한국컴퓨터정보학회 동계학술대회 논문집 제32권 제1호 (2024. 1)

생성형 인공지능에 대한 보안 이슈와 대응 방안

육세영[O], 강아름[*]
[O]배재대학교 정보보안학과,
[*]배재대학교 정보보안학과
e-mail: 2184023@pcu.ac.kr[O], armk@pcu.ac.kr[*]

Security Issues and Countermeasures for Generative Artificial Intelligence

Se Young Yuk[O], Ah Reum Kang[*]
[O]Dept. of Information Security, Pai Chai University,
[*]Dept. of Information Security, Pai Chai University

● 요 약 ●

4차 산업 혁명의 시작으로 인공지능이 빠르게 발달함에 따라 현재 생성형 인공지능이 주목받고 있다. 이에 따라 딥보이스 기술과 딥페이크 기술을 활용하여 다양한 범죄가 발생하고 있어 관련 사례와 이를 해결하기 위해 진행 중인 연구에 대해서 조사하였다. 딥보이스와 딥페이크를 탐지하는 연구는 지속되고 있지만 관련 기술이 상용화되어 있지 않아 범죄를 예방하기에는 부족한 실정이다. 범죄에 악용되는 속도가 빨라지고 있는 만큼 더 많은 연구가 신속하게 이루어져야 한다.

키워드: 생성형 인공지능, 딥보이스, 딥페이크

I. Introduction

4차 산업혁명의 시작으로 인공지능은 빠르게 발전하고 있다. 규칙 기반 시스템과 논리에 기반을 둔 기호 인공지능(Symbol Artificial Intelligence)을 시작으로 현재는 대규모 언어 모델과 같은 생성형 인공지능(Generative Artificial Intelligence)이 주목받고 있다. 인공지능의 발전으로 편의성 증대, 생산성 향상, 새로운 기술 및 서비스 개발 같은 긍정적인 영향을 가져왔다. 하지만 개인정보 노출, 일자리 변화 및 보안 위협 등의 문제점도 가져왔다. 본 논문에서는 2장에서 생성형 인공지능의 문제점에 대해서 기술하고 3장에서는 현재 문제를 해결하기 위해 이루어지고 있는 연구에 대해서 다룰 예정이다.

II. Preliminaries

1. 딥보이스(Deep Voice) 기술

인공지능의 발달로 딥보이스 기술이 등장하면서 전문적인 지식 없이도 누구나 간편하게 음성을 조작할 수 있게 되었다[1]. 특정 사람의 목소리를 모방하여 가짜 목소리를 생성하는 것을 악용으로 다양한 범죄가 발생하고 있다. 대표적인 사례 중 하나가 보이스피싱으로 인한 금융 사기이다.

실제 사례로는 대표적으로 두 가지가 있다. 캐나다에서 아들의 목소리를 모방하여 가짜 아들 목소리 만들었다. 이를 이용하여 아들이 교통사고를 내서 수감되었다며 아들의 목소리를 들려주었고 이에 부모는 속아 약 2021만원을 보이스피싱범에게 입금하였다. 중국에서는 지인을 사칭에 급하게 자금이 필요하다고 전화를 하여 이에 보이스피싱범에게 약 1억 7500만원을 송금한 사례가 발생하였다.

2. 딥페이크(Deepfake) 기술

인공지능 기술의 발전으로 딥러닝을 활용하여 이미지나 영상을 쉽고 정교하게 조작하는 기술이 등장하였다. 이러한 기술을 악의적으로 활용하는 경우를 딥페이크 기술이라고 부른다[2]. 딥페이크 기술은 영상 분야에서 특수효과나 과거 재현 등에서 유용하게 사용된다. 하지만 이를 악용한 부분도 있다. 예를 들면 가짜 뉴스 생성 및 음란물 제작이다.

실제 사례로는 대표적으로 세 가지로 옷 벗기기 사이트 제작, 아동 성 착취물 생성 및 나체 사진을 제작해 협박하는 사례가 있다. 옷 벗기기 사이트는 딥페이크 기술을 사용하여 특정 개인의 얼굴과 몸을 합성하여 옷을 벗은 것처럼 만들어 처음에는 옷을 입고 있다가 점점 벗기는 사이트이다. 아동 성 착취물 생성은 어린이의 몸을 노출시

키거나 성적 행위가 포함된 것을 딥페이크 기술을 통해 만드는 것이다. 나체 사진을 제작해 협박하는 사례로는 개인의 얼굴을 가지고 가짜 나체 사진을 제작하여 이를 소셜 미디어에 유포한다고 협박하며 돈을 갈취하려는 사례이다.

III. Deepfake Detection

인공지능의 발전으로 딥보이스 기술과 딥페이크 기술이 발전함에 따라 보이스피싱과 성범죄 등의 문제에 직면해 있다. 이를 해결하기 위해서 딥보이스 기술을 이용한 음성 변조를 구분하는 기술과 딥페이크 기술을 활용한 이미지나 영상을 식별할 수 있는 기술을 개발하고 있다.

한승우 외는 일반 음성과 딥보이스를 구분하기 위해 Mel-Spectrogram과 MFCC을 활용한 시스템을 제안하였다[3].

이대현과 문종섭은 Bidirectional Convolutional LSTM과 어텐션 모듈을 결합해 딥페이크를 탐지하는 모델을 개발하였다. 이는 기존에 제안된 모델보다는 정확도가 향상되었지만 적절한 하이퍼파라미터를 설정하지 못한다면 기존 모델보다 성능이 떨어진다는 한계점이 있었다[4].

손석빈 외는 RGB 채널 기반 분석과 Gray 채널 기반 분석을 통해 딥페이크 탐지의 효율성을 실험하였다. 이에 Gray 채널 기반 분석이 더 효과적인 방법이라는 결과를 얻었다[5].

이와 같이 최근에 딥보이스와 딥페이크를 탐지하는 연구가 이루어지고 있다. 그러나 연구에서 사용된 실험 환경을 그대로 현실 환경에 적용하는 것에 어려움을 겪고 있어 아직 상용화되지 못한 실정이다.

IV. Conclusions

본 논문에서는 생성형 인공지능이 발전함에 따라서 나타나는 문제점에 대해서 조사하였다. 조사 결과 딥보이스와 딥페이크 기술이 범죄에 이용하는 사례가 증가하고 있음을 확인하였다. 이를 해결하기 위해서 일반 음성과 딥보이스를 구분하는 기술 및 딥페이크를 탐지하는 기술이 연구되고 있다. 그러나 현재까지 이러한 기술이 상용화되지 못한 실정이며 악용 사례는 계속 증가하고 있어 신속한 연구가 필요하다.

ACKNOWLEDGEMENT

Following are results of a study on the "Convergence and Open Sharing System" Project, supported by the Ministry of Education and National Research Foundation of Korea.

REFERENCES

[1] Sowoon Kim, Sungtaek Lee, "Development of Voice Phishing Damage Prevention Service Misusing Deep," The Journal of Korean Institute of Communications and Information Sciences, Vol. 47, No. 10, pp. 1677-1685, Oct 2022.

[2] Anupama Chadha, Vaibhav Kumar, Sonu Kashyap, Mayank Gupta, "Deepfake: An Overview," Proceedings of Second International Conference on Computing, Communications, and Cyber-Security, Vol. 203, pp. 557-566, May 2021.

[3] Seung-Woo Han, Seong-Hun Han, Seong-Min You, Dong-Ho Song, Chang-Jin Seo, "A Study on the Development of Deep Learning-Based Deep Voice Detection System Using Mel-Spectrogram and MFCC," The Transaction of The Korean Institute of Electrical Engineers, Vol. 72P, No. 3, pp. 186-192, Sept 2023.

[4] Dae-hyeon Lee, Jong-sub Moon, "A Method of Detection of Deepfake Using Bidirectional Convolutional LSTM," Journal of The Korea Institute of Information Security and Cryptology, Vol. 30, No. 6, pp. 1053-1065, Dec 2020.

[5] Seok Bin Son, Hee Hyeon Jo, Hee Yoon Kang, Byung Gul Lee, Youn Kyu Lee, "A Comparative Study on Deepfake Detection using Gray Channel Analysis," Journal of Korea Multimedia Society, Vol. 24, No. 9, pp. 1224-1241, Sept 2021.