

Jak poznat digitálně upravený či uměle vytvořený obsah a jak mu porozumět

Snadno dostupné a uživatelsky přívětivé modely umělé inteligence nám mohou pomoci při učení i tvorbě obsahu. Mohou ale také **zvýšit riziko, které nepravdivé informace představují** pro otevřenou společnost a demokratický diskurz. Je důležité, aby se veřejný prostor nezahltil nepravdivými informacemi založenými na digitálně upraveném nebo uměle (umělou inteligencí) vytvořeném obsahu.

Problém částečně řeší nové technologie, např. iniciativy pro prokazování původu obsahu a detekční software. **Technická řešení ale nejsou ani zdaleka dokonalá. Proto potřebujeme nezávislé fact-checkery, kteří společnosti poskytují ověřené informace.**

Nabízíme stručný přehled toho, co profesionální, nezávislí fact-checkeři dělají, aby rozpoznali a vyvrátili nepravdivé informace vytvořené pomocí umělé inteligence, a toho, co si z nich můžete vzít i vy.

Umělá inteligence na vzestupu

Nepravdivé informace vygenerované umělou inteligencí dnes tvoří pouze mizivé procento obsahu, který ověřují profesionální, nezávislí fact-checkeři. Mnohem větší část jejich práce tvoří digitálně upravený obsah.

Interní průzkum mezi členy EFCSN ale ukázal, že **většina fact-checkerů se shoduje na tom, že význam digitálně upraveného či uměle vytvořeného obsahu bude jen narůstat**. A [nedávné příklady](#) v souvislosti s evropskými volbami tuto prognózu podporují.

VYSVĚTLIVKA: Pojem *digitálně upravený* označuje jakýkoli obsah, který byl výrazně upraven za účelem změny původního vyznění, a to včetně úprav pomocí nástrojů umělé inteligence. To nezahrnuje úpravy zlepšující kvalitu či srozumitelnost.

Pojem *vytvořený umělou inteligencí* (nebo *uměle vytvořený*) označuje obsah, který byl vytvořen pomocí nástroje umělé inteligence.



Technologie se vyvíjejí rychle, ale nelze se spoléhat pouze na ně.

Odborníci na AI a profesionální fact-checkeři se shodují, že **nástroje pro detekci umělé inteligence na rozpoznání či vyvrácení uměle vytvořeného či digitálně upraveného obsahu nestačí.**

Odborníci doporučují seznámit se před použitím detektoru s fungováním generativní AI a systému pro její rozpoznání. Když fact-checkeři rozumí, jak se modely AI učí, a spojí to s trochou statistiky, mohou rozeznat silné i slabé stránky nástroje a jeho šanci na úspěch. **Přesto mohou být tyto nástroje užitečným odrazovým můstkem.**

Iniciativy pro **prokazování původu obsahu**, jako např. specifikace C2PA, mohou pomoci s ověřením zdroje a původu mediálního obsahu. Ani vodoznaky a specifikace ale nejsou zcela nepřekonatelnou překážkou.

Jak nás ovlivňují dezinformace vytvořené pomocí AI

„Kdykoliv na obsah zareagujete instinktivně, podařilo se mu obejít váš rozum.“

- Christine Dugoin*

PSYCHOLOGIE: Vlivové operace jsou často navrženy tak, aby využívaly kognitivní zkreslení.

Pochopení, jak zkreslení ovlivňují nejen vás, ale i vaše publikum, vám může pomoci v boji proti dezinformacím.

ÚČEL: Proč by někdo zneužíval AI k tvorbě a šíření dezinformací? Jak chtějí ovlivnit skutečný svět?

- Chtějí rozšířit svůj vliv na novou zemi či komunitu?
- Chtějí se vyhnout odhalení a zahltit fact-checkery mnoha variantami podobných tvrzení?
- Chtějí ovlivnit myšlení či přesvědčení za pomoci celé sítě falešných profilů?

* Christine Dugoin je výzkumná pracovnice v oblasti informačního vlivu na Sorbonně.

Vyvracení nepravdivých informací vyžaduje komplexní přístup a detailní pochopení problematiky

Když tedy detektory nefungují, co ano? Je důležité pochopit nejen obsah tvrzení, ale i jeho kontext. Profesionální fact-checkerů skvěle ovládají potřebné investigativní dovednosti. Mají několik tipů.

„Detektory nikdy nefungují zcela spolehlivě – a nepočítám s tím, že někdy budou.“
– Henk van Ess**



OVĚŘIT ZDROJ: Lze ověřit jeho identitu? Jaká témata probírá a sdílí? Kdo na obsah reaguje? Jaký vliv může mít obsah na čtenáře?



ZJISTIT VĚROHODNOST: Informace si samostatně ověřujte u důvěryhodného zdroje, jako jsou např. odborníci se zkušenostmi v dané oblasti. Dává obsah smysl, je v souladu s vašimi znalostmi?



Využívejte techniky **FORENZNÍ ANALÝZY MÉDIÍ** k doplnění tradiční investigativní žurnalistiky a dokumentárního výzkumu. Může jít například o extrakci dat, geolokaci, biometrické údaje nebo analýzu vzorců.



UČIT SE & PŘIZPŮSOBIT: Tvůrci nepravdivých informací využívající umělou inteligenci se neustále přizpůsobují. Přizpůsobte se také.

SDÍLEJTE SVOU PRÁCI

Experti doporučují sdílet nejen vyvrácené tvrzení, ale také transparentní analýzu a odkazy na zdroje. Čtenářům to může pomoci lépe pochopit váš postup i kompletní narativ.

Občas je analýza důležitější než to, zda obsah vytvořila AI.

** Henk van Ess je odborníkem na zpravodajství z ověřených zdrojů a metody ověřování faktů.

Stručná příručka: Co dělat, co nedělat a na co se zaměřit

Následující indicie vám napoví, že je obsah digitálně upravený nebo vytvořený pomocí AI. Spolu s dalšími tipy z této příručky (kontext, investigativní metody a detekční software) vám mohou pomoci odhalit pravdu za tím, co vidíte.

Text

- Často (ale ne vždy) **ovládá gramatiku lépe** než člověk.
- Pravděpodobně používá **příliš formální jazyk nebo složitou strukturu**, obzvláště v kontextu sociálních sítí.
- **Nadužívá příslovce a přídavná jména.**
- Zcela postrádá emoce, humor, sarkasmus nebo obrazný jazyk.
- Mohou **chybět detaily** (jména, data, místa) nebo originální myšlenky.
- A hlavně: Jsou informace v textu pravdivé?

Video

- Nepoužívejte detektor obsahu vytvořeného pomocí AI na jednotlivé snímky z videa.
- Zaměřte se na **mimiku a pohyb obličeje**, například mrkání, a na to, jestli pohyb úst odpovídá zvukové stopě.
- Mohou se vyskytovat **ostré přechody a stříhy**.

Zvuk

- **Porovnejte podezřelou nahrávku s autentickým vzorem** pomocí nástrojů, které rozeznají odlišnosti v řečových vzorcích, dýchání či intonaci.
- Při použití detektorů se vyhněte nekvalitním vzorkům, které mají v pozadí ruch či statický šum.
- Může mít **nepřirozené či strojové vzorce řeči**, chybí pauzy či přirozené dýchání.

Obrázky

- Hledejte **nepřirozené detaily**: dokonalá pleť, rozmazané pozadí, nepřirozená krása, zvláštní světlo a další neobvyklé znaky, např. přebytečné prsty.
- **Hledejte vodoznak** běžného generátoru obrázků.
- Soustředte se na detaily: dávají smysl? Hodí se do kontextu?
- Při používání detektorů **používejte rané verze obrázků nebo verze s vysokým rozlišením, ne několikrát předsílené soubory**.