9. 7. 2024

Vezměte, prosíme, na vědomí, že text článku odpovídá platné právní úpravě ke dni publikace.

Jak na implementaci aktu o umělé inteligenci (nejen) ve firemním prostředí

Implementace evropského aktu o umělé inteligenci (AI Act)[1] vyžaduje, aby adresáti norem provedli několik důkladných kroků k zajištění shody s novými právními požadavky. V tomto článku jsou prakticky vysvětleny některé klíčové procesní kroky, na které je třeba se zaměřit, aby příprava na implementaci tohoto aktu byla úspěšná. Je nutné zdůraznit, že každá příprava musí být individualizovaná, aby zohlednila konkrétní specifika vaší instituce, použitých AI systémů a také interních opatření.

1. Pochopení právního rámce

Všechny dotčené osoby musí pečlivě prostudovat text AI Act, aby pochopili jeho požadavky, definice a klasifikace AI systémů vč. regulovaných AI modelů, a identifikovali klíčové požadavky, resp. určili, které části aktu jsou relevantní pro konkrétní typy AI systémů nebo modelů, které daná osoba vyvíjí nebo používá, nebo se na jejich zavedení teprve připravuje. AI Act přitom není jednoduchý na pochopení ani pro právníky, protože řada požadavků vyžaduje znalosti technických pojmů, fungování umělé inteligence, specifik konkrétního případu (např. jaký je tok dat, kde jsou ukládána nebo odkud se bere výpočetní výkon), a souvisejících horizontálních (nařízení GDPR či směrnice NIS 2) i sektorových předpisů (např. nařízení DORA).

Nutno podoktnout, že ačkoli ustanovení AI Actu budou postupně nabíhat, celkovou analýzu prozatím znemožňuje nedokončený legislativní proces u souvisejících právních předpisů jako jsou pravidla pro civilní odpovědnost způsobenou umělou inteligencí,[2] u které se mj. rozšiřuje seznam právních aktů EU, za jejich porušení bude možné vést řízení o hromadných žalobách napříč EU/EHS.[3] Finální podoba těchto ustanovení může být přitom klíčová při nastavování technických a organizačních opatření (např. jakým způsobem nakládat k předpokládanou příčinnou souvislostí u újmy způsobenou AI jako je otázka vysvětlitelnosti příslušného systému vs. princip černé skříňky). Je však jisté, že za výstupy AI bude vždy odpovědný jejich provozovatel anebo samotný uživatel, protože se nepředpokládá, že by umělá inteligence samotná byla odpovědná za případnou újmu.[4]

Postupné nabíhání povinností například znamená, že určité AI systémy, které byly umístěny na trh nebo uvedeny do provozu 36 měsíců od nabytí platnosti AI Actu, budou muset uvést svoji činnost do souladu ke konci roku 2030 (čl. 111), zatímco ostatní AI systémy již od data nabytí účinnosti. Nejdříve, a tedy nejrychleji je nutné uvést do souladu zakázané AI praktiky, a to již do 6 měsíců od schválení AI Actu.

2. Vytvoření implementačního týmu

Pro úspěšné zvládnutí implementačního procesu lze doporučit sestavení odborného týmu složeného z právníků, compliance a IT specialistů, specialistů na ochranu dat, kybernetickou bezpečnost, případně řízení rizik, kontinuity činností, a dalších relevantních odborníků s ohledem na potenciální přínosy a rizika, kterým je nutné čelit. Např. AI Act vyžaduje postupné budování znalostí (gramotnosti) o umělé inteligenci příslušnými zaměstnanci, kteří s nimi přijdou do styku, takže zapojení personálního oddělení při zajištění AI vzdělávání zde bude jistě výhodou. Každý člen týmu

by měl mít definovánu roli a odpovědnosti pro zajištění shody podle svých znalostí a kompetencí.

3. GAP analýza, hodnocení rizik / dopadů a plán implementace

Možná nejdůležitějším krokem přípravy je řádně provedená GAP analýza, která popíše aktuální stav, a dále seznam požadavků (výstupů), které je nutné zvládnout v konkrétním termínu (plán implementace). Možná si myslíte, že se na vás AI Act nevztahuje, nebo s ním přijdete do styku jen minimálně. GAP analýza by vám měla ukázat, jak na tom skutečně jste. Realita může být totiž často odlišná. Například některé doplňky AI systémů nebo modelů jsou dnes již běžně dostupné v běžných aplikacích jako je Adobe Acrobat (třeba že v beta verzi), vaši zaměstnanci mohou využívat „open source" řešení přes webové rozhraní nebo aplikace v mobilním telefonu, a nemusíte se nutně vyhnout ani sofistikovanému útoku na vaše informační aktiva při použití AI systémů. Používání AI se zkrátka nevyhneme a je jen otázka, jak na tento fakt budeme reagovat. Kromě hodnocení přínosů se budete muset ptát, jaká jsou rizika spojená s AI Actem a využíváním AI systémů, a jak je úspěšně řídit.

a. **Jaká je moje role?**

V rámci AI Act je nutné zjistit všechny vaše role, se kterými AI Act operuje. Nejvíce povinností se vztahuje na (malé) poskytovatele a uživatele AI systémů / modelů, ale můžete být také v postavení výrobce, dovozce, zplnomocněného zástupce, a to vše podle faktických činností, které vykonáváte ve smyslu definic uvedených v čl. 2.

b. **Je moje činnost regulovanou činností AI Actem?**

Každá regulovaná činnost má vymezený set právních požadavků, přičemž v některých případech se na vás regulace nemusí vztahovat vůbec (jde-li výhradně o osobní použití AI modelů, např. LLM), anebo konkrétní kontext využití spadá do výjimky (např. výhradně za účelem výzkumu nebo vývoje kromě testování na reálných datech subjektu údajů).

c. **Mapování AI systémů a modelů**

Provedena by měla být identifikace a katalogizace všech AI systémů / modelů, které záměrně používáte, vyvíjíte, nebo hodláte integrovat do stávajících systémů a procesů, resp. také těch, které v rámci výkonu své činnosti tolerujete.

V prvním kroku je nutné určit, zda daný AI systém nebo regulovaný AI model vůbec vyhovuje definici podle AI Act. Pokud ne, jste mimo rozsah regulace a povinnosti AI Act se na vás nevztahují. Evropská komise si však vyhradila právo usměrňovat intepretaci definic v AI Actu a měnit některé definiční prvky prováděcími nařízeními, pokud by účelem bylo obcházení rozsahu regulace. Zejména definice regulovaného AI modelu je poměrně vágní, takže určení rozsahu může být složitou otázkou.

Myslet budete muset také na to, že určité změny systémů nebo modelů mohou vést k tomu, že se později do rozsahu AI Actu dostanete (takže se nevyhnete např. evidenci potenciálních AI systémů, průběžnému monitoringu nebo ověřování shody s rozsahem AI Actu).

Naproti tomu mnoho systémů se komerčně prezentuje jako AI, nicméně pro rozsah regulatorních povinností bude vždy rozhodující definice uvedená v AI Act, nikoli komerční název produktu či služby. Ty mohou postrádat často esenciální definiční prvky jako je určitá míra autonomie, vyspělosti či schopnost adaptability, která překračuje pouhou exekuci určitých posloupných kroků, pokud jsou naplněny předem specifikované podmínky na principu IFTTT, což může být případ většiny chatbotů, jejichž rozhodovací proces je designován na principu „rozhodovacího stromu" (decision tree) a nikoli neuronových sítí.

d. **Hodnocení dopadů využití AI a celkový kontext**

Ať už používáte AI systémy pro konkrétní účely nebo AI modely s obecným použitím, vždy budete muset zmapovat konkrétní způsob využití a jeho celkový kontext jako jsou dopady na systémy, procesy nebo na dotčené osoby jako jsou zákazníci, zaměstnanci a jiné fyzické osoby (např. dodavatelé), zda zapojení AI se týká vašeho klíčového procesu, resp. zda AI systém má vykonávat některé rozhodovací činnosti, které by jinak vykonával člověk,[5] a jaké jiné dopady má zapojení AI na vaše činnosti (např. závislost na technologii z hlediska zajištění kontinuity činností, dojde-li k výpadku nebo přerušení činnosti nejrůznějšími vlivy). Toto hodnocení pak využijete při přípravě plánu implementačních opatření, resp. jeho adekvátnosti.

Významné také může být, zda AI budete využívat pouze vy (např. vašimi zaměstnanci), anebo jej umožníte využívat také vašimi zákazníky, dodavateli nebo jinými obchodními partnery, anebo v rámci koncernu, kde se například mimo pozice uživatele můžete také dostat do pozice poskytovatele se všemi souvisejícími povinnosti.

e. **Geografický rozsah**

Pro určení rozsahu vašich povinností je také klíčové zjistit geografický rozsah použití AI Actu, tj. zda vaše činnost se týká oblasti EU/EHS, resp. zda výstupy AI systémů nebo regulovaných AI modelů jsou využity v EU/EHS, přestože máte sídlo mimo tento region. Pakliže tato souvislost nenastane, můžete být zcela mimo rozsah AI Actu. Neměla by však nastat situace, že AI systémy a operace zpracování dat poběží na cloudu v datových centrech umístěných mimo EU/EHS, aby se obcházelo aplikaci AI Actu.

f. **Hodnocení rizik**

Na základě evidence AI systémů / modelů by mělo být provedeno hodnocení rizik pro každý AI systém nebo regulovaný model podle kritérií stanovených v AI Act s ohledem na dotčené osoby vedle standardního hodnocení finančních, operačních a reputačních rizik.[6] Každý by měl minimálně zohlednit zakázané AI praktiky v interních procesech, vyškolit k tomu zaměstnance a tuto aktivitu adekvátně zdokumentovat.

Jiným případem jsou vysoce rizikové AI systémy (čl. 6 a násl.), které jsou dovolené při splnění poměrně přísných opatření včetně získání deklarace konformity, CE známky a registrace u příslušného orgánu. AI Act umožňuje, abyste (pro vymezené rizikové AI systémy) nemuseli některé povinnosti plnit za podmínky, že prokážete nižší rizikový dopad na dotčené fyzické osoby, anebo že výstupy systému nemají materiální dopad na rozhodování (např. vykonává pouhou úzce vymezenou část procesu). V takovém případě je nutné se pouze registrovat u příslušného orgánu (čl. 49 odst. 2), a na jeho žádost mu předložit vaše vlastní hodnocení (čl. 6 odst. 4), které je oprávněn přezkoumat, resp. přehodnotit, že kritéria pro výjimku nesplňujete. Vedle práva na soukromí se budou hodnotit také další základní lidská práva jako je právo na život, zdraví, ochranu majetku, lidskou důstojnost, ochranu před diskriminací atp.

Ostatní AI systémy budou muset plnit méně přísné podmínky (zejména v oblasti transparence a poskytování informací podle čl. 50), nicméně AI Office a členské státy se mohou snažit podporovat tvorbu kodexů chování, které mají podnítit dobrovolnou aplikaci požadavků na rizikové AI systémy (čl. 95) včetně těch, které do režimu AI Act doposud nespadají.

Zvláštní požadavky platí pro regulované AI modely s obecným použitím, u kterých větší penzum povinností dopadne na ty, které působí systemické riziko, a které se budou muset notifikovat u AI Office.

4. Vývoj a implementace interních opatření, procesů a kontrol

Dalším krokem je analýza současného stavu, ze které by měla vyplynout konkrétní doporučení pro vývoj a implementaci interních opatření, procesů a kontrol, které hodláte zavést nebo pozměnit. AI Act totiž může mít vliv na celou řadu vašich interních procesů jako jsou dopady na řízení produktu, distribuce, zákaznického servisu, kybernetickou bezpečnost, IT/ICT change management, nákup IT služeb a jejich monitorování (vč. pravidel pro regulovaný outsourcing), M&A aktivity či dopady na data governance aj.

Z pohledu řízení právních rizik může být významné, jakým způsobem je rozdělena odpovědnost mezi poskytovatele a uživatele, pokud jde o případné nároky na náhradu újmy a další nároky dotčených osob jako je ochrana soukromí nebo vlastnických vč. autorských práv. V tomto ohledu je důležitý obsah rizikové analýzy týkající se základních práv (čl. 49 a příloha VIII) nebo technické specifikace či dokumentace, která má představovat jakýsi návod k použití vysoce rizikového AI systému nebo regulovaného AI modelu (čl. 17 a příloha IV; čl. 53 a příloha XI), na kterou musí uživatel reagovat, protože ten sám v konečném důsledku určuje podmínky konkrétního užití a zpravidla je také pod jeho výhradní kontrolou. Jinými slovy za funkcionalitu nese odpovědnost poskytovatel, zatímco za použití pod jeho výhradní kontrolou by měl být odpovědný samotný uživatel, což implicitně znamená, že by měl znát i limity a omezení daného AI systému / regulovaného AI modelu jako je kvalita vstupních dat nebo vyspělost, přesnost a spolehlivost použitého modelu.

Výsledkem tohoto cvičení by mohl být konkrétní seznam způsobů využití AI v konkrétním procesu (aktivitě) instituce, u které je předem definovaný set opatření k řízení (např. právních) rizik v závislosti na postavení instituce a specifickým rizikům, kterým čelí sama nebo i dotčené osoby. Například je nutné vyjasnit, jakým způsobem bude zapojení AI ověřováno nebo testováno před spuštěním, a jak bude probíhat monitoring a následná kontrola (vstupních dat, modelu, výstupů) po jeho spuštění.

5. Technické, organizační a personální opatření

a. **AI strategie**

Přestože AI Act stanoví povinnost strategie týkající se AI compliance pouze u vysoce rizikové AI systému (čl. 17), lze její přijetí doporučit také u „méně rizikových“ AI systémů jako výchozí bod, u kterého si rozhodující orgán nebo osoba ujasní, ve kterých činnostech zapojení AI a s jakými parametry dává smysl, a kde nikoli (s ohledem na přínosy, zákonné podmínky, rizika, a případně finanční a jiná omezení jako jsou potřebné AI znalosti).[7]

b. **Governance, schvalování a další role**

Jiným důležitým bodem je řádně nastavené governance a schvalování zapojení AI do činnosti instituce.[8] U méně rizikových případů, kde nehrozí riziko hromadných žalob lze uvažovat, že by o zapojení AI mohli rozhodovat vedoucí zaměstnanci, jinak by pravidlem mělo být, že rozhodující vliv by měl mít statutární orgán s ohledem na limitovanou odpovědnost zaměstnanců, riziko hromadných žalob a maximální výměry sankce za porušení AI Actu (až 35 milionů EUR nebo 7% celosvětového obratu), či možnosti členů statutárního orgánu sjednat pojištění odpovědnosti z výkonu své funkce. Neměli byste přitom zapomenout, že zapojení AI systému by mohlo zajímat vaši pojišťovnu, která pojišťuje vaši odpovědnost z důvodu zvýšení pojistného rizika (§ 2790 občanského zákoníku), a přehodnotit rozsah krytí z pojištění kybernetických rizik.

U veřejných institucí je vůbec otázkou, jestli by o zapojení AI do některých esenciálních činností státu, které dopadne na základní práva občanů, měl vůbec rozhodovat o proporcionalitě a zapojení

AI vedoucí této instituce (exekutivy či justice), jak bylo např. judikováno v rozhodnutí ve věci E – kasa, ale spíše zákonodárce, který zohlední test proporcionality a případné záruky ochrany základních práv (např. proti zneužití přístupu k údajům či garance efektivní nezávislé kontroly).[9] Neproporcionální by bylo neomezené automatizované kontrolování evidenčních čísel aut na cestách za účelem ochrany před jakýmkoli nebezpečím.[10]

Z pohledu instituce by každý AI systém či model zapojení do interního procesu měl mít osobu odpovědnou za její používání jako je vedoucí příslušného útvaru, kterému náleží kompetence ohledně svěřené činnosti s podporou dalších odborných funkcí, jsou-li dostupné jako je právní nebo IT/ICT podpora, a dále compliance, ochrana informační bezpečnosti, či řízení rizik, přičemž celkový design by měl být ověřován nezávislým auditem, je-li to vyžadováno.

c. **Zavedení interních politik a postupů**

Jedním z klíčových nástrojů pro řádnou implementaci opatření v oblasti AI Act je vytvoření, aktualizace a implementace interních politik a procedur pro zajištění souladu s AI Act. Ty mohou být velmi závislé na stavu AI gramotnosti v dané instituci ať už jde o koncové uživatele AI systémů / regulovaných modelů, osob testujících a ověřujících jejich funkčnost před jejich zapojením[11] a v oblasti následného monitoringu, či osob, které nad AI systémem / modelem budou vykonávat dohled (např. tak, aby nesprávnost výstupu dokázali včas identifikovat, interpretovat a zasáhnout). K tomu může sloužit například vytvoření seznamu povolených AI systémů / modelů, který daný uživatel může používat při své činnosti s ohledem na absolvovaný základní nebo rozšířený trénink v oblasti AI či dosažení určitého vzdělání v relevantní oblasti jako je funckionalita AI, etické, právní, regulatorní a jiná (např. reputační) rizika. AI systém nebo model je zkrátka nástroj, se kterým by měli pracovat lidé, kteří k tomu mají určité kvalifikační předpoklady a schopnosti.

Pro vývojáře AI a především poskytovatele je nanejvýš důležité mít zabezpečen řádný systém řízení kvality[12] a bezpečnosti dat a systémů, který AI Act doplňuje nebo rozšiřuje jako jsou techniky, postupy a systematické činnosti týkající se designu AI systému či modelu, jeho kontroly, verifikace, anebo testování či validace dat. To obzvláště důležité tam, kde určité obchodní procesy jsou vysoce regulovány jako je oblast vývoje produktů, jeho distribuce a následný servis po uzavření smlouvy.[13]

Dalším nástrojem by mohla být tvorba nových politik na užití AI, aktualizace smluv s konkrétními dodavateli AI řešení nebo tvorba vzorových smluvních doložek v rámci řízení vztahů s novými dodavateli IT/ICT služeb podporující vzájemnou komunikaci a spolupráci při exekuci potřebných opatření.

Poměrně značný dopad může mít zapojení AI do oblasti ochrany soukromí subjektu údajů, pokud by docházelo k automatizovanému zpracování nebo profilování[14] zaměstnanců, zákazníků a jiných fyzických osob s ohledem na zajištění práva na informační sebeurčení. V takovém případě je nutné zajistit provedení alespoň hodnocení dopadu na ochranu dat (DPIA), aktualizaci privacy notice, a zajištění oblasti řízení souhlasu subjektu údajů,[15] včetně odpovídajícího technického zabezpečení dat.

d. **Compliance a etický design**

V rámci implementace bude muset každá instituce stanovit jasná pravidla a postupy pro etické používání AI včetně ochrany osobních údajů, spravedlnosti, transparentnosti a odpovědnosti. Doporučit lze například přijmout nebo aktualizovat Etický kodex, ve kterém mohou být definovány etické zásady a hodnoty, které bude instituce dodržovat při vývoji a nasazení AI systémů, a nad uplatňováním principů a zásad bude bdít např. etická komise. Dále mohou být přijata specifická pravidla při vývoji, testování (např. vzorků) a nasazování AI systémů, ve kterých se budou aplikovat

principy etického designu, které zahrnují ochranu soukromí, spravedlnost a zabránění diskriminaci.

e. **Zajištění transparentnosti a vysvětlitelnosti**

Z pohledu implementace technických opatření pro zajištění transparentnosti a sledovatelnosti AI systémů bude nutné zajistit, že algoritmy používané v AI systémech jsou transparentní a že je možné vysvětlit jejich rozhodovací procesy, a že jsou vytvořeny adekvátní mechanismy pro sledování a dokumentaci všech kroků při vývoji a nasazení AI systémů. U vysoce rizikových AI systémů budou vždy povinné logy důležitých událostí jednak z preventivních důvodů, jednak z důvodů rychlé nápravy a možnosti včas zasáhnout.

Je tedy potřeba vytvořit přiměřený „audit track" s ohledem na povahu AI systému, který bude podpořen příslušnou technickou dokumentací. Měla by existovat vždy jedna zodpovědná osoba, která bude schopna zodpovídat případné dotazy orgánům dohledu.

Z pohledu dotčené osoby bude muset být zřejmé, že je předmětem automatizovaného zpracování, že komunikuje s AI, že výstup daného procesu byl zpracován AI, anebo že konkrétní výstup je svou povahou „deep fake".[16] Případně bude žádoucí také poskytnout upozornění, že výstupy AI systému mohou obsahovat chyby a uživatel bude muset správnost výstupu zkontrolovat.

f. **Zajištění AI znalostí**

Důležitou součástí implementace je také proces řízení znalostí v oblasti AI, které by se měli zařadit např. již v oblasti náboru zaměstnanců, u nichž v popisu práce nebo jeho obsahu by mělo být určitá činnost dotýkající se AI systému, přičemž taková osoba by měla absolvovat školení týkající se AI odpovídající dané funkci. Zaměstnavatel také může vytvořit a distribuovat vzdělávací materiály a příručky, které publikuje na svém intranetu nebo v tištěných brožurách. Také současní zaměstnanci, pokud se jich činnost AI dotýká, by měli absolvovat školení o nových požadavcích a postupy, které musí dodržovat. Rozšířeny by měla být současná školení týkající se informační bezpečnosti, pokud jde o sofistikované útoky za použití AI technologií, resp. týkající se ochrany dat a GDPR.

g. **Zajištění lidského dohledu**

Každá instituce bude muset zajistit lidský dohled nad regulovaným AI systémem, kde je nutné určit osoby odpovědné za dohled nad AI a stanovit postupy pro testování AI systémů a validace dat, resp. modelů, přičemž tyto postupy pravidelně hodnotit, či spíše (mezi řádky regulace) zdokonalovat.

h. **Zajištění soukromí a data governance**

Při použití AI systémů často dochází ke zpracování velkých dat, anebo automatizovanému zpracování určité typologie dat, která mohou být obalena různými vrstvami hodnot či práv chráněných právním řádem. Instituce by se proto měla seznámit s příslušnými datovými toky a povahou dat použitých pro trénování modelu, i případných vstupních dat, která bude používat pro generování výstupů při použití AI modelu. AI systémy / modely jsou trénovány na určitém typu dat, a není bez významu, že již samotný tréninkový proces AI by měl být v souladu s právními předpisy, nikoli až samotné zapojení nebo použití při konkrétní činnosti.

Konkrétně vstupní nebo o tréninková data mohou být chráněna autorským právem jako jsou nejrůznější literární díla,[17] hudební produkty, snímky a fotografie,[18] ostatní zvukové či audiovizuální záznamy, anebo počítačový program. U jazykových modelů se proto výslovně vyžaduje přijmout a aplikovat copyright policy, což neznamená, že v ostatních případech mohou být autorská práva porušována. U vstupních dat za soulad s právem autorským ponese odpovědnost především uživatel vč. následného použití výstupů AI, pokud vkládá fotografie nebo jiná autorská díla do AI

modelu k dalšímu zpracování.

Je nabíledni, že samotný model je tak kvalitní jako jsou kvalitní data, na kterých byl vycvičen, resp. vstupní data (vč. promptů). V případě ochrany práv subjektů údajů podle GDPR mohou být AI modely trénovány na skutečných anebo na syntetických datech subjektu údajů, která nejsou skutečná, ale mohou se například skládat ze zcela nových a umělých datových bodů bez vzájemných vztahů s původními daty. U skutečných dat je nutné získat souhlas subjektu údajů již v rámci tréninku a testování AI modelů, a také, pokud vstupní data mají být předmětem automatizovaného zpracování AI systémů. Pokud instituce chce například používat některý AI systém vytvořený mimo EU/EHS, bude muset počítat s tím, že tréninkový proces AI modelu pravděpodobně nebyl již od počátku v souladu s GDPR,[19] protože EU standard ochrany soukromí je příliš vysoký na to, aby se na něj poskytovatelé AI systémů mimo EU dostatečně připravili ve svých domovských státech, což by mělo být předmětem posouzení souladu.[20]

Jiným příkladem může být ochrana databází, ať už jde o metody data scraping[21] nebo použití jiných nástrojů, které se používají pro extrakci strukturovaných dat z různých zdrojů. Pro vytěžování databází je nutný souhlas majitele databáze nebo udělení licence,[22] případně další podmínky předání a jiné zpracování údajů jako je souhlas autora či subjektu údajů.

Někdy se může využívat web scraping, který se zaměřuje na získávání především nestrukturovaných dat z webových stránek, který může čelit právním výzvám spojených s podmínkami použití webových stránek anebo autorským právům k dílům zde publikovaným. Zvláštní režim může platit pro tzv. otevřená data (vč. jejich případného napojení na API) v rámci volného pohybu neosobních údajů[23] i osobních,[24] kde specifickou roli mohou hrát např. evropská datová centra,[25] anebo speciální úpravy týkající se sdílení dat (např. v oblasti identifikace klienta).[26]

Je nutné také zjistit, zda AI model si může vámi vložená data používat v rámci svého dalšího výcviku či nikoli, což má velký význam z pohledu ochrany vašich informačních aktiv,[27] případně zda vámi zadané údaje ihned vymaže a neuchovává si ani jejich kopie. Vedle již výše zmíněných může jít například o ochranu obchodního tajemství, bezpečnostních anebo jiných citlivých údajů. Kromě smlouvy týkající se vyloučení následného uchovávaní nebo jiného zpracování vámi poskytnutých údajů by měly existovat přiměřené záruky (opatření) proti zneužití interních opatření.

i. **Zajištění nediskriminace**

Tréninková data nemusí být dostatečně diverzifikována a mohou nezřídka obsahovat zkreslení nebo omezení na určitou skupinu, kde AI model může zpochybnit výsledky pro jiné skupiny. Tato datová nedokonalost, resp. nedostatečné propojení, interpretace či vysvětlení významu dat s konkrétním kontextem pak může vést k tomu, že AI systém obsahuje nebo vytvoří skryté předsudky, které povedou například k preferenci určité skupiny nebo naopak k jejich vyloučení.[28] To může záviset na tom, jak se například jazykové modely vyrovnávají s nejistotou ohledně výsledku v rámci strojového učení.[29]

Aby se předešlo nespravedlivé diskriminaci, měla by být zapojena určitá opatření jako je testování a validace výsledků při vývoji modelu, resp. průběžné monitorování výsledků při použití AI sytému a vytvořením specifických antidiskriminačních kontrol (např. validaci modelu nebo výsledků pomocí expertů na AI).

j. **Zajištění bezpečnosti**

Instituce by měly zajistit, že AI systémy splňují požadavky na kybernetickou bezpečnost a ochranu osobních i neosobních údajů včetně odpovídajících bezpečnostních standardů v daném ekonomickém

sektoru nebo činnosti. Vyšší nároky budou kladeny na osoby, které jsou v rozsahu směrnice NIS 2 a nařízení DORA včetně jejich dodavatelů IT/ICT služeb.

k. **Technologická opatření**

Také by měly být implementovány potřebné technické změny v AI systémech, jako jsou zajištění transparentnosti, sledovatelnosti a bezpečnosti v rámci celého životního cyklu AI systému. Transparentnost (vysvětlitelnost) AI systémů je klíčová pro důvěru veřejnosti a dohledové orgány, ale i pro správné použití jejich uživateli, či zajištění řádného provozu jejich poskytovateli, kteří by měli udržovat technickou dokumentaci, která popisuje, jak systém funguje, jaké algoritmy používá a jaké jsou jeho výstupy.

Primárně by měly být použity modely, které jsou vysvětlitelné. To znamená, že jejich rozhodovací procesy jsou pochopitelné a lze je vysvětlit uživatelům i dohledovým orgánům. Dále by měly být zavedeny mechanismy pro auditování AI systému (např. záznamy o eventech a jejich příčinách), které umožní dohledovým orgánům sledovat jeho činnost a ověřit, zda nedochází k diskriminaci nebo jiným problémům.

Existovat by měl také jednoduchý mechanismus pro rychlé znemožnění dalších operací AI systému („switch-off button“) neboli možnost lidského zásahu, u kterého by mělo být myšleno také na náhradní (někdy také pohotovostní) plán, jak bude zajištěn další provoz dotčených aktivit, pokud AI systém bude dočasně nebo trvale vyřazen z provozu.

l. **Připravenost na incidenty**

V neposlední řadě by instituce měly vytvořit plán reakce na incidenty spojené s AI systémy, včetně mechanismů pro rychlé řešení a hlášení incidentů, a to jak interní, tak externí hlášení.

6. Dokumentace a záznamy

Instituce se nevyhnou ani povinnosti vést, shromažďovat anebo udělit přístup k určité dokumentaci nebo záznamům týkající se AI systémů. V rámci vedení záznamů lze doporučit vytvořit a udržovat podrobný záznam o všech AI systémech, včetně hodnocení rizik a kroků podniknutých k zajištění shody, aby se prokázala dostatečná robustnost přijatých opatření.

V rámci dokumentace je také klíčové řádné zaznamenat dokumentaci testovacích procesů a výsledků prokazujících bezpečnost a účinnost AI systémů, případně výsledky monitoringu, kontrol, auditu, ostatní validace (modelu nebo dat), resp. průběžné aktualizace procesů a postupů na základě nových zjištěných nedostatků či změny regulace.

7. Komunikace a spolupráce s regulačními orgány

V rámci procesu implementace bude nutná příprava a podávání potřebných zpráv a dokumentace regulačním orgánům, které si mohou vyžadovat jako je posouzení souladu s AI Actem a jeho vysvětlení (čl. 20 a 21), dále technická dokumentace, přístup k logům (čl. 21), či písemná smlouva mezi poskytovatelem a uživatelem AI systému.

Příprava zahrnuje také spolupráci s dozorovými orgány, ať už jde o plnění registračních povinností (např. v EU databázi pro testování na reálných datech), notifikačních povinností, podstoupení testování v sandboxu (čl. 57 a násl.), či hlášení vzniklých incidentů. Aktivní spolupráce zahrnuje dále poskytnutí požadovaných informací a spolupráce při inspekcích či auditech.

Podstoupení testování v AI regulatorním sandboxu by mělo mít určité výhody, protože výstupem by

měla být závěrečná zpráva o absolvování programu v sandboxu, jejíž účinkem by mělo být, že orgán dohledu by v rozsahu testovacího plánu a na základě podmínek účasti neměl mít možnost ukládat následně sankce. To se však netýká vzniku náhrady újmy, které by měly být plně nahraditelné i přes úspěšné absolvování této administrativní procedury. Závěrečná zpráva se samozřejmě bude týkat jen stavu v době účasti v programu, takže jakékoli podstatné změny systému, aktualizace či upgrade by již z této výjimky neměly těžit a bylo by nutné provést nové posouzení konformity.

Závěr

Připravenost na implementaci evropského aktu o umělé inteligenci vyžaduje komplexní a systematický přístup. Dodržování výše uvedených kroků může dané instituci pomoci zajistit shodu s novými regulačními požadavky a minimalizovat rizika spojená s použitím AI technologií.

AI Act je horizontální regulace, která dopadá na všechny ekonomické sektory jak ve sféře soukromé, tak také pro orgány veřejné moci, což zvedá některé otázky spojené s proporcionalitou regulace a jejich účinnému uplatňování. Důležitý a rozhodující bude přístup dohledových orgánů k interpretaci a uplatňování požadavků. Příliš restriktivní přístup by mohl mít rdousící efekt na inovace, zatímco příliš benevolentní přístup by mohl znamenat, že některé obavy spojené s umělou inteligencí se mohou projevit velmi záhy. Větší problémy s implementací mohou pocítit hlavně ty instituce, které nemají dostatečné finanční nebo lidské zdroje či vlastní kapacity (např. malí živnostníci) na splnění všech podmínek regulace AI Actu, což by jim mělo ulehčit splnění podmínek konformity ve zjednodušeném režimu.

Klíčovým faktorem bude rovněž otázka národní adaptace AI Actu jako je určení kompetentních orgánů. Pro regulované AI modely bude mít dohledové pravomoci přímo Evropská komise, zatímco AI systémy budou dohlíženy na národní úrovni. Pro finanční služby by to měla být příslušným orgánem Česká Národní Banka, jak deklaruje AI Act pro vysoce rizikové AI systémy (čl. 74 odst. 6) v souvislosti s výkonem regulovaných činností. Pro ostatní AI systémy se může zákonodárce rozhodnout odlišně (recitál 158). Na evropské úrovni by mělo být zachováno sektorové rozdělení EU institucí pro dohled nad finančními institucemi: EBA, ESMA, EIOPA. V některých státech tuto roli obecně vykonává úřad pro ochranu osobních údajů, ale na konečnou podobu dohledového modelu si budeme muset ještě chvíli počkat.

Mgr. Robert Šimek Ph.D.

---

[1] Návrh nařízení Evropského parlamentu a rady, kterým se stanoví harmonizovaná pravidla pro umělou inteligenci (Akt o umělé inteligenci): k dispozici >>> zde.

[2] Návrh směrnice Evropského a Parlamentu a Rady o přizpůsobení pravidel mimosmluvní občanskoprávní odpovědnosti umělé inteligenci (směrnice o odpovědnosti za umělou inteligenci): k dispozici >>> zde.

[3] Známá je v oblasti AI systému například hromadná žaloba na společnost HireVue, která bez informovaného souhlasu subjektů údajů (zájemců o pracovní místo) o nich zpracovávala (biometrické) osobní údaje: k dispozici >>> zde.

[4] To se týká například používání nejrůznějších nástrojů pro zákaznický servis nebo poskytování

informací (např. chatboti). Pokud tyto nástroje poskytnout zákazníkovi špatnou informaci anebo radu, pak instituce bude odpovědná za případnou újmu stejně jakoby tuto nesprávnou informaci nebo radu poskytl člověk, jehož (deliktní) jednání se instituci přičítá.

[5] Např. společnost CIGNA byla obviněna, že její algoritmus odmítá stovky tisíc nároků pacientů, čímž se snižuje přístup ke zdravotní péči: k dispozici >>>zde.

[6] V rámci řízení rizik lze využít normu ISO/IEC 23894:2023 Information technology – Artificial intelligence – Guidance on risk management.

[7] Řádný system řízení je rovněž podstatou normy ISO/IEC 42001:2023 Information Technology – Artificial intelligence – Management system.

[8] V této souvislosti lze využít normu ISO/IEC 38507:2022 Information technology – Governance of IT – Governance implications of the use of artificial intelligence by organizations.

[9] Srov. nález Ústavního soudu Slovenské republiky ze dne 10.11.2019, sp. zn. III. PL. ÚS 25/2019, ve kterém byl hodnocen systém e-kasa v souladu s Ústavním pořádkem, kde návrhu nevyhověl v části, pokud jde o přístup a zpracování údajů ze strany finanční správy pro účely kontroly povinností vyplývajících ze strany zvláštního zákona, nikoli však pro sbírání údajů a automatizované analytické posuzování rizikovosti podnikatelů, který se neobejde bez potřebného legislativního ukotvení včetně řádného testu proporcionality. Navíc, z údajů identifikujících konkrétní kupující a z údajů o jimi nakoupeném zboží a službách bylo lze vyčítat možnost vytvářet individuální profily kupujících týkající se každodenních návyků, jejich sociální postavení a finanční možnosti (dle ceny zboží nebo služby), údaje týkající se jejich zdravotního stavu (například podle údajů o nákupu léků), náboženské zaměření (dle nákupu zboží s náboženským tématem) nebo údaje o intimním životě těchto lidí (podle nákupu zboží nebo služeb intimního charakteru).

[10] Podle rozhodnutí Spolkového ústavního soudu Německa (BVerfG, sp. zn. 1 BvR 142/15, bod 104).

[11] Chybovost AI systému může mít nemalé finanční dopady na vznik povinnosti hradit kompenzace, např: >>> zde.

[12] Pro nastavení řízení kvality lze využít normu ISO/IEC 25059:2023 Software engineering – Systems and Software Quality Requirements and Evaluation (SQuaRE) – Quality model for AI systems.

[13] Např. ESMA Public statement on the use of Artificial Intelligence (AI) in the provision of retail investment services:k dispozici >>> zde.

[14] Srov. rozsudek Soudního dvora EU ze dne 7. prosince 2023 ve věci C-634/21 ohledně automatizované stanovení hodnoty pravděpodobnosti týkající se schopnosti osoby splnit v budoucnu platební závazky („SCHUFA“): „automatizované stanovení hodnoty pravděpodobnosti založené na

osobních údajích týkajících se osoby a její schopnosti splnit v budoucnu platební závazky, které provádí obchodní informační agentura, představuje „automatizované individuální rozhodování“ ve smyslu tohoto ustanovení, pokud na této hodnotě pravděpodobnosti rozhodujícím způsobem závisí skutečnost, zda třetí osoba, které je uvedená hodnota pravděpodobnosti sdělena, uzavře smluvní vztah s touto osobou, bude z něj plnit či jej ukončí.“

[15] Například italský úřad pro ochranu osobních údajů (Garante) spatřoval porušení nařízení GDPR u společnosti Uber, která zpracovával osobní údaje zákazníků bez jejich souhlasu: k dispozici >>> zde.

[16] Například pro pojišťovny bude velmi složité bojovat se stále více sofistikovanými pojistnými podvody: k dispozici >>> zde , zde.

[17] Konkrétně New York Times v roce 2023 zažaloval společnosti OpenAI a Microsoft, že při tréninku chatbotů využili bez souhlasu nakladatele články, které byly určeny k tréninku chatbotů ohledně poskytování informací čtenářům: k dispozici >>> zde.

[18] Například společnost Getty Images zahájila spor se společností Stability AI (poskytovateli nástroje Stable Diffusion), která zneužila fotografie k tréninku AI modelu: k dispozici >>> zde, zde.

[19] Příkladem může být dočasný zákaz zpracování osobních údajů společnosti OpenAI nabízející AI model ChatGPT uložený italským úřadem pro ochranu osobních údajů (Garante): k dispozici >>> zde.

[20] European Data Protection Board v této souvislosti přijal dne 3. června 2024 stanovisko týkající se zajištění souladu ochrany dat evropských institucí při využívání generativních AI systémů:k dispozici >>> zde.

[21] Například společnost Clearview AI byla obviněna francouzským úřadem pro ochranu osobních údajů (CNIL) z masivního porušování soukromí poté, co vytěžovala selfie z internetu a použila z nich osobní údaje k vytvoření nástroje pro rozpoznávání obličeje. Bylo zjištěno, že společnost Clearview porušila řadu požadavků stanovených GDPR a francouzským zákonem o ochraně osobních údajů: k dispozici >>> zde.

[22] Podle směrnice (EU) 96/9/ES ze dne 11. března 1996 o právní ochraně databází.

[23] Např. podle směrnicek dispozici >>> zde, nebo nařízení (EU) 2023/2854 ze dne 13. prosince 2023 o harmonizovaných pravidlech pro spravedlivý přístup k datům a jejich využívání (nařízení o datech).

[24] Např. návrh nařízení Evropského parlamentu a Rady o rámci pro přístup k finančním údajům (FIDA): k dispozici >>>zde.

[25] Např. podle návrhu nařízení Evropského parlamentu a Rady o evropském prostoru pro zdravotní data (EHDS): k dispozici >>>zde.

[26] Podle návrhu nařízení Evropského parlamentu a Rady, kterým se mění nařízení (EU) č. 910/2014, pokud jde o zřízení rámce pro evropskou digitální identitu (eIdas 2.0): k dispozici >>>zde.

[27] Společnost Samsung zakázala svým zaměstnancům používat nástroj ChatGPT po tom, co unikl citlivý kód, který byl následně viditelný pro ostatní uživatele: k dispozici >>>zde.

[28] Například ve státě New York tamní orgán dohledu zjistil, že AI systémy umožnovaly stanovovat přesněji cenu pojištění, nicméně prediktivní model obsahoval také předsudky, které vedly k diskriminaci zákazníků: k dispozici >>> zde.

[29] Příkladem diskriminace může být případ společnosti Workday, která při screeningu a hodnocení zaměstnanců využívala faktory jako je věk, rasa nebo invalidita: k dispozici >>>zde.

© EPRAVO.CZ – Sbírka zákonů, judikatura, právo | www.epravo.cz

Další články:

- Zápis ochranné známky bez komplikací. Klíčem k úspěchu je kvalitní předběžná rešerše
- Zneužití práva na přístup podle GDPR
- Byznys a paragrafy, díl 31. - létající pořizovatel ve světle nového stavebního zákona
- Právní povaha sítě elektronických komunikací – režim náhrady škody
- Náhrada ušlého nájemného při předčasném ukončení nájemní smlouvy na nebytové prostory
- Jak fungují plánovací smlouvy v reálných situacích (2. díl)
- Nejvyšší soud a forma smlouvy o smlouvě budoucí: krok zpět v ochraně právní jistoty?
- „Za každou kauzou je živý příběh“
- Přehnaná, nebo důvodná prevence? Zajištění a utvrzení závazků v praxi
- Spoluvlastnictví a správa společné věci
- Doručování soudních písemností ze zahraničí do ČR