*, str. 55.

¹⁰⁵ KURKI, *A Theory of Legal Personhood*, str. 96.

¹⁰⁶ *Ibidem*, str. 176.

¹⁰⁷ *Ibidem*.

¹⁰⁸ *Ibidem*.

¹⁰⁹ *Ibidem*, str. 180.

¹¹⁰ *Ibidem*.

¹¹¹ „Pokud by tedy cílem umělé inteligence bylo nashromáždit co největší bohatství, zvažila by – podobně jako ideální *Homo economicus* v ekonomické teorii – potenciální výhody a nevýhody určitého nezákonného jednání (např. insider trading) a jednoduše by zvolila možnost s největší střední hodnotou. Právní odpovědnost umělé inteligence by pak mohla být ospravedlněna na základě odstrašení, protože by snížila střední hodnotu nežádoucího chování.“ *Ibidem*.

¹¹² *Ibidem*.

¹¹³ „Vezměme si čtyřčlennou rodinu Smithových, která si rezervovala ubytování v malém hotelu, který má pouze tři čtyřlůžkové apartmány. Než se Smithovi přijdou ubytovat, přijde dvanáctičlenná rodina Frankových a požádá je, zda by si mohli rezervovat celý hotel, protože by rádi strávili noc na stejném místě. Hotel není schopen Smithovy kontaktovat, ale majitel se zeptá v menším hotelu naproti přes ulici, zda mají volný pokoj pro čtyři osoby. Odpověď je kladná, a tak jsou manželé Smithovi přemístěni do druhého podniku a dokonce dostanou 30% slevu. Vyhrájí všichni: Smithovi dostanou slevu, Frankovi mohou přespát pod jednou střechou a oba hotely získají více zákazníků. Bez ohledu na to byla porušena smluvní povinnost, protože Smithovi (respektive ten, kdo provedl rezervaci) s touto změnou nesouhlasili.“ *Ibidem*.

mýlit v obsahu práva – stejně jako se to stává lidem, kteří jsou následně sankcionováni.¹¹⁴ Umělá inteligence by v takovém případě tedy mohla být sankcionována také.¹¹⁵

Jenže pokud tedy přistoupíme na to, že by umělá inteligence měla být za porušení povinnosti sankcionovaná, jakou by taková sankce měla mít podobu? Jako hlavní nástroj Kurki uvádí ekonomické sankce; uvědomuje si však, že by pak umělá inteligence musela mít možnost vlastnit majetek.¹¹⁶ Je ale nutné, aby umělá inteligence vlastnila majetek? Neexistují nějaké další sankce? Kurki se domnívá, že existují. I kdyby totiž disponovala třeba jen čistě zatěžující právní osobností (tedy by měla status podobný otrokům), je možné ji sankcionovat jednoduše tím, že ji vypneme.¹¹⁷ Dá se to ale brát jako relevantní sankce, když se nejedná o živou bytost? Není nutný i nějaký zájem o sebe sama? Podle Kurkiho to relevantní sankcí je a není třeba ani nějaký sebezájem – umělá inteligence si totiž dokáže uvědomit, že by v případě vypnutí byly zmařeny její cíle, proto se takové sankci bude chtít vyhnout.¹¹⁸

Kurki se dále zaměřuje na otázku, zda by neměl být za jednání umělé inteligence odpovědný spíš ten, kdo ji naprogramoval. Zde však narážíme na problém toho, na jakých principech je umělá inteligence vlastně postavená. Nejedná se jenom o to, že ji někdo naprogramuje skrze sadu instrukcí. Jak jsem již rozebíral ve čtvrté kapitole, umělá inteligence se sama učí, je schopná se přizpůsobit prostředí a na základě různorodých vstupních dat tvořit vlastní výstup. Když pracujeme s velkým množstvím takových dat, tak je výsledek nepředvídatelný. Kurki upozorňuje, že tato značná míra autonomie umělé inteligence je pak komplikací pro to, abychom za její jednání činili odpovědným jejího tvůrce.¹¹⁹ Uvádí také, že právě díky zmíněnému procesu strojového učení jsou sankce dobře použitelné právě na samotnou umělou inteligenci, jelikož umělá inteligence bude sankce brát v úvahu při svém rozhodování.¹²⁰

Odpovědnost tvůrce umělé inteligence uznává Kurki v případech, kdy by úmyslně vytvořil umělou inteligenci cíleně pro to, aby skrze ni například páchal trestnou činnost.¹²¹

Jenže co když to bylo neúmyslně? Co když za takovým účelem vlastně vytvořená vůbec nebyla a začala ji páchat sama od sebe? Má tvůrce nést odpovědnost za její jednání, když z něj třeba profituje? Je obtížné říct, jak dlouho z něj bude profitovat a jak dlouho bude umělá inteligence svého tvůrce poslouchat. Kurki v tomto ohledu poznamenává, že umělá inteligence se sklonem porušovat normy může časem zjistit, že její tvůrce je překážkou k cílům.¹²² Ve výsledku tedy i samotný tvůrce může v budoucnu potřebovat to, aby byla umělá inteligence odpovědná za své jednání. Dávalo by tedy smysl tvůrce sankcionovat?

Kurki upozorňuje, že odpovědnost umělé inteligence by mohla sloužit stejnému účelu jako trestní odpovědnost otroků na Jihu USA před občanskou válkou (kteří měli, optikou balíkové teorie, čistě zatěžující právní osobnost). Trestní odpovědnost otroků tehdy totiž chránila jak majitele otroků, tak třetí osoby před zločiny otroků.¹²³

Debaty o právní odpovědnosti v souvislosti s umělou inteligencí mají dnes i praktickou podobu. Momentálně probíhají velké debaty například ohledně právní odpovědnosti za nehody způsobené samořízenými vozidly. Tyto debaty však nesměřují primárně na právní odpovědnost umělé inteligence, ale na právní odpovědnost řidiče, výrobce samotného auta, výrobce samotné technologie pro samořízení a výrobce konečné kombinace auta a samořídící technologie.¹²⁴ S filozofickým sporem o práva robotů a jejich právní osobnost se takové debaty tedy mívají.

Závěr

Nejprve jsem uvedl přístupy, které ohledně pojmu umělé inteligence rozlišují počítačovní vědci, následně jsem se zabýval filozofickým pojetím, které do filozofie vnesl John Searle. Ten rozlišoval dva přístupy – *weak AI*, který chápe umělou inteligenci jako užitečnou simulaci lidské mysli, a *strong AI*, podle kterého je umělá inteligence, která dokonale napodobí lidskou mysl, sama o sobě také myslí. Poté jsem uvedl obecně přijímaný praktický přístup k pojmu umělé inteligence, který nemá přesah do metafyzických otázek o lidské

¹¹⁴ Ibidem, str. 181.

¹¹⁵ Ibidem.

¹¹⁶ Ibidem.

¹¹⁷ Ibidem.

¹¹⁸ Ibidem.

¹¹⁹ Ibidem.

¹²⁰ Ibidem.

¹²¹ Ibidem.

¹²² Ibidem.

¹²³ Ibidem, str. 181–182.

¹²⁴ Viz TARPLEY, P. – JANSMA, S. D. *Autonomous vehicles: The legal landscape in the US*. 2016.

mysli. Tento přístup rozlišuje úzkou umělou inteligenci (schopnou plnit úlohy, které mu zadal člověk), umělou obecnou inteligenci (která se dokáže chovat nerozeznatelně od člověka) a umělou superinteligenci (která dokáže člověka předstihnout). Přesných definic je však mnoho, proto jsem uvedl, které dvě klíčové charakteristiky vždy naplňuje pojem umělá inteligence používaný ve vědeckých studiích – řešení velmi složitých úkolů a přizpůsobení se prostředí.

Následně jsem se zabýval pojmem robot, kdy jsem uvedl jeho původ ve sci-fi literatuře a odlišné podoby, kterých nabývá v různých pojetích. Zmínil jsem širokou definici Neila M. Richardse a Williama D. Smarta, kteří definují robota jako zkonstruovaný systém, který se po fyzické i mentální stránce jeví jako aktér, ale v biologickém smyslu není živý. Robot v tomto pojetí je kombinací umělé inteligence a nějakého fyzického projevu, přičemž výhodou této definice je, že hledí na vnější charakteristiky, nikoliv na to, co se děje uvnitř. To nám (stejně jako zmíněný praktický přístup k pojmu umělé inteligence) umožňuje s pojmem *robot* operovat nezávisle na metafyzických úvahách.

Poté jsem uvedl čtyři základní komponenty subjektivních práv, pro které se vžilo označení hohfeldovské incidenty. Mezi ně patří privilegium, nárok, pravomoc a imunita. Jednotlivé teorie subjektivních práv se pak snaží zjistit, jaké jsou mezi těmito incidenty vztahy a co mají společného. Zabýval jsem se poté dvěma základními teoriemi – volní a zájmovou. Volní teorie zdůrazňuje práva jako ochranu voleb a je často kritizovaná jako neinkluzivní vůči těm, kteří volby činit nemohou. K řešení problému s neinkluzivitou se nabízí zájmová teorie – ta zdůrazňuje práva jako ochranu zájmů, které mají silnou hodnotu.

Následně upozorňuji, že i když může být volní teorie vnímaná jako neinkluzivní vůči zvířatům, novorozencům či lidem ve vegetativním stavu, v případě práv robotů tomu může být jinak. Součástí umělé inteligence je totiž její schopnost se učit, díky čemuž v dnešní době zvládá spoustu věcí, které by dříve byly nepředstavitelné. V budoucnu je dokonce možné, že bude umělá inteligence vykonávat většinu manažerských rozhodnutí místo lidí. Pokud práva mají plnit funkci ochrany volby, pak by je měli mít i roboti. Vnímám však jako úskalí volní teorie, když s její pomocí můžeme dojít k závěru, podle kterého umělé entitě přisoudíme vyšší status než živé.

Dále jsem uvedl, že ve prospěch práv robotů se dá argumentovat i v duchu zájmové teorie. Stačí totiž zdůraznit zájmy lidí a vůbec se nemusíme zabývat reálnými schopnostmi robotů. Zmínil jsem paralelu se zákazem týrání zvířat, který má kromě zvířat samotných za cíl chránit společenský řád, stejně jako přísnější trestání vraždy novorozence oproti vraždám dospělých. Pokud přijmeme silnější právní ochranu robotů kvůli tomu, že taková ochrana poslouží lidem, aby se jim ve společnosti lépe fungovalo, pak ale pokládám za problematické takovou úpravu nazývat právy robotů. Kromě toho, že nemají žádné zájmy, nejsou oproti lidem a zvířatům ani živí, natož aby cítili bolest.

Jak je vidět, obě teorie – jak volní, tak zájmová – mohou sloužit jako základ pro další argumentaci v oblasti práv robotů. Domnívám se tedy, že stojí za to se tématu dále věnovat a pokoušet se teorie aplikovat na nové výzvy, které nám přináší technologický rozvoj. Ne nutně proto, abychom jimi argumentovali ve prospěch práv robotů, ale třeba i proto, abychom ukázali, že jejich aplikací docházíme k pochybným závěrům, a tím je zároveň problematizovali.