SÄRTRYCK UR *STRATEGISK UTBLICK 8*

Fejkade nyhetsbilder och verklig motståndskraft

Niclas Wadströmer, David Gustafsson och Patrik Thunholm

Främmande makt kan idag skapa och sprida virtuella bilder och filmsekvenser som belägg för fejkade nyheter, bilder som är svåra att skilja från foton av verkligheten. Framstegen inom artificiell intelligens har gjort att nära på vem som helst med en persondator kan skapa skenbart verkliga filmsekvenser. Dessutom har den digitala informationsmiljön förändrat medielandskapet och förutsättningarna för att kunna sprida information. Detta är en ny utmaning för totalförsvaret i arbetet med att motarbeta främmande makts möjligheter att bedriva informationspåverkan. Samtidigt ska totalförsvaret värna den fria åsiktsbildningen som är grunden för det demokratiska samhället.

Informationspåverkan

Föreställ dig att det dyker upp ett videoklipp i ditt flöde på sociala medier från en presskonferens med en makthavare som redogör för en allvarlig händelse. Klippet har redan delats tusentals gånger innan det når ett mediehus som börjar undersöka dess sanningshalt. En nyhetsreporter får kontakt med makthavaren som med bestämdhet förnekar att presskonferensen ens har ägt rum. Reportern menar att bild och ljud i videoklippet är bevis för att presskonferensen har ägt rum och undrar varför makthavaren förnekar framträdandet. Makthavaren får också frågor om hur det kommer sig att klippet har delats i hens sociala medier och att budskapet dessutom har spridits med dennes officiella e-postadress. Förvirringen är total. Efter en tid står det klart att det viralt spridda nyhetsklippet är fejkat och att makthavarens digitala kommunikationsvägar har utnyttjats för att förstärka vilseledningen. Ändå fortsätter spekulationerna i sociala medier om videoklippets sanningshalt. Går det grävande mediehuset verkligen att lita på? Långt senare avslöjas att den som producerat den fejkade videon och hackat kontona har velat minska den allmänna tilltron till nyhetsförmedling genom att grumla bilden av vad som är verkligt och vad som är fejkat.

Ovanstående scenario illustrerar hur Sverige skulle kunna påverkas och manipuleras eller utsättas för kognitiva påfrestningar, av en enskild aktör men också av en främmande makt, utan att landet befinner sig i krig.

På Youtube publiceras dagligen runt hundratusen videoklipp. Bland dessa finns många exempel på att man både kan manipulera verkliga foton och filmer, och kan skapa foton och filmer som ser verkliga ut men som är helt datorgenererade. Bland dessa finns ett videoklipp med en datorgenererad nyhetsuppläsare som knappt går att skilja från en verklig person. Ett annat exempel är en video med vad som förefaller vara den förre amerikanske presidenten Barack Obama som gör ett oväntat uttalande. I det fallet har någon läst in en text, gjort om ljudet så att det låter som Barack Obama och ändrat ansiktsrörelserna i en redan existerande film med den före detta presidenten så att de stämmer med det inlästa talet.

Forskare på företaget Nvidia har samlat in många tusentals porträttbilder från internet och använt maskininlärning för att skapa ett program som kan generera högupplösta porträttbilder som ser ut att vara foton av verkliga personer men där ingen av personerna finns i verkligheten. Dessa videoklipp är resultatet av de senaste årens framsteg inom artificiell intelligens (AI), vilka har gjort det möjligt att producera rörlig media där valfritt budskap kan framföras av valfri person på valfri plats.

FOI Memo 6750
Projekt Nr I181103
Godkänd av Mike Winnerstig

Augusti 2019

Manipulerade bilder

Det finns många historiska exempel på manipulerade bilder av verkligheten. Man har länge kunnat retuschera fotografier även om det har krävts skickliga konstnärer för att kunna göra retuscheringen trovärdig.

Filmbranschen har länge kunnat göra tecknade animerade filmer och med datorns intåg har dessa blivit allt mer verklighetstrogna. Nu finns exempel på spelfilmer där man har skapat virtuella bilder av personer som inte vill eller kan vara med vid inspelningen.

Virtuella bilder är bilder som ger sken av att vara fotografiska avbildningar av verkligheten men som helt eller delvis är skapade i en dator. Bilden ser ut att visa något som finns i verkligheten men är en chimär skapad i en dator och det är svårt att skilja dessa skapade virtuella bilder från fotografiska bilder av verkligheten.

Det krävs fortfarande stor kompetens och stora resurser för att göra spelfilmer med datorgenererade personer. Men teknik för att skapa och manipulera bilder med ansikten blir samtidigt allt mer lättillgänglig. Nu kan man enkelt ladda ner en applikation som kan byta ut ett ansikte mot ett annat. Det räcker med en vanlig persondator och kräver inte särskilt djupa kunskaper för att med överraskande god kvalitet lyckas.

“Det krävs fortfarande stor kompetens och stora resurser för att göra spelfilmer med datorgenererade personer. Men teknik för att skapa och manipulera bilder med ansikten blir samtidigt allt mer lättillgänglig.”

Förändrat medielandskap

Den digitala informationsmiljön har blivit en viktig arena för krigföring. Gamla föreställningar om att krig utspelas mellan arméer på ett slagfält har minskat i relevans och en viktig uppgift för det psykologiska försvaret är att kunna identifiera, analysera och möta påverkanskampanjer från främmande makt. För att lyckas med detta uppdrag krävs bland annat kunskap om vad som är möjligt att göra med modern teknik.

En majoritet av svenskarna är dagligen uppkopplade och med stöd av mobilteknik som smarta telefoner har medievanorna genomgått stora förändringar. Vi kan numera röra oss mellan en traditionell roll som tittare, lyssnare och läsare, till en mer aktiv roll där vi själva producerar och distribuerar innehåll. Denna utveckling har samtidigt gjort det enklare för främmande makt att använda nya psykologiska tekniker för att påverka våra uppfattningar, inställningar och uppträdande i syfte att nå specifika målsättningar. Målsättningarna skulle kunna handla om att påverka den allmänna opinionen och det demokratiska beslutsfattandet genom att underminera beslutsförmåga eller manipulera uppfattningar.

I de fall en främmande makt vill förstärka sin egen position och samtidigt försvaga Sverige och svenska intressen kan denna tänkas sprida information av varierande sanningshalt i informationsmiljön. Det kan handla om händelser som aldrig har inträffat, till exempel övergrepp och misshandel i kända miljöer eller polisbrutalitet mot minoriteter – allt i syfte att göda polarisering och splittra ett land inifrån. Gemensamt för dessa hot är dels att de riktas mot eller utnyttjar värden som är sårbara till sin natur, såsom demokrati och yttrandefrihet, dels att antagonisterna ofta använder ny teknik.

Den nya tekniken gör det möjligt att enkelt producera fejkade nyhetsbilder som sedan kan vara svåra eller omöjliga att spåra tillbaka till den som har skapat dem. Det kan till exempel handla om anonym spridning över internet eller att spridningen sker med hjälp av en kapad digital identitet. Sammantaget är möjligheten till förnekbarhet hög. Den senaste tidens utveckling av maskininlärning har bidragit till att rörliga bilder nu måste betraktas med skepsis. I framtiden kommer uttrycket ”att lägga orden i någons mun” att få en mer bokstavlig mening.

Maskininlärning

Maskininlärning, ett delområde inom artificiell intelligens, har tagit rejäla kliv framåt med artificiella neurala nätverk (ANN). ANN är en struktur för datorprogram som har inspirerats av biologiska hjärnor. ANN lär sig från många exempel av in- och utdata istället för att programmeras med explicita

regler för hur utdata fås från indata. *Deep learning* är ANN som har förmåga att representera information i hierarkiska nivåer. Google, Amazon, Facebook och andra aktörer med tillgång till mycket stora mängder bilder med vidhängande bildtexter kan lära sina datorer att inte bara förstå bilders innehåll utan också att redigera och skapa virtuella bilder. Det går att automatiskt ändra en landskapsbild från vinter till sommar. Man kan göra om en ritad skiss till en bild som ser ut att vara en fotografisk bild av ett verkligt landskap. Det går att byta ut ett ansikte, lägga till och ta bort en person. Man kan skapa syntetisk film där en person framträder som är mycket svår att skilja från en autentisk film.

Maskininlärning innebär att man låter en maskin lära sig från exempel. Säg att man vill skapa ett program som skapar porträttbilder. Man ger då datorn ett stort antal exempel på porträttbilder och låter maskinen hitta särdrag som är typiska för porträttbilder och sedan kan maskinen skapa slumpmässiga bilder med dessa särdrag. Innan det fanns maskininlärning hade en ingenjör fått ta reda på särdrag som är typiska för porträttbilder och skriva ett program som skapar slumpmässiga bilder med de givna särdragen. Med moderna maskininlärningsalgoritmer är datorn i många fall bättre än ingenjören på att hitta relevanta särdrag.

Generativa metoder ger verkligare bilder

Generativa metoder kan skapa syntetiska bilder, text och ljud, det vill säga saker som är helt och hållet skapade i datorn, utan en verklig förlaga. År 2014 introducerades *Generative adverserial networks* (GAN), en ny generativ metod, som medförde ett genombrott inom generering av verklighetstrogna bilder. GAN är en vidareutveckling av *Deep learning* som har fått en mycket stor spridning inom AI. GAN tränas genom motvalls träning. Två olika nätverk tävlar mot varandra – ett generativt nätverk genererar bilder av en bestämd sort och ett klassificeringsnätverk lär sig att särskilja riktiga bilder från genererade bilder. Det generativa nätverket förbättras genom att beakta egenskaper i de genererade bilderna som klassificeringsnätverket använder för att särskilja de genererade bilderna från verkliga bilder. Klassificeringsnätverket fungerar som sparringpartner till det generativa nätverket och genom en iterativ process blir de båda allt bättre. Resultatet är att det genereras bilder som blir allt svårare att separera från de verkliga bilderna. I många fall lyckas GAN-nätverket skapa bilder som en lekman inte kan särskilja från riktiga fotografiska bilder och som kan vara svåra för expertis att särskilja.

Den här förmågan blir allt mer tillgänglig och lättanvänd. Man kan hämta programkod på internet och det behövs inte någon avancerad kunskap för att använda den. I många fall finns det program och färdigtränade modeller för generering av olika typer av bilder eller filmer fritt tillgängliga för nerladdning från internet. Där det tidigare krävdes mycket stora resurser, som bara stater har, för att skapa material av det här slaget till påverkanskampanjer räcker det idag med en liten grupp personer. Det är också värt att notera att det är privata företag som har publicerat de bästa forskningsresultaten. Här saknas det insyn i vilken förmåga företagen har och vad de väljer att hålla hemligt och för vilka syften de använder tekniken.

Totalförsvaret

I denna nya tekniska och digitala miljö återstartar Sverige totalförsvaret. Liksom under det kalla kriget är ett psykologiskt försvar och motståndskraft mot samhällshotande informationspåverkan en förutsättning för att totalförsvaret ska fungera. Utan samhällelig motivation fungerar ingen del av totalförsvaret – inte heller Försvarsmakten. Betydelsen av psykologiska försvarsåtgärder har ökat i jämförelse med det kalla kriget. Detta eftersom en fientlig statlig aktör nu kan uppnå mål som förr krävde en militär operation, genom att bland mycket annat använda sig av påverkan inom den digitala informationsmiljön. Mot denna bakgrund är det viktigt att totalförsvarets psykologiska delar följer teknikutvecklingen och tar fram metoder som ger verklig motståndskraft mot den som vill skada oss med falska bilder.

En motståndare som ägnar sig åt informationspåverkan ser till att identifiera och utnyttja sårbarheter systematiskt. Sårbarheter finns inom flera olika områden. Det moderna mediesystemet har flera sårbarheter i förhållande till bland annat ny teknik, nya journalistiska affärsmodeller och en tilltagande mängd nyhetskällor på nätet. Även opinionsbildningen har blivit mer sårbar i takt med den digitala informationsmiljöns framväxt. Med internets tillkomst är det lättare än någonsin att fabricera sociala bevis och att väcka

ilska, provocera eller uppröra. Kognitiva sårbarheter kan uppstå till följd av att den mänskliga hjärnan tenderar att ta genvägar och inte är konstruerad för att hantera all den mängd information den ibland utsätts för. Här kan informationspåverkan utnyttja tankemönster och information om oss för att påverka uppfattningar, beteende och beslutsfattande.

Verklig motståndskraft

Totalförsvaret ska skydda grundläggande värden som vår demokrati och vårt självbestämmande mot angrepp från främmande makt som vill skada oss. Sveriges motståndskraft mot dessa hot är avhängig av hur digitalt kompetent, tekniskt kunnigt och snabbfotat det psykologiska försvaret är. Detta kommer att ställa krav inte bara på myndigheter, utan också på medborgare, politiker och teknikleverantörer.

Verklig motståndskraft får vi ytterst genom att var och en har ett kritiskt förhållningssätt till bilder och det som bilden säger. Det behövs dessutom en allmänt ökad kunskap om vilka möjligheter som finns att manipulera och förfalska bilder. Utöver detta krävs forskning om forensiska metoder för att avslöja fejkade bilder och system för att ursprungsmärka bilder så att den som tar del av en bild kan veta vem som är avsändare och upphovsman.

För vidare läsning

Niklas H. Rossbach, Psykologiskt försvar - avgörande för svensk försvarsförmåga, Strategisk utblick 7, 2017, FOI-R--4454--SE.

Tove Gustavi, Jörgen Karlholm, Daniel Oskarsson, Tam Beran, Oscar Björnham, Erik Gudmundson, Hugo Heden, Henrik Karlzén, Johannes Lindblom, Jonas Nordlöf, Ioana Rodhe, Teodor Sommestad, Peter Svenmarck och Markus Svensson, Försvarsnära tillämpningar av Artificiell Intelligens, 2019, FOI-R--4707--SE.

FOI
Totalförsvarets forskningsinstitut
SE-164 90 Stockholm

Telefon: +46 8 5550 3000
www.foi.se