

Hlasové deepfakes nedokáží rozpoznat lidé ani bezpečnostní systémy, útoků přibývá

30.1.2024 - Václav Koníček | Vysoké učení technické v Brně

Umělá inteligence se vyvíjí rychle a deepfake nahrávky hlasu si tak může doma a ve vysoké kvalitě vytvořit téměř každý

Hlasový deepfake je podmnožinou syntetické řeči – jde o automaticky vytvořené médium s využitím hlubokých neuronových sítí.

„Lidé neočekávají, že se s deepfake hlasem mohou setkat, a jsou tak schopni ignorovat i chyby nebo horší kvalitu nahrávky. V ohrožení jsou všichni uživatelé telefonu a sociálních sítí. Otevírají se tak možnosti vishingových útoků, což je kombinace deepfake hlasu a phishingu, na velké množství lidí,“ dodává výzkumník a upozorňuje, že vhodnou ochranou může být právě zvyšování obecného povědomí o problematice.

V ohrožení jsou podle něj všichni, kdo používají telefon, počítač či mají účet na sociálních sítích. Častý případ útoku typu social engineering je například vyzrazení interních informací ve firmách pomocí telefonátu. „Zazvoní telefon a ozve se váš kolega z jiné pobočky. Zná ty správné formulace i slova a tváří se, že mu nefunguje počítač a potřebuje, abyste se za něj podívali do systému a třeba mu sdělili přístupové údaje,“ popisuje Firc.

Deepfakes možnosti těchto social engineering útoků rozšiřují. Syntetické nahrávky si dnes ve vysoké kvalitě dokáží doma vytvořit i lidé bez velkých technických znalostí. A systémy hlasové biometrie, které ověřují identitu volajících do bank či call center, nedokáží spolehlivě rozeznat syntetickou nahrávku od reálné lidské řeči. „Testoval jsem dva komerčně rozšířené systémy hlasové biometrie a potvrdilo se, že ani ony nedokáží rozeznat pravou nahrávku od umělé,“ uvádí výzkumník.

„Jediné, co zatím odborníci odhalili, je, že u deepfake nahrávek se vyskytuje více energie ve vyšších frekvencích, zatímco u lidské řeči je tato energie rozložená lineárněji,“ vysvětluje Anton Firc | Autor: Václav Koníček

Největším problémem podle něj je, že ani vývojáři biometrických systémů nemají k dispozici metodiku, jak odolnost systémů proti deepfake útokům testovat. „Existují modely, detektory deepfakes, založené na neuronových sítích, které jsou schopné detekovat, zda se v nahrávce vyskytují anomálie, které se v běžné řeči nenacházejí, a vyhodnotit, jestli je pravá nebo syntetická. Je ale velmi náročné vysvětlit, na základě čeho se tyto modely opravdu rozhodují. Jediné, co zatím odborníci odhalili, je, že u deepfake nahrávek se vyskytuje více energie ve vyšších frekvencích, zatímco u lidské řeči je tato energie rozložená lineárněji,“ upozorňuje výzkumník a dodává, že detekce a správné testování deepfakes je teprve v začátcích.

Zatímco aktuálně jsou terčem útoků především banky a soukromé firmy, do budoucna mohou na díry v kybernetické bezpečnosti doplatit i běžní lidé.

„Jedna slovenská banka je vám pouze na základě hlasové verifikace ochotna vydat kreditní kartu. Vzhledem k tomu, že úniky dat jsou běžné a není problém koupit si něčí osobní informace, bude s pomocí deepfake nahrávek hlasu velmi snadné o kreditku zažádat na jiného člověka. A navíc, umělá inteligence se vyvíjí tak rychle, že brzy budeme schopni tyto útoky automatizovat a zapojit do nich jazykové modely, jako je ChatGPT. V nejhorším scénáři tak může vzniknout armáda umělých telemarketérů, kteří budou volat starším lidem a předstírat, že jsou například členové rodiny, měli

autonehodu a potřebují okamžitě poslat peníze,” nastiňuje Firc možné scénáře zneužití deepfake nahrávek v budoucnu.

Anton Firc se deepfakes poprvé zabýval v diplomové práci, v níž zkoumal odolnost hlasové biometrie vůči deepfake hlasu | Autor: Václav Koníček

Problematikou deepfakes v rámci kybernetické bezpečnosti se začalo zabývat i Ministerstvo vnitra, které podalo výzvu na bezpečnostní výzkum, na kterém Anton Firc (za skupinu Security@FIT) spolupracuje se skupinou Speech@FIT a společností Phonexia. Cílem je vyvinout nástroje, které dokáží uměle vytvořené nahrávky spolehlivě identifikovat.

<https://www.zvut.cz/tema/-f38144/hlasove-deepfakes-nedokazi-rozpoznat-lide-ani-bezpecnostni-systemy-utoku-pribyva-d250717>