31. 7. 2024

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

(Ne)označování obsahu vytvořeného generativní AI

Umělá inteligence (AI) přináší společnosti nepochybně celou řadu výhod, zejména nám může pomáhat s výzkumem, s repetitivními úkoly a v neposlední řadě také s tvorbou obrovského množství syntetického obsahu, mnohokrát s vynaložením pouze minimálního úsilí v podobě jednoduchého promptu[1]. To, co se může zdát jako zcela pozitivní má však i určitou negativní stránku. Autoři v tomto svém článku v úvodu rozeberou problematiku obsahu tvořeného generativní AI a nastíní současnou právní úpravu označování obsahu vytvořeného pomocí generativní AI a budoucnost této úpravy v podobě Aktu o umělé inteligenci[2].

Problematické aspekty obsahu tvořeného generativní AI

Jak již bylo naznačeno výše, generativní AI[3] umožňuje jejich uživatelům i za použití jednoduchých promptů generovat velké množství syntetického obsahu, tj. nejrůznější texty, obrázky a videa. Často jsou tyto obecné systémy AI[4] dokonce bezplatné (např. Perplexity, ChatGPT apod.). S postupným technologickým pokrokem navíc bývá čím dál těžší rozeznat obsah tvořený lidským autorem od obsahu generovaného obecným systémem AI. Na jednu stranu je pozitivní, že i lidé bez potřebných odborných znalosti dokáží za minimálního úsilí vytvořit obsah srovnatelný s obsahem vytvořeným programátorem či grafickým designérem. Na druhou stranu však existují i negativní aspekty tohoto fenoménu, které mohou mít zásadní dopady na fungovaní naší společnosti.

Jeden z nejzávažnějších negativních aspektů se týká zcela jistě kyberbezpečnosti. Asi každý z dnešních uživatelů internetu již čelil nějaké formě phishingových útoků[5], které však až doposud nebyly příliš přesvědčivé (typicky se mohlo jednat o podvodný e-mail plný gramatických chyb). Ovšem s příchodem AI je možné očekávat stále komplexnější a přesvědčivější phishingové útoky, zejména v podobě tzv. spear-phishingu[6]. Podvodníci, kteří shromažďují prolomené údaje získané pomocí kybernetických útoků, mohou pomocí technologie umělé inteligence tyto údaje přečíst a vytvořit vysoce účinný spear-phishingový nástroj. Jako příklad lze uvést situaci, kdy podvodník získá pomocí kybernetického útoku některá vaše data a ví, že navštěvujete určitou internetovou stránku. Následně vám pošle gramaticky bezchybný e-mail, v němž se vydává za provozovatele této stránky, který potřebuje ověřit některé údaje o vašem účtu. Tímto způsobem tak může získat další data a informace (zejména pak vaše osobní či bankovní údaje), či provést další kybernetický útok.[7]

Dalším negativním aspektem spojeným s obsahem generativní AI je problematika tzv, „deep fake“ obsahu[8]. Tento obsah se s příchodem AI stává čím dál tím více dokonalejší a bezchybný. Zároveň je značně rozšířen okruh osob, které mohou tento obsah tvořit. Pomocí AI je například možné replikovat hlas (případně i podobu ve formě videa) určité celebrity, politika apod. Zneužívání „deep fake“ obsahu pak může v konečném důsledku představovat i riziko pro demokratické zřízení naší společnosti, kdy používání takovéhoto obsahu se stává součástí hybridní války vedené ze strany Ruska či dalších autokratických zemí za účelem ovlivnění výsledků demokratických voleb v ostatních zemích.[9]

Posledním problémem, který zde zmíníme, je nízká míra transparentnosti, kdy uživatelé obecného systému AI nezveřejní, že obsah byl vytvořen tímto systémem. Příkladem může být grafický designer

(např. fyzická osoba podnikající), který při tvorbě díla neuvede, že dílo vytvořil pomocí obecného systému AI a objednatel díla se pak může mylně domnívat, že se jedná o dílo vytvořené člověkem, tj. že vytvoření tohoto díla stálo příslušnou osobu určité větší úsilí a že je tudíž spojeno i s určitými kvalitami jeho autora.

Výše uvedené problematické aspekty pak svědčí o potřebě obsah generovaný obecnými systémy AI určitým způsobem regulovat. Další část článku v této souvislosti pojednává o současném stavu regulace a jejím možném budoucím vývoji.

Současná právní úprava

V současné době zatím český právní řád neobsahuje účinnou právní úpravu vztahující se k povinnostem uživatelů k označování obsahu tvořeného pomocí generativní AI. První legislativní úpravu tak představuje již zmíněný Akt o umělé inteligenci, který bude popsán podrobněji v další části tohoto článku. Povinnost označovat obsah tvořený pomocí generativní AI je tak aktuálně upravena především smluvní dokumentací jednotlivých AI nástrojů, či provozovatelů příslušných internetových stránek, platforem nebo aplikací.

V níže uvedené tabulce můžeme nalézt přehled povinností uživatelů vybraných obecných systémů AI:

Obecný systém AI	**Úprava označování obsahu generativní AI**[10]
ChatGPT	**Sharing & publication policy** (Zásady sdílení a zveřejňování) *„Uveďte, že obsah je vytvořen umělou inteligencí způsobem, který by žádný uživatel nemohl rozumně přehlédnout nebo špatně pochopit.“*[11] **Terms of use** (Podmínky použití) *„Co nemůžete dělat. Naše služby nesmíte používat k žádné nezákonné, škodlivé nebo zneužívající činnosti. Například máte zakázáno:* *[…]* *„prezentovat, že výstup byl vytvořen člověkem, i když tomu tak není.“*[12]
stability.ai	**Acceptable Use Policy** (Zásady přijatelného užívání) *„Souhlasíte s tím, že nebudete používat technologii Stability AI ani nedovolíte, aby ji používali jiní:* *[…]* *k úmyslnému klamání nebo uvádění ostatní v omyl, včetně použití technologie Stability AI v souvislosti s:* *[…]* *předstíráním nebo uváděním lidí v omyl, že použití technologie Stability AI nebo jednotlivé výstupy jsou generovány člověkem“*[13]

	Terms of Service (Podmínky poskytování služeb)
Midjourney	*„Buďte opatrní při sdílení. Sdílení vašich výtvorů mimo komunitu Midjourney je v pořádku, ale zvažte, jak by se na váš obsah mohli dívat ostatní.“*[14]

Z uvedené smluvní dokumentace vybraných obecných systémů AI tedy vyplývá, že explicitní povinnost označovat, že byl obsah tvořen generativní AI mají například uživatelé ChatGPT, avšak toliko v případech, kdy chtějí tento svůj obsah sdílet s veřejností (např. prostřednictvím sociálních sítí). U ostatní vybrané smluvní dokumentace takovouto povinnost nenalezneme.

Zároveň je však nutné podotknout, že někteří poskytovatelé obecných systémů AI[15] začali u těchto používat technická řešení, která umožňují vkládat metadata, která indikují, že obsah byl vytvořen pomocí umělé inteligence. Příkladem je možné uvést společnost OpenAI, která pro obecný model AI DALL·E 3 (model sloužící ke generování především obrazového obsahu) začala používat C2PA standardy, s jejichž pomocí může jakákoliv třetí osoba zjistit, že se jedná o obsah vytvořený pomocí AI. Předmětná metadata je však možné kdykoliv odstranit.[16]

Velice proaktivní je také přístup společnosti Meta Platforms, Inc. (společnost provozující např. sociální sítě Facebook, Instagram a Threads), která začala označovat synteticky vytvořený obsah pomocí svého obecného systému AI štítkem „Vytvořeno s AI“[17]. V případě, že příslušný obsah obsahuje indikátory, že byl vytvořen pomocí systému AI, začala tato společnost navíc sama tento obsah označovat uvedeným štítkem. Zároveň po uživatelích určitých svých platforem (např. Instagram) společnost META vyžaduje, aby realistický obsah vytvořený pomocí obecného systému AI tímto štítkem sami označovali.[18],[19]

Akt o umělé inteligenci

Problematika transparentnosti je řešena samostatnou kapitolou IV Aktu o umělé inteligenci, respektive článkem 50 tohoto Aktu. Stanovuje se povinnost poskytovatelů systémů AI (např. OpenAI) vytvářejících syntetický zvukový, obrazový, video nebo textový obsah označovat tyto výstupy strojově čitelnou formou, tak aby bylo možné zjistit, že tyto byly uměle vytvořeny. Poskytovatelé těchto systémů mají relativně volný výběr ohledně konkrétní podoby technického řešení, tj. jedná se o určitou formu technologické neutrality. Poskytovatelé těchto systémů však musí zajistit, aby předmětná technická řešení byla účinná, interoperabilní, a spolehlivá. Bod 133 preambule Aktu o umělé inteligenci pak uvádí možná technická řešení v podobě vodoznaků, identifikace metadat, kryptografických metod apod. Pokud však systém AI bude používán pouze např. pro editaci textu, nebude nutné výše uvedené splnit.[20]

AI akt řeší také transparentnost ve vztahu k „deep fake“ obsahu. Subjekty zavádějící systém AI (například obchodní společnost podnikající v oboru vydavatelské činnosti), které tvoří „deep fake“ obsah, musí zveřejnit, že obsah byl uměle vytvořen. Tato povinnost se však nevztahuje na fyzické osoby. V případě, že „deep fake“ obsah představuje umělecké, tvůrčí či satirické vyjádření, je stanovena povinnost takovýto obsah zveřejnit toliko vhodným způsobem, který nebrání zobrazení nebo užívání díla.[21]

Výše uvedená právní úprava bude účinná 24 měsíců ode dne vstupu Aktu o umělé inteligenci v platnost. Autoři tohoto článků hodnotí velice pozitivně, že se evropský zákonodárce rozhodl upravit v Aktu o umělé inteligenci i problematiku transparentnosti včetně povinnosti označování obsahu tvořeného generativní AI. Navržené řešení představuje dobrý kompromis mezi řešením současných

negativních aspektů obsahu tvořeného generativní AI (viz první část tohoto článku) a přílišnou regulací omezující obchodní činnost a vývoj v oblasti AI.

Závěr

Autoři úvodem svého článku poukázali na některé negativní aspekty obsahu tvořeného generativní AI, které odůvodňují potřebu určité právní regulace této oblasti. Následně nastínili současnou právní úpravu označování obsahu tvořeného generativní AI a budoucnost této úpravy v podobě Aktu o umělé inteligenci, který v části věnující se problematice transparentnosti zhodnotili jako přínosnou regulaci. Závěrem autoři apelují na čtenáře, aby byli obezřetní a důsledně prověřovali obsah, se kterým se seznámili prostřednictvím internetu, tak aby snížili riziko kybernetických útoků, zneužití osobních údajů a manipulace skrze obsah vytvořený pomocí systémů AI.

Mgr. Michal Hašan,
advokát

Mgr. Bc. Karel Pelikán,
advokátní koncipient

Doležal & Partners s.r.o., advokátní kancelář

Koliště 1912/13
602 00 Brno

Růžová 1416/17
110 00 Praha

tel.: +420 543 217 520
e-mail: office@dolezalpartners.com

[1] Prompt je možné charakterizovat jako pokyn pro konkrétní systém AI k vytvoření určitého výstupu daného systému.

[2] Z důvodu, že Akt o umělé inteligenci nebyl v době psaní tohoto článku vyhlášen v Úředním věstníku Evropské unie vychází tento článek ze znění Aktu o umělé inteligenci dle Legislativního usnesení Evropského parlamentu ze dne 13. března 2024 o návrhu nařízení Evropského parlamentu a Rady, kterým se stanoví harmonizovaná pravidla pro umělou inteligenci (Akt o umělé inteligenci) a mění určité legislativní akty Unie, dostupné >>>zde.

[3] Pro účely tohoto článku se generativní AI rozumí obecný systémem AI dle Aktu o umělé inteligenci.

[4] Obecný systém je Aktem o umělé inteligenci definován jako: „systém AI založený na obecném modelu AI, který je schopen sloužit různým účelům, a to jak pro přímé použití, tak pro integraci do jiných systémů AI“

[5] Národní úřad pro kybernetickou a informační bezpečnost (NÚKIB) definuje phishing, jako: „techniku sociálního inženýrství, jejímž cílem je získat od uživatelů jejich důvěrné, nejčastěji přihlašovací, údaje. Spočívá v rozesílání podvodných e-mailů, zpráv na sociálních sítích nebo sms zpráv, které se mohou maskovat třeba jako oznámení, že je kapacita naší e-mailové schránky naplněna, nebo jako výhodná nabídka. Zprávy působí důvěryhodným dojmem, využívají například název známé instituce či firmy včetně jejího loga, barev či jiných grafických prvků a většinou obsahují odkaz na podvodné internetové stránky.“ Srov. Spear-phishing a jak se před ním chránit [online]. NÚKIB [cit. 11.06.2024] >>> zde.

[6] Dle NÚKIB: „Spear-phishing je personalizovaná forma podvodných e-mailů (phishingu), která cílí na konkrétní osobu nebo skupinu osob. Základem útoku je tzv. sociální inženýrství, tedy techniky manipulace oběti k tomu, aby se chovala způsobem, který není v jejím zájmu. V kontextu kybernetické bezpečnosti jde většinou o snahu získat od cílové oběti konkrétní informace (např. heslo) nebo uživatele přesvědčit ke stažení přílohy obsahující malware (obecné označení pro škodlivý software) a jejímu otevření. Srov. Spear-phishing a jak se před ním chránit [online]. NÚKIB [cit. 11.06.2024] >>> zde.

[7] How AI is changing phishing scams [online]. Microsoft [cit. 11.06.2024] >>> zde.

[8] Akt o umělé inteligenci definuje „deep fake“ obsah jako „obrazový, zvukový nebo video obsah vytvořený nebo manipulovaný umělou inteligencí, který se podobá existujícím osobám, objektům, místům nebo jiným subjektům či událostem a který by se dané osobě mohl nepravdivě jevit jako autentický nebo pravdivý.“

[9] Doppelgängers and deepfakes: How Russian trolls are meddling in the world’s second-biggest democratic vote [online]. CNN [cit. 15.06.2024] >>> zde.

[10] Níže uvedená vybraná ustanovení smluvní dokumentace jsou aktuální ke dni 15.06.2024 a jsou volně přeložena autory tohoto článku.

[11] Sharing & publication policy [online]. OpenAI [cit. 15.06.2024] >>> zde.

[12] Europe Terms of Use [online]. OpenAI [cit. 15.06.2024] >>>zde.

[13] Acceptable Use Policy [online]. Stability AI [cit. 15.06.2024] >>> zde.

[14] Terms of Service [online]. Midjourney [cit. 15.06.2024] >>> zde.

[15] Poskytovatelem dle Aktu o umělé inteligenci je: „fyzická nebo právnická osoba, veřejný orgán, agentura nebo jiný subjekt, který vyvíjí systém AI či obecný model AI nebo nechává vyvíjet systém AI či obecný model AI a uvádějí je na trh nebo které uvádějí systém AI do provozu pod svým vlastním jménem nebo ochrannou známkou, ať už za úplatu, nebo zdarma."

[16] C2PA in DALL·E 3 [online]. OpenAI [cit. 15.06.2024] >>> zde.

[17] Označování obsahu vygenerovaného AI na Instagramu [online]. Instagram [cit. 18.06.2024] >>> zde.

[18] Our Approach to Labeling AI-Generated Content and Manipulated Media [online]. Meta [cit. 18.06.2024] >>>zde.

[19] Labeling AI-Generated Images on Facebook, Instagram and Threads [online]. Meta [cit. 18.06.2024] >>> zde.

[20] Článek 50 odst. 2 Aktu o umělé inteligenci.

[21] Článek 50 odst. 4 Aktu o umělé inteligenci.

Další články:

- Posouzení shody dle AI Act - zkušenosti z praxe
- Začínají soudy zohledňovat náklady podnikatelů při plnění právních povinností v oblasti e-commerce?

- Byznys a paragrafy, díl 35: Ručení za dluhy z podnikání u OSVČ a s.r.o.
- Bezpilotní systémy vlastní konstrukce v kategorii Specific: regulatorní požadavky a praktické aspekty
- Nefungující rozsah péče o dítě. Cesta přes využití terapie a dalších opatření podle ustanovení § 503 zákona o zvláštních řízeních soudních
- De iure traktor, de facto nákladní vozidlo, už ne tolik výhodná dualita
- Digitální důkazy z webu v soudním řízení: jak doložit, co bylo online zveřejněno?
- Pokuta 32 mil. EUR pro Dacia/Renault – evropské soutěžní úřady tvrdě došlapují na no-poaching. Měla by Vaše společnost být na pozoru?
- Rozdělení společného jmění manželů v případech výdělečné činnosti pouze jednoho z manželů
- Oběť znásilnění má nárok na peněžitou satisfakci
- Digitalizace AML povinností: jak technologie mění plnění povinností pro tisíce povinných osob