

Pravidla pro označování deep fake a transparentnost AI obsahu

Čl. 1 Úvod

1. Tato příloha stanoví zvláštní pravidla pro označování obsahu představujícího deep fake a pravidla pro uplatnění výjimky z povinnosti označování textů vytvořených nebo manipulovaných pomocí systému AI. Jedná se o text, jenž je zveřejněn za účelem informování veřejnosti o záležitostech veřejného zájmu podle čl. 50 odst. 4 a 5 nařízení o umělé inteligenci.
2. Není-li v této příloze stanoveno jinak, použijí se pravidla pro označování obsahu vytvořeného nebo upraveného pomocí systému umělé inteligence podle směrnice Pravidla pro užívání umělé inteligence v Českém rozhlasu a jejích ostatních příloh.
3. Pro způsoby užití uvedené v těchto pravidlech platí, že u nich byla posouzena rizika a dopady v souladu se směrnicí Pravidla pro užívání umělé inteligence v Českém rozhlasu.

Čl. 2 Výklad pojmů pro účely směrnice AI

1. Deep fake je obrazový, zvukový nebo video obsah vytvořený nebo manipulovaný umělou inteligencí, který se podobá existujícím osobám, objektům, místům, subjektům či událostem a který by se dané osobě mohl nepravdivě jevit jako autentický nebo pravdivý.

Příklady deep fakes:

*Obrázek se dvěma profesionálními fotbalisty před budovou připomínající fotbalový stadion upravený AI;
Video generované AI, které obsahuje AI vygenerované zobrazení influencera v reklamním nebo propagačním kontextu;*

Video vygenerované AI, na kterém je zobrazena osoba připomínající politika, která pronáší projev před publikem;

Zvuk generovaný AI zahrnující klonování hlasu pravidelných moderátorů zpravodajského podcastu a hostujícího partnera diskutujícího o některých zpravodajských událostech

Příklady, které nepředstavují deep fakes:

Obrázek scény se sfingou letící nad Eiffelovou věží generovaný AI;

Karikatura již existujícího obrázku zobrazujícího historickou událost vygenerovaná AI;

Video myši vygenerované AI, které se v rámci reklamní kampaně hádají lidskou řečí o nejlepší druh sýra;

Rozhlasové vysílání upravené AI, kdy dochází k úpravě technických zvukových parametrů (např.: normalizace úrovně hlasitosti, redukce šumu, komprese zvuku apod.) beze změny skutečně pronesených slov nebo způsobu mluvení.

Příklady jsou převzaty z Guidelines Evropské komise k čl. 50 nařízení o umělé inteligenci.

2. Text vytvořený nebo manipulovaný pomocí systému AI, jenž je zveřejněn za účelem informování veřejnosti o záležitostech veřejného zájmu (dále také jako „text ve veřejném zájmu“), se rozumí takový text zveřejněný neurčitému počtu nesouvisejících čtenářů s cílem

informovat veřejnost o záležitostech relevantních pro společnost jako celek, ať už na národní, regionální nebo mezinárodní úrovni.

Čl. 3 Označování deep fake

1. Deep fake musí být označen způsobem stanoveným touto směrnicí AI a jejími ostatními přílohami a dalšími pokyny či manuály upravujícími označování obsahu generovaného nebo upravovaného AI.
2. Pro deep fake platí vedle obecných pravidel označování tato zvláštní pravidla:
 - a) označení musí být jasně a zřetelně vnímatelné a rozlišitelné nejpozději v okamžiku prvního zpřístupnění obsahu uživateli,
 - b) u vizuálního obsahu vytvořeného nebo manipulovaného umělou inteligencí musí být označení umístěno přímo v obsahu, případně formou ekvivalentní alternativy k zabudování přímo do obsahu (a to např. překryv uživatelského rozhraní, který se fyzickým osobám jeví jako součást obsahu), na vhodném místě, kde nejsou překrývající se prvky (např. pravý horní roh);
 - c) u audiovizuálního obsahu musí být označení uvedeno na začátku obsahu; je-li to s ohledem na délku, způsob užití nebo způsob dalšího šíření vhodné nebo technicky možné, zobrazí se rovněž opakovaně v průběhu přehrávání, zejména po přerušeních,
 - d) u zvukového obsahu musí být na začátku obsahu uvedeno zvukové upozornění; u delších zvukových záznamů se zvukové upozornění podle technických možností opakuje v přiměřených intervalech, zejména po přerušeních; je-li k dispozici obrazovka, bude kromě zvukového upozornění poskytnuto také doplňkové vizuální upozornění v podobě ikony nebo ekvivalentního označení (např. pokud je k dispozici obrazovka v automobilu nebo na displeji smartphonu).
 - e) je-li to technicky možné, zachová se označení i při dalším šíření obsahu prostřednictvím digitálních služeb Českého rozhlasu a třetích stran.
3. Označení deep fake se provádí stejným označením jako u ostatního obsahu vytvořeného nebo upraveného pomocí umělé inteligence podle této směrnice AI a ostatních příloh. Tato příloha stanoví pouze zvláštní pravidla pro jeho umístění, dobu zobrazení a četnost.
4. Je-li deep fake součástí zjevně uměleckého, tvůrčího, satirického, fiktivního nebo obdobného díla nebo programu, povinnost transparentnosti podle tohoto článku se omezuje na zveřejnění informace o existenci obsahu vytvořeného nebo manipulovaného pomocí systému AI vhodným způsobem, který nebrání zobrazení ani užívání díla v souladu se čl. 4 této přílohy.

Čl. 4 Deep fake v uměleckých, tvůrčích, satirických, fiktivních a obdobných dílech

1. Je-li deep fake součástí zjevně uměleckého, tvůrčího, satirického, fiktivního nebo obdobného díla nebo programu, označení se provede způsobem stanoveným touto směrnicí AI, jejími přílohami a souvisejícími manuály upravujícími označování obsahu vytvořeného nebo upraveného pomocí systému umělé inteligence, přičemž z hlediska jeho umístění, doby zobrazení a četnosti se provede tak, aby nebránilo běžnému zobrazení, užívání ani vnímání díla.
2. Označení musí být i v těchto případech jasně vnímatelné, rozlišitelné a zpřístupněné uživateli nejpozději při prvním seznámení s obsahem. Musí být zobrazeno nebo jinak zpřístupněno po dobu dostatečnou k jeho běžnému vnímání.

3. Není-li s ohledem na povahu díla/pořadu vhodné umístění označení přímo v obsahu, může být označení umístěno jiným vhodným způsobem v jeho bezprostřední souvislosti tak, aby bylo uživateli zřejmé a přístupné nejpozději při prvním seznámení s obsahem, zejména:
 - a) v popisu pořadu nebo obsahu,
 - b) v úvodních nebo závěrečných titulcích,
 - c) jako součást uživatelského rozhraní digitální služby,
 - d) v doprovodných informacích nebo jiným obdobným vhodným způsobem.
4. Pro účely tohoto článku se za deep fake, který je součástí zjevně uměleckého, tvůrčího, satirického, fiktivního nebo obdobného díla nebo programu, považuje deep fake, který je součástí obsahu, jehož umělecká, tvůrčí, satirická, fiktivní nebo obdobná povaha je zřejmá zejména s ohledem na jeho styl nebo formát a kontext, ve kterém je prezentován. Výjimka podle tohoto článku se neuplatní, slouží-li takový obsah primárně k informování veřejnosti nebo ke komerčním účelům a je-li jako takový rozpoznatelný.

Čl. 5 Text zveřejňovaný za účelem informování veřejnosti

1. Text vytvořený nebo manipulovaný systémem AI a zveřejněný za účelem informování veřejnosti o záležitostech veřejného zájmu podléhá podle nařízení o umělé inteligenci povinnosti označení v souladu s obecnými pravidly pro označování obsahu vytvořeného nebo upraveného systémem AI podle čl. 4. této přílohy, zejména jasným, rozlišitelným a vnímatelným způsobem. Tato povinnost se neuplatní, jsou-li splněny obě tyto podmínky:

- a) text prošel před zveřejněním lidskou redakční kontrolou,
- b) za jeho zveřejnění nese redakční odpovědnost Český rozhlas.

Vzhledem k tomu, že veškerý textový obsah zveřejňovaný Českým rozhlasem prochází před zveřejněním redakční kontrolou v rámci standardních redakčních postupů podle Kodexu Českého rozhlasu a Žurnalistických zásad, budou tyto podmínky v naprosté většině případů splněny a text se neoznačuje.

2. Odpovědný editor nebo šéfredaktor v rámci kontroly podle odstavce 1 písm. a) před zveřejněním ověří zejména

- a) věcnou správnost,
- b) úplnost,
- c) soulad s Kodexem Českého rozhlasu a Žurnalistickými zásadami.

Samotná formální nebo pravopisná kontrola textu ani výlučně automatizovaná kontrola nepostačuje k naplnění podmínky podle odstavce 1 písm. a).

3. Je-li systém AI použit k obsahové úpravě nebo doplnění textu až po provedení kontroly podle odstavce 1 písm. a), považuje se podmínka podle odstavce 1 písm. a) za nesplněnou a takto upravený text musí být před zveřejněním označen v souladu s obecnými pravidly pro označování obsahu vytvořeného nebo upraveného systémem AI podle této přílohy, ledaže je před jeho zveřejněním provedena nová úplná kontrola podle odstavce 1 písm. a).

4. Pro účely uplatnění výjimky podle odst. 1 tohoto článku platí, že Český rozhlas při zveřejňování textů vytvořených nebo manipulovaných pomocí systémů umělé inteligence využívá své

stávající redakční postupy, systém lidské kontroly a redakční odpovědnosti podle Kodexu Českého rozhlasu, Žurnalistických zásad a dalších interních předpisů. Tyto postupy se považují za splnění podmínek výjimky podle čl. 5 odst. 1 této přílohy, a tedy za splnění požadavků na lidskou redakční kontrolu podle čl. 50 odst. 4 nařízení o umělé inteligenci, zahrnují-li v konkrétním případě ověření podle čl. 5 odst. 2 této přílohy.

Čl. 6 Redakční odpovědnost

1. Český rozhlas vede interní dokumentaci prokazující uplatnění výjimky podle čl. 5 odst. 1 této přílohy.
2. Dokumentace obsahuje zejména
 - a) určení organizačního útvaru, funkce nebo osoby (příslušného vedoucího na 1. nebo 2. stupni řízení) vykonávající redakční odpovědnost,
 - b) popis redakčního postupu zajišťujícího lidskou redakční kontrolu textů před zveřejněním,
 - c) určení odpovědných osob vykonávajících redakční kontrolu.
3. Požadavek na dokumentaci podle tohoto článku je naplněn tím, že příslušná pravidla jsou stanovena touto Směrnicí, zákonem o ČRo, Kodexem Českého rozhlasu, Žurnalistickými zásadami a interními předpisy upravujícími redakční kontrolu a odpovědnost (např. organizačním řádem); samostatná evidence jednotlivých případů redakční kontroly textu se nevyžaduje. V případech postoupených podle čl. 9 odst. 3 této přílohy se výsledek posouzení zaznamená formou zápisu pracovní skupiny Strategického výboru k využití AI v ČRo.

Čl. 7 Interní procesy pro označování deep fake

1. Český rozhlas vede dokumentaci o způsobu označování deep fake podle čl. 3 a 4 této přílohy.
2. Dokumentace obsahuje zejména
 - a) používaný způsob označování podle této přílohy, směrnice AI a jejich ostatních příloh a souvisejících manuálů,
 - b) pravidla umístění označení podle jednotlivých typů obsahu,
 - c) určení odpovědných osob za správné označení,
 - d) postup kontroly správnosti označení,
 - e) postup nápravy v případě chybného nebo nesprávného označení.
3. Požadavek na dokumentaci podle tohoto článku je naplněn tím, že příslušná pravidla jsou stanovena touto Směrnicí, touto přílohou, Kodexem Českého rozhlasu, Žurnalistickými zásadami, Manuálem pro označování obsahu vytvořeného nebo upraveného za pomoci systémů AI a souvisejícími interními pracovními postupy jednotlivých redakcí a útvarů; samostatná evidence jednotlivých označených výstupů se nevyžaduje, ledaže ji pro konkrétní typ obsahu stanoví příslušná příloha č. 1 až 4 směrnice AI (např. evidence promptů v redakčním systému atd.).

Čl. 8 Transparentnost

1. Český rozhlas zveřejňuje způsobem umožňujícím dálkový přístup informace o:

- a) používaných způsobech a podobě označování, jejich umístění podle druhu obsahu (obrazový, audiovizuální, zvukový, textový) a popis
 - b) významu jednotlivých označení.
2. Označení podle této přílohy nenahrazuje splnění dalších povinností vyplývajících z právních předpisů ani interních předpisů Českého rozhlasu.

Čl. 9 Odpovědnost

1. Za správné označení deep fake a textu ve veřejném zájmu odpovídá příslušný vedoucí na 1. nebo 2. stupni řízení.
2. Za posouzení, zda obsah naplňuje znaky deep fake nebo zda jde o text ve veřejném zájmu podle první věty odst. 1 čl. 5 této přílohy, odpovídá příslušný vedoucí na 1. nebo 2. stupni řízení.
3. V případě pochybností se věc předloží Pracovní skupině Strategického výboru k využití AI v ČRo.