Kodex správné praxe pro transparentnost obsahu vytvořeného umělou inteligencí

- navazuje na článek 50 Aktu o AI
- cílem je usnadnit rozpoznání obsahu vytvořeného nebo upraveného AI, posílit důvěru veřejnosti a sjednotit pravidla pro označování AI generovaného obsahu
- naznačuje standard transparentního využívání AI v Evropské unii.

Poskytovatelé AI systémů

Mají zajistit, aby AI výstupy byly:

- strojově označené,
- detekovatelné jako plně generované nebo asistované,
- efektivní (jasný původ obsahu)
- interoperabilní,
- odolné vůči úpravám a útokům,
- spolehlivé.

Neexistuje jediná technologie, která by vše vyřešila. Proto se má používat kombinace více metod současně.

Povinné vrstvy:

- metadata (informace o souboru)
- neviditelný vodoznak
- otisky obsahu a logování

Má být možné dohledat:



odkud obsah vznikl | jakými kroky prošel | co vytvořil člověk a co AI | jaké další AI zásahy proběhly

Další důležité zásady pro poskytovatele:

- neztrácet ani nemařit existující AI označení,
- podporovat společné evropské standardy,
- pravidelně testovat řešení,
- školit zaměstnance,
- spolupracovat s dohledovými orgány

Kodex správné praxe pro transparentnost obsahu vytvořeného umělou inteligencí

Nasazovatelé AI

Organizace a osoby, které AI obsah zveřejňují

Hlavní princip:
Pokud vytvoříte deepfake (obraz, video, audio), nebo zveřejníte AI generovaný/manipulovaný text k informování veřejnosti, musíte to jasně označit.

Obsah plně generovaný AI vs. obsah generovaný s asistencí AI

Pravidla pro deepfake

Video
- označení při prvním zobrazení, příp. po celou dobu přehrávání,
- doplněné textovým nebo zvukovým upozorněním.

Obrázky
- viditelná ikona, nesmí být skrytá.

Audio
- upozornění na začátku,
- podcasty a delší formáty s opakovaným upozorněním.

Výjimka pro uměleckou a kreativní tvorbu
je možné použít méně rušivé označení, aby nebyl narušen zážitek, označení ale musí stále existovat.

Pravidla pro AI texty

Týkají se textů zveřejněných za účelem informovat veřejnost o věcech veřejného zájmu.

Označit se nemusí, pokud:
- proběhla lidská redakční kontrola,
- existuje odpovědná osoba nebo redakce, která za text nese odpovědnost.

Jinak se označit musí, např.:
- automaticky generované zprávy,
- public affairs texty,
- politické či společenské komentáře generované AI.

Označování obsahu vytvořeného s využitím AI

AI může být užitečným pomocníkem při tvorbě textů, prezentací, analýz, grafiky i dalších dokumentů. Současně však platí, že každý zaměstnanec veřejné správy odpovídá za transparentní a důvěryhodné využívání AI. Občan i kolega by měli mít možnost přiměřeně rozpoznat, jakou roli AI při vzniku obsahu sehrála.

Rozlišujeme dvě základní situace:

Obsah plně generovaný AI

vytvořen převážně nebo zcela AI, člověk provedl pouze minimální nebo formální úpravy

Typicky:

- zadám pokyn a AI vytvoří celý článek,
- AI připraví celou prezentaci a já opravím překlepy,
- AI vytvoří obrázek, grafiku nebo video bez mých významných změn,
- AI připraví analýzu, kterou převezmu bez větších zásahů.

Obsah vytvořený s asistencí AI

člověk je hlavním autorem a AI sloužila jako pomocný nástroj

Typicky:

- AI navrhne osnovu dokumentu,
- AI zkrátí nebo shrne text,
- AI provede jazykovou korekturu,
- AI navrhne několik variant formulací,
- AI pomůže s překladem.

já obsah zásadně dopracuji, přepracuji nebo odborně doplním

Položte si jednoduchou otázku:

Kdybych odstranil práci AI, zůstal by zde stále můj vlastní dokument?

Ano → AI asistovaný obsah

Ne → plně generovaný AI obsah

Nejde o to, kolik minut jsem na textu strávil, ale **kdo je skutečným autorem obsahu a myšlenkového sdělení**.

Materiály určené občanům nebo veřejnosti

U obsahu zveřejňovaného navenek je vhodné postupovat transparentněji.

Označení se použije zejména, pokud:

- byl text, obrázek, video nebo zvuk vytvořen plně AI,
- AI významně ovlivnila obsah nebo sdělení,
- může vzniknout dojem, že obsah vytvořil člověk bez pomoci AI,
- jde o informační materiály, články, kampaně, prezentace, audiovizuální obsah nebo komunikaci s veřejností.

Příklady označení:

Obsah převážně vytvořený AI

Tento materiál byl vytvořen s významným využitím umělé inteligence a následně zkontrolován odpovědnou osobou.

Obsah vytvořený s asistencí AI

Tento materiál byl připraven zaměstnanci úřadu s využitím nástrojů umělé inteligence. Finální obsah byl ověřen a schválen odpovědnou osobou.

Deepfake

Pokud byl vytvořen nebo upraven obraz, video nebo hlas osoby pomocí AI tak, že může působit jako autentický, musí být tato skutečnost zřetelně uvedena.

Interní práce uvnitř úřadu

U běžných pracovních dokumentů není cílem zaměstnance administrativně zatěžovat. Ve většině případů není nutné označovat každý e-mail, poznámku nebo pracovní koncept vytvořený s pomocí AI.

Označení **zpravidla není nutné**:

- při brainstormingu,
- při přípravě prvních návrhů textů,
- při korekturách,
- při sumarizaci podkladů pro vlastní potřebu,
- při práci s koncepty či pracovními poznámkami.

Označení **je vhodné**, pokud:

- dokument bude sloužit jako významný podklad pro rozhodování,
- bude předáván dalším útvarům,
- obsahuje rozsáhlé AI generované části,
- může vzniknout nejasnost ohledně autorství nebo původu obsahu.

Praktické označení interních dokumentů:

Například poznámkou:

Při zpracování tohoto dokumentu byla využita umělá inteligence pro návrh struktury, sumarizaci podkladů nebo jazykovou úpravu textu.

nebo

Návrh textu byl vytvořen s využitím AI a následně odborně posouzen a upraven autorem.

Co platí vždy, bez ohledu na to, zda je obsah označen:

- odpovědnost za výstup nese vždy zaměstnanec nebo organizace, nikoliv AI,
- výstupy AI musí být věcně ověřeny,
- AI nesmí samostatně rozhodovat o právech a povinnostech občanů,
- využití AI nesmí být utajováno tam, kde by mohlo ovlivnit důvěru,
- každý zaměstnanec by měl být schopen vysvětlit, jak byla AI při vzniku materiálu využita.

Kdy interní materiály označovat a kdy ne

Čím více se obsah blíží veřejné komunikaci a čím větší dopad může mít na občany nebo rozhodování, tím vyšší má být transparentnost ohledně využití AI.

Situace	Označit?	Doporučení
Brainstorming nápadů pro sebe	**Ne**	Běžné pracovní využití AI
Návrh interního e-mailu kolegům	**Ne**	Pokud jde o běžnou komunikaci
Korektura vlastního textu pomocí AI	**Ne**	AI pouze upravuje jazyk
Shrnutí dlouhého dokumentu pro vlastní práci	**Ne**	Slouží pouze jako pracovní pomůcka
Analytický podklad pro vedení vytvořený převážně AI	**Ano**	Uvést využití AI a provést odbornou kontrolu
Interní prezentace pro více útvarů s rozsáhlým AI obsahem	**Ano**	Doporučuje se stručná poznámka o využití AI
Návrh metodiky nebo interního předpisu generovaný AI	**Ano**	Důležitý dokument s vlivem na rozhodování
Podklad pro správní rozhodnutí nebo významné manažerské rozhodnutí	**Ano**	Vhodné evidovat využití AI a zachovat kontrolní stopu

Pokud AI pomohla pouze Vám, označovat zpravidla nemusíte. Pokud AI významně vytvářela obsah, který bude číst nebo používat někdo další, označení je vhodné.