Kommittén för teknologisk innovation & etik

Komet kommenterar

Deepfakes, manipulerade filmer – korta faktablad om aktuell teknik

Komet kommenterar 2020:27, publicerad 2020-11-04

Kort om manipulerade filmer gjorda med AI-teknik (eng. deepfakes) – för beslutsfattare och andra som är nyfikna på hur aktuell teknik påverkar samhället.

Kommenterad rapport

Davis, R. **Deepfakes**. *Tech factsheets for policymakers*. Ed. Jayanti A. Belfer Center for Science and International Affairs, Harvard Kennedy School, Spring 2020[1]

Korta faktablad om aktuell teknik

Belfer Center vid Harvard University ger ut en serie faktablad om aktuella teknikområden. Serien är riktad till politiska beslutsfattare i USA i syfte att ge överblick och förståelse av ny teknik. Komet Kommenterar gör en svensk uppföljning av serien.

Belfer Center for Science and International Affairs är del av Harvard Kennedy School of Government. Belfer arbetar bland annat med hur ny teknik kan komma till nytta i samhället.

Komet:s kommentarer

- Våren 2018 ställdes en fråga i Sveriges riksdag om vilka åtgärder regeringen planerar för att deepfakes (manipulerade videofilmer) och annan bild- och ljudmanipulation inte ska skada tilltron till rörliga medier.[2] Ansvarig minister svarade att den snabba teknikutvecklingen ställer allt högre krav på såväl enskilda individer som samhällets aktörer. Förutom god informationssäkerhet behöver de som företräder samhället vara källkritiska, hålla sig informerade och låta bli att sprida oriktig information. Därtill lyfte ministern fram arbetet med att stärka den digitala kompetensen bland medborgarna och påtalade att manipulation av bild kan vara straffbart som förtal.
- Sveriges Utbildningsradio, UR, har gjort en film om hur deepfakes fungerar och vilka utmaningar de för med sig.[3] Den vänder sig i första hand till ungdomar och illustrerar hur tekniken kan användas för desinformation som i värsta fall kan undergräva människors tillit till sanningen, när något som man tror är sant och äkta i själva verket är manipulerat.
- Den brittiska regeringen har tagit fram en rapport om deepfakes och vilseledande information.[4] En slutsats är att det inte räcker med lagstiftning för att hantera risker, det krävs även investeringar i ny teknik för att avslöja manipulationer. I likhet med svenska initiativ ser den brittiska regeringen behov av utbildningssatsningar för att höja kunskap om deepfakes bland medborgarna.
- Exempel på svensk forskning inom området är utveckling av en app som kan avslöja deepfakes[5].

Länkar

1. www.belfercenter.org/sites/default/files/2020-10/tappfactsheets/Deepfakes.pdf
2. https://www.riksdagen.se/sv/dokument-lagar/dokument/skriftlig-fraga/deepfakes_H511938
3. https://urplay.se/program/216095-var-digitala-planet-i-fokus-deepfakes
4. www.gov.uk/government/publications/cdei-publishes-its-first-series-of-three-snapshot-papers-ethical-issues-in-ai/snapshot-paper-deepfakes-and-audiovisual-disinformation
5. www.lth.se/article/app-som-avsloejar-deepfake-under-utveckling/

Sammanfattning av originalrapporten

Det engelska ordet deepfake är en kombination av deep learning (djupinlärning) och fake (påhittat). Det som avses är användning av artificiell intelligens för att förändra film eller andra visuella medier utan att förändringen ska märkas, så att filmen ser ut att vara autentisk.

Deepfakes kan vara olika tekniskt sofistikerade och ha varierande tillämpningar.

Författarna beskriver en skala från enkla förändringar (eng. cheap / shallow fakes) till deepfakes. Just deepfakes skiljer ut sig genom att manipulationen görs via djupinlärning. Enkla förändringar görs i stället genom någon form av bild- eller videoredigeringsprogram.

Tabellen visar olika tekniker, från enkla förändringar till deepfakes. Översättning av tabell i den amerikanska rapporten.

Kort om tekniken

Deepfakes bygger på djupinlärning, dvs artificiell intelligens som använder algoritmer inspirerade av hjärnans funktion (så kallade artificiella neurala nätverk).

Den mest använda metoden är en generativ modellering som kallas Generative adversarial networks, GAN. Den använder två neurala nätverk, där det ena skapar en helt ny bild varvid det andra tar över och bedömer sannolikheten att bilden är äkta. Det första nätverket får återkoppling om hur troligt det är att bilden uppfattas som verklig. Detta pågår tills det inte går att skilja den datorgenererade bilden från verkligheten. Nätverken tränas upp var för sig, det första på att skapa bilder och det andra på att bedöma äkthet. Ju större mängd data de tränat på, desto bättre blir förfalskningen.

En annan variant av generativa modeller är variationsautokodare (eng. variational autoencoder, VAE). Här arbetar två nätverk tillsammans. Ett av dem sammanfattar alla data som matas in, sedan tar det andra vid och försöker återskapa de ursprungliga data. Nätverken tränas på gemensamma data (t.ex. hundratals foton av en filmstjärna), till dess att inmatade och återskapade data stämmer överens. Därefter kan det återskapande nätverket justeras så att något läggs till i bilden (t.ex. att filmstjärnan har ett ärr på kinden). Genom att kombinera två uppsättningar VAE kan t.ex. filmstjärnans kropp få en politikers ansikte.

Klassifikation	Teknik för att förändra filmen
Enkla förändringar (eng. cheap fakes, shallow fakes) ↓ Deepfakes	*Förändrad kontext.* Ett befintligt filmklipp omnämns på ett felaktigt sätt, eller visas ur sitt sammanhang.
	Dubbelgångare. En skådespelare som är lik en verklig person hyrs in för att spela in ett filmklipp.
	Snabba upp och sakta ner. Filmhastigheten ändras på ett sätt som förändrar innebörden i det som filmats, eller hur man uppfattar det som skildras.
	Byta ut ansiktet. Animeringstekniken rotoskopi utgår från filmsekvenser, ritar av konturerna i varje filmruta och skapar animerad film. Tekniken kan användas för att få en illusion av att ansiktet är utbytt.
	Läppsynkronisering. Genom att synkronisera läpprörelserna i en filmsekvens med ljud som spelats in i ett helt annat sammanhang framstår det filmade uttalandet som autentiskt.
	Ersätta ansiktet. På digital väg överlagra bilden av en persons ansikte på någon annan.
	Syntetisk talproduktion. En verklig persons röst efterhärmas på konstgjord väg.
	Återskapande av ansikte och röst. Förändring av ansiktsdrag, mimik eller röst via bild- och ljud-behandling.
	Skapa ett helt nytt ansikte. Generera nya bilder som överlagras på en verklig persons ansikte med hjälp av avancerad mjukvara (utgår inte från någon befintlig verklig bild, helt nya bilder skapas digitalt).

Om Komet Kommenterar
Komet kommenterar aktuella internationella rapporter som rör regelverk, teknikutveckling och innovation.
Syftet är att ge ett svenskt perspektiv, sätta information i ett sammanhang och göra underlaget lätt tillgängligt.

Kommittén för teknologisk innovation och etik (Komet) arbetar för att skapa goda förutsättningar för innovation och konkurrenskraft samtidigt som utvecklingen och spridningen av ny teknik sker tryggt, säkert och med ett långsiktigt samhällsperspektiv.

Detta informationsblad är sammanställt av kommitténs sekretariat.
Läs gärna mer på vår webb: kometinfo.se
e-post: n.komet@regeringskansliet.se