Ta nejhorší „dezinformace“. Rakušanovo vnitro platí výzkum věci, která budí hrůzu

- CZ24 News | 12. ledna 2024

ČESKO: Kremelský, bruselský či jakýkoliv jiný narativ jako největší ohrožení demokracie? Podle expertů budeme za pár let s nostalgií vzpomínat na časy, kdy dezinformace měly pouze podobu příběhu se specifickými akcenty. Mnohem větší výzvou se má stát výplod umělé inteligence zvaný deepfake. Falešné video, vypadající jako reálné, má prý potenciál skutečně ovlivňovat lidi a měnit svět. Rakušanovo vnitro proto platí výzkum českých odborníků, jak tomuto nebezpečí čelit. Někteří ovšem vyjadřují pochybnosti, zda je to vůbec možné.

Nedávno jsme psali o projektu CEDMO, plným názvem Středoevropská observatoř digitálních médií, výzkumné akci koordinované Václavem Moravcem, která se snaží přijít na to, jak ochránit společnost před dezinformacemi.

Součástí tohoto výzkumu, na který kromě milionů z veřejných prostředků věnovala milion eur třeba společnost Google, je i technologický výzkum možností obrany před deepfakes a dalšími výzvami rozvíjející se technologie umělé inteligence.

Deepfake. Otrocky by se to spojení dalo přeložit jako hloubkový podvrh, reálně se používá pro pokročilé mediální manipulace. Takové, které jsou i pro experty komplikovaně odhalitelné. Spojováno je především s neobyčejně realistickými úpravami audiovizuálních záznamů, ve kterých dokážou uměle vytvořit falešnou, ale důvěryhodně vypadající obrazovou i zvukovou stopu. Kvalita tohoto počítačového zpracování se přitom neustále zvyšuje. První zdařilý experiment s pornografickým videem celebrit se internetem virálně rozšířil v roce 2017, od té doby možnosti tohoto fixlování s videem pokročily o mnoho kroků.

Podle mediální analytičky Ireny Ryšánkové, se kterou ParlamentníListy.cz hovořily, je třeba se tímto tématem zabývat mediálně, vědecky i legislativně. Deepfakes totiž představují mnohem větší nebezpečí než vyhledávání „narativů“, které je oblíbeným tématem bádání a stejně oblíbeným dotačním titulem v současnosti.

Jako příklad uvedla aktuálně příběh z války v Gaze, kde byla jako zpravodajská zveřejněna fotografie dospělého muže arabského vzezření, nesoucího několik dětí na rukou a klopýtajícího přes sutiny města srovnaného se zemí. Pronikla i do západního tisku. Obrazová informace přitom byla kompletně vytvořena generativní umělou inteligencí DALL·E 3. A to tak dokonale, že pokud nejste na místě a neznáte-li detailně tvary budov či podobu muže na snímku, nerozeznáte falzum. Čímž se zásadně liší od dosavadních primitivních „fejků“ s fotografiemi užívanými doposud, jako že se vytáhla válečná fotografie z úplně jiného bojiště a přidal se k ní popisek.

CELÝ TEXT ZDE.

Stejným způsobem lze dnes podle Ryšánkové falšovat video, třeba s pokorným přiznáním politika k pedofilii. Stačí několik fotografií dotyčného, ze kterých uděláte trojrozměrný portrét a rozhýbete ho jako figurku v digitální hře. Pak si seženete jeho hlas a jeho projevy. Abyste vytvořili jeho hlasem úplně jiný projev. Protože stejným způsobem jako foto lze dnes falšovat audio – stačí k tomu hlasový vzorek o úplně jiném obsahu o rozsahu několika minut. A stejným způsobem lze falšovat text jakéhokoliv autora jen na základě dostatečně velkého vzorku jeho textu, z něhož budou jasné

markanty, kterých se autor v minulosti dopustil a bude se dopouštět i nadále.

Téma umělé inteligence a deepfakes je v posledních dnech diskutováno velmi intenzivně. I Česká televize mu ve středu večer věnovala aktuální díl svého pořadu Bilance a ukázala, jak je už s dnešními technologiemi snadné vložit ministru Ivanu Bartošovi do úst slova, která nikdy neřekl.

Tématem se v poslední době začínají intenzivně zabývat i výzkumníci. Moravcova observatoř se prezentuje multidisciplinárním týmem, kde o obraně proti deepfakes společně bádají experti na společenské vědy i na technické řešení. Kromě expertů z ČVUT jsou to třeba i experti ze Slovenska.

Výzkumu se věnují i další česká akademická pracoviště, například Vysoké učení technické v Brně.

„Téma deepfakes je skloňováno již řadu let v ČR, Evropě i dalších regionech. Vyšetřovány jsou podvody nebo pokusy o podvod v řádu až milionů dolarů. S rozvojem umělé inteligence klesají technologické bariéry a zvyšuje se dostupnost technologií. Ty mohou být zneužity, na což reaguje AI Act připravovaný na evropské úrovni. Důležitost a význam výzkumu deepfakes tedy reaguje zejména na postupné zvýšení dostupnosti technologií," objasnili výzkumníci pro ParlamentníListy.cz svůj vědecký zájem.

Zatímco dříve bylo podobných případů méně s vysokou škodou, nyní je podle nich třeba reagovat na rostoucí četnost a cílení deepfakes na širší veřejnost.

Bilance ČT zmínila i hrozby v podobě šmejdských praktik, kdy budou falešná videa spojena třeba s již dnes užívanými fintami k získání údajů o bankovních účtech, které budou následně vybíleny.

VUT Brno se účastní i programu bezpečnostního výzkumu ČR na roky 2021–2026, který organizuje Ministerstvo vnitra.

„Hlavním cílem Programu SECTECH je prostřednictvím mobilizace potenciálu podnikového sektoru, zejména začínajících, malých a středních podniků, k participaci na vývoji a transferu nových bezpečnostních technologií podpořit dosažení technologické a technické úrovně, která umožní jednotlivým složkám bezpečnostního systému ČR získávat, osvojovat si, udržovat a rozvíjet specifické schopnosti pro zajištění bezpečnosti státu a jeho občanů," stojí ve výzvě, na kterou se přijímaly žádosti během roku 2023.

„VUT Brno se tématu deepfakes věnuje jako důležitému tématu ovlivňujícímu výzkum i společnost. Věnujeme se výzkumu jak obrazové části (podvržený obrázek nebo video), tak zvukové části (podvržený hlas jiného mluvčího)," vysvětluje pro ParlamentníListy.cz mluvčí Vysokého učení technického Petr Kubíček.

Aktuálně se podle něj výzkumníci věnují výzkumu nástrojů k boji proti hlasovým deepfakes. „Na realizaci projektu na fakultě informačních technologií spolupracují hned dvě výzkumné skupiny, a to Security@FIT z pohledu bezpečnosti a Speech@FIT z pohledu technologií rozpoznávání informací v hlasu. Do projektu je zapojena i společnost Phonexia s.r.o., která se dlouhodobě věnuje vývoji řešení hlasové biometrie," doplňuje mluvčí.

Projekt podle něj reaguje na aktuální hrozbu umožňující díky rozvoji systému modifikace a syntézy hlasu útoky v podobě cílené úpravy hlasových výstupů umělou inteligencí.

„Projekt přispěje k odhalování pokusů o páchání kriminality a zvýší povědomí odborné veřejnosti o tématu i dostupných nástrojích," dodává.

VUT Brno je v tomto směru dlouhodobě výzkumným partnerem řady institucí bezpečnostní komunity.

„Mnohé výzkumné týmy VUT Brno, například zmíněný tým Speech@FIT pod vedením prof. Černockého má zkušenosti i z úspěšné realizace projektů financovaných z amerických programů institucí DARPA a IARPA," doplňuje.

Pokud jde o projekt spolupráce s Ministerstvem vnitra, podle mluvčího „aktivity projektu jsou zaměřeny na zvýšení odolnosti hlasových biometrických systémů, vytvoření nástrojů umožňující boj proti škodlivému použití hlasových deepfakes a vyhodnocení schopnosti lidí rozpoznat hlasové deepfakes".

Spolupracují na něm dva týmy zabývající se výzkumem z bezpečnostního i technologického pohledu.

„O výše zmíněný projekt projevilo zájem Policejní prezidium ČR v roli uživatelské organizace. Vnímáme zde důležitost spolupráce kvůli nasměrování k praktickým výstupům, ale i dlouhodobý zájem českých autorit o adopci posledních generací technologií a moderních přístupů," vysvětluje mluvčí Kubíček.

Mediální expert Petr Žantovský ale pro ParlamentníListy.cz vyjádřil obavu, že z výzkumu hlasových deepfakes se v českých podmínkách stane spíše kšeft pro kamarády ministra, který ale snaze o zastavení tohoto fenoménu příliš nepřispěje.

Deepfakes neboli hlubinná manipulace podle mediálního experta skutečně je problém. Vzpomíná, jak již před několika lety dokázali tito experti rozebrat na nanočástice záznam hovořícího Baracka Obamy a vložit mu do úst úplně jiná slova. A protože lidé médiím důvěřují, je zde skutečně neomezený potenciál jejich manipulace.

„Je to skutečně velké nebezpečí. Vypustili jsme technologii, která dokáže komukoliv dát do úst cokoliv, aniž je to rozpoznatelné," varuje. Džin už přitom prý byl z láhve vypuštěn.

O nebezpečí deepfakes se podle Žantovského ve světě reálně diskutuje už přes pět let, jejich vývoj běží ještě mnohem déle. Víra, že dnes dokážeme přijít na to, jak tohoto džina vrátit zpátky do láhve, je podle Žantovského spíše pošetilá.

„Já jsem pro každý výzkum, který může lidstvo posunout, ale tady mi přijde, že to ten rozjetý vlak už zastavit nedokáže. Vzpomeňte si, jak to tu dopadlo, když šmludlové kolem pana Klímy chtěli napsat zákon proti dezinformacím a skončili na tom, že ani ten pojem dezinformace nedokázali definovat," připomíná s pochybnostmi, zda s takovou úrovní expertů dokážeme bojovat s jednou z největších technologických výzev současnosti.

Proto výzkum organizovaný vnitrem vnímá spíše jako klasickou „vyrobenou agendu", ve které spíše než o reálné výsledky jde o ukousnutí z koláče veřejných peněz.

Konkrétně smlouva mezi MV a VUT, podepsaná 3. 1. 2024, je uzavřena na maximální výši podpory v hodnotě 25 170 479 korun.

Na dotazy ParlamentníchListů.cz, jak má být výzkum využit pro posílení bezpečnosti země, Ministerstvo vnitra bohužel nereagovalo.

„Projekt reaguje na aktuální hrozbu, kdy se s rozvojem systému na modifikaci a syntézu hlasu množí útoky na systémy a osoby pomocí umělé inteligence a cílené úpravy hlasového výstupu text-to-speech na konkrétní hlas. Projekt přispěje k odhalování pokusu o páchání kriminality," vymezuje se ve smlouvě cíl výzkumu.

Podle expertů z VUT Brno ale je kromě technologického výzkumu nezbytná i osvěta mezi veřejností.

ZDROJ