OFFPRINT FROM *STRATEGIC OUTLOOK 8*

Fake news images and genuine resilience

Niclas Wadströmer, David Gustafsson and Patrik Thunholm

Nowadays, foreign powers can create and distribute virtual images and videos as 'evidence' of fake news. Such images are difficult to distinguish from genuine photos. Advances in artificial intelligence have made it possible for anyone with a computer to create what look like real videos. Moreover, the digital information environment has altered the media landscape and criteria for distribution of information. This presents the Swedish total defence with a new challenge in its efforts to counter opportunities for foreign powers to engage in information influence operations. The total defence also has to safeguard the freedom to form opinions, which is at the very heart of democratic society.

Information influence operations

Imagine a video clip appearing in your social media feed that shows a press conference where a person in authority is describing a serious event. This clip has already been shared thousands of times before it reaches a media house that starts to investigate its authenticity. A reporter gets in touch with that person in authority, who firmly denies that the press conference has even taken place. The reporter believes that what he sees and hears in the clip are proof that the press conference has taken place, and wonders why the person is denying everything. He also asks the person how the clip came to be shared on his social media channel, and how the message was distributed using his official email address. There is complete and utter confusion. After a while, it becomes clear that this viral news clip is fake and that the digital communication pathways of the person in authority have been exploited to underpin the deception. Nevertheless, social media continues to speculate about the authenticity of the video. Can the media house investigating the clip really be trusted? Much later, it is revealed that the person who produced the fake video and hacked the accounts wanted to reduce public confidence in the reporting of news by blurring the lines between what is genuine and what is fake.

The scenario above illustrates how Sweden could be influenced and manipulated or exposed to cognitive stress by a single individual; but also by a foreign power, without the country being at war.

Hundreds of thousands of videos are published on YouTube every day. These include many examples of how actual photos and videos can be manipulated, and how photos and videos can be completely generated by a computer but still look real. These examples include a video clip showing a computer-generated newsreader who is barely distinguishable from a real person. Another example is a video showing what appears to be former US President Barack Obama making an unexpected statement. In this case someone has recorded a text, transformed the audio so that it sounds like Barack Obama and altered the facial movements in an existing video of the former President so that they match the words recorded.

Researchers at Nvidia have collected many thousands of profile pictures from the Internet and used machine learning to create a program that can generate high-resolution profile pictures that appear to be photos of real people – but none of these people actually exist. These video clips are the result of recent advances in artificial intelligence (AI) which have made it possible to produce moving media, allowing any message to be presented by any person in any location.

FOI Memo 6870
Project Nr I181103
Approved by Mike Winnerstig

October 2019

FOI

Manipulated images

There are many historic examples of manipulated images of reality. Retouching images has been possible for years, although skilled artists were required to make the photos credible.

The film industry has been making animated films for years, and these have become increasingly lifelike as computers have become more widespread. Now we have feature films involving the creation of virtual images of people who are unwilling or unable to participate in the actual filming.

Virtual images are images that look like photographic depictions of reality, but which were actually created entirely or at least partly by a computer. The image appears to show something that genuinely exists, but it is an illusion created by a computer. And it is difficult to distinguish these virtual images from actual photos of real things.

A great deal of expertise and extensive resources are still needed to make feature films using computer-generated characters. That said, technology for creating and manipulating images containing faces is becoming increasingly accessible. Apps that can replace one face with another can now be downloaded with ease. All that is needed is an ordinary computer. Not even particularly in-depth knowledge is required to produce images of surprisingly good quality.

> “A great deal of expertise and extensive resources are still needed to make feature films using computer-generated characters. That said, technology for creating and manipulating images containing faces is becoming increasingly accessible.”

Altered media landscape

The digital information environment has become an important arena for warfare. The relevance of the old notions, that war is played out between armies on battlefields, has diminished and it is important that the psychological defence tis able to identify, analyse and address influence operations from foreign powers. To succeed in this, information on what can be achieved with modern technology is required.

Most Swedish people are online every day, and their media habits have undergone major changes with the help of mobile technology such as smartphones. Nowadays we can take on the traditional role of viewer, listener and reader, or we can switch to a more active role where we produce and distribute our own content. At the same time, this development has made it easier for foreign powers to use new psychological techniques to influence our perceptions, attitudes and behaviours in order to achieve specific objectives. These objectives could include influencing public opinion and democratic decision-making by undermining decision-making capacity or manipulating opinions.

If a foreign power wanted to reinforce its own position while also weakening Sweden and Swedish interests, it could potentially distribute information of varying authenticity in the information environment. They may report incidents that have never occurred – such as abuse and assaults in known surroundings, or police brutality against minorities – with a view to encouraging polarisation and dividing a country from within. One thing these threats all have in common is that they are aimed at or exploit values vulnerable by their very nature, such as democracy and freedom of expression; along with the fact that antagonists often use new technology.

New technology makes it possible to produce fake news images with ease, rendering it difficult or impossible to trace them back to whoever created them. For instance, images may be distributed anonymously over the Internet or using a stolen digital identity. All in all, the potential for deniability is high. Moving images now have to be regarded with scepticism as a result of the recent development of machine learning. In future, ‘putting words into somebody’s mouth’ will take on an even more literal meaning.

Machine learning

Machine learning, a subfield of artificial intelligence, has made great advances with artificial neural networks (ANNs). ANNs are a software structure inspired by biological brains. ANNs learn from many examples of input and output data, instead of being programmed with explicit rules on how to obtain output data from input data. Deep learning are ANNs with the ability to represent information in hierarchical layers. Google, Amazon, Facebook and other stakeholders with access to enormous numbers of images with associated captions can teach computers not only to understand the image content, but also to edit and create virtual images. It is possible to change a landscape photo from winter to summer automatically. A sketch drawn by a person can be converted into an image that looks like a photo of a real landscape. It is possible to replace a face, add or remove a person. It is possible to create synthetic video featuring a person that is very difficult to tell apart from an authentic video.

Machine learning involves allowing a machine to learn from examples. You may want to create a program to make portrait pictures, for instance. To do this, you give the computer a large number of examples of portrait pictures and allow the machine to identify special features that are typical for portrait pictures. The machine can then create random images containing these features. Before machine learning came into being, an engineer would have had to identify special features typical for profile pictures and then write a program that created random images with the specified features. Modern machine learning algorithms mean that in many cases, computers are better than engineers when it comes to identifying relevant features.

Generative methods produce more lifelike images

Generative methods can create synthetic images, text and audio – media content created entirely in the computer, that is – without an actual original to work from. Generative Adversarial Networks (GAN), introduced in 2014, is a new generative method which resulted in a breakthrough in the generation of lifelike images. GAN are a further development of Deep learning, which is widely used in AI applications. GAN is trained by means of adversarial training. Two different networks compete against one another: a generative network generates images of a particular type, and a classification network learns to tell the difference between genuine images and generated images. The generative network is enhanced by taking into account features in the generated images that the classification network then uses to tell the generated images apart from genuine images. The classification network acts as a sparring partner for the generative network, and they both go on improving by means of an iterative learning process. As a result, it is becoming increasingly difficult to tell generated images apart from real images. In many cases, the GAN succeeds in creating images that the average person would be unable to tell apart from genuine photos, and even experts may find it difficult to tell the difference.

This capability is becoming increasingly accessible and easy to use. Program code can be downloaded from the Internet, and advanced knowledge is not needed in order to use it. Programs and pre-trained models for generating images or videos of different types are often freely available to download from the Internet. Nowadays, a small group of people will suffice to create materials of this kind for influence operations, and enormous resources – of the kind that are the preserve of states – are no longer needed. It is also worth noting that private companies have actually published the best research outcomes. There is a lack of transparency here in respect of corporate capabilities and what they choose to keep secret, along with the purposes for which they use this technology.

The total defence

Sweden is restarting its total defence in view of this new technological and digital environment. The Swedish total defence concept encompasses both civil and military defence in a whole-of-society approach to security. As during the Cold War, psychological defence and resilience to information influence operations that threaten society are a prerequisite if total defence is to work. Without social motivation, no part of total defence will function – and nor will the Swedish Armed Forces. Compared to the Cold War, psychological defence measures have taken on greater importance as hostile state actors can now achieve objectives that previously required military operations, simply by using the influence within the digital information environment. Given this fact, it is important for the psychological aspects of the total defence to follow technological development and devise methods that provide genuine resistance to any party wishing to harm us with false images.

Any opponent engaging in information influence operations systematically ensures that vulnerabilities are identified and exploited. There are vulnerabilities in a variety of areas. The modern media system has a number of vulnerabilities in relation to factors such as new technology, new journalistic business models and the increasing number of online news sources. Advocacy efforts have also become more vulnerable as the digital information environment has emerged. With the advent of the Internet, it is easier than ever to fabricate social evidence and incite anger, provocation or upset. Cognitive vulnerabilities may occur due to the way in which the human brain tends to take shortcuts: it is not designed to handle the vast amount of information that it sometimes receives. In this regard, information influence operations can utilise thought patterns and information about us to influence perceptions, behaviours and decision-making processes.

Genuine resilience

The total defence aims to protect fundamental values such as democracy and our self-determination from attacks by foreign powers intent on harming us. Sweden's resilience to these threats is dependent on the agility, digital competence and technical capability of our psychological defence system. This will make demands not only of authorities, but also of citizens, politicians and technology providers.

Ultimately, genuine resilience will be achieved when each and every person adopts a critical approach to images and what they tell us. There is also a need for a general increase in awareness of the opportunities available for manipulating and forging images. Furthermore, research into forensic methods is required so that fake images and systems can be revealed, allowing labels to be attached indicating the origins of images so that anyone viewing such images knows where they originated.