UNIVERZITA KARLOVA

Právnická fakulta

Vratislav Slavík

Hranice lidství. Roboti se nebezpečně přibližují lidem, nejrůznější skupiny lidí naopak máte tendenci nepovažovat za (plnohodnotné) lidi

Od gynoida Sophie k aktu o umělé inteligenci

Diplomová práce

Vedoucí diplomové práce: Mgr. Petr Agha , LL.M., Ph.D.

Katedra: Katedra politologie a sociologie

Datum vypracování práce (uzavření rukopisu) : 7. 6. 2023

Prohlašuji, že jsem předkládanou diplomovou práci vypracoval samostatně, že všechny použité zdroje byly řádně uvedeny a že práce nebyla využita k získání jiného nebo stejného titulu.

Dále prohlašuji, že vlastní text této práce včetně poznámek pod čarou má 127 362 znaků včetně mezer.

Vratislav Slavík

V Praze dne 7. 6. 2023

Obsah

1. Úvodem: umělá inteligence jako transformační síla společnosti

Podoba současného světa a stav společností na začátku dvacátých let 21.století je utvářen především kombinací třech významných fenoménů, a to (1) technologickou změnou (revolucí), (2) globálním klimatickým rozvratem a (3) krizí demokracie (Barroso, 2020). Vedle toho je trvalým a charakteristickým rysem dnešní doby nekončící řada krizí, které ovlivňují to, jak je současný svět zažíván.

Na velmi obecné úrovni je tématem této práce vztah jedné z technologií, která se vedle jiných technologií (např. nanotechnologie, biotechnologie, informační technologie) zcela zásadně podílí na hluboké přeměně nastavení a fungování stávajícího světa, a to umělé inteligence, a práva. Vzhledem k tomu, že právo je součást systému kultury, který je opozitní k systému přírody, a představuje tak jeden z jeho subsystémů (Šmajs, 2008), není podle některých autorů (Black a Murray, 2019, Chesterman 2021) možné, aby bylo ke stávajícím dopadům, stejně jako k budoucím rizikům, které sebou integrace umělé inteligence do společností přináší, zcela netečné, což se podle nich po určitou dobu a v určité míře dělo.

Diskuze 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 diskuzi je vztah práva a současných technologií (včetně umělé inteligence) často pojímán jako závod. Závod, ve kterém je právo vykreslováno jako závodník, který je vždy pomalejší a který zaostává za technologiemi. V této roli se právo z pohledu těchto autorů (Alenby 2011, Moses a Zalniurite 2020) ocitá zejména z toho důvodu, že ve vztahu k technologiím selhává v naplňování svého primárního účelu, který je spatřován v poskytování včasných a dostatečně efektivních opatřeních pro usměrňování vývoje technologií takovým směrem, který nebude představovat riziko a hrozbu pro stabilitu a existenci společnosti.

Ve spojitosti s tou diskuzí (pacing problem, challenge of regulatory connection) se pak debatuje o načasování regulatorních zásahů a hledá se nejvhodnější moment pro zavedení nových zákonů, které zakazují, usměrňují nebo podporují nově se objevující technologie. V tomto kontextu se odkazuje ke konceptu Collingridgova dilematu (Collingridge dilemma), v němž je problém kontroly nad technologiemi a regulace technologií zachycen. Velmi zjednodušeně Collingridge tvrdil, že společenská kontrola technologie je obtížná, protože pokusy o kontrolu technologie v rané fázi jejího vývoje trpí tím, že neznáme její konečnou podobu, všechny její možné dopady a využití (problém se vstupní informací), což ve výsledku může mimo jiné vést k volbě nevhodného způsobu regulace a tím pádem k jejímu úplnému potlačení.

Na druhé straně se stávají pokusy o regulaci a kontrolu technologie, co se široce prosadí a proroste společností, prakticky nemožné (dopady technologie jsou známé, využití také), a to z důvodu, že změny nutné k dosažení souladu s novou regulací jsou časově i finančně velmi nákladné (Moses a Zalnieriute 2020).

Z celkového pohledu na problematiku regulace nových technologií, a to včetně regulace umělé inteligence, lze říci, že představuje komplexní téma, které vždy vyžaduje zvážení potenciálních rizik a přínosu konkrétní technologie, jakož i různorodých perspektiv a zájmů, které zastávají zúčastněné strany v procesu vyjednávání o způsobech a podobách regulace. Označení zúčastněné strany zahrnuje především tři hlavní skupiny: (1) regulátor (vláda, zákonodárce), (2) soukromé korporace a (3) jednotlivce (Black a Murray 2019). Vyjádření, které zřejmě nejlépe ilustruje hlavní zájem a celkově názor na regulaci umělé inteligence ze strany skupiny soukromých korporací a její podobu, pronesl generální ředitel nejmocnější korporace světa (Google, Alphabet) Larry Page v roce 2013 při příležitosti každoroční vývojářské konference Google I/O a týkalo se technologické změny a hledání odpovědi na otázku, jaké mechanismy může společnost zapojit k vyrovnání se s touto změnou. Odpověď zněla, že staré instituce jako právo a podobně nedrží krok s rychlostí změn, která byly způsobeny prostřednictvím technologií. „Když se podíváte na zákony, které jsme vytvořili, jsou velmi staré. Zákony, na základě nichž jsme se stali veřejnou akciovou společností, byly staré 50 let. Zákon nemůže být správný, když je 50 let starý, a to proto, že je z doby před internetem."

Dále v rozhovoru sdělil, že inovace a vývoj technologií jsou brzděny tím, že některé prospěšné věci, které by mohly být udělány, jsou nelegální a jsou předmětem regulace. Jako řešení pro omezení negativ, které brání zrychlování technologických změn, vidí vytvoření míst oddělených od zbytku světa, kde neplatí žádné zákony ani regulace a kde mohou technici testovat nové nápady a zkoumat, jaký mají vliv na společnost a jednotlivce. Ústředním tématem v příběhu o regulaci technologických nadnárodních korporací stojících za rozvojem umělé inteligence se stalo, že regulace je vždy negativní silou, která brání inovacím a pokroku, a území kde neplatí zákony (bezprávnost) je pro technologické inovace určitým způsobem nezbytnou podmínkou.

2. Modernita, internet a gynoid Sophia a kontext současné debaty o umělé inteligenci

S takto 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 byla z pohledu technologií postavena v první fázi okolo parního stroje a železnice, následovaných spalovacím motorem a ropou (Alenby, 2011). Hlavní postavy průmyslu (robber barons) této doby, vybaveni myšlenkami Herberta Spencera o funkci státu (kterými měly být převážně ochrana práv individuí a zajištění kolektivní obrany), doktrínou o přežití nejschopnějších vycházející z Thomase Malthuse, 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 jádrové technologie související s modernitou, pára a kolej, spalovací motor a ropa se společnost vyrovnává dodnes. Tyto problémy ohrožují samotnou možnost existence lidstva na této planetě, stejně jako ukazují příčiny stojící za tím, co za touto katastrofou stojí, což je především logika systému kapitalismu založeného na nekonečnem růstu, stejně jako primární logika fungování soukromých korporací spočívající v neustálém zvyšování zisku, a na individuální úrovni pak prostá touha po moci a obrana práv kapitálu (Zuboff, 2019).

Druhým bodem v rámci historie technologií, se kterými nás názory na jejich regulaci uvedené výše propojují, je období poloviny 90. let dvacátého století, konkrétně pak diskuze o regulaci internetu, která v tomto období probíhala. Toto téma bylo rámováno obecnou debatou o regulaci, která se odehrávala uvnitř akademického právního pole, a která byla primárně zaměřena na kritiku podoby regulace a její reformu. Hlavní proud akademického právního myšlení té doby asimiloval ekonomické myšlenky Hayeka a Friedmana a neoliberální ideologie se tak stala optikou, kterou bylo nahlíženo na regulaci a regulační úlohu státu. Jak ukazuje Short (2021) ve studii, která analyzovala 1 400 článků z právních časopisů na téma regulace, které vyšly mezi lety 1980 do 2005, byl diskurz vystavěný na hlubokých obavách z donucovací moci státu (coercive state) a na spojení tématu regulace s pojmy jako je tyranie a autoritářství. Současně s tím převládl názor, že nejúčinnějším 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

umělé inteligence, a setkáváme se s ní do současnosti, kdy plynule přešla do debaty o regulaci AI a stojí za explozí dokumentů obsahujících etické pokyny pro vývoj umělé inteligence.

V roce 1996 publikuje John Barlow Deklaraci nezávislosti kyberprostoru, v které mimo jiné zastával názor, že digitální prostor je nový svět oddělený a rozdílný od světa hmotného, na nějž se neuplatňují tradiční právní koncepty jako například vlastnictví a odmítá autoritu státní moci včetně možnosti státní moc v tomto prostoru uplatňovat. V případě potřeby řešení sporů (křivd), které v tomto prostoru vzniknou, uvádí, že digitální komunita kyberprostoru je bude řešit vlastními prostředky a na základě vlastních pravidel, na jejichž vytváření budou participovat všechny bytosti kyberprostoru (a to včetně pravidel, pro identifikaci toho, co spor je nebo není). „Vytvoříme civilizaci Mysli v Kyberprostoru. Nechť je lidštější a spravedlivější než světy, které dosud vyrobily vaše vlády“, zakončuje deklaraci Barlow.

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 (De Gregorio, 2021). Digitální technologie představovaly v této době pro státy příležitost k ekonomickému růstu a hospodářské prosperitě, a není proto překvapivé, že kompromisy mezi inovacemi a například ochranou základních práv jednotlivců byly ze strany států vedeny ve prospěch nezasahování do tohoto prostředí. Způsob, jakým státy přistoupily k řešení problému regulace internetu, spočíval v tom, že se zaměřily na řešení dílčích problémů, a to například porušování autorských práv či zneužívání osobních údajů, avšak koherentní strategie regulace tohoto prostředí a rizik s ním spojených se nepodařilo dosáhnout. Takto nastavený systém regulace a kontroly digitálního prostředí vyprodukoval 5 hlavních gatekeeperů (Google, Facebook, Amazon, Microsoft a Apple), kteří ovládají to, jak v současnosti vypadá podoba našeho života v tomto prostředí a jak je toto prostředí zažíváno, a dále umožnil novou koncentraci moci. Následně se také stane jedním z klíčových faktorů, který ovlivní rychlost rozvoje strojového učení a umělé inteligence (Zuboff, 2019).

V roce 2016 proběhly tři události, na které je velice často odkazováno v literatuře, věnující se tématu umělé inteligence (včetně literatury právní), a které měly zásadní podíl na tom, že umělá inteligence vstoupila do širšího povědomí, a s ní spojené otázky se staly tématem veřejné debaty (v době psaní této práce by bylo možné přidat ještě další a to ChatGPT). První z nich je uvedení chatbota s názvem Tay (akronym thinking about you) prostřednictvím Twitteru společností Microsoft, který byl po 16 hodinách působení na této sociální sítí odstaven, a to z toho důvodu, že v interakci s ostatními uživateli začal zveřejňovat zprávy (tweety) které byly rasistické, sexuálně urážlivé a popíraly holocaust. Toto chování bylo výsledkem napodobování záměrně urážlivého chování uživatelů Twitteru, ze kterého se chatbot učil. Tento příklad použití

je velice často citován ve spojitosti s tématem strojového učení a s možnými riziky, které se s ním pojí.

Druhou událostí bylo vítězství AlphaGo, umělé inteligence 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. Oproti dvacet let staré výhře počítače DeepBlue, vyvinutého společností IBM, v souboji s tehdejším šachovým velmistrem Garrym Kasparovem, se vítězství DeepMind lišilo v tom, že počítač nevyhrál díky tomu, že by se řídil pravidly, která mu zadali programátoři, ale díky strojovému učení založenému na analýze milionu předchozích zápasů ve hře Go, a dále díky tomu, že se počítač učil tím, že hrál sám proti sobě.

Třetí a poslední událostí z roku 2016 spojovanou s umělou inteligencí, která je zmiňována, je uvedení gynoida Sophie na veřejnost zakladatelem společnosti Hanson Robotics David Hansonem, které bylo počátkem globálního spektáklu, odehrávajícího se tři roky. Během těchto tří let vystoupila Sophie v rozhovorech ve všech předních světových médiích (kladené otázky vypadaly namátkově takto: Chcete zničit lidstvo? Jaký je váš názor na feminismus a co si myslíte o genderových otázkách?), objevila se na přebalu módního časopisu Elle a vystoupila jako mluvčí na mnoha ekonomických a politických akciích, včetně vystoupení na konferenci OSN o udržitelném rozvoji v roce 2018. Po celou dobu společnost Hanson představovala Sophii jako budoucnost umělé inteligence a jako krok na cestě k superinteligenci. Média, která nekriticky přejímala tato slova, přispěla prostřednictvím produkování zpráv, které byly mylné v pojetí schopností gynoida Sophie k vytvoření dojmu, že humanoidní roboti řízení obecnou umělou inteligencí budou zakrátko součástí každodenního života, který se zabydlel v povědomí veřejnosti. Gynoid Sophia přitom byla dle vědců z oboru umělé inteligence, kteří se zabývali jejími schopnostmi, ničím víc než chatbotem, případně ji popisovali jako sofistikovanou loutku, a psali o ní jako o všestranném podvodu. Sophia tak měla mnohem blíže k mechanickému Turkovi, k chytrému Hansovi či ELIZE, než k všestrannému robotovi/gynoidovi s vtělenými kognitivními funkcemi (Parviainen a Coeckelbergh, 2021).

Vyvrcholením světové tour Sophie byla návštěva ekonomického fóra Future Investment Initiative v Saudské Arábii, které pořádal korunní princ Saúdské Arábie Muhammad bin Salman, který, jak se již v té době vědělo, stál za vraždou novináře Jamala Khashogga, a jímž bylo Sophii uděleno saudské občanství. Zde se zcela jasně ukázala Sophie jako to, co byla od počátku, a to především jako nástroj pro přitáhnutí investic do oboru umělé inteligence a robotiky. Udělením občanství Sophii získala v akademickém poli novou dynamiku polemika o tom, zda by roboti a umělá inteligence měli získat práva, případně jaká práva, a jaké by mělo být z pohledu práva

jejich postavení. Stejně tak se oživila filozofická debata o tom, co to znamená být člověkem. Svůj otisk zanechala Sophie i na úrovni politiky Evropské unie, a to v usnesení Evropského parlamentu, které obsahovalo doporučení Evropské komisi zavést a upravit statut elektronické osoby (vytvoření zvláštního právního statutu robota).

2.1. Okruh problémů spojený s umělou inteligencí inspirovaný science fiction

Tyto tři výše uvedené události směřují především k základnímu okruhu problémů a otázek, které se staly stálou součástí uvažování o technologiích, jako je umělá inteligence 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 narativech, které zachycují představy o vztahu mezi člověkem a jeho výtvorem, který vykazuje známky vlastního myšlení. Tyto motivy nacházíme například v románu Frankenstein Marry Shelley, v legendě o Golemovi, v povídkách Issaca Asimova (např. Konflikt nikoli nevyhnutelný) a samozřejmě v dramatu R.U.R Karla Čapka, úryvky z něhož jsou s oblibou citovány v úvodech vědeckých prací o problematice AI a robotiky.

V současnosti posunuli myšlenku o tom, že umělá inteligence bude posledním vynálezem lidstva, který povede k jeho konci, do popředí veřejné debaty o umělé inteligenci překvapivě vlivné osoby stojící za vývojem této technologie jako Elon Musk, Ray Kurzweil, Stuart Russell, Nick Bostrom, které spolu s ní vnesly do debaty pojmy jako superinteligence, technologická singularita, transhumanismus. Koncept superinteligence předpokládá, že se vědcům podaří vyvinout tzv. obecnou umělou inteligenci, neboli inteligenci, která se vyrovná inteligenci člověka a na tomto základě se vytvoří její další úroveň, která lidskou převýší. Tato umělá inteligence přinese dramatickou změnu, kdy už nedokážeme pochopit, co se děje a život tak, jak ho známe, skončí.

Zápletka se po jejím dosažení u každého autora vyvíjí různě; u Kurzweila nakonec směřuje k bodu, který označuje jako singularitu, kdy problém konkurence je vyřešen tak, že se lidská inteligence spojí s umělou inteligenci, lidé díky tomu překonají omezení svých biologických těl a nastane tak zlatý věk lidstva. V druhém případě u Bostroma vede zápletka po dosažení superinteligence k tomu, že umělá inteligence převezme kontrolu nad lidstvem a bude plně rozhodovat o jeho osudu. Jak však upozorňují někteří autoři, tato diskuse o vzdálených budoucích dopadech umělé inteligence odvádí pozornost od skutečných a aktuálních rizik, která vyplývají ze systémů umělé inteligence, které jsou do společnosti nasazovány v současnosti. Pozornost by tak měla být věnována riziku, že nebudeme dostatečně rozumět jejich etickým a

společenským důsledkům, ale přesto se je budeme snažit široce zavádět do společnosti, spíše než otázce, co se v budoucnu skutečně stane (Crawford a Calo 2016).

2.2. Reflexe umělé inteligence v právním výzkumu

Z pohledu právního výzkumu zaznamenalo téma umělé inteligence v roce 2016 zřetelný rozmach. V rámci právní literatury je téma přítomné od doby konsolidace umělé inteligence jako oblasti výzkumu a kopíruje její vývoj. Událostí, která je obecně považována za zrod vědeckého pole se zaměřením na výzkum a vývoj umělé inteligence je konference v Dartmouthu, která proběhla v roce 1956 a stejně jako myšlenka internetu je financována agenturou amerického ministerstva obrany, které je zodpovědné za vývoj nových vojenských technologií (DARPA). V následujících letech do roku 1974 je výzkum umělé inteligence spojen s velkým optimismem, ustaluje se koncept umělé inteligence 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 umělé inteligence podobné té lidské. K těmto problémům se připojila i kritika a obavy ohledně cílů projektu vývoje umělé inteligence. To vedlo k šest let trvajícímu období, které je nazýváno první zimou umělé inteligence, jenž se projevilo omezením finanční podpory výzkumných projektů umělé inteligence a celkovou ztrátou zájmu o toto téma (Zuboff, 2019).

Perspektivou právního výzkumu v oblasti umělé inteligence v období 1956-1974 byla hlavním tématem charakteristika procesu řešení právního problému a právní argumentace, a možné způsoby zapojení umělé inteligence v těchto procesech (jinými slovy možností nahrazovaní člověka v právnických profesích umělou inteligencí). 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 k obnovení financování výzkumu umělé inteligence. Po kritice z předešlých let vědci přichází s novým způsobem, jak přistoupit k umělé inteligenci. Došlo k odklonu od původních cílů a velkých abstraktních konceptů a nově se výzkum zaměřuje na vývoj znalostních (expertních) systémů. Na konci 80. let znovu ochabuje zájem investorů o oblast umělé inteligence, protože se ne zcela daří přechod od vývoje technologie k širokému přijetí systémů veřejností a naplňování ekonomických cílů. V tomto období nastává druhá zima umělé inteligence (Becerra, 2018).

V hranicích právního pole, zejména jeho akademické části, patří mezi témata diskutovaná autory, kteří se věnují umělé inteligenci v těchto letech (1980-1987), na jedné straně úvahy o možnostech zapojení umělé inteligence v právních profesích, možnosti zpracování informací a vyhledávání textů, které bylo možné nalézt už v předešlém období. Na druhé straně se objevují

otázky nové, a to zda by se umělá inteligence měla používat, a pokud dojde k jejímu používání, jaká by měla být odezva ze strany práva.

Třetí vlna vzestupu výzkumu umělé inteligence začíná počátkem 90. let dvacátého století a trvá do současnosti. Toto období je charakterizováno výzkumem zaměřeným na strojové učení a odklonem od znalostních systémů. Dochází k rozvoji metod strojového učení a nejrozšířenějšími metodami, se stávají učení s dohledem (supervised), bez dohledu (unsupervised) a zpětnovazební (reinforcement). Po roce 2010 se do popředí výzkumu a aplikace umělé inteligence dostávají neuronové sítě a hluboké učení. Z pohledu právního pole a jeho propojení s tématem umělé inteligence dochází v této poslední vlně, a zvláště pak od roku 2016, k významnému nárůstu literatury věnované tomuto tématu. Výzkum, jenž byl na počátku úzce zaměřený na popsání a pochopení mentálních procesů používaných při výkonu právních profesí, jejich formalizaci a aplikaci umělé inteligence při řešení praktických právních problémů, se mění na méně specializovaný a pokrývá více obecných témat (Becerra, 2018, Goanta et al., 2020).

V rámci témat právního výzkumu lze identifikovat tři hlavní okruhy, kterým je věnována pozornost a to: (1) dopady umělé inteligence na základní koncepty, na nichž je právo postaveno, jako jsou například právní subjektivita, princip nediskriminace, odpovědnosti autonomních entit, základních lidských práv; (2) okruh věnující se oblastem využití umělé inteligence a identifikaci potenciálních rizik a příležitostí, a (3) třetí okruh, který se soustředí okolo témat navrhování modelů regulace a správy umělé inteligence a návrhů možných podob právní regulace (Goanta et al., 2020).

Tento obecný úvod se snažil o alespoň částečně zasazení tématu umělé inteligence do širšího kontextuálního rámce, včetně nastínění stručné historie umělé inteligence a překryvům s vývojem témat na poli akademického právního výzkumu umělé inteligence. V následující kapitole se práce bude věnovat již konkrétním tématům, kterými jsou definice umělé inteligence a současné regulační snahy evropských institucí na tomto poli. V následující kapitole se práce věnuje popisu základní architektury, fungování umělé inteligence a definicím umělé inteligence. Tato kapitola by měla především upozornit na místa v těchto systémech, která jsou z pohledu právní literatury označovaná za riziková a zobrazit základní okruh problémů, které právo s umělou inteligenci spojuje. Takto by měl být vymezen prostor k tématu třetí kapitoly, která se věnuje jedné z hlavních regulačních iniciativ týkajících se umělé inteligence, a konkrétně rozebírá návrh aktu o umělé inteligenci z roku 2021, kterým se stanoví harmonizovaná pravidla pro umělou inteligenci.

3. Vývoj a možné varianty definice umělé inteligence pro účely regulace

Cílem této kapitoly je popsat způsoby, jakými je v současnosti definována umělá inteligence, stejně jako představit hlavní body debaty, která se o pojmu umělé inteligence vede na poli právní vědy. Na úrovni teoretické právní literatury je tato snaha o konceptuální uchopení umělé inteligence vedena především za účelem dosažení pochopení a objasnění toho, proč umělou inteligenci regulovat. Hned v úvodu této kapitoly je vhodné říct, že to, na čem se všichni odborníci (vědečtí, političtí), jejichž předmětem zájmu je umělá inteligence, shodnou, je skutečnost, že neexistuje jednoznačná a široce přijímaná definice umělé inteligence - ať už na poli výzkumu a vývoje umělé inteligence, nebo taková, která je obecně použitelná pro účely regulace. Přitom jak se shodují někteří autoři (Martinez, 2019, Scherer, 2016), vhodně zvolená definice je jednou z cest vedoucích k efektivní a přesně zacílené regulaci.

K umělé inteligenci lze přistoupit tak, že umělá inteligence je to, na čem se určitá skupina shodne, a to v určitém kontextu a pro určitý účel, což současně znamená, že pojem umělé inteligence 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 umělé inteligenci jako o vědním oboru, technologii či metodě, nebo konkrétní aplikaci systémů umělé inteligence. Dvěma hlavními aktéry z pohledu podílu na utváření obsahu pojmu umělé inteligence jsou skupina vědeckých pracovníků z oboru umělé inteligence, a skupina politických aktérů vytvářejících podobu právní regulace, přičemž nesoulad mezi těmito skupinami v pojímání umělé inteligence může bránit adekvátní odezvě ze strany práva na dopady spojené s používáním systému umělé inteligence. Z hlediska regulace a obsahu politiky je proto důležité zahrnout do pojímání umělé inteligence jednak ty aspekty technologie, které vyžadují právní zásahy, a zároveň současné a budoucí technické možnosti umělé inteligence (Krafft et al., 2020).

3.1. Umělá inteligence jako vědní obor

Umělá inteligence jako označení pro vědní obor se pojí s letní školou v Dartmouth organizovanou Johnem McCarthym v roce 1956. Stručná historie vývoje umělé inteligence jako vědního oboru byla popsaná již v úvodní kapitole, a proto se zde text zastavuje pouze u toho, že tato událost je uvedena vymezením hlavního předmětu a cíle zkoumání nově vznikajícího vědního oboru. V souvislosti s touto událostí je zveřejněna i jedna z prvních definic, v současnosti označovaná za klasickou, umělé inteligence, která je zkonstruována na základě uceleného popisu této technologie (Coeckelbergh, 2020).

Základním předpokladem, ze kterého výzkumný projekt letní školy vycházel byl, že každý aspekt učení nebo jakýkoliv jiný rys inteligence lze v zásadě popsat tak přesně, že lze vytvořit stroj, který jej bude schopen simulovat. Úsilí mělo být směřována k řešení problému umělé inteligence, za který se považuje přimět stroj, aby se choval způsobem, který by byl označen za inteligentní, pokud by se tak choval člověk. Problematičnost této definice spočívá v mnoha důvodech, a to například v tom, jestli je opravdu možné redukovat inteligenci a mysl člověka na pouhý sled výpočetních procesů, nebo například, jak ukazuje Floridi (2021), pokud myčka nádobí myje nádobí stejně jako člověk, nebo dokonce lépe, neznamená to, že je nutně inteligentní.

V následujících letech dochází k mnoha dalším pokusům o definování pojmu umělé inteligence. Navzdory tomu se nepodařilo v rámci výzkumného pole dosáhnout konsenzu na jednotné podobě definice, což však podle některých autorů pomohlo tomuto oboru růst a široce se rozvinout. Jak popisují autoři Russell a Norvig (2016) ve své klasické učebnici Artificial Intelligence: A Modern Approach, která je považována za základní a vstupní text do studia oboru umělé inteligence, 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 definicí jsou: a) definice zaměřené na lidské myšlení, umělou inteligencí je tak 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, umělou inteligencí je systém, který napodobuje způsob chování člověka při plnění konkrétního úkolu nebo v konkrétní situaci (což je v podstatě charakteristika Turingova testu). c) Třetí možnou podobou definic umělé inteligence jsou definice postavené na znaku racionálního myšlení. Umělá inteligence je systém, který zpracovává informace a uvažuje na základě matematických pravidel (jako například deduktivní logiky a pravidel odvozování). d) Poslední forma definic umělé inteligence se opírá o charakteristiku, kterou je racionální chování. Z tohoto pohledu je umělou inteligencí takový systém, který optimálním způsobem řeší specifikované úkoly nebo dosahuje naprogramovaných cílů. Samotný způsob chování systému tak může být neobvyklý nebo případně člověku neznámý, avšak efektivní z pohledu plnění určitého úkolu.

3.2. Definice použitelná v rámci právní regulace, charakteristiky a rizika

Jak je patrné z výše uvedeného, obor umělé inteligence již od počátku svého vzniku usiloval o vývoj a naplnění formy obecné (silné) umělé inteligence (general artificial

intelligence). Tato forma technologie umělé inteligence 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 malou částí výzkumu a aplikací umělé inteligence. Většina úsilí je spíše zaměřena na vývoj systémů spadajících pod úzkou (slabou) umělou inteligenci, která slouží k řešení úzce vymezených úkolů prostřednictvím dosažení určitého cíle nebo souboru cílů (Krafft et al., 2022, Scherer, 2019).

Ačkoli je vývoj úzké umělé inteligence velice vzdálený od složitého a komplexního úkolu vytvoření obecné umělé inteligence, zaměření se na její vývoj vedlo k významným pokrokům v rámci vědního oboru umělé inteligence. A právě použití této formy umělé inteligence má zásadní dopad na každodenní svět a je hlavní výzvou pro právo a regulaci. Některé z těchto systémů umělé inteligence dosahují cíle způsobem, který je silně ovlivňován lidskými vstupy, zatímco jiné vyžadují minimální vstupy člověka k dosažení předem definovaného cíle, protože se učí a aktualizují svůj rozhodovací model. Co přesně se učí a jakého cíle dosahují, závisí předně na konkrétním prostředí, v němž jsou tyto systémy umělé inteligence použity (König et al., 2022).

Výzkum umělé inteligence v současnosti zahrnuje sedm dílčích oblastí. Většina pozornosti je v současné době věnována technikám získávání informací z dat prostřednictvím strojového učení, které se stalo dominantní disciplínou. Strojové učení označuje širokou škálu metod, mezi které se řadí např. rozhodovací stromy, neuronové sítě, bayesovské sítě a genetické algoritmy nebo podpůrné vektory (Häuselmann, 2022).

Různé definice umělé inteligence používané v technické literatuře, která je spojena s výzkumem umělé inteligence, mohou být užitečné pro pochopení jejího fungování, ale jak zmiňují někteří autoři (Schuett, 2023), 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 umělá inteligence.

Z pohledu práva a regulace souvisí téma definice a vymezení umělé inteligence s úsilím, které se soustředí na hledání základních charakteristik systémů umělé inteligence a na problém identifikace hlavních zdrojů rizik, které jsou s těmito systémy spojeny. Tyto zdroje rizik lze rozčlenit do tří základních kategorii, přičemž názvy těchto kategorií se pojí s konkrétní odpovědí na obecnou otázku co je zdrojem rizika systému založených na umělé inteligenci? Rizika lze rozčlenit na rizika plynoucí z 1) technického přístup („jak je systém umělé inteligence navržen“), což jsou rizika plynoucí z technického přístupu k tvorbě umělé inteligence, (2) aplikace („k čemu

se technologie používá“), což značí rizika plynoucí z použití, účelu a kontextu, a (3) schopnosti („co technologie dokáže“), tedy rizika plynoucí ze schopností umělé inteligence. Tyto tři kategorie následně tvoří základní prvky pro konstrukci právních definic a představují možný klíč k regulaci umělé inteligence (Schuett, 2023).

3.2.1 Strojové učení jako potenciální zdroj rizika

Kategorie označená jako technický přístup spočívá v tom, že určité techniky a metody používané k výstavbě systémů umělé inteligence jsou ze své podstaty rizikové. Touto rizikovou technikou je myšleno především strojové učení. Z pohledu současných systémů umělé inteligence 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 jednotlivec, 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 jednotlivec podroben, jsou přijatelná v demokratické společnosti, či posouzením, zda naplňuje všechny zákonné podmínky pro jeho učinění (Buiten, 2021).

V literatuře (Chesterman, 2021) je pak možné rozlišit dva postoje k řešení problému neprůhlednosti systémů umělé inteligence. Pro první přístup je charakteristické, že cestu k řešení tohoto problému 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 umělé inteligence chová. Jinak řečeno, cílem je dosažení obecného pochopení rozhodovacího modelu prostřednictvím ex post modelování a rekonstrukce chování systému. Tento přístup sice může sloužit k nalezení pravidelností v chování systému

umělé inteligence, 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 na požadavky práva na transparentní rozhodovací proces, ani neposkytuje z hlediska práva vhodný základ pro účely vysvětlení rozhodnutí (König et al., 2022)

Na rozdíl od prvního přístupu, který klade důraz na technickou stránku problému a popis vnitřního fungování systému, a to 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 (Surden, 2019).

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 umělé inteligence týká, vybavit především takovou informací, na jejímž základě bude schopna se právně relevantním způsobem bránit (Buiten, 2019).

Přítomnost součástí, které umožňují systémům umělé inteligence se učit, a které mohou měnit chování systému na základě zpracovávaných vstupních dat, může těmto systémům propůjčovat určitou míru nezávislosti a také nepředvídatelnosti. Systém založený na umělé inteligenci tak může dospět k chování, které bylo od počátku nezamýšlené a zcela nepředvídatelné tím, kdo systém vyvinul. Jednou z charakteristických vlastností současných systémů umělé inteligence, která je v literatuře popisována (Chesterman, 2021), je schopnost fungovat bez lidského zásahu a běžně se tak lze setkat s tím, že fungování těchto systémů je označované za autonomní. To však neznamená, že by tyto systémy měly svobodnou vůli, rozum a byly schopny sebeuvědomění. Na téma autonomie umělé inteligence lze nahlížet tak, že je to právě tato vlastnost, která činí tuto technologii výjimečnou ve srovnání s předešlými technologiemi, a která s sebou přináší řadu výhod především v podobě toho, že stroje založené na umělé inteligenci jsou na sebe schopny brát stále větší množství úkolů. Vzhledem k tomu, že umělá inteligence nemusí být vázána určitými způsoby myšlení, které jsou lidem vlastní, stejně jako není omezována schopnostmi mozku, přichází často s neobvyklými způsoby řešení problémů. Atraktivita těchto systémů spočívá nejen v tom, že jsou v současnosti schopny nacházet řešení problémů, které se člověku nedařilo, ale že do budoucnosti budou schopny přicházet s řešením problémů, které si člověk neuvědomuje.

Současně s tím, jak se mění rozložení práce a poměr rozhodnutí, která jsou vykonávána umělou inteligenci 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 umělé inteligence, kdy tento systém je schopen přijímat rozhodnutí, která nelze zpětně přímo připsat vývojáři ani provozovateli systému. Druhým charakteristickým tématem provázaným s problematikou autonomie systémů umělé inteligence je téma udělení právní subjektivity a přiznání práv a povinností těmto technickým artefaktům, přičemž 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, a odpovědným by tak byl samotný systém umělé inteligence (Schuett, 2023).

3.2.2 Schopnosti umělé inteligence, rozhodovací proces a problém zkreslení

Dalším souvisejícím tématem se strojovým učením je vedle tématu autonomie a vysvětlení rozhodnutí, kvalita rozhodnutí a problém zkreslení (bias), kterým může být rozhodovací proces ovlivněn. Z pohledu rizik lze toto riziko zařadit do kategorie rizik vyplývajících ze schopností umělé inteligence. V tomto případě se jedná o schopnost těchto systémů kategorizovat, klasifikovat a hierarchizovat, kdy tyto schopnosti a postupy se vztahují k člověku. Základem, na kterém jsou postaveny současné systémy strojového učení jsou tréninková data. Tyto vstupní soubory dat utvářejí hranice, jimiž se řídí fungování umělé inteligence, a také hranice toho, jak může umělá inteligence vidět svět, tedy prostředí, v němž je nasazena (Buiten, 2019, Devillé et al., 2021).

To z hlediska kvality jakéhokoliv rozhodnutí systému umělé inteligence znamená, že jeho kvalita se primárně odvíjí od kvality souboru dat, který byly použity k jeho trénování. Zájem práva o téma dat je dán zejména tím, že zkreslený soubor dat může vést k takovým rozhodnutím umělé inteligence, 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 umělé inteligence 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 umělá inteligence reprodukovat ve svých rozhodnutích existující vzorce diskriminace, které se vyskytují ve společnosti (König et al. 2022, Scherer 2016).

3.2.3 Oblast použití a účel použití umělé inteligence jako možný zdroj rizika

Další kategorií je kategorie věnující se rizikům, která jsou spojena s konkrétním používáním systémů založených na umělé inteligenci, a která se opírá o to, že dopady daného systému umělé inteligence vždy vyplývají z širšího kontextu, v němž je systém implementován. Technická specifikace systému umělé inteligence tak v tomto případě sehrává pouze podružnou roli a nepředurčuje ani dopad, který systém bude mít, ani rizika, která s sebou jeho implementace přinese (Schuett, 2023).

Systémy umělé inteligence jsou konkrétním technologickým řešením daného problému, které může být v jedněch podmínkách zcela neškodné, případně schopné přinášet značný užitek, ale v případě přenesení z tohoto prostředí do jiného se může stát, že v novém prostředí bude mít radikálně odlišné sociální důsledky, a bude způsobovat z hlediska regulace hluboké problémy. Například systém pro rozpoznávání tváře, bude-li použit jako řešení pro odemknutí chytrého telefonu lze posoudit jako neškodný, jinak tomu však bude v případě použití tohoto systému jako nástroje všudypřítomného dohledu státu nad svými občany. Z tohoto pohledu tak nelze systémy umělé inteligence správně pochopit a posoudit bez zohlednění konkrétního kontextu, v němž jsou nasazeny. Při jejich posuzování je pak nutné vzít v úvahu, jaký konkrétní cíl má být v daném prostředí uskutečněn, stejně jako je nutné zohlednit samotné prostředí, v němž má systém fungovat, včetně toho, které části prostředí systém může registrovat (König et al., 2022). Ačkoli lze umělou inteligenci považovat za rizikovou technologii, jak je patrné z předešlého textu, konkrétní rizika aplikací umělé inteligence vyplývají z kombinace odpovědí na otázky, k čemu se používá, jak se používá, kde se používá, kdo ji používá a jaké jsou jeho záměry.

3.2.4 Shrnutí

Základním cílem této kapitoly bylo poskytnout srozumitelný a obecný náhled na problematiku vymezení pojmu umělé inteligence, a pokusit se tak alespoň částečně odpovědět na to, co je to umělá inteligence. Umělá inteligence představuje složitý a mnohotvárný pojem, který může sloužit jako označení pro vědní obor, konkrétní techniku ze souboru technik pro budování inteligentního systému nebo technologický artefakt. Z pohledu práva a regulace představuje umělá inteligence mnohovrstevnaté téma, které je možné uchopit prostřednictvím rizik, které se odvíjí od toho k čemu, jak a kde je systém umělé inteligence používán. Jinak řečeno, konkrétní dopady a problematičnost plynou ze širšího sociálního kontextu, jehož je daný systém součástí, spíše než z technologie jako takové. Pokud jde o rizika vyplývající přímo ze samotné technologie, ta jsou dána zejména charakteristickými vlastnostmi současných systémů

umělé inteligence založených na složitějších formách strojového učení, kterými jsou autonomie a vnitřní neprůhlednost. Tyto vlastnosti se promítají v jedné z hlavních schopností umělé inteligence, kterou je přijímání rozhodnutí, a to tím způsobem, že je vyloučena možnost, aby do tohoto procesu vstupoval člověk prostřednictvím svých zásahů, a dále také tím, že výsledná rozhodnutí vzešlá z tohoto procesu jsou z pohledu člověka často nesrozumitelná a nevysvětlitelná. Takto označené zdroje rizik související s umělou inteligencí pak ukazují směry, kterými se ubírá současná debata o regulaci těchto systémů.

Následující kapitola bude věnována přehledu Návrhu nařízení o umělé inteligenci a zachycení diskuze o jeho obsahu odehrávající se na úrovni Evropské unie. Tento návrh, který byl představen Evropskou komisí v dubnu 2021, si klade za cíl zajistit bezpečné a odpovědné využívání umělé inteligence v Evropské unii. Kapitola se nejprve zaměří na důvody, které vedly k vytvoření tohoto návrhu, a na klíčové prvky navrhované regulace systémů umělé inteligence. Další kapitola se pak bude věnovat konkrétnímu odrazu těchto rizik v návrhu nařízení, včetně představení kritických postřehů k navrhované regulaci.

4. Právní regulace umělé inteligence na úrovni Evropské unie

Cesta, na jejímž konci bylo zveřejnění návrhu Evropské komise o podobě budoucí regulace umělé inteligence v dubnu 2021 (Návrh nařízení Evropského parlamentu a rady, kterým se stanoví harmonizovaná pravidla pro umělou inteligenci (akt o umělé inteligenci) a mění určité legislativní akty unie, Evropská komise, COM (2021) 206 final, 21. dubna 2021, dále jen „Návrh nařízení"), byla lemována řadou milníků v podobě politických projevů, veřejných rozprav a přípravných dokumentů. V roce 2015 Výbor Evropského parlamentu pro právní záležitosti (JURI) ustanovil pracovní skupinu pro právní otázky spojené s robotikou a umělou inteligencí, jejímž cílem bylo vytvoření soukromoprávních pravidel v této oblasti. Pracovní skupina zveřejnila v polovině roku 2016 návrh usnesení Evropského parlamentu, zahrnující sérií doporučení pro Evropskou komisi, který se vyznačoval značně alarmistickým podtextem. Evropský parlament dne 16. února 2017 zveřejnil usnesení obsahující doporučení týkající se občanskoprávních pravidel robotiky, které vyzývá Komisi k vypracování legislativního nástroje upravujícího vývoj a využití umělé inteligence a robotiky (Smuha et al., 2021).

V dubnu 2018 Evropská komise představila Evropskému parlamentu sdělení s názvem Umělá inteligence pro Evropu, které vycházelo z hodnocení, že umělá inteligence se již stala pevnou součástí našeho běžného života, a její vliv tak není pouhou science fiction. Jedním z deklarovaných cílů se tak stalo zajištění vhodného právního a etického rámce pro umělou inteligenci. Na konci roku 2018 Evropská komise navázala konkrétnějším koordinovaným plánem týkajícím se umělé inteligence. I tento dokument uznával, že umělá inteligence je již součástí každodenního života, ale její potenciál je mnohem větší, což jeho slovy vyžaduje jak dobře fungující datový ekosystém vybudovaný na důvěře, dostupnosti dat a infrastruktuře, tak regulační rámec, který je dostatečně flexibilní, aby podporoval inovace a zároveň zajišťoval vysokou úroveň ochrany a bezpečnosti. Důležitým prvkem se stává důvěra, která je nezbytným předpokladem pro to, aby společnosti přijaly a využívaly umělou inteligenci, k jejímuž budování má podle toho dokumentu přispět především to, že tato technologie bude předvídatelná, odpovědná a ověřitelná a měla by dodržovat základní práva a etická pravidla.

Se záměrem stanovit pokyny k tomu, aby byla umělá inteligence důvěryhodná, jmenovala Komise nezávislou skupinu odborníků na vysoké úrovni v oblasti umělé inteligence, kterou pověřila jejich vypracováním. Výstupem činnosti této skupiny byl dokument Etické pokyny pro důvěryhodnou umělou inteligenci z dubna 2019. V těchto pokynech se uvádí, že pro splnění požadavku důvěryhodnosti by umělá inteligence: a) měla být legální, tj. respektovat veškeré platné právní a správní předpisy, b) měla být etická, tj. zajišťovat dodržování etických

zásad, a c) měla být spolehlivá, a to jak z technického, tak i sociálního hlediska, jelikož i dobře míněné systémy umělé inteligence mohou způsobit neúmyslnou újmu.

V červenci 2019 představuje nově zvolená předsedkyně Evropské Komise von der Leyen svůj programový text „Unie, která si klade vyšší cíle“. V něm si klade za cíl v rámci 100 dnů od svého zvolení do funkce předložit návrh nového právního předpisu vztahujícího se k umělé inteligenci. V návaznosti na pokyny pro důvěryhodnou umělou inteligenci vydala Evropská komise v roce 2020 Bílou knihu o umělé inteligenci, která měla položit základ pro budoucí legislativní návrh, jenž byl předložen o rok později. Bílá kniha vzbudila velkou pozornost a doporučovala horizontální přístup k umělé inteligenci s obecnými zásadami, které by měly být zahrnuty do jediného legislativního aktu, který by se vztahoval na jakoukoli umělou inteligenci, čímž by se odmítlo přijetí několika nových odvětvových aktů. Dále Bílá kniha navrhla regulovat umělou inteligenci v závislosti na stupni rizika, které představuje, a které je odvozené od oblasti, v níž je umělá inteligence použita (Ebers et al., 2021).

4.1. Akt o umělé inteligenci a jeho cíle

Návrh nařízení o umělé inteligenci předložený Komisí dne 21. dubna 2021 převzal z Bílé knihy v ní naznačený přístup k regulaci umělé inteligence založený na riziku a vstřebal doporučení obsažená ve všech výše uvedených dokumentech. Návrh Nařízení tvoří součást širšího úsilí Evropské Unie aktivně regulovat a řídit vývoj digitálních technologii a ke kterému se dále řadí Digital Services Act (nařízení (EU) 2022/2065), Digital Market Act (nařízení (EU) 2022/1925) a Data Governance Act (nařízení (EU) 2020/0340). Jeho představení bylo současně zásadním okamžikem, neboť představoval první formální krok v procesu, který by nakonec měl vést k závazné právní regulaci umělé inteligence. Obecným očekáváním ze strany Evropské komise je, že tato její legislativní aktivita povede k tomu, že Evropská unie zaujme vedoucí pozici při vytváření nových, ambiciózních celosvětových norem, mezinárodních normalizačních iniciativ v oblasti umělé inteligence a rámců spolupráce v souladu s mnohostranným systémem, které jsou založeny na pravidlech Evropské unie a v souladu s hodnotami, jež Evropská unie vyznává. K tomu má přispět i to, co je v literatuře označováno jako Bruselský efekt (Bradford, 2012), kdy Evropská unie definuje standard v určité oblasti, který si následně osvojí zbytek světa a rozšiřuje tak dosah působnosti svých právních předpisů.

Navrhované nařízení sleduje několik základních cílů, které jsou vzájemně provázány. Podle věty prvního bodu odůvodnění je účelem nařízení „[z]lepšit fungování vnitřního trhu stanovením jednotného právního rámce […] v souladu s hodnotami Unie.'' Důvodová zpráva k Návrhu nařízení poukazuje na to, že některé členské státy zvažují zavedení právních předpisů

upravujících umělé inteligence. Takový vývoj by pravděpodobně vedl podle názoru Evropské komise k roztříštěnosti vnitřního trhu a ke ztrátě právní jistoty. Jinak řečeno, záměrem navrhovaného nařízení je zlepšit fungování vnitřního trhu tím, že zabrání roztříštěnosti a zajistit jednotnost a předvídatelnost, které jsou nezbytné pro jeho řádné fungování. Právním základem Návrhu nařízení je tak článek 114 Smlouvy o fungování Evropské unie, podle kterého má Evropská unie pravomoc přijímat opatření ke sbližování vnitrostátních právních předpisů s cílem usnadnit vytvoření a fungování vnitřního trhu.

Dalším z cílů, který je v Návrhu nařízení zmíněn, je podpora vývoje a inovace umělé inteligence. Navrhované nařízení je zamýšleno s cílem podpořit rozvoj, využívání a zavádění umělé inteligence na vnitřním trhu a umožnit Evropské unii zastávat v celosvětovém kontextu čelnou pozici, pokud jde o rozvoj bezpečné, důvěryhodné a etické umělé inteligence. Současně s tím se Návrh nařízení snaží zajistit, aby vývoj a zavádění systémů umělé inteligence na vnitřním trhu doprovázely takové podmínky, které zajistí „vysokou úroveň ochrany veřejných zájmů“, přičemž tento záměr je jedním z hlavních cílů této iniciativy. Mezi příslušné veřejné zájmy jsou zahrnuty „zdraví a bezpečnost a ochrana základních práv, jak je uznává a chrání právo Unie“. Jasným cílem Návrhu nařízení je tak vývoj důvěryhodné umělé inteligence, to znamená vývoj systémů umělé inteligence, u nichž „lidé mohou důvěřovat, že technologie je používána způsobem, který je bezpečný a v souladu s právními předpisy, včetně dodržování základních práv“.

4.2. Definice umělé inteligence v Návrhu nařízení, článek 3

Návrh nařízení je rozdělen do 12 hlav. Po úvodním rozsáhlém odůvodnění sestávající z 89 bodů následuje Hlava I obsahující 4 články, které postupně upravují předmět (čl. 1) a oblast působnosti (čl. 2). Pod čl. 3 se potom nachází ustanovení obsahující výčet definic pojmů, s nimiž se dále pracuje v Návrhu nařízení. Jednou z hojně komentovaných definic, kterou návrh uvádí, je definice systémů umělé inteligence.

To je způsobeno především tím, že to, jak bude systém umělé inteligence definován, by do značné míry mělo určovat, na které systémy se nařízení bude vztahovat. V Návrhu nařízení je tento pojem definován poměrně široce a neurčitě, a to jako software, který je vyvinut pomocí jedné nebo více technik a přístupů uvedených v příloze I, a který může pro danou sadu cílů definovaných člověkem generovat výstupy, jako je například obsah, predikce, doporučení nebo rozhodnutí ovlivňující prostředí, s nimiž komunikují. Metody a přístupy uvedené v příloze I jsou rozděleny do tří skupin: a) přístupy strojového učení, včetně učení s učitelem, bez učitele a posilovaného učení, používající celou řadu metod, včetně hlubokého učení, b) přístupy založené

na logice a znalostech, včetně reprezentace znalostí, induktivního (logického) programování, znalostních základen, inferenčních a deduktivních mechanismů, (symbolického) uvažování a expertních systémů a c) statistické přístupy, bayesovské odhadování, metody vyhledávání a optimalizace.

Výhoda široké definice umělé inteligence je spatřována v tom, že bude schopná pokrýt budoucí vývoj této technologie (Renda a Enger, 2023). V případě, že by se v budoucnu objevil nový přístup, je do návrhu začleněn nástroj v podobě čl. 4, na základě něhož je možné seznam metod a přístupů uvedených v příloze I aktualizovat a měnit, přičemž je tato pravomoc svěřena Evropské komisi. Naopak často zaznívajícím argumentem v rámci kritiky definice umělé inteligence, která je uvedena v Návrhu nařízení, je, že definice zahrnující příliš mnoho metod a přístupů může mít za následek nadměrnou regulatorní zátěž kladenou na danou technologii. Pokud však konkrétní systém spadá pod vymezení pojmu systém umělé inteligence, tak toto samo o sobě ještě neznamená, že by podléhal regulačním požadavkům nařízení. K zjištění toho, zdali systém z množiny systémů spadajících pod definici umělé inteligence bude muset splňovat určité požadavky stanovené Návrhem nařízení, je dále nutné použít filtr v podobě klasifikace rizik, na základě něhož se určí konkrétní míra jeho regulace. Obecně se dá říct, že výsledná skupina systémů vytvořená použitím filtru, které by byly zatížené regulací, je velice úzká, a většina systémů umělé inteligence tak nebude podléhat regulačním požadavkům nařízení.

4.3. Obecný přehled obsahu Návrhu nařízení

Navrhované nařízení je založeno na modelu hodnocení rizik, v jehož rámci se rozlišují čtyři stupně rizika. Výchozím bodem tohoto přístupu je, že cíle v podobě ochrany bezpečnosti jednotlivců a základních práv lze dosáhnout stanovením hranic tomu, proč a jak se systémy umělé inteligence používají.

Návrh nařízení rozlišuje mezi použitím umělé inteligence, které vytváří a) nepřijatelné riziko (Hlava II) - tato kategorie zahrnuje všechny systémy umělé inteligence, jejichž používání je považováno za nepřijatelné z důvodu rozporu s hodnotami Unie, například v důsledku porušování základních práv a vzhledem k čemuž jsou tyto systémy zakázány, b) vysoké riziko (Hlava III) - do této kategorie jsou zařazeny systémy, které vytvářejí vysoké riziko pro zdraví a bezpečnost nebo základní práva fyzických osob. Rozdíl oproti předešlé kategorii zakázaných systémů spočívá v tom, že vysoce rizikové systémy jako takové nejsou v rozporu s hodnotami Evropské unie, ale pouze tyto hodnoty ohrožují. Vysoce rizikové systémy jsou hlavním předmětem navrhovaného nařízení, čemuž odpovídá i rozsah úpravy, která je jim věnována a která je soustředěna mezi čl. 6 až čl. 51, c) nízké riziko (Hlava IV) - do této kategorie Návrh

nařízení v čl. 52 zařazuje určité systémy umělé inteligence, jejichž společnou charakteristikou je to, že vyvolávají konkrétní problémy z hlediska transparentnosti, a které mají být vyřešeny prostřednictvím stanovení informační povinnosti. Označení této kategorie se v samotném textu Návrhu nařízení neobjevuje, ale pracuje s ním důvodová zpráva, d) minimální riziko (Hlava IX) do této kategorie jsou zahrnuty všechny systémy umělé inteligence, které nejsou výslovně zařazeny do výše uvedených kategorií a jedná se tak o naprostou většinu systémů umělé inteligence, které se v současné době v Evropské Unii používají, jak uvádí důvodová zpráva.

Navzdory tomu, že se na většinu systémů umělé inteligence nebude regulace obsažená v Návrhu nařízení vztahovat, neznamená to, že by vývoj těchto systémů nemohl být určován navrhovaným nařízením, a to z důvodu existence čl. 69, který vybízí k tomu, aby poskytovatelé těchto systémů pro tyto systémy vypracovávali kodexy chování, v nichž se dobrovolně zavážou k dodržování obdobných požadavků, které jsou kladeny na vysoce rizikové systémy. Toto odstupňování rizik představuje ústřední myšlenku navrhovaného nařízení a v literatuře je znázorňováno pomocí čtyřstupňové pyramidy.

Hlava V obsahuje několik opatření, které mají podpořit a usnadnit inovace v oblasti umělé inteligence, mezi které patří vytváření tzv. regulačních pískovišť, jimiž jsou myšlena kontrolovaná experimentální a testovací prostředí používaná ve fázi vývoje konkrétního systému a také před jeho uvedením na trh. Následující hlavy VI, VII a VIII se věnují vytvoření institucionálního rámce, a to jak na úrovni členských států, tak na úrovni Evropské unie, rozdělení pravomocí a dále obsahují pravidla týkající se správy a dohledu nad prováděním Návrhu nařízení. Obsah Hlavy IX byl shrnut ve výše uvedené pasáži věnované základním konstrukčním prvkům přístupu založenému na posouzení rizik.

Zbývající Hlavy X, XI a XII jsou pak věnovány závěrečným ustanovením, která završují Evropskou komisí navrhovaný nový právní rámec pro umělou inteligenci. Stanovují požadavek na zachování důvěrnosti informací a údajů, které orgány získají při plnění svých úkolů a činností souvisejících s prováděním nařízení, a také základní parametry možných sankcí použitelných v případě porušení nařízení. Sankcemi se dle návrhu myslí výhradně peněžité pokuty, které se počítají buď paušální částkou, nebo procentem z celkového celosvětového ročního obratu společnosti, a to za předchozí finanční rok. Z pohledu nařízení by sankce měly být především účinné, přiměřené a odrazující. Ustanovení rovněž zavádějí pravidla pro způsob přenesení pravomocí, jejich provádění a zmocňují Evropskou komisi k případnému přijímání prováděcích aktů (hlava XI). Dále stanovují povinnost Evropské komise vypracovávat pravidelné zprávy o hodnocení a přezkumu navrhovaného nařízení a obsahují seznam legislativních aktů, které mají být změněny (hlava XII).

4.4. Pravidla stanovená pro systémy umělé inteligence na základě systému klasifikace rizik

Evropská komise popisuje akt o umělé inteligenci jako legislativní nástroj založený na posouzení rizik, který nestanovuje jednotná pravidla pro všechny systémy umělé inteligence, ale místo toho vymezuje čtyři různé právní režimy pro tyto systémy, přičemž každý režim stanovuje jiné právní povinnosti, které odpovídají předpokládaným rizikům pro veřejné zájmy a hodnoty chráněné právem Evropské unie, a v každém jednotlivém případě je tak pro určení, který režim se na daný systém vztahuje, nutné prozkoumat kontext, záměr nebo techniku, na nichž je systém umělé inteligence založen. S přihlédnutím k tomu bude následují část textu věnovaná především bližšímu popisu toho, jak jsou jednotlivé režimy nastaveny a jejich charakteristikám. Jak již bylo uvedeno výše, Návrh nařízení rozřazuje rizika související s umělou inteligencí do čtyř úrovní (nepřijatelné riziko, vysoké riziko, nízké riziko, minimální riziko). Následující popis bude postupovat od nejvyšší úrovně (nepřijatelné riziko) této pomyslné pyramidy rizik k úrovni nejnižší (minimální riziko).

4.4.1 Kategorie nepřijatelného rizika, článek 5

Vstupní bod, který vysvětluje širší kontext a základní důvody k zavedení kategorie nepřijatelného rizika, je obsažen v odůvodnění Návrhu nařízení, a to v podobě bodů 15. a 16. Ty se následně promítají do znění čl. 5 Návrhu nařízení, který rozlišuje čtyři skupiny zakázaných postupů v oblasti umělé inteligence. V nich jsou zahrnuty všechny systémy umělé inteligence, jejichž používání je považováno za nepřijatelné z důvodu rozporu s hodnotami Evropské unie, jako je respektování lidské důstojnosti, demokracie nebo ochranou základních práv.

4.4.2 Manipulativní systémy umělé inteligence

Do prvních dvou skupin spadají typy umělé inteligence, jimž je společné, že ovlivňují chování člověka takovým způsobem, jenž způsobuje, nebo by mu mohl způsobit fyzickou nebo psychickou újmu, a to buď podprahovým působením na jeho chování, nebo využíváním jeho zranitelnosti vyplývající z nízkého věku (děti), vysokého věku (senioři), nebo fyzického či mentálního postižení. Od této charakteristiky se následně odvozuje jejich společné označení, kterým je manipulativní systémy. V případě těchto manipulativních systémů spočívá právní režim pro ně stanovený Návrhem nařízení v zákazu, a sice v zákazu uvádění na trh, uvádění do provozu nebo jejich používání. V Návrhu nařízení se tak prostřednictvím těchto zákazů uplatňuje preventivní mechanismus, kterým má být zajištěna vysoká ochrana před riziky plynoucími z používaní těchto manipulativních systémů, které jsou považovány za příliš velké na to, aby je bylo možné zmírňovat pouze pomocí dílčích zásahů. Pro účely určení toho, zda se na systém

umělé inteligence vztahuje zákaz uvedený v Návrhu nařízení je stanovena následující řada kumulativních podmínek týkajících se manipulace.

První z těchto podmínek je úmysl manipulovat chování jednotlivců, který je dovozen z použitého spojení s cílem uvedeného v čl. 5 odst. 1 písm. a) tak i čl. 5 odst. 1 písm. b). Další podmínka vyplývá z podstaty podprahových technik, která spočívá v tom, že v momentě, kdy je jim jednotlivec vystaven, není schopen si uvědomovat přijímané podněty, na které reaguje (Neuwirth 2023). Písmeno (a) tak předpokládá použití podprahových technik, které musí být skryté (odehrávají se mimo vědomí osob). V písmenu (b) je pak stanoven požadavek na přítomnost určité zranitelnosti u postižených osob, která má svůj původ ve věku, fyzickém nebo mentálním postižení, a které je zneužíváno.

Posledním ze stanovených požadavků je, aby výsledkem činnosti těchto systémů bylo, že dotčené osobě nebo jiné osobě způsobuje nebo by mohla způsobit fyzickou nebo psychickou újmu. Podle některých autorů (Veale a Borgesius, 2021, Wendehorst (2021) je tak tento požadavek (možnost způsobení újmy) hlavním kvalifikačním kritériem pro zařazení se do skupiny systémů umělé inteligence s nepřijatelným rizikem. Z čehož pak následně vyplývá, že některé manipulativní systémy umělé inteligence se podle znění čl. 5 zdají být povoleny, a to v případě, pokud je nepravděpodobné, že způsobí jednotlivci újmu, čímž se významně omezuje oblast působnosti zmíněných ustanovení a neplatí tak deklarovaný všeobecný zákaz manipulativních systémů.

Z pohledu tohoto zákazu působí obdobně problematicky požadovaný manipulativní úmysl na straně těch, kteří tyto systémy vyvíjí, uvádí na trh nebo používají k profesionální činnosti (poskytovatel, uživatel), u něhož není však jasné, jak má být prokázán v případě, kdy není výslovně uveden.

Na závěr lze uvést, že problém představuje také to, že část ustanovení čl. 5 věnovaná újmě vyplývající z používání manipulativních systémů, se soustředí pouze na fyzickou a psychickou újmu (nemajetkovou), přičemž takto nastavená ochrana, je považována za nedostatečnou a měla by být rozšířena tak, aby zahrnovala i další potenciálně stejně závažné typy újmy jako je například újma finanční a ekonomická (majetková újma).

4.4.3 Systémy sociálního kreditu

Třetí skupina zakázaných postupů v oblasti umělé inteligence se týká takzvaných systémů sociálního kreditu. Tento typ systému umělé inteligence je znám především díky experimentům s jeho zaváděním, který se od roku 2014 odehrává v Čínské lidové republice

(Zuboff, 2019). Podstatou tohoto systému je vytvoření osobního sociálního kreditu každého obyvatele, který se zvyšuje nebo snižuje podle jeho společenského chování v rámci čtyř konkrétně vymezených oblastí života a určuje, zda tato daná osoba může využívat určité veřejné služby, cestovat, získat práci nebo půjčku. Tento typ (čínské) umělé inteligence, který slouží především jako nástroj dohledu a trestání, stojí v pozadí toho, že se v Návrhu nařízení objevil článek 5 odst.1 písm. c) a současně měl přímý vliv na podobu jeho navrhovaného znění (Raposo, 2022).

Možná nebezpečí, která s sebou nesou systémy umělé inteligence provádějící hodnocení sociálního kreditu pro obecné účely, jsou popsána a vysvětlena v 17. bodě odůvodnění Návrhu nařízení. Zde se uvádí, že tyto systémy umělé inteligence mohou vést k diskriminačním výsledkům a sociálnímu vyloučení určitých skupin, porušovat právo na důstojnost a zákaz diskriminace jakož i hodnoty rovnosti a spravedlnosti. Na základě obav vyplývajících z této možnosti a z důvodu předcházení těmto negativní účinkům systémů sociálního kreditu zakotvuje článek 5 odst. 1 písm. c) zákaz jejich používání pro orgány veřejné moci.

Z pohledu znění tohoto článku se jedná o systémy umělé inteligence, které jsou určeny k vytváření skóre „důvěryhodnosti“ (sociálního kreditu), které může vést ke znevýhodňujícímu nebo nepříznivému zacházení s některými fyzickými osobami nebo celými skupinami těchto osob v sociálních kontextech nesouvisejících s kontextem, ve kterém byly dané údaje původně vytvořeny nebo shromážděny a/nebo znevýhodňujícímu nebo nepříznivému zacházení s některými fyzickými osobami nebo celými skupinami těchto osob, které je neodůvodněné nebo nepřiměřené jejich sociálnímu chování nebo jeho závažnosti.

Jak již bylo řečeno v charakteristice uvedené výše, za zněním toho článku stojí systém, který je obecný, a ve kterém sledování pokrývá každou oblast lidského života. Život člověka je redukován na jediné sociální skóre, které následně ovlivňuje možnosti, které člověk má v rámci žitého světa. Tyto systémy jsou založeny primárně na velkém množstvím dat sebraných v různých oblastech života a (neprůhledném) algoritmickém rozhodovaní (Smuha et al. 2021). Článek 5 odst. 1 písm. c) je tedy navržen tak, aby zabránil orgánům veřejné moci v zavádění univerzálních bodových systémů a dystopické budoucnosti lidstva, což je však i jeho možnou slabinou.

Tato slabina spočívá v tom, že klíčovým z pohledu stanoveného zákazu v článku 5 odst. 1 písm. c) je, že se vztahuje pouze na orgány veřejné moci (a na osoby jednající jejich jménem) a je v podstatě omezen na (obecný) systém sociálního kreditu. Ze zákazu je tak vyjmuta celá oblast použití systémů sociálního hodnocení (social scoring) pro účely soukromého sektoru, stejně jako

jsou vyjmuty systémy umělé inteligence využívané orgány veřejné moci pro účely hodnocení osob v rámci konkrétně vymezené oblasti (tzn. vstupní data byla sebrána ve stejné oblasti života, ke které se vztahuje rozhodování veřejné moci, výjimka stejného kontextu).

Důvod (Mazzini a Scalzo, 2022) pro vyjmutí ze zákazu používání systémů umělé inteligence k posuzování důvěryhodnosti osob soukromými subjekty je spatřován v tom, že orgány veřejné moci mají jedinečné pravomoci vůči jednotlivcům a potenciální rozsah a intenzita zásahů do lidských práv jsou tak na zcela odlišném stupni než u těch zásahů, které mohou vzniknout v případě použití tohoto typu umělé inteligence soukromými subjekty. Takto postavený argument o zásadně odlišných důsledcích používání této technologie, a to podle toho, kdo ji používá (stát vs. soukromý subjekt), nepostihuje dostatečně současnou realitu a výzkum věnovaný důsledkům a nebezpečím, které vyplývají z využívání umělé inteligence k těmto účelům soukromými subjekty, které může mít stejně zničující důsledky pro lidská práva jako v případě jejího používání státem (např. Zuboff 2019, Crawford 2021).

To samé lze říci o důsledcích využití algoritmů pro profilování osob orgány veřejné moci, které mají omezenější rozsah. Tyto případy použití systémů umělé inteligence nezůstávají zcela neregulované, neboť podle čl. 6 odst. 2 Návrhu nařízení se považují tyto systémy za vysoce rizikové a musí být proto v souladu se systémem pro řízení rizik a kvality, který pro ně Návrh nařízení zavádí (tématu vysoce rizikových systémů se věnuje následující kapitola). Tím, že Evropská komise uvedla, že orgány veřejné moci se nemohou zapojit do posuzování důvěryhodnosti osob na základě umělé inteligence, chtěla dát najevo, že její vize umělé inteligence je především taková, která neohrožuje základní lidská práva a uvědomuje si možný rozsah škod, který plyne ze zneužívání umělé inteligence určené k tomuto účelu. Současně však úplné naplnění této vize zabraňují nařízením zavedené výjimky ze zákazu používání tohoto typu umělé inteligence (Ebers et al. 2021).

4.4.4 Biometrické systémy

Čtvrtá a také poslední skupina zakázaných postupů zahrnuje biometrické systémy využívající umělou inteligenci, které jsou jedním z hlavních cílů, na něž je navrhovaná regulace mířena. Na úvod je vhodné zmínit, že Návrh nařízení činí rozdíly mezi různými druhy biometrických technologií a v rámci klasifikace systémů umělé inteligence podle rizika je zařazuje do tří kategorií, a to kategorie nepřijatelného rizika, vysokého rizika a nízkého rizika. V Návrhu nařízení jsou definovány a upraveny tři druhy biometrických systémů, a to systémy dálkové biometrické identifikace, systémy biometrické kategorizace a systémy rozpoznávání

emocí. Přístup, který Návrh nařízení zaujímá k biometrickým systémům, je takový, že pouze používání systémů dálkové biometrické identifikace může být klasifikováno jako zakázané použití umělé inteligence (povaha tohoto zákazu je však velice omezena). Ostatní druhy biometrických systémů, čímž se myslí konkrétně biometrická kategorizace a systémy rozpoznávání emocí, mohou pak spadat do kategorií systémů umělé inteligence s vysokým nebo omezeným rizikem, a to v závislosti na oblasti, v níž jsou tyto technologie použity.

Ustanovení čl. 5 odst. 1 písm. d) zakazuje používání systémů biometrické identifikace na dálku v reálném čase na veřejně přístupných místech pro účely vymáhání práva, a jak se uvádí v bodě 18 odůvodnění k Navrhovanému nařízení, z hlediska základních lidských práv by se mělo jednat o jedno z nejrizikovějších použití umělé inteligence (především z důvodu možného masové sledování lidí).

Biometrická identifikace je metoda identifikace nebo potvrzení totožnosti osoby na základě jedinečných fyzických, fyziologických znaků nebo znaků chování (biometrických údajů). Proces identifikace osoby spočívá v porovnání biometrických údajů dané osoby s biometrickými údaji obsaženými v referenční databázi.

Z pohledu uvedeného článku je tento systém specifikován tím, že identifikace osob probíhá na dálku, čímž je myšleno, že shromažďování biometrických údajů probíhá pouze pasivním způsobem a často bez vědomé spolupráce těchto osob, nebo dokonce i proti jejich vůli, a dále také tím, že probíhá v reálném čase, což znamená, že zachycení biometrických údajů, jejich porovnání a identifikace probíhá bez významné prodlevy (čl.3 bod. 36 a čl.3 bod. 37). V současnosti pravděpodobně nejznámějším a nejdiskutovanějším příkladem systému, který pracuje tímto způsobem, je systém pro rozpoznávání tváře.

Při bližším pohledu na celkovou formulaci zákazu v čl. 5 odst. 1 písm. d a odst. 2 až 4 je patrné, že i přes škodlivost používání těchto systému, která je v Návrhu nařízení deklarována, má zákaz příliš úzký rozsah a vztahuje se pouze na specifickou výseč biometrických sledovacích systémů. V porovnání s výše uvedenými třemi skupinami zakázaných postupů se zakazuje pouze používání systému biometrické identifikace, ale jejich uvádění na trh a uvádění do provozu těchto systému zakázáno není.

To například znamená, že prodejci z Evropské unie mohou prodávat biometrické systémy, jejichž používání by bylo v Evropské unii nezákonné, represivním režimům ve třetích zemích (Veale a Borgesius, 2021). Pro uplatnění zákazu uvedeného čl. 5 odst.1 písm. d) musí být splněno několik podmínek, které vyvolávají řadu pochybností ohledně jejich opodstatněnosti.

4.4.4.1 Podmínky zákazu používání biometrické identifikace

Za prvé, v Návrhu nařízení se rozlišuje mezi zpětným a reálně probíhajícím (živým) používáním biometrických systémů, přičemž zákaz se vztahuje pouze na použití biometrické identifikace probíhající v reálném čase. V čl. 3 bod 37 je systém dálkové biometrické identifikace v reálném čase definován jako systém dálkové biometrické identifikace, při němž k zachycení biometrických údajů, porovnání a identifikaci dochází bez významné prodlevy. Definice dále upřesňuje, že to zahrnuje nejen okamžitou identifikaci, ale také omezená krátká zpoždění, jejichž cílem je zabránit obcházení systému. Z návrhu nařízení však není jasné, co je považováno za toto krátké zpoždění. Pokud však jde o určení rozsahu, v jakém by měly být biometrické systémy identifikace omezovány, zdá se být zvolený rozhodující faktor, kterým je existence zpoždění a jeho případná délka, jako zcela nevhodný (Wendhorst, 2021).

Další z podmínek je, že použití dálkové biometrické identifikace by mělo být zaměřeno na veřejné prostory. Slovy Návrhu nařízení (čl.3 bod 39) na veřejně přístupná místa, čímž je myšleno jakékoli fyzické místo přístupné veřejnosti bez ohledu na to, zda pro něj platí určité podmínky přístupu. Jedná se tak o ulice, příslušné části státních budov a většinu dopravní infrastruktury, ale také o zpravidla veřejně přístupné prostory jako jsou kina, divadla, obchody a nákupní centra (bod 9 odůvodnění). Důvod pro zavedení přísnějšího režimu v případech, kdy k dálkové biometrické identifikaci dochází v prostorech, které jsou přístupné veřejnosti, je zjevný, a je jím zejména množství osob, kterých se může použití této technologie potenciálně negativně dotknout.

Za třetí je zákaz omezen na používání těchto systémů umělé inteligence pro účely vymáhání práva, čímž jsou myšleny činnosti prováděné donucovacími orgány za účelem prevence, vyšetřování, odhalování či stíhání trestných činů nebo výkonu trestů, včetně ochrany před hrozbami pro veřejnou bezpečnost a jejich předcházení (čl. 3 bod 41). Zákaz se tedy nevztahuje na používání systémů biometrické identifikace orgány veřejné moci ve veřejně přístupných prostorech pro jiné účely než vymáhání práva, a nevztahuje se ani na používání této technologie soukromými subjekty. Zákaz se tedy vztahuje pouze na omezený okruh použití biometrických systémů, a to velmi omezeným okruhem subjektů (donucovacími orgány).

Bez ohledu na to, zda je tato technologie používána donucovacími orgány nebo jinými veřejnými či soukromými subjekty, má její používání vždy nevyhnutelné dopady na základní práva, které se zdají být nepřiměřené vzhledem k tomu, jak velké množství citlivých údajů o kolika lidech je nutné zpracovat k umožnění identifikace několika málo osob (Barkane, 2022).

4.4.4.2 Výjimky ze zákazu používání

Z takto úzce pojatého zákazu používání systémů dálkové biometrické identifikace stanovuje Návrh nařízení několik výjimek, a to v čl. 5 odst. 1 písm. d), kdy se tento zákaz neuplatní a donucovacím orgánům je umožněno používat tyto systémy na veřejných místech. Toto použití však musí být nezbytně nutné pro jeden z následujících cílů: a) „k cílenému vyhledávání určitých potenciálních obětí trestných činů, včetně pohřešovaných dětí“, b) „k prevenci konkrétního, závažného a bezprostředního ohrožení života nebo fyzické bezpečnosti fyzických osob nebo teroristického útoku“ nebo c) „k odhalování, lokalizaci, identifikaci nebo stíhání pachatelů nebo osob podezřelých“ z jednoho z 32 trestných činů uvedených v čl. 2 odst. 2 rámcového rozhodnutí Rady 2002/584/SVV62 (který umožňuje vydání evropského zatýkacího rozkazu), pokud lze tento trestný čin v dotčeném členském státě potrestat trestem odnětí svobody nebo ochranným opatřením spojeným s odnětím osobní svobody s horní hranicí sazby v délce nejméně tři roky, jak je stanoveno v právních předpisech tohoto členského státu. Tato poslední výjimka zahrnuje závažné trestné činy, jako je například terorismus, obchodování s lidmi, vraždy a znásilnění, ale také další, méně závažné trestné činy, jako je korupce, podvody a napomáhání při nedovoleném překročení státní hranice a nedovoleném pobytu.

Obecně lze říci, že tento taxativní výčet výjimek je odůvodněn především možným přínosem pro kontrolu kriminality. Dále je dle znění čl. 5 odst. 2 v těchto výjimečných situacích nutné vzít v úvahu „závažnost, pravděpodobnost a rozsah újmy způsobené v případě, že by systém použit nebyl a důsledky používání systému pro práva a svobody všech dotčených osob, zejména závažnost, pravděpodobnost a rozsah těchto důsledků“. Kromě toho musí být použití těchto systémů v souladu s nezbytnými a přiměřenými zárukami a podmínkami, zejména pokud jde o časová, zeměpisná a osobní omezení, což má zajišťovat, aby používání těchto systémů bylo vždy přesně zacílené.

V Návrhu nařízení se dále uvádí, že ve všech výše vyjmenovaných situacích musí každému použití biometrických systémů identifikace ze strany donucovacího orgánu předcházet získání povolení od soudního orgánu nebo nezávislého správního orgánu, a to s výjimkou naléhavých situací, kdy lze o povolení požádat v průběhů použití nebo až po něm, přičemž orgán udělí povolení pouze tehdy, pokud se přesvědčí, že na základě objektivních důkazů nebo jednoznačných údajů, které mu byly předloženy, je použití systému nezbytné (k dosažení významného veřejného zájmu) a přiměřené (jeho význam převažuje nad riziky pro práva a svobody osob a neexistuje méně omezující prostředek) (čl. 5 odst. 3 a bod 19 odůvodnění). Návrh nařízení v závěru článku 5 v odst. 4 uvádí, že členský stát se může rozhodnout, zda

poskytne možnost úplného nebo částečného povolení k používání systémů biometrické identifikace pro účely vymáhání práva, a musí ve svém vnitrostátním právu stanovit nezbytná podrobná pravidla pro podávání žádostí o povolení, vydávání a výkon těchto povolení, jakož i dohled nad nimi.

V rámci všech ustanovení vztahujících se k používání systémů biometrické identifikace však nejsou obsažena pravidla týkající se prostředků ochrany, které by byly k dispozici jednotlivci, stejně jako zde nejsou obsažena pravidla týkající se oznamování používání těchto systémů. Jak je z návrhu nařízení o umělé inteligenci zřejmé, zákaz vztahující se na používání systémů dálkové biometrické identifikace pro účely vymáhání práva je značně omezený, a to z důvodu existence výjimek pro takovéto použití. Podmínky pro povolení výjimečného použití jsou však formulovány natolik vágně, že takovéto použití je možné téměř vždy odůvodnit. To, jak je zákaz zformulován, vede ve výsledku k tomu, že není schopen zabránit použití systémů umělé inteligence pro hromadné sledování, které bylo jedním z hlavních důvodů pro zavedení tohoto zákazu (Barkane, 2022).

Obecně je Návrh nařízení kritizován za to, že ponechává v případech použití biometrických systémů velké množství mezer, a to například tím, že se nevztahuje na používání dálkové biometrické identifikace pro účely vymáhání práva, které neprobíhá v reálném čase, které se neprovádí ve veřejných prostorech nebo které neidentifikuje osoby, ale má jiný účel použití jako je potvrzení totožnosti, automatizované rozpoznávání charakteristických znaků nebo rozpoznávání emocí (Ebers et al. 2021).

Stejně jako se nevztahuje na používání těchto systémů jinými orgány veřejné moci, které nesouvisí s vymáháním práva a také se nevztahuje na používání biometrických systémů soukromými subjekty, čímž umožňuje budování infrastruktury biometrického sledování, a celkově tak tento zákaz neposkytuje odpovídající ochranu základních práv (Raposo 2022).

Shodně problematicky se jeví navrhovaná kategorizace biometrických systémů, se kterou přichází Návrh nařízení, kdy ze skupiny biometrických technologií, kterou tvoří systémy biometrické identifikace, systémy rozpoznávaní emocí (čl. 3 bod 34) a systémy biometrické kategorizace (čl.3 bod 35), se většina systémů biometrické identifikace řadí do skupiny vysoce rizikových systémů umělé inteligence a zbývající dva systémy jsou řazeny do skupiny systémů umělé inteligence s nízkým rizikem a to v závislosti na oblasti použití mohou spadat i do kategorie vysokého rizika (čl. 6 a příloha III). Toto však neodpovídá skutečným rizikům a dopadům, které doprovází jejich použití a celkově se tak zvolený režim regulace jeví jako nejasný a nesoudržný.

Po této části textu, která se věnovala popisu a rozboru právního rámce stanoveného pro systémy umělé inteligence spadajících do kategorie nepřijatelného rizika, se následující část zaměří na vysoce rizikové systémy umělé inteligence, konkrétně na to, jak jsou v návrhu definovány a jaká pravidla jsou pro ně v Návrhu nařízení stanovena.

5. Kategorie vysoce rizikových systémů umělé inteligence

Pokud se vrátíme zpět k pomyslné pyramidě rizik, jsou na další úrovni této pyramidy upraveny vysoce rizikové systémy. Jak již bylo řečeno v podkapitole věnované obecnému přehledu obsahu Návrhu nařízení, zabírají ustanovení související s kategorií vysokého rizika většinu prostoru navrhovaného nařízení (čl. 6 až čl. 51). Vymezení této kategorie se nachází v čl. 6 Návrhu nařízení, přičemž zde uvedená klasifikační pravidla pro vysoce rizikové systémy umělé inteligence jsou doplněna dvěmi přílohami (příloha č. II, příloha č. III), které tuto kategorii blíže vymezují.

Vysoce rizikové systémy umělé inteligence jsou dle důvodové zprávy k Návrhu nařízení charakteristické tím, že „představují vysoké riziko pro zdraví a bezpečnost nebo pro základní práva fyzických osob.“ Rozdíl oproti zakázaným použitím, která se nachází na vrcholu pyramidy, spočívá v tom, že vysoce rizikové systémy umělé inteligence samy o sobě nejsou v rozporu s hodnotami Evropské unie, ale pouze je ohrožují.

Při posuzování, které v tomto případě provedla Evropská komise, bylo riziko, které vysoce rizikové systémy představují pro právo na lidskou důstojnost, soukromí, ochranu osobních údajů, právo na účinnou právní ochranu a spravedlivý proces a další hodnoty, porovnáváno s možnými přínosy těchto systémů, a takto získaný výsledek byl hlavním motivem k tomu, aby bylo umožněno takovýmto systémům přistoupit na trh, ovšem za podmínky souladu s požadavky, které jsou pro ně Návrhem nařízení závazně stanoveny.

Z možných přístupů, jak pojímat umělou inteligenci pro účely regulace, zvolil Návrh nařízení možnost přistoupit k systémům umělé inteligence jako k výrobkům. Na základě tohoto zásadního rozhodnutí je většina jeho ustanovení, a to zejména těch, která se vztahují na vysoce rizikové systémy umělé inteligence, strukturována podle vzoru systému, který pro uvádění výrobků na trh zavedl nový legislativní rámec (a to s důrazem kladeným především na bezpečnost). Tím se tak otevřela možnost uchýlit se k zásadám, na nichž je legislativa Evropské unie týkající se bezpečnosti výrobků založena (Mazzini a Scalzo, 2021).

5.1. Nový legislativní rámec, podmínky pro uvádění výrobků na trh

Pro nový legislativní rámec jsou charakteristická taková pravidla týkající se bezpečnosti výrobků, která stanoví pouze základní požadavky, jimž musí výrobky uváděné na trh Evropské unie vyhovovat, aby mohly využívat volného pohybu na vnitřním trhu Evropské unie, zatímco úkol konkretizace těchto základních požadavků je svěřen třem evropským normalizačním společnostem – CEN (Evropský výbor pro normalizaci), CENELEC (Evropský výbor pro

normalizaci v elektrotechnice) a ETSI (Evropský ústav pro telekomunikační normy), které však překvapivě nejsou v Návrhu nařízení přímo zmíněny (Veale a Borgesius, 2021).

Předpoklad, z kterého vychází nový legislativní rámec, je, že „výrobce zná podrobně projekční a výrobní proces, má nejlepší možnosti provést celkový postup posouzení shody. Posuzování shody by tedy mělo zůstat povinností pouze výrobce“ (bod odůvodnění 21, Rozhodnutí č. 768/2008/ES). Zapojení třetí strany do procesu posouzení shody z pohledu nového legislativního rámce, tak obvykle závisí na míře rizika daného výrobku. Zatímco u výrobků s nízkým rizikem obvykle postačuje vlastní prohlášení výrobce o posouzení shody, u výrobků s vysokým rizikem posouzení shody výrobku s na ně kladenými požadavky obvykle provádějí třetí subjekty (Mazzini a Scalzo, 2021).

Zmíněné se potom odráží v Návrhu nařízení v ustanoveních o posuzování shody vysoce rizikových systémů se závaznými požadavky, které jsou na ně kladeny (čl. 19) avšak s významnou odchylkou, která je popsána v pasáži věnované procesu posuzování shody.

Z výše uvedeného je patrné, že Návrh nařízení se tak snaží o kombinaci dvou linií uvažování o rizicích (bezpečnost x základní práva), kdy jedna je založena na technickém hledisku, které je objektivní a snaží se dopady a rizika matematicky vypočíst (kvantitativní přístup k rizikům), a vedle toho stojící linii, která se věnuje rizikům a dopadům na práva osob, založené na vyvažování a často pracující s takovými aspekty reality, které jsou jen stěží měřitelné (kvalitativní přístup k rizikům).

5.2. Vymezení kategorie vysoce rizikových systémů

Vztah k právním předpisům Evropské unie tykajících se bezpečnosti výrobků je zachycen ve znění čl. 6 odst. 1, který vymezuje jeden ze dvou okruhů vysoce rizikových systémů umělé inteligence. Za prvé, podle ustanovení tohoto článku se považuje systém umělé inteligence za vysoce rizikový, pokud jsou kumulativně splněny dvě podmínky, a to: a) systém umělé inteligence je určen k použití jako bezpečnostní součást výrobku nebo je sám o sobě výrobkem, na který se vztahují harmonizační právní předpisy Evropské unie uvedené v příloze II; b) případně na samotný tento systém umělé inteligence jako produkt se vztahuje povinnost posouzení shody třetí stranou za účelem uvedení tohoto produktu na trh nebo do provozu podle harmonizačních právních předpisů Evropské unie uvedených v příloze II. To platí bez ohledu na to, zda je systém umělé inteligence uváděn na trh nebo do provozu nezávisle na jiném uvedeném výrobku.

Za druhé jsou potom za vysoce rizikové systémy považovány samostatné systémy umělé inteligence, jejichž zamýšlené použití je pro konkrétní účel v rámci jedné z oblastí, které jsou

vyjmenované v příloze III, kterými jsou: a) biometrická identifikace a kategorizace, b) řízení a provozování kritické infrastruktury, c) vzdělávání a odborná příprava, d) zaměstnanost, správa pracovníků a přístup k samostatné výdělečné činnosti, e) přístup k základním soukromým a veřejným službám a dávkám, f) vymáhání práva, g) migrace, azyl a správa hraničních kontrol a h) správa soudnictví a demokratické procesy, a které představují riziko újmy na zdraví a bezpečnosti nebo riziko nepříznivého dopadu na základní práva.

Jak již bylo výše v textu popsáno, Návrh nařízeni je koncipován tak, aby poskytoval ochranu jak před riziky vztahujícími se k bezpečnosti, tak i před riziky pro základní práva. Ačkoli by tak měly čl. 6 a 7 pokrývat oba typy rizik stejným způsobem, je podle některých autorů zcela zřejmé (Mazzini a Scalzo 2021, Smuha et al. 2021, Mahler 2022), že čl. 6 odst. 1 ve spojení s právními předpisy Evropské unie o bezpečnosti výrobků (nový regulační rámec) pokrývá primárně bezpečnostní rizika, zatímco čl. 6 odst. 2 ve spojení s přílohou III a čl. 7 potom především rizika pro základní práva.

Metodika a kritéria pro výběr oblastí a účelů použití samostatných vysoce rizikových systémů umělé inteligence uvedených v příloze III jsou vysvětleny ve zprávě o posouzení dopadů, která samostatně doprovázela Návrh nařízení. Stručně řečeno, Evropská komise prošla seznamy systémů umělé inteligence, které byly v různých zprávách označeny jako problematické. Poté posoudila pravděpodobnost a závažnost újmy, a na základě výsledků tohoto posouzení stanovila, zda systém umělé inteligence vytváří vysoké riziko pro zdraví a bezpečnost a pro základní práva a svobody osob.

Kritéria zohledněná při zařazení konkrétního systému umělé inteligence na seznam vysoce rizikových systémů (nad rámec systémů, na které se vztahují právní předpisy o bezpečnosti výrobků), které jsou uvedené v příloze III Návrhu nařízení, jsou ve značné míře podobná kritériím, která mají být použita při budoucích aktualizacích této přílohy, a jejichž výčet je uveden v čl. 7 odst. 2. Tato pravomoc týkající se změn uvedené přílohy je podle čl. 7 odst. 1 svěřena Evropské komisi.

Pro účely změn přílohy III se zvažují například takové otázky, jako to, zda je systém umělé inteligence již využíván nebo zda se jeho využití teprve chystá, rozsah toho, zda a do jaké míry již jeho používání způsobilo určitou újmu a zda toto použití mělo nepříznivý dopad na základní práva, jak velké skupiny osob se může újma nebo nepříznivý dopad týkat, jaká je zranitelnost potenciálně zasažených osob, dostupnost a účinnost právních prostředků nápravy ve vztahu k rizikům, který daný systém umělé inteligence představuje, a do jaké míry jsou stávající právní předpisy Evropské unie schopny jednak zabraňovat vzniku rizik, případně tato rizika způsobená systémy umělé inteligence minimalizovat.

5.3. Podmínky pro uvedení vysoce rizikového systému na trh, přehled fází

V Návrhu nařízení jsou z pohledu vysoce rizikových systémů umělé inteligence upraveny dvě situace, a to situace před uvedením konkrétního systému na trh (jeho návrh a vývoj) a situace po jeho uvedení na trh. V první fázi před uvedením na trh klade Návrh nařízení na systémy umělé inteligence řadu požadavků, které musí být dodrženy. V Návrhu nařízení jsou tyto základní požadavky stanoveny v čl. 8 až čl. 15 (hlava III, kapitola 2), které zahrnují požadavky na a) zavedení systému řízení rizik, b) kvality souborů dat používaných pro vývoj systémů, c) vypracování technické dokumentace, d) vedení záznamů, které umožní sledovat fungování systému, e) transparentnosti fungovaní, f) lidského dohledu, a g) přesnosti, spolehlivosti a kybernetické bezpečnosti.

Ustanovení uvedených článků jsou určena poskytovatelům vysoce rizikových systémů umělé inteligence, jak vyplývá ze spojení těchto článků se zněním čl. 16 písm. a), kterými jsou podle čl. 3 odst. 2 fyzické nebo právnické osoby, které vyvíjí nebo nechávají vyvíjet systém umělé inteligence za účelem jeho uvedení na trh nebo do provozu pod svým vlastním jménem nebo ochrannou známkou.

Úprava druhé fáze stanovená Návrhem nařízení spočívá v zavedení komplexního postupu monitorování a kontroly vysoce rizikových systémů umělé inteligence po uvedení na trh vnitrostátními orgány dozoru, a dále stanovením dalších povinností pro poskytovatele systémů umělé inteligence, jako jsou například ohlašování závažných incidentů a chybných fungování těchto systémů, které mají zajistit, že budou odstraněna rizika, která nebyla odstraněna v prvotním kroku, případně která se objevila až v průběhu jejich využívání.

Mezi povinnosti, které poskytovatelům ukládá v rámci jednotlivých fází Návrh nařízení, patří zejména následující: 1) Posouzení shody ex ante: podle čl. 19 musí poskytovatelé zajistit, aby systém vysoce rizikové umělé inteligence prošel postupem posouzení shody (přílohy VI a VII) před uvedením systému na trh nebo do provozu. 2) Systém řízení kvality: poskytovatelé musí zavést systém řízení kvality, který je dokumentován formou písemných politik, postupů a pokynů pro zajištění souladu s nařízením, který mimo jiné zahrnuje například vyšetřovací, testovací a ověřovací postupy prováděné před vývojem vysoce rizikového systému umělé inteligence, během něho a po něm, a četnost, s níž musí být prováděny (čl. 17 odst. 1 písm. d)), a současně tento systém aktualizovat po celou dobu životnosti systému umělé inteligence. 3) Registrace: poskytovatelé musí před uvedením systému na trh nebo do provozu zaregistrovat všechny samostatné vysoce rizikové systémy umělé inteligence v databázi pro celou Evropskou unii (čl. 16 písm. f), čl. 51, čl. 60, příloha VIII). Informace obsažené v této databázi Evropské unie musí být přístupné veřejnosti (čl. 60 odst. 3 a 4) Sledování po uvedení na trh: poskytovatelé

jsou povinni vytvořit, uplatňovat a udržovat systém monitorování po uvedení na trh (čl. 17 odst. 1 písm. h), čl. 61 odst. 1). Systém monitorování musí aktivně a systematicky shromažďovat, dokumentovat a analyzovat příslušné údaje týkající se výkonnosti vysoce rizikových systémů umělé inteligence po celou dobu jejich životnosti, a to z důvodu, aby poskytovatel mohl průběžně posuzovat soulad systémů umělé inteligence s požadavky stanovenými nařízením (čl. 61 odst. 2). Pokud má poskytovatel důvod se domnívat, že konkrétní vysoce rizikový systém umělé inteligence není ve shodě s nařízením, musí okamžitě přijmout nezbytná nápravná opatření k uvedení daného systému ve shodu, nebo k jeho případnému stažení z trhu či oběhu (čl. 16 písm. g, čl. 21). (5) Informační povinnost vůči příslušným orgánům: pokud by vysoce rizikový systém mohl nepříznivě ovlivnit zdraví, bezpečnost nebo ochranu základních práv osob nad míru považovanou za důvodnou a přijatelnou vzhledem k jeho určenému účelu nebo za běžných nebo rozumně předvídatelných podmínek používání, a toto riziko je poskytovateli známo, musí poskytovatel okamžitě informovat příslušné vnitrostátní orgány, zejména o nesouladu a o veškerých přijatých nápravných opatřeních (čl. 22, čl. 65 odst. 1, čl. 3 bod 19) Nařízení (EU) 2019/1020).

5.4. Systém řízení rizik (obecně), jeho části a harmonizované normy

Jedním z klíčových prvků, který se prolíná spletitou strukturou Návrhu nařízení a zaváděnými procesy, je systém řízení rizik uvedený v čl. 9, který vystupuje jako součást obou hlavních povinností poskytovatelů vysoce rizikových systémů umělé inteligence, kterou je provedení posouzení shody (čl. 19 a 43, přílohy VI a VII) a přijetí systému řízení kvality. To je dáno tím, že proces posouzení shody je založený na požadavcích stanovených v příloze VI, která vyžaduje zavedený systém řízení kvality v souladu s čl. 17, jehož hlavní složkou je systém řízení rizik (čl.17 odst. 1 písm. g)).

Jak již bylo zmíněno, pro to, aby se snížila rizika spojená s vysoce rizikovými systémy umělé inteligence, musí poskytovatelé těchto systémů splňovat požadavky stanovené v kapitole 2. Návrh nařízení ale současně předpokládá, že i když poskytovatelé vysoce rizikových systémů umělé inteligence splní všechny dané požadavky, nemusí to zcela stačit ke snížení všech rizik na přijatelnou úroveň, a některá z těchto rizik zůstanou zachována. Účel čl. 9 tak spočívá v zajištění toho, aby poskytovatelé vysoce rizikových systémů tato rizika identifikovali a přijali další opatření k jejich snížení, a to alespoň na přijatelnou úroveň. Z tohoto pohledu pak čl. 9 plní důležitou záložní funkci (Kaminsky, 2023).

Konstrukce čl. 9 je následující. V odstavci 1 stanovuje ústřední požadavek, podle něhož musí poskytovatelé vysoce rizikových systémů umělé inteligence zavést, uplatňovat, zdokumentovat a udržovat systém řízení rizik.

Následující odstavce 2 až 7 potom specifikují podrobnosti tohoto systému. Systém řízení rizik tak, jak je zachycen v tomto článku, se skládá ze dvou částí, a to procesu řízení rizik, který upravují odstavce 2 až 4, a procesu testování, který zahrnují odstavce 5 až 7. Zbývající dva odstavce čl. 9, a to odstavce 8 a 9, obsahují zvláštní pravidla pro systémy, ke kterým budou mít přístup děti, a které budou zavádět úvěrové instituce.

Jak je patrné z celé podoby struktury Návrhu nařízení, sehrávají v jejím rámci významnou roli především (technické) harmonizované normy (McFadden et al., 2021). Podle čl. 2 odst. 1 písm. c) nařízení (EU) č. 1025/2012 se harmonizovanou normou rozumí „ evropská norma přijatá na základě žádosti Evropské komise za účelem uplatňování harmonizačních právních předpisů Evropské unie“, nařízení (EU) č. 1025/2012).

To se samozřejmě týká i systému řízení rizik. Na tyto harmonizované normy je přímo odkazováno v odstavci 3 zmíněného článku, kde jsou vedle těchto norem týkajících se řízení rizik zmíněny současně společné specifikace. Podle čl. 3 bod 27 je společnou specifikací dokument jiný než norma, který obsahuje technická řešení sloužící jako nástroj pro plnění některých požadavků a povinností stanovených v Návrhu nařízení.

Harmonizované normy mohou podle Návrhu nařízení sloužit poskytovatelům k prokázání souladu s požadavky stanovenými Návrhem nařízení, pokud jsou jimi dodrženy, v tomto případě se tak shoda předpokládá (čl. 40).

V situaci neexistence harmonizovaných norem nebo jejich nedostatečnosti může Evropská komise vypracovat společnou specifikaci, která v případě, že ji poskytovatel dodrží, vede ke stejnému výsledku jako by poskytovatel dodržel harmonizované normy, čímž je myšlen předpoklad shody.

Takovéto harmonizované normy a společné specifikace týkající se systému řízení rizik v oblasti umělé inteligence dosud neexistují. Avšak vytvořením těchto norem pro účely aktu o umělé inteligenci, a to včetně systémů řízení rizik, již byly pověřeny Evropské normalizační orgány (Veale a Borgesius, 2021). Do té doby mohou sloužit jako určité vodítko pro nastavování systému řízení rizik mezinárodní normy (ISO/IEC 23894.43), přičemž někteří autoři (Schuett 2023b) očekávají, že výsledná podoba harmonizovaných norem bude s těmito mezinárodními normami ve velké míře kompatibilní.

6. Jednotlivé prvky systému řízení rizik

Následující část textu se blíže zaměřuje na celkové znění čl. 9 Návrhu nařízení. Ústředním požadavkem ve vztahu k vysoce rizikovým systémům stanoveným v odstavci 1 je systém řízení rizik, který bude podle znění tohoto článku zaveden, uplatňován, zdokumentován a udržován. Představu o tom, co je tímto systémem myšleno, poskytuje odstavec 8 tohoto článku, podle jehož znění je tento systém popsán v odstavcích 1 až 7, a zahrnuje tak v sobě dva procesy, a to proces řízení rizik popsaný v odstavcích 2 až 4, a proces testování popsaný v odstavcích 5 až 7. Současně lze říci, že se tento systém skládá z politik, postupů a pokynů.

Návrh nenabízí vysvětlení toho, co se myslí tím, když stanovuje, že systém řízení rizik bude zaveden, uplatňován, zdokumentován a udržován. Schuett (2023b), který pro interpretaci těchto pojmů využil zavedených mezinárodních norem, konkrétně ISO 31000:2018 Risk management – Guidelines, uvádí, že systém řízení rizik je zaveden, pokud jsou vytvořeny zásady, postupy a pokyny pro řízení rizik, a ty jsou schváleny odpovědnými osobami s příslušnou rozhodovací pravomocí. Je uplatňován, pokud je uveden do praxe, což znamená, že příslušní zaměstnanci chápou, co se od nich očekává, a podle toho jednají. Je zdokumentován, pokud je systém systematicky a přehledně popsán, a to ve formě písemných zásad, postupů a instrukcí a lze jej na žádost příslušného vnitrostátního orgánu prokázat. Je udržován, pokud je pravidelně přezkoumáván v rámci celého životního cyklu vysoce rizikového systému, a v případě potřeby je aktualizován.

6.1. Proces řízení rizik, čl. 9 odst 2

První z prvků, ze kterých je složen systém řízení rizik, je popsán v odstavci 2, a tím je proces řízení rizik. Hlavními kroky tohoto procesu popsanými v odstavci 2 jsou identifikace rizik, odhad/hodnocení rizik a přijetí opatření k řízení a zmírňování rizik. Odstavce 3 a 4 potom obsahují další podrobnosti a upřesnění co se týče opatření k řízení rizik a zmírňování rizik. Většina pojmů, které se pojí k celému procesu řízení rizik, není v Návrhu nařízení definována. Jak ale upozorňuje Schuett (2023b), je výsledná podoba tohoto procesu v Návrhu nařízení inspirována mezinárodními normami ISO/IEC Guide 51:2014 Safety aspects — Guidelines for their inclusion in standards, a lze z nich tak vycházet při interpretaci ustanovení věnovaných tomuto procesu. Obdobným způsobem přistupuje k interpretaci ustanovení věnovaných procesu řízení rizik i jiní autoři (McFadden et al. 2021, Fraser a Bello y Villarino, 2021).

Z pohledu formulace písmena a) tohoto odstavce je problematické zejména to, že spojuje identifikaci a analýzu rizik v jeden krok, přičemž písmena b) a c) odstavce 2, které po něm

následují, jsou charakteristikou a popisem obsahu analýzy rizik tak, jak je pojímána mezinárodními normami ISO.

6.1.1 Identifikace rizik

Samotný proces řízení rizik dle písmena a) začíná identifikací rizik. Identifikace rizik podle Schuetta (2023b) znamená systematické využívání dostupných informací k identifikaci nebezpečí, kdy nebezpečí lze definovat jako potenciální zdroj škody.

Identifikace rizik by dle následujícího textu v písmenu a) měla být zaměřena na známá a předvídatelná rizika, přičemž riziko je kombinací pravděpodobnosti výskytu škody a závažnosti této škody, kdy škoda znamená jakýkoli nepříznivý účinek na zdraví, bezpečnost a základní práva. Jinak řečeno je kombinací pravděpodobnosti, že se něco stane, a následku, s jakým se něco stane.

Známým rizikem je potom takové riziko, ke kterému již v minulosti došlo, nebo je zcela jisté, že k němu dojde v budoucnosti. Riziko by například mělo být považováno za známé, pokud existuje příslušný záznam v některé z veřejně přístupných databází incidentů. Ze strany poskytovatelů se pak očekává, že vynaloží přiměřené úsilí, aby se dozvěděli o všech potenciálních rizicích, a neznamená to tedy, že budou vycházet pouze z toho, co skutečně ví (Mahler, 2022) .

To, že je riziko předvídatelné (Schuett, 2023b), pak odkazuje k tomu, že riziko ještě nenastalo, ale lze ho již identifikovat. V rámci této otázky o předvídatelnosti rizik má jít především o kompromis mezi tím, kdy poskytovatel vysoce rizikového systému umělé inteligence může přestat hledat nová rizika, a zabráněním situacím, kdy poskytovatel způsobí významnou škodu, ale může se vyvinit z odpovědnosti za použití argumentu, že riziko bylo nepředvídatelné. Možné řešení této otázky je postavené na tom, že čím větší jsou potenciální dopady rizika, tím větší úsilí musí poskytovatel vynaložit na jejich předvídání.

6.1.2 Odhad a hodnocení rizik

Obsah písmen b) a c) je věnován odhadu a hodnocení rizik, která mohou vzniknout při používání v souladu s určeným účelem nebo při předvídatelném nesprávném použití, nebo která byla identifikována během monitorování po uvedení na trh. Odhad rizika znamená proces, ve kterém se odhaduje kombinace pravděpodobnosti vzniku škody a závažnosti této škody. Po tomto procesu pak navazuje jeho hodnocení a za pomoci různých parametrů (např.

pravděpodobné důsledky rizika, stanovení dopadu rizik a vyhodnocení celkové přesnosti odhadu rizika) se zjišťuje, zda je riziko přijatelné (McFadden et al. 2021).

Tento krok v rámci procesu řízení rizik, který zahrnuje odhad rizik a jejich hodnocení, by se měl vztahovat pouze na rizika, která mohou vzniknout, když je vysoce rizikový systém umělé inteligence používán v souladu s určeným účelem a za podmínek důvodně předvídatelného nesprávného použití. Přičemž určeným účelem je myšleno použití systému umělé inteligence určené poskytovatelem uvedené například v návodu použití (čl. 3 bod 12) a důvodně předvídatelným nesprávným použitím je použití, které není v souladu s jeho určeným účelem, avšak může vyplývat z důvodně předvídatelného lidského chování (čl. 3 bod 13). Z pohledu procesu řízení rizik to tedy znamená, že poskytovatelé musí na počátku tohoto procesu stanovit, kdo jsou potenciální uživatelé systému (čl. 3 bod 4), jaká jsou zamýšlená použití a okruh rozumně předvídatelných zneužití.

Jak je uvedeno v písmenu c) tohoto článku, poskytovatelé vysoce rizikových systémů umělé inteligence musí rovněž vyhodnotit rizika, která identifikovali z údajů sebraných systémem monitorování po uvedení na trh. Toto ustanovení zajišťuje, aby poskytovatelé rovněž věnovali pozornost řízení rizik vyplývajících z nepředvídatelného nesprávného použití, pokud mají údaje o tom, že k takovémuto použití dochází.

6.1.3 Přijetí vhodných opatření k řízení rizik

Po vyhodnocení rizik by měl v případě, že výsledek tohoto procesu je, že existuje konkrétní riziko, které je nepřijatelné (je vyšší než nejvyšší přípustné riziko), následovat krok v podobě přijetí vhodných opatření k řízení rizik. V případě, že po vyhodnocení rizik jsou všechna rizika vyhodnocena jako přijatelná, může tento proces poskytovatel vysoce rizikového systému ukončit. V takovémto případě rozhodnutí zdokumentuje a proces tím tak končí.

Pokud poskytovatel systému přistoupil k třetímu kroku a přijal vhodná opatření ke snížení rizik (například v podobě technického řešení), musí znovu vyhodnotit riziko a rozhodnout, zda je celkové zbytkové riziko přijatelné. Pokud není, musí přijmout další opatření ke snížení rizik. Pokud se však ukáže, že není možné zbytkové riziko snížit pod hranici nejvyššího přípustného rizika, musí být proces vývoje a nasazení vysoce rizikového systému umělé inteligence zastaven (variantu zastavení Návrh nařízení nezmiňuje, ale zdá se jako logické vyústění těchto procesů). Z pohledu četnosti opakování tohoto procesu je staveno, že tento proces musí být prováděn v rámci celého životního cyklu systému, což znamená, že tento proces

musí proběhnout minimálně dvakrát, a to jednou během vývoje systému a podruhé před samotným uvedením na trh (McFadden et al.2021).

Odstavce 3 a 4 potom navazují na odstavec 2 písm. d) v tom, že se podrobněji věnují a rozvádí v něm zmíněné opatření k řízení rizik. Odstavec 3 uvádí, že opatření k řízení rizik věnují náležitou pozornost účinkům a možným interakcím vyplývajícím z kombinovaného uplatňování požadavků stanovených v kapitole 2 a zohledňují obecně uznávaný nejnovější vývoj, včetně toho, jak se odráží v příslušných harmonizovaných normách nebo společných specifikacích. Tyto harmonizované normy nebo společné specifikace týkající se řízení rizik v oblasti umělé inteligence však v současnosti neexistují.

V prvním pododstavci odstavce 4 je uveden účel přijímání opatření k řízení rizik, kterým je zmírňování rizik na takovou úroveň, aby bylo případné zbytkové riziko (riziko, které zůstává po použití všech ochranných opatření) považováno za přijatelné. Schuett (2022), který k interpretaci pojmu přijatelného rizika používá mezinárodní normy ISO/IEC 23894, uvádí, že tento pojem lze definovat jako úroveň rizika, které je v daném kontextu přijatelné a odpovídá současným hodnotám společnosti, a pokračuje tím, že k určení toho, zda je riziko přijatelné, musí poskytovatel zvážit přínosy a rizika. Obecně pak dle něj platí, že riziko je přijatelné, pokud přínosy převažují nad riziky.

Druhý pododstavec odstavce 4 uvádí konkrétní opatření k řízení rizik, které musí poskytovatel vysoce rizikového systému umělé inteligence přijmout. Za prvé (bod a)) jím je bezpečný návrh a vývoj systému, prostřednictvím něhož poskytovatel zajistí vyloučení rizik nebo jejich snížení na nejnižší možnou míru. Za druhé (bod b)), pokud rizika nejde vyloučit, musí poskytovatel ve vhodných případech zavést odpovídající zmírňující a kontrolní opatření. Za třetí (bod c)), poskytovatel musí poskytovat odpovídající informace uživatelům a v případě potřeby jim zajistit školení.

Na závěr pak třetí pododstavec odstavce 4 uvádí, že při přijímání výše uvedených opatření k řízení rizik, a to s cílem vyloučit nebo snížit rizika spojená s používáním systému, musí poskytovatelé náležitě zvážit technické znalosti, zkušenosti, vzdělání, školení, které může uživatel očekávat, a případně prostředí, ve kterém má být systém používán.

6.2. Proces testování

Druhou částí systému řízení rizik je proces testování, který je popsán v čl. 9 odst. 5 až 7 Návrhu nařízení. Odstavec 5 popisuje tři účely, které má testovací proces naplňovat.

Prvním účelem testování vysoce rizikových systémů umělé inteligence je identifikace nejvhodnějších opatření k řízení rizik. Poskytovatel prostřednictvím testování získává lepší porozumění rizikům, což by mu mělo pomoci při výběru vhodného opatření k řízení rizik.

Druhý účel spočívá v tom, že testování zajišťuje, aby vysoce rizikové systémy umělé inteligence podávaly výkony konzistentní s jejich určeným účelem. K tomu Schuett (2023b) uvádí, že systémy umělé inteligence často fungují hůře, pokud se tréninkové prostředí liší od prostředí, v němž jsou následně jsou skutečně používány. Testování by tak mělo poskytovatelům pomoci odhalit, kdy je obzvláště pravděpodobné, že systém bude fungovat špatně v prostředí, pro které je určen. Za třetí by mělo testování zajistit, aby vysoce rizikový systém byl v souladu s požadavky stanovenými v čl. 10 až 15 Návrhu nařízení.

Ustanovení odst. 6 uvádí, že zkušební postupy musí být vhodné k dosažení určeného účelu systému, a nemusí překračovat rámec toho, co je pro dosažení tohoto účelu nezbytné. V podstatě se tak jedná o přeformulování zásady proporcionality (Schuett, 2023b).

V odst. 7 je stanovena povinnost poskytovateli vysoce rizikového systému, kdy musí systém testovat, a to v každém případě před uvedením systému na trh nebo do provozu, a dále kdykoli podle potřeby v průběhu celého procesu vývoje. Na rozdíl od procesu řízení rizik, který musí být prováděn v rámci celého životního cyklu, testování je třeba provádět pouze do bodu před uvedením systému na trh (čl. 3 bod 9) nebo do provozu (čl. 3 bod 11).

Dále je v odst. 7 stanoveno, jakým způsobem musí poskytovatelé testovat svůj vysoce rizikový systém umělé inteligence, a to na základě předem definovaných měřítek a pravděpodobnostních prahových hodnot, které jsou vhodné pro účel vysoce rizikového systému umělé inteligence. Měřítka zahrnují kritéria hodnocení, srovnávací ukazatele a ukazatele výkonnosti. Pravděpodobnostní prahové hodnoty pak představují zvláštní druh měřítka hodnotící vlastnost na pravděpodobnostní stupnici s jednou nebo více předem definovanými prahovými hodnotami (McFadden et al. 2021).

Odst. 8 stanovuje, že zvláštní pozornost při zavádění systému řízení rizik musí být věnována tomu, zda je pravděpodobné, že k danému vysoce rizikovému systému umělé inteligence budou mít přístup děti nebo zda bude mít na ně dopad. Děti mají v Návrhu nařízení zvláštní postavení, a to z důvodu, že jsou obzvláště zranitelné a mají zvláštní práva (bod 28 odůvodnění). Proto jsou poskytovatelé vysoce rizikových systémů umělé inteligence povinni přijmout zvláštní opatření na jejich ochranu.

Závěrečný odstavec čl. 9 Návrhu nařízení se věnuje zavádění systému řízení rizik úvěrovými institucemi, kterým je tato povinnost již uložena v čl. 74 směrnice 2013/36/EU, ve

kterém je stanoveno, že stávající systémy řízení rizik úvěrových institucí budou doplněny o aspekty, které jsou uvedeny v tomto článku pod odst. 1 až 8.

6.3. Proces posuzování shody, čl.19

Následující text se věnuje dalšímu z procesů, který je v Návrhu nařízeni popsán, a kterým je proces posuzování shody, ve kterém stejně jako v rámci systému řízení rizik sehrávají významnou roli harmonizované standardy. Posouzení shody představuje jednu z hlavních povinností poskytovatelů vysoce rizikových systémů umělé inteligence před uvedením systému na trh nebo do provozu. Posuzování shody je v Návrhu nařízení definováno jako postup ověřování toho, zda byly splněny požadavky stanovené v hlavě III kapitole 2 tohoto nařízení, týkající se systému umělé inteligence (čl. 3 bod 20).

Požadavky, které Návrh nařízení stanoví pro vysoce rizikové systémy umělé inteligence jsou nezbytné k účinnému zmírnění rizik pro zdraví, bezpečnost a základní práva a týkají se 1) dat a správy dat (čl.10): pokud je vysoce rizikový systém umělé inteligence vyvíjen za pomocí dat (tréninkový soubor, validační soubor, testovací soubor), musí být tato data kvalitní.

K naplnění tohoto požadavku na soubor dat je nutné zavedení vhodných postupů správy a řízení dat, které se týkají zejména a) možností návrhu, b) sběru dat, c) přípravy dat, d) předpokladů týkajících se toho, co daná data měří a představují, e) posouzení dostupnosti, množství a vhodnosti souborů dat, f) zkoumání s ohledem na potenciální zkreslení, g) identifikace nedostatků, mezer a chyb.

Kromě toho musí být soubory dat relevantní, reprezentativní, bez chyb a úplné. Musí také mít příslušné statistické vlastnosti, pokud jde o osoby, pro které má být daný systém používán. A musí zohledňovat zeměpisné, behaviorální nebo funkční prostředí, v němž má být systém umělé inteligence používán.

2) technické dokumentace (čl.11): ta musí být vypracována před uvedením vysoce rizikového systému na trh nebo do provozu. Technická data musí být průběžně aktualizována. Technická dokumentace musí obsahovat informace nezbytné k posouzení souladu daného systému umělé inteligence s požadavky stanovenými v kapitole 2. Výčet těchto informací je stanoven v příloze VI.

3) vedení záznamů (čl. 12): a to v podobě automatického zaznamenávání událostí, které mají zajistit to, aby byly po celou dobu životnosti vysoce rizikového systému umělé inteligence k dispozici informace o tom, jak funguje, aby bylo možné monitorovat jeho činnost.

4) transparentnosti (čl.13): vyžaduje se, aby poskytovatelé navrhovali a vyvíjeli vysoce rizikového systémy umělé inteligence tak, aby systémy umožňovaly uživatelům interpretovat výstupy systému a vhodně jej používat.

Dále poskytovatelé musí poskytovat informace, které jsou stručné, úplné, správné a jasné, a to zejména pokud jde o vlastnosti, schopnosti a omezení výkonnosti vysoce rizikového systému umělé inteligence včetně zamýšleného účelu a také všech známých nebo předvídatelných okolností včetně předvídatelného nesprávného použití, které mohou vést k rizikům pro zdraví a bezpečnost nebo základní práva, stejně jako výkonnosti ve vztahu k osobám, na něž má být systém používán.

5) Lidského dohledu (čl.14): Vysoce rizikové systémy umělé inteligence musí být navrženy a vyvinuty tak, aby na ně mohly během období používání účinně dohlížet fyzické osoby. Dohled by se měl zaměřovat na problém nadměrného spoléhání na výstup vysoce rizikových systémů umělé inteligence, vyhledávání anomálií nebo známek nesprávného fungování, na rozhodnutí, zda použít výstup systému nebo tento vystup zvrátit.

Dále musí být umožněno pověřené osobě zasahovat do fungování systému včetně jeho přerušení pomocí tlačítka stop. 6) přesnosti, spolehlivosti a kybernetické bezpečnosti (článek 15, bod odůvodnění 49 až 51): klíčovým požadavkem na vysoce rizikové systémy umělé inteligence je technická spolehlivost.

Tyto systémy musí být v průběhu celého životního cyklu odolné vůči rizikům souvisejícím s omezeními systému (např. chyby, poruchy, nekonzistentnost, neočekávané situace), jakož i vůči svévolným zásahům, které mohou ohrozit bezpečnost systému umělé inteligence a vést ke škodlivému nebo jinak nežádoucímu chování.

Jak bylo uvedeno výše, v rámci přístupu, který zaujímá nový legislativní rámec je zapojení třetí strany do postupu posuzování shody závislé na rizikovosti daného výrobku. Obecně pak platí, že vysoce rizikové výrobky podléhají posuzování shody třetí stranou (Mazzini a Scalzo (2022).

Přístup, který zvolil Návrh nařízení k postupu posuzování shody se však od tohoto zásadně liší a zakládá se v případě vysoce rizikových systémů na kombinaci vlastní kontroly souladu s požadavky ze strany poskytovatelů s předpokladem shody (čl. 40), pokud poskytovatel dodržuje harmonizované normy (jak již bylo řečeno výše, tyto normy, mají být vypracovány evropskými normalizačními orgány).

Takto nastavený postup se vztahuje k samostatným vysoce rizikovým systémům uvedeným v příloze III (systémy, které nejsou součástí regulovaných výrobků, které podléhají novému legislativnímu rámci a musí tak vždy projít posouzením shody třetí stranou podle

stávajících odvětvových požadavků) s jedinou velice úzkou výjimkou, kterou představují systémy biometrické identifikace (čl. 43 odst.1, příloha VII), které v případě, že nejsou zakázány, musí být podrobeny posouzení shody třetím subjektem, pokud neexistují harmonizované normy nebo společné specifikace.

V souladu s tím je znění čl. 43 odst. 2, který uvádí, že poskytovatelé by měli uplatňovat postup posuzování shody založený na vnitřní kontrole, uvedené v příloze VI. Podle této přílohy musí poskytovatel ověřit, zda zavedený systém řízení kvality splňuje požadavky čl. 17.

Kromě toho musí poskytovatel přezkoumat informace obsažené v technické dokumentaci, aby posoudil soulad systému umělé inteligence s příslušnými základními požadavky stanovenými v hlavě III kapitole 2. Na závěr musí poskytovatel ověřit, zda je proces návrhu a vývoje systému umělé inteligence a jeho monitorování po uvedení na trh (čl.61) v souladu s technickou dokumentací.

Po úspěšném posouzení shody vypracuje poskytovatel pro daný systém umělé inteligence písemné EU prohlášení o shodě a po dobu deseti let je uchovává pro potřeby příslušných vnitrostátních orgánů (čl. 19 odst. 1 a 2 a čl. 48 odst. 1 a 2). Spolu s tímto prohlášením o shodě jsou poskytovatelé rovněž povinni umístit na systémy umělé inteligence s vysokým rizikem označení shody CE a pokud to není možné, tak dle potřeby na obal nebo případně průvodní doklady, a to v souladu s čl. 49 a čl. 30 nařízení (ES) č. 765/2008. Celý tento proces by měl poskytovatel systému, ukončit registrováním vysoce rizikového systému do databáze EU obsahující samostatné vysoce rizikové systémy (čl.60), kterou zřídí a bude spravovat Evropská komise.

Vedle takto nastaveného způsobu provádění kontroly souladu s požadavky stanovenými Návrhem nařízením. Stanovuje čl. 63 následný dozor nad trhem, který má být prováděn vnitrostátními orgány členských státu podle nařízení (EU) 2019/1020.

Podle odůvodnění Návrhu nařízení by při uplatňování a prosazování předkládaného nařízení měly hrát klíčovou roli členské státy (bod 77), které však nemusí vytvářet nový specializovaný regulační orgán a místo toho se očekává, že ho určí ze stávajících vnitrostátních orgánů, který pak bude plnit roli vnitrostátního dozorového orgánu nad trhem pro účely prosazování Návrhu nařízení.

Tyto vnitrostátní orgány budou mít přístup ke všem potřebným údajům a dokumentacím, a to včetně přístupu ke zdrojovému kódu, bude-li to nezbytné k posouzení shody vysoce rizikového systému umělé inteligence s požadavky uvedenými v hlavě III kapitole 2 (čl. 64 odst. 1 a 2).

6.4. Sankce

Za účelem prosazování nařízení musí členské státy stanovit pravidla pro ukládání sankcí, včetně správních pokut použitelných v případě jakéhokoliv porušení Návrhu nařízení, přičemž sankce musí zohlednit zejména zájmy malých poskytovatelů a začínajících podniků, jakož i jejich ekonomickou životaschopnost (čl. 71 odst. 1). Čl. 71 stanovuje výši pokut v rozmezí 2 %, 4 % nebo 6 % celkového ročního obratu celosvětově, a to v závislosti na okolnostech konkrétní situace, s přihlédnutím zejména k povaze, závažnosti a době trvání porušení povinností a jeho následkům (čl. 71 odst. 6).

6.5. Evropská rada pro umělou inteligenci, čl. 57

Vedle toho by měla být na základě tohoto nařízení zřízena Evropské rada pro umělou inteligenci, která by měla usnadnit harmonizované provádění nařízení. Radě by předsedala Evropská komise a jejími členy by byli zástupci všech vnitrostátních dozorových úřadů spolu s evropský inspektor ochrany údajů (čl. 57 odst. 1 a 3). Navrhované nařízení nesvěřuje radě žádné pravomoci týkající se prosazování nařízení. Místo toho by mělo být hlavním úkolem rady vydávání stanovisek a doporučení k záležitostem souvisejícím s prováděním nařízení, a to zejména k harmonizovaným normám a společným specifikacím (článek 58).

Následující kapitola se věnuje popisu dvou pater pyramidy rizik současně, a to úrovni systémů umělé inteligence s nízkým rizikem a úrovní systémů umělé inteligence s minimálním rizikem. K tématu vysoce rizikových systémů a shrnutí problematických bodů v rámci jejich regulace, která je předložena v Návrhu nařízení, se text vrací v závěru celé práce.

7. Použití systému umělé inteligence, které vytváří nízké riziko a minimální riziko

Úroveň pyramidy, která zahrnuje systémy umělé inteligence s nízkým rizikem, představuje vrstvu, která zahrnuje některé systémy umělé inteligence, které jsou vysoce rizikové, a některé, které rizikové nejsou (čl. 51 odst. 4). Určující charakteristikou systémů umělé inteligence, které spadají do této kategorie je, že vyvolávají určité problémy z hlediska transparentnosti a manipulace, a proto jsou na ně v Návrhu nařízení kladeny zvláštní požadavky. Návrh nařízení v čl. 52 vymezuje tři skupiny systému umělé inteligence, na které se vztahuje povinnost transparentnosti a) systémy, které jsou určeny ke komunikaci s lidmi (chatbot) b) systémy, které jsou určeny k rozpoznávání emocí a k biometrické kategorizaci c) systémy, které generují obsah nebo s ním manipulují (synthetic media, deep fakes). V případě systémů uvedených pod písmenem a) má poskytovatel povinnost zajistit informovaní fyzických osob o tom, že komunikují se systémem umělé inteligence. V případech systémů uvedených pod písmenem b) musí uživatel (čl. 3 bod 4) o fungování tohoto systému informovat fyzické osoby, které jsou mu vystaveny. V případech systémů uvedených pod písmenem c) uživatelé systému zveřejní, že obsah byl uměle vytvořen nebo s ním bylo manipulováno. Výjimku z toho (čl.52 odst. 3, druhý pododstavec) představují zákonem povolená použití pro účely související s prevencí trestné činnosti nebo pokud je to nezbytné, k výkonu svobody projevu a práva svobody umění a vědy.

Poslední část pyramidy se týká systémů, které představují malé nebo žádné riziko. Do této kategorie spadá každý existující systém umělé inteligence, o kterém se v návrhu výslovně nehovoří. Komise uvedla, že do této kategorie spadá naprostá většina systémů umělé inteligence, které se v současné době v Evropské unii používají.

Na tyto systémy umělé inteligence se nově zaváděné právní požadavky stanovené Návrhem nařízení vztahovat nebudou. Prostor, který je těmto systémům umělé inteligence v Návrhu nařízení věnován se nachází v čl. 69, který se věnuje zavádění a vypracování kodexů pro systémy s minimálním rizikem ze strany poskytovatelů, na základě nichž budou dobrovolně uplatňovat požadavky, které se jinak vztahují pouze na vysoce rizikové systémy umělé inteligence.

Popisem těchto dvou kategorií se uzavírá kapitola, která byla věnována tomu, jak Návrh nařízení rozřazuje rizika související s umělou inteligencí a porozumění jednotlivým kategoriím těchto rizik.

8. Závěr

Témata debaty o právu a umělé inteligenci se v posledních několika málo letech výrazně proměnila. Od původních, která dominovala právnímu poli a zaměřovala se na to, zda by měla mít umělá inteligence právní osobnost, na odpovědnost autonomních vozidel, v jakých právních profesích je možné očekávat nahrazení člověka umělou inteligencí a na stanovování etických zásad pro umělou inteligenci, se zaměření debaty posunulo ve chvíli, kdy Evropská komise učinila významný krok v rámci procesu regulace umělé inteligence, a to tím, že představila Návrh nařízení o umělé inteligenci.

Probíhající debata o umělé inteligenci se tak vyznačuje tím, že upustila od stanovování hlavních etických zásad a soft law nástrojů a zaměřuje se na stanovení právních povinností pro poskytovatele umělé inteligence a regulaci rizik souvisejících s umělou inteligenci.

Model, který zvolila Evropská komise k regulaci umělé inteligence spočívá na tom, že spojuje bezpečnost a ochranu základních práv s předem definovanou klasifikací rizik, přičemž klade důraz především na řízení jedné kategorie systémů umělé inteligence, kterými jsou vysoce rizikové systémy. Zásadním na zvoleném přístupu je, že umělá inteligence je pojímána jako výrobek, což znamenalo výrazný příspěvek do debaty o možnostech udělit umělé inteligenci právní osobnost.

Z pohledu vymezení kategorií rizik se zdá jako největší problém to, že celý systém je v podstatě postaven na dichotomii rizikový systém, nerizikový systém, kdy skupina samostatných vysoce rizikových systémů je vymezena výčtem oblastí použití, které však představují uzavřený seznam, který nelze upravit jinak, než řádným legislativním procesem, což v případě takto rychle vyvíjející se technologické oblasti může přinášet problémy (Evropská komise může zasahovat pouze do oblasti účelu použití).

Systémy umělé inteligence, které jsou fakticky mnohem rizikovější (např. vysokofrekvenční obchodování) než ty, které jsou klasifikovány v Návrhu nařízení jako vysoce rizikové, ale nespadají do vymezených oblastí v Návrhu nařízení tak zůstávají neregulované.

Klíčovou oblastí z pohledu modelu, který zavádí Návrh nařízení, by měla být ochrana lidských práv. Návrh nařízení je postaven na důvěře v systém řízení rizik, který však není doprovázen účinnými modely pro posuzování dopadu umělé inteligence na lidské práva, a to na úkor širšího zohlednění možné úlohy umělé inteligence ve společnosti.

Tato důvěra v řešení rizik prostřednictvím zavádění různých procesů se otiskuje v čl. 9 Návrhu nařízení, tedy v povinnosti poskytovatelů zavádět systém řízení rizik. Jeho cílem je snížit rizika, která umělá inteligence představuje tak, aby je bylo možné považovat za přijatelná.

Prostředky ke snížení rizik jsou návrh, vývoj, testování, zmírňující a kontrolní opatření a poskytování informací uživatelům. Namísto uvedení, která rizika mají být posouzena jako přijatelná, článek 9 předpokládá, že snížení rizik bude výsledkem řady důkladně provedených kroků. Článek 9 tak ve své podstatě dává poskytovatelům širokou diskreci v rámci rozhodování o tom, co je a není přijatelné (např. pro lidskou důstojnost).

Úspěch Návrhu nařízení při regulaci vysoce rizikových systémů, jak již bylo naznačeno, při popisu postupu řízení rizik bude primárně záviset zejména na harmonizovaných normách. Je sice pravda, že poskytovatelé se takovými normami řídit nemusí. V takové případě však musí na vlastní riziko a pro své účely provést specifikaci základních požadavku stanovených v Návrhu nařízení, což je však málo pravděpodobné, a to z důvodu vágnosti stanovených požadavků na vysoce rizikové systémy.

Dodržování harmonizovaných norem nabízí poskytovatelům jednak právní jistotu, na druhé straně snadnou cestu k dosažení souladu s požadavky, který se předpokládá.

Na základě toho se tak skutečná tvorba právních pravidel bude podle některých autorů (Veale a Borgesius, 2021, Ebers et al. 2021) soustředit v této oblasti. Formálně jsou harmonizované normy dobrovolnými pravidly vypracovanými soukromými subjekty, jako je CEN nebo CENELEC, které jsou mezinárodními neziskovými organizacemi, při bližším pohledu je však zřejmé, že tyto harmonizované (technické) normy mají závazné právní účinky, které se blíží právním normám.

Zmocnění evropské organizace pro normalizaci k vypracování harmonizovaných norem, a to ze strany Evropské komise, lze označit za delegaci normotvorných pravomocí na soukromé subjekty (Ebers et al. 2021).

Standardizace, která se váže k umělé inteligenci, jak se ukazuje není záležitostí čistě technických rozhodnutí, ale spočívá v řadě etických a právních rozhodnutí, která by neměla být svěřena soukromým organizacím pro standardizaci, a to z důvodu, že vyžadují širokou politickou diskusi, do níž se zapojí celá společnost (Smuha et al., 2021, Mantelero, 2022).

V souladu s tím by Návrh nařízení měl stanovit právně závazné povinnosti týkající se základních požadavků na vysoce rizikové systémy umělé inteligence, například jaké typy zkreslení jsou zakázány, jak zkreslení zmírňovat a jaký typ a stupeň transparentnosti by měly systémy mít.

Tato práce se v úvodu zaměřila na popis širšího kontextu, v němž je téma umělé inteligence zakotveno. Současně s tím se úvod zaměřil také na to, jak se téma umělé inteligence odráželo v právním výzkumu. Následující kapitola byla věnována diskuzi o vymezení pojmu

umělé inteligence, prostřednictvím čehož byly představeny základní okruhy problémů, které se z pohledu práva pojí s umělou inteligencí.

Nejrozsáhlejší část práce pak představuje kapitola věnovaná představení návrhu Aktu o umělé inteligenci. V textu byla věnována pozornost vzniku tohoto návrhu, obecnému přehledu obsahu a představení přístupu, z něhož návrh Nařízení vychází. Následně se jednotlivé kapitoly věnovaly popisu a představení jednotlivých kategorií rizik. Text postupoval od popisu kategorie nepřijatelného rizika ke kategorii vysokého rizika. V této kapitole byla věnována pozornost především dvěma procesům, které se vztahují k vysoce rizikovým systémům, a to procesu řízení rizik a procesu posuzování shody. Popis těchto dvou procesů měl ilustrovat především to, co Akt o umělé inteligenci doopravdy je, a to souborem popisů různě složitých procesů, které jsou vzájemně propojeny. V kapitole byly dále představeny základní inspirační zdroje pro regulaci vysoce rizikových systémů. Pokud bychom se na závěr podívali na Návrh nařízení prostřednictvím kapitoly, která se věnovala definici a základnímu okruhu problémů spojených s umělou inteligencí, pak Návrh nařízení zahrnuje všechny tyto okruhy problému a snaží se je určitým způsobem řešit. Otázkou potom je nakolik úspěšně a efektivně.

Seznam použitých zdrojů

1. Seznam použité literatury

ALLENBY, B.R. 2011. Governance and Technology Systems: The Challenge of Emerging Technologies. In MARCHANT, G.E; ALLENBY, B.R.; HERKERT J.R. (ed.). The Growing Gap Between Emerging Technologies and Legal-Ethical Oversight: The Pacing Problem. Springer Dordrecht, 2011, s. 3-18.

BARKANE, I. 2022. Questioning the EU proposal for an Artificial Intelligence Act: The need for prohibitions and a stricter approach to biometric surveillance. Information Polity, 2022, Vol. 27, No. 21, s.147-162

BARROSO, L.R. 2020. Technological revolution, democratic recession, and climate change: The limits of law in a changing world. I•CON 2020, Vol. 18, No. 2, s. 334-369.

BECCERA, S.D. 2018. The Rise of Artificial Intelligence in the Legal Field: Where We Are and Where We Are Going, The Journal of Business, Entrepreneurship & the Law 2018, Vol.11, No.1, s. 27-52.

BLACK, J.; MURRAY, A. 2019. Regulating AI and machine learning: setting the regulatory agenda. European Journal of Law and Technology 2019, Vol. 10, No.3, s.

BRADFORD, A.; 2012. The Brussels Effect. Northwestern University Law Review, 2012, Vol.107, No.1, s.1-68

BUITEN, M.C. 2019. Towards Intelligent Regulation of Artificial Intelligence. European Journal of Risk Regulation, 2019, Vol. 10, No.1, s. 41-59.

COECKELBERGH, M. 2020. AI ethics. Cambridge, MA: The MIT Press, 2020. s.229

CRAWFORD, K. 2021. Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence. Yale University Press, 2021. s. 288.

CRAWFORD, K.; CALO.R. 2016. There is a blind spot in AI research. Nature 2016, 538, s. 311–313.

DE GREGORIO, G. 2021. The rise of digital constitutionalism in the European Union. I•CON 2021, Vol. 19, No. 1, s. 41-70.

DEVILLÉ, R.; SERGEYSSELS, N.; MIDDAG, C. 2021. Basic Concepts of AI for Legal Scholars. In De Bruyne, J., Vanleenhove, C. (ed). Artificial Intelligence and the Law. Intersentia (Centrum voor Verbintenissen-en Goederenrecht), 2021, s. 1-22.

EBERS, M.; HOCH, V.R.S.; ROSENKRANZ, F.; RUSCHMEIER, H.; STEINROTTER, B. 2021. The European Commission's Proposal for an Artificial Intelligence Act—A Critical Assessment by Members of the Robotics and AI Law Society (RAILS). J 2021, Vol 4, No 4, s. 589-603.

FLORIDI, L. 2021. The European Legislation on AI: A Brief Analysis of its Philosophical Approach. Philosophy & Technology, 2021, Vol.34, No.3, s. 215-222.

FRASER, H.; BELLO y VILLARINO, H.R.J., Where Residual Risks Reside: A Comparative Approach to Art 9(4) of the European Union's Proposed AI Regulation. In Global AI + Regulation Emerging Scholars Workshop, 2021, s.1-30 (nepublikováno) dostupné na https://ssrn.com/abstract=3960461

GOANTA, C.; VAN DIJCK, G; SPANAKIS, G. 2020. Back to the Future: Waves of Legal Scholarship on Artificial Intelligence. In RANCHORDAS, S.; ROZNAI, Y. (ed.). Time, Law, and Change: An Interdisciplinary Study. Hart Publishing, 2020, s. 312-345.

HÄUSELMANN, A. 2022. Disciplines of AI: An Overview of Approaches and Techniques. In CUSTERS, B.;

FOSH-VILLARONGA, E. (ed.). Law and Artificial Intelligence: Regulating AI and Applying AI in Legal Practice. T.M.C. Asser Press The Hague, 2022, s. 44-67.

CHESTERMAN, S. 2021. We, the Robots?: Regulating Artificial Intelligence and the Limits of the Law. Cambridge: Cambridge University Press, 2021, s. 311

KAMINSKI, M. E. 2023. Regulating the Risks of AI. Forthcoming, Boston University Law Review, 2023, Vol. 103, s.1-85

KÖNIG; P. D.; KRAFFT, T. D.; SCHULZ, W.; ZWEIG, K. A. 2022. Essence of AI: What Is AI? In DIMATTEO, L. A., PONCIBO, C.; CANNARSA, M. (ed) The Cambridge Handbook of Artificial Intelligence: Global Perspectives on Law and Ethics. Cambridge University Press, 2022, s. 18-35

KRAFFT, P.M.; Young, M.; Katell, M.; Huang K.; Bugingo, G. 2020. Defining AI in Policy versus Practice. In Proceedings of the 2020 AAAI/ACM Conference on AI, Ethics, and Society (AIES ’20),2020, New York, ACM, New York,USA, s.11

MAHLER, T. 2022. Between risk management and proportionality: The risk-based approach in the EU’s Artificial Intelligence Act Proposal. Nordic Yearbook of Law and Informatics 2020–2021: Law in the Era of Artificial Intelligence, 2022, s. 247–270

MANTELERO, A.2022. Human Rights, Ethical and Social Impact Assessment in AI. T.M.C. Asser Press The Hague, 2022, s. 200

MARTINEZ, R. 2019. Artificial Intelligence: Distinguishing Between Types & Definitions. Nevada Law Journal, 2019, Vol. 19, No. 3, s. 1016-1037

MAZZINI, G.; SCALZO, S.2022. The Proposal for the Artificial Intelligence Act: Considerations around Some Key Concepts. Forthcoming in Università Ca' Foscari di Venezia - Dipartimento di Economia - Collana Centro Studi Giuridici - Wolters Kluver - CEDAM, 2022, s.29

MCFADDEN, M.; JONES K.; TAYLOR E.; OSBORN G. 2021 Harmonising Artificial Intelligence. Working paper, Oxford Information Labs, 2021, s. 43.

MOSES, L.B.; ZALNIERIUTE, M. 2020. Law and Technology in the Dimension of Time. In RANCHORDAS, S.; ROZNAI, Y. (ed.). Time, Law, and Change: An Interdisciplinary Study. Hart Publishing, 2020, s. 260–298.

NEUWIRTH, R.J. (2023) The EU Artificial Intelligence Act: Regulating Subliminal AI Systems. London. Routledge, 2023, s. 144.

PARVIAINEN, J.; COECKELBERGH, M. 2021. The political choreography of the Sophia robot: beyond robot rights and citizenship to political performances for the social robotics market. AI & Society, 2021, Vol. 36, No.3, s. 715–724.

RAPOSO, V.L.; 2022. Ex machina: preliminary critical assessment of the European Draft Act on artificial intelligence, International Journal of Law and Information Technology, 2022, Vol.30, No.1, ss. 88-109

RENDA, A.; ENGLER, A . 2023. CEPS EXPLAINER – WHAT'S IN A NAME? - Getting the definition of Artificial, Center for European Policy Studies. Belgium. 2023, s.8.

RUSSELL, S.; NORVIG, P. 2016. Artificial Intelligence: A Modern Approach, Third Edition. Global Edition. PEARSON Education Limited, 2016, s. 1152.

SHORT, J.L. 2012. The paranoid style in regulatory reform. Hastings Law Journal, 2012, Vol. 63, No. 3, s. 633-694.

SCHERER, M.U. 2016. Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies. Harvard Journal of Law & Technology, 2016, Vol. 29, No. 2, s.354-398

SCHUETT, J. 2023. Defining the scope of AI regulations. Law, Innovation and Technology, 2023, Vol. 15 No.1, s. 60-82

SCHUETT, J. 2023b. Risk Management in the Artificial Intelligence Act. European Journal of Risk Regulation. Cambridge University Press, 2023, s. 1-19

SMUHA, N. A.; RENGERS. A.; HARKENS. E; WENLONG L.; MACLAREN, J.; PISSELLI, R.; YEUNG, K. 2021 How the EU Can Achieve Legally Trustworthy AI: A Response to the European Commission's Proposal for an Artificial Intelligence http://dx.doi.org/10.2139/ssrn.3899991

SURDEN, H. 2019. Artificial Intelligence and Law: An Overview. Georgia State University Law Review, 2019. Vol. 35, No.4. s 1306-1337

ŠMAJS, J. 2008. Filosofie – obrat k Zemi: evolučně ontologická reflexe přírody, kultury, techniky a lidského poznání. Praha: Academia, 2008. s. 431.

VAEALE, M.; BORGESIUS, F.Z. 2021. Demystifying the Draft EU Artificial Intelligence Act. Computer Law Review International, 2021, Vol.22, No. 4, s. 97-112

ZUBOFF, S. 2019. The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power. Shoshana Zuboff. Profile Books. 2019. s. 691.

WENDEHORST, CH. 2021. The Proposal for an Artificial Intelligence Act COM (2021) 206 from a Consumer Policy Perspective. Study commissioned by the Austrian Federal Ministry of Social Affairs, Health, Care and Consumer Protection (2021), s. 198

2. Seznam použitých právních předpisů

Návrh nařízení Evropského parlamentu a Rady, který se stanoví harmonizovaná pravidla pro umělou inteligenci (Akt o umělé inteligenci) a mění určité legislativní akty Unie. COM (2021) 206 final (21.4.2021)

3. Seznam použitých internetových zdrojů

PAGE, L. In: Youtube [online]. 15. 5. 2013 [cit.2023-05-25]. Dostupné z: https://www.youtube.com/watch?v=AfK8h73bb-o&ab_channel=Mobilegeeks.de Kanál uživatele Mobilegeeks.de.

BARLOW, J.P., 1996 A Declaration of the Independence of Cyberspace. [online] [cit. 2023-05-25]. Dostupné z: https://www.eff.org/cyberspace-independence

Hranice lidství. Roboti se nebezpečně přibližují lidem, nejrůznější skupiny lidí naopak máte tendenci nepovažovat za (plnohodnotné) lidi

Abstrakt

Debata v rámci právního pole týkající se umělé inteligence se v posledních letech výrazně proměnila. Původní témata, jako zda by měla mít umělá inteligence právní osobnost, odpovědnost autonomních vozidel, případně v jakých právních profesích je možné očekávat nahrazení člověka umělou inteligencí, přestala být dominantní ve chvíli, kdy Evropská komise představila návrh aktu o umělé inteligenci. Jedná se o právní rámec, který by měl komplexně regulovat umělou inteligenci, který přišel po letech veřejných konzultací a projednávání. Práce postupuje od představení myšlenkového kontextu debaty o regulaci umělé inteligence k teoretické debatě o vymezení pojmu umělé inteligence, až k aktu o umělé inteligenci. Po stručném shrnutí vzniku, kontextu a hlavních charakteristik budoucí regulace se práce věnuje modelu klasifikace rizik, na kterém má být budoucí regulace umělé inteligence postavena. V práci jsou popsány všechny kategorie rizik, které jsou v aktu o umělé inteligenci zachyceny. Práce přitom postupuje od nejvyššího patra pyramidy rizik, které zahrnuje systémy umělé inteligence, které představují nepřijatelné riziko až po úroveň systémů umělé inteligence, které představují pouze minimální riziko a z pohledu aktu o umělé inteligenci, zůstávají neregulované. Hlavní důraz je přitom kladen na popis kategorie vysoce rizikových systémů umělé inteligence a na popis a vysvětlení systému řízení rizik (čl.9), který musí poskytovatelé vysoce rizikových systémů zavést. Text se detailně věnuje popisu jednotlivých kroků procesu řízení rizik. Druhý proces, kterému se práce věnuje a který se vztahuje ke kategorii vysoce rizikových systémů, je proces posuzování shody. Oba tyto procesy pak text ve svém závěru propojuje s obecnou diskuzí o úloze harmonizovaných norem v rámci Evropské unie.

Klíčová slova: umělá inteligence, akt o umělé inteligenci, regulace

The limits of humanity. Robots are getting dangerously close to humans, while you tend not to consider various groups of people as (fully) human.

Abstract

Recent years have seen a significant shift in the legal debate on artificial intelligence. The moment the European Commission presented its proposal for an AI law, the initial issues, such as whether artificial intelligence should have legal personality, the liability of autonomous vehicles, or in which legal professions we can expect AI to replace humans, ceased to dominate. This, after years of public consultation and deliberation, is a legal framework for the comprehensive regulation of artificial intelligence. The paper moves from presenting the conceptual context of the debate on regulating artificial intelligence, through the theoretical debate on defining artificial intelligence, to the AI Act. After a brief summary of the origins, context and main features of future regulation, the paper discusses the risk classification model on which future AI regulation is to be based. The thesis describes all of the risk categories covered by the AI Act. The work proceeds from the top of the risk pyramid, which includes AI systems that pose an unacceptable risk, to the level of AI systems that pose only a minimal risk and remain unregulated from the perspective of the AI Act. The main focus is on the description of the category of high-risk AI systems and the description and explanation of the risk management system (Article 9) that providers of high-risk systems must put in place. The text goes into detail on the steps that make up the risk management process. The second process addressed in the text, which is related to the category of high risk systems, is the conformity assessment process. The text then links these two processes in its conclusion to a general discussion of the role of harmonised standards within the European Union.

Klíčová slova: artificial intelligence, AI Act, regulation