DEEPFAKES: A TECHNOLOGY OF SYMBIOTIC PARADOX

a manifesto by haneu(l) bak

Deepfakes is an artificial intelligence technique for human image synthesis. The jargon, which almost sounds like a nickname or a slang, is a combination of the phrase “deep learning” and the word “fake.” Deepfakes uses a type of machine learning called Generative Adversarial Network, or GAN, in order to study and replicate human image in video formats. This technical replication includes not only the general facial features such as contour and shapes, but also the subjects’ gestures or movements on their faces.

Deepfakes technology has been in use for various media contents for decades. For instance, films such as <Fight Club> 1999 has a brief but crucial scene in which the main character faces his inner self who looks exactly like him, thanks to the deepfakes technology. The use of deepfakes in mainstream media continues to increase and evolve to this day. On the other hand, the development of deepfakes technology, naturally, resulted in its greater accessibility among the general public. Various corporations, individuals, or certain platform users have started developing and sharing deepfakes interfaces that are easy to use. Such accessibility brought in a phenomenon in which unidentified/unidentifiable online users create countless number of deepfakes videos, including fake news and fake pornographies with detrimental impact on the replicated individuals in the videos as well as the misled, misinformed viewers. Deepfakes pornography cause tremendous trauma on the victims, and deepfakes news create significant impacts on the possible outcomes of political events or elections. But more than 90% of deepfakes contents are pornographic, and about 99% of them feature facial features of women in the entertainment industry. It is clear that deepfakes technology, since its accessibility, has been used primarily for malicious and petty pornographic contents.

Despite the prevalence of deepfakes contents with powerful, negative impacts, most countries and international governments do not have sufficient legal solutions. Because deepfakes is not only a new technology but ever-evolving AI technology, the stringent set of legislation does not follow up quickly. An exceptional, pioneering example is the state of California in the Unites States of America. California’s deepfake laws ban production of any pornographic or political deepfakes contents. However, such a law is highly abstract, lacks any tangible methods to prevent the production of such malicious contents. This legal limitation shows the inevitable gap between passing a bill and enforcing the law, leaving both the court and the public still unclear on how to properly react to the harmful deepfakes contents. The Californian legislature also allows residents to sue anyone who uses deepfake technology to place them in pornographic material without consent. This law does not promise any feasible methods of prevention either, and the bill seems to fall under the already-existing legal item of defamation. This dilemma of legal necessity and the apparent uselessness of it leaves us confused and frustrated, invoking the need to react collectively to the technological evolution.

Thankfully, there is deepfake detection technology capable of identifying deepfakes contents, although its success might be fleeting. Using the same machine learning technology used by deepfakes softwares, GAN, deepfakes detection technology analyses people’s slightest facial expressions and demeanors, and uses the data to catch deepfake videos based on their lack of personal features. It detects the lack of subtle habitual expressions of people that the current deepfakes AI technology cannot completely replicate. Some deepfake detection softwares are more than 90% accurate, and have been successful at relieving us with the promise of being able to tell what is real and what is not.

But deepfakes technology is constantly evolving and its ability to replicate improving. And it is inevitable that detection technology rather helps providing information about what deepfakes technology is missing and how it can be improved, especially because they both use the same source of data, GAN. Deepfakes and its detection technologies will continue to evolve, teaching and outrunning one another in turn, almost in a symbiotic way. The public will have to fear for the complete inability to tell what video is real and what not, and depend on the ever-changing technology of deepfake detection. In the near future, deepfakes contents can be even more refined and prevalent while the detection technology stays behind. If tech companies monopolize the detection technology or governments put insufficient effort in accessorizing it, the public will be deprived of the right to know what contents are trustworthy.

There is a metaphysical irony in the inevitably symbiotic relationship between deepfakes and its detection technology. Deepfakes detection technology is both the only reliable solution to and an important resource for further developing deepfakes softwares. Such an irony emphasizes that in the emerging future of deepfakes, what requires our attention is the close relationship between the counteracting technologies in order to divest meaningless competition and destruction. It is pointless and even dangerous to let deepfakes and its detection technologies separate and competing, because these technologies will be ever-evolving as long as the AI technology keeps advancing. The detection technology can be incredibly refined and advanced enough to catch any deepfakes contents, but it will soon be the source for deepfakes softwares further advancements, and more importantly, would not prevent malicious contents from being produced in the first place. In the endless technological competition, there will continuously be distortions, victims, new laws, and new controversies. There may be incredible developments in the human image synthesis technologies, but the assurance of truth will not be offered.

The inherently ironic existence, growth, and evolution of deepfakes technology brings up an interesting paradox that penetrates many other high technologies. How do we accept the inevitable oxymoron? How should we think about such a technology so complex?