

# 인공지능 기반 화자 식별 기술의 불공정성 분석\*

신 나 연\*, 이 진 민\*\*, 노 현\*\*, 이 일 구\*\*\*

## 요 약

Covid-19으로 인한 디지털화는 인공지능 기반의 음성인식 기술을 급속하게 발전시켰다. 그러나 이 기술은 데이터셋이 일부 집단에 편향될 경우 인종 및 성차별과 같은 불공정한 사회적 문제를 초래하고 인공지능 서비스의 신뢰성과 보안성을 열화시키는 요인이 된다. 본 연구에서는 대표적인 인공지능의 CNN(Convolutional Neural Network) 모델인 VGGNet (Visual Geometry Group Network), ResNet(Residual neural Network), MobileNet을 활용한 편향된 데이터 환경에서 정확도에 기반한 불공정성을 비교 및 분석한다. 실험 결과에 따르면 Top1-accuracy에서 ResNet34가 여성과 남성이 91%, 89.9%로 가장 높은 정확도를 보였고, 성별 간 정확도 차는 ResNet18이 1.8%로 가장 작았다. 모델별 성별 간의 정확도 차이는 서비스 이용 시 남녀 간의 서비스 품질에 대한 차이와 불공정한 결과를 야기한다.

## Analysis of unfairness of artificial intelligence-based speaker identification technology

Shin Na Yeon<sup>\*</sup>, Lee Jin Min<sup>\*\*</sup>, No Hyeon<sup>\*\*</sup>, Lee Il Gu<sup>\*\*\*</sup>

## ABSTRACT

Digitalization due to COVID-19 has rapidly developed artificial intelligence-based voice recognition technology. However, this technology causes unfair social problems, such as race and gender discrimination if datasets are biased against some groups, and degrades the reliability and security of artificial intelligence services. In this work, we compare and analyze accuracy-based unfairness in biased data environments using VGGNet (Visual Geometry Group Network), ResNet (Residual Neural Network), and MobileNet, which are representative CNN (Convolutional Neural Network) models of artificial intelligence. Experimental results show that ResNet34 showed the highest accuracy for women and men at 91% and 89.9% in Top1-accuracy, while ResNet18 showed the slightest accuracy difference between genders at 1.8%. The difference in accuracy between genders by model causes differences in service quality and unfair results between men and women when using the service.

**Key words : Speaker Identification, Biased dataset, AI Fairness, CNN, VoxCeleb1**

접수일(2022년 10월 31일), 게재확정일(2023년 01월 02일)

\* 성신여자대학교 미래융합기술공학과(주저자)

\*\* 성신여자대학교 미래융합기술공학과(공동저자)

\*\*\* 성신여자대학교 미래융합기술공학과(교신저자)

★ 이 논문은 2022년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원(No. 2020R1F1A1061107)과 2022년도 정부(산업통상자원부)의 재원으로 한국산업기술진흥원의 지원(P0008703, 2022년 산업혁신인재성장지원사업), 과학기술정보통신부 및 정보통신기획평가원의 ICT혁신인재 4.0 사업의 연구결과로 수행되었음 (IITP-2022-RS-2022-00156310).

## 1. 서 론

스마트 기기가 빠르게 발달하고 보편화되면서 다양한 형태의 빅데이터가 생산되었다. 데이터가 많아지면서 중요한 데이터를 선별하고 활용하는 기술도 함께 발전되었다. 특히 인공지능 기술은 빅데이터 기술과 함께 텍스트와 이미지, 영상, 음성 등 다양한 형식의 데이터를 학습하여 이미지 검색, 취향에 맞는 영화 추천, 손글씨 인식, 음성 검색 등 다양한 분야에 활용되고 있다[1][2].

코로나19 팬데믹 상황으로 인해 언택트 애플리케이션의 수요가 증가하면서 음성 데이터를 활용하는 기술도 활성화되고 있다. 음성을 활용한 기술은 가정뿐만 아니라 의료, 언론, 교통 등 각종 전문적인 분야에서 접목되고 있다. 특히 세계 음성 인식 시장 규모가 꾸준히 성장하여 2020년부터 2027년까지의 성장률이 연평균 10.8%에 이를 것이라고 한다[3]. 구글과 애플 등 주요 글로벌 IT 기업에서도 인공지능 기반 음성 인식 서비스를 제공하고 있으며, 향후 음성 인식 기술은 인공지능 기술과 함께 고부가가치 산업에 접목되어 보편화될 것으로 전망된다[4].

음성신호의 인식 기술은 음성 인식과 화자 인식 기술로 분류할 수 있다. 음성 인식(Speech Recognition)은 화자의 음성을 인식하여 문자 데이터로 변환하는 기술이다. 이는 AI(Artificial Intelligence) 스피커의 주된 기능 중 하나이며, 청소기, 냉장고 등 가전제품을 제어하는 기술로써 활용된다. 그러나, AI 스피커의 음성인식 기술이 프로그래머블한 지능을 가지면서 프라이버시 침해 위협의 가능성이 문제되고 있다. 2017년 1월 텍사스주에서 발생한 사례로 아마존 닷컴에서 개발한 스마트 스피커인 아마존 에코에게 장난감과 쿠키 주문을 요구한 결과, 알림 없이 18만원 상당의 해당 상품이 배송되었다[5]. 이는 불특정된 대상의 발화를 모두 명령으로 인식하여 발생한 문제이며, 사용자 서비스에 대한 권한의 범위가 제한적으로 규정되지 않은 측면에서 보안 문제가 이어질 수 있다. 이 사건을 계기로 아마존 에코에 화자를 인식할 수 있는 화자 인식(Speaker Identification) 기술이 추가되었다.

화자 인식에는 임의의 화자가 입력한 음성을 사용하여 등록된 화자 중에 가장 유사한 발성 화자를 찾

아내는 기술인 화자 식별(Speaker Identification)과 사전에 화자가 특정될 때 발성된 음성이 특정 화자의 목소리인지를 판단하여 발성 화자가 제시 화자인지 검증하는 화자 검증(Speaker Verification)이 있다. 화자 식별 및 화자 검증을 하기 위해서는 발성된 음성의 문장 형식, 종류에 제한이 없는 문장 독립과 미리 정해진 문장으로 발성한 화자의 음성을 대상으로 수행하는 문장 종속이 있다. 이러한 음성 인식 기술은 인공지능과 결합하여 화자 인식 시스템에 적용될 수 있다.

한편 인공지능은 데이터와 알고리즘에 편향이 존재할 수 있다. 이러한 편향은 인공지능 공정성에 영향을 미친다. 종래[6]의 연구 문헌에서는 인공지능 공정성에 대해 다양하게 정의를 내리고 있지만, 본 연구에서는 인공지능이 데이터 특유의 부분집합이나 알고리즘에 의존하지 않고 차별 없이 결과를 도출하는 것으로 정의한다. 인공지능 기술은 데이터셋에 대한 의존도가 높으므로 학습된 데이터셋이 성별, 국적 혹은 인종 등과 같은 민감한 특정 정보를 가지는 데이터의 비율이 다르면, 편향(bias)되고 부정확한 결과를 도출할 수 있으며 인종 또는 성차별 등의 사회 문제가 야기될 수 있다[7]. 인공지능이 차별적인 판단을 하는 것을 방지하려면 학습에 사용되는 입력 데이터가 공정하게 선별되어야 한다. 그러나, 데이터를 정제하는 과정에서 자의성(arbitrariness)이 개입될 수 있으며, 이는 인공지능이 편향성을 갖는 결과를 가져올 수 있다.

편향성의 원인은 세 가지로 나눌 수 있다. 첫 번째는 목표 변수를 정의할 때, 특정 일부 집단이 과잉 혹은 과소 대표되거나 배제되는 경우이다. 두 번째로, 데이터를 전처리하고 레이블링(labeling)할 때, 적절하지 않은 평가를 하는 경우이다. 특히, 데이터 수집 과정에서 발생한 문제가 해결되지 않으면 측정 편향과 회상 편향이 발생할 수 있고 인공지능은 전혀 다른 결론을 내릴 수 있다. 마지막으로, 특징 선택(feature selection) 단계에서, 인종적 편향이 나타날 수 있는데, 이는 여러 특징을 상호 비교 연결함으로써 새롭게 조합되거나 민감한 개인정보를 인공지능이 편향적인 데이터들을 학습하는 경우이다. 이는 의사결정 단계에서 특정 방향으로 결론을 유도하게 되는 결과를 얻을 수 있다. 즉, 데이터 전처리 및 학습 과정에서 다양한 편

향이 개입될 수 있다[8].

본 연구에서는 음성 데이터셋인 VoxCeleb1을 사용하여 데이터셋의 분포를 살펴본다. 그리고 대표적인 컨볼루션 신경망(Convolutional Neural Network, CNN) 모델들을 대상으로 해당 데이터셋을 학습시킨 후 성별에 따른 Top1-accuracy, Top5-accuracy를 구하여 AI 공정성 관점으로 모델별에 따른 차이를 비교 및 분석한다.

본 연구는 다음과 같이 구성된다. 2장에서는 배경 기술과 관련 연구를 분석한다. 3장에서는 VoxCeleb1 데이터셋 기반의 공정성 관련 연구를 설명한다. 4장에서는 VoxCeleb1 데이터셋과 각 대표적인 CNN 모델들을 사용하는 실험 환경에 대해서 다루고, 마지막으로 5장과 6장에서는 실험 결과에 관한 고찰 및 결론을 서술한다.

## 2. 배경기술 및 관련연구

본 장에서는 VoxCeleb을 이용한 화자 인식과 데이터 편향 및 공정성에 대해 알아보고자 화자 인식 선행연구 동향과 인공지능 정확도 및 공정성에 대해 분석하고자 한다.

초기의 화자 인식 방식은 가우시안 혼합 모델(Gaussian Mixture Model, GMM) 이후에 서포트 벡터머신(Support Vector Machine, SVM)이 사용되었다. 이후에는 CNN 형태의 딥러닝 방식으로 알고리즘이 발전되면서 딥러닝 모델을 활용하여 화자 인식의 성능을 개선하려는 변화가 나타났다. [9]에서는 대표적인 CNN인 VGGNet(Visual Geometry Group Network)과 ResNet(Residual neural Network)에 self-attention 계층을 추가하여 화자 식별 모델을 만들었고, 이는 음성 식별 작업에서의 성능을 향상시켰다. 또한 [10]에서는 VoxCeleb을 이용하여 Open-set Speaker Recognition의 손실함수에 대한 평가를 했다. 또한, vanilla triple loss로 훈련된 네트워크가 대부분의 AM-Softmax 및 AAM-Softmax 훈련된 네트워크에 비해 성능이 더 우수하다는 것을 증명하였다. 상술한 종래의 연구들은 주로 정확도 측정 및 성능 향상에만 초점을 두어 진행되었고 딥러닝 및 기계학습의 정확도가 공

정성(Fairness)에 미치는 영향에 대해서는 다루지 않았다.

[11]에서는 화자 검증 과정에서 특정한 성별과 국가의 데이터가 입력 값일 때, 성능 저하가 크고 모든 단계에 편향이 존재한다는 결과를 확인하였다. 또한, 화자 검증의 성능 격차를 정량화하기 위한 평가 프레임워크와 화자 인식의 편향을 해결하는 방안을 제안했다. 그러나, 공개적으로 사용 가능한 벤치마크 모델을 블랙박스 예측 변수로 사용하여 제한된 범위로 실험을 진행했기 때문에 모든 화자 검증 모델에 대한 일반적인 평가로 간주하기 어렵다는 한계점이 있다. 또한 [11]에서는 데이터 그룹 간의 격차를 줄이기 위해서 GFN(Group-adapted Fusion Network) 아키텍처로 심층 화자 인식의 알고리즘 공정성 탐색을 하며, 공정한 화자 인식 모델과 적절한 평가 프레임워크의 필요성을 강조하였다. 본 연구에서는 VoxCeleb1 데이터셋에서 성별 균형을 조작하여 불균형 훈련 데이터가 화자 검증 모델의 그룹 공정성에 미치는 영향을 분석하였다. 인공지능 화자 식별의 불공정성은 국가, 나이 또는 성별에 따라 공정하지 않은 서비스를 제공할 수 있음을 의미한다. 해당 연구는 화자 인식의 훈련 데이터셋에서 성별 간 불균형한 그룹 표현이 모델 불공정으로 이어질 수 있다는 직접적인 증거를 제공한다는 점에서 기여점이 있다. 하지만, ResNet 모델에만 적용되어 다른 딥러닝 모델에서는 동일한 결과를 확인할 수 없으며, 훈련된 GFN을 VoxCeleb 외의 데이터셋에는 적용할 수 없다는 한계점이 있다.

## 3. 실험 결과와 분석

### 3.1 실험환경

본 연구의 실험 환경은 Colab(Colaboratory) Pro 환경에서 오픈소스 머신러닝 프레임워크인 PyTorch를 사용하여 구현하였다. 그리고 VoxCeleb : a large-scale speaker identification dataset[12]의 VGG 기반 화자 식별 모델인 VGGVox를 학습 모델로 활용했다. VGGVox는 음성 데이터를 짧게 나눈 후 단편화된 음성 데이터를 STFT(Shorttime fouriertransform, 국소

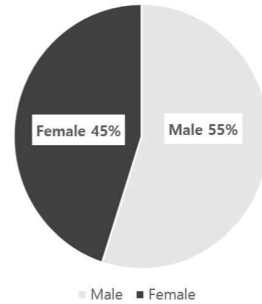
푸리에 변환) 기반의 전처리 방법을 사용하여 시간 도메인 음성 데이터를 푸리에 변환하여 분석한다.

실험에서는 사용된 비교 모델로써 VGGNet[13], ResNet18, ResNet34, MobileNetV1, MobileNetV2이며, 음성을 푸리에 변환하여 시각화했을 때 이를 학습할 수 있는 CNN(Convolutional Neural Network)을 사용한다. VGG에 의해 개발된 모델인 VGGNet 네트워크의 깊이를 최적화하여 CNN 성능을 개선했다. ResNet은 잔차 대표(residual representation) 함수를 학습함으로써 깊은 신경망을 가질 수 있다. 이전 계층의 입력을 다음 계층으로 전달하기 위해 잔차 연결(skip connection)을 사용한다. MobileNet은 깊이 분리형 컨볼루션 (Depthwise Separable Convolution)을 활용한 경량화 네트워크로 효율적인 연산이 가능하고, 모바일 기기에 적용할 수 있도록 구조를 단순하게 설계했다.

본 연구에서 사용한 데이터셋은 VoxCeleb1으로 YouTube에 업로드된 동영상에서 1,251명의 유명인에 대한 1,000,000개 이상의 발언의 집합이다. 해당 데이터셋은 Kaggle에서 제공하고 있는 훈련 데이터셋인 dev(development dataset) 및 테스트 데이터셋인 test(test dataset)을 다운로드해 사용하였으며, 화자의 데이터로는 다양한 분포의 인종, 악센트, 직업, 나이를 포함하고 있다. 해당 데이터셋은 표 1과 같이 데이터셋 순서를 0부터 시작하는 index로 표현하고 VoxCeleb1의 목소리의 식별값인 VoxCeleb1 ID, 화자를 구분하는 VGGFace1 ID와 함께 매칭되어 있었다. 성별은 남성을 m(male)으로 여성을 f(female)로 표기하여 분류되었고, 국가와 성별도 함께 라벨링 되어 있다.

<표 1> VoxCeleb1 데이터셋

index	VoxCeleb1 ID	VGGFace1 ID	Gender	Nationality	Set
0	id10001	A.J._Buckley	m	Ireland	dev
1	id10002	A.R._Rahman	m	India	dev
2	id10003	Aamir_Khan	m	India	dev
3	id10004	Aaron_Tveit	m	USA	dev
4	id10005	Aaron_Yoo	m	USA	dev
...	...	...	...	...	...
1246	id11247	Zachary_Levi	m	USA	dev
1247	id11248	Zachary_Quinto	m	USA	dev
1248	id11249	Zack_Snyder	m	USA	dev
1249	id11250	Zoe_Saldana	f	USA	dev
1250	id11251	Zulay_Henao	f	USA	dev



(그림 1) VoxCeleb1 성별에 대한 비율

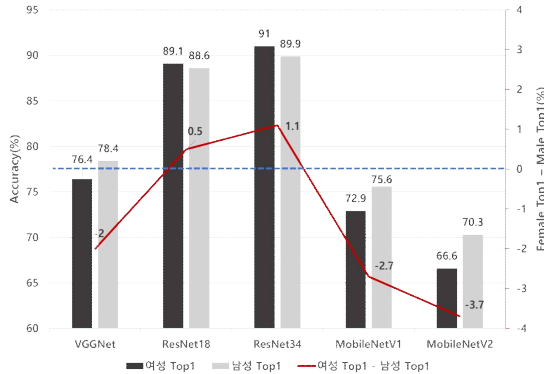
VoxCeleb1에서 제공하는 음성데이터를 성별에 따라 분류할 경우 그림 1과 같이 여성은 45%(561개), 남성은 55%(690개)로 남성의 데이터가 약 10% 정도 더 많은 비율을 차지했다.

각 모델은 전체 데이터셋을 학습한 횟수인 에포크(epoch)를 최대 500으로 설정하여 실험하였다. 또한, 학습 조기 종료(early stopping) 기법을 적용하여 모델의 성능이 일정한 지점부터 증가하지 않을 때 학습이 중지되도록 하였다. 또한, 학습률을 0.01로 설정하고, 배치 사이즈를 50으로 설정하였다.

### 3.2 실험결과

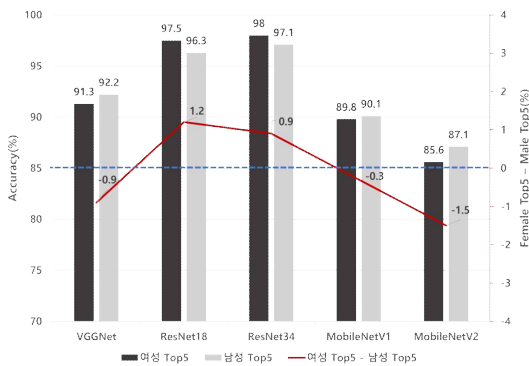
평가지표로는 정확도인 Top1-accuracy와 Top5-accuracy를 사용하였다. Top1-accuracy는 모델의 예상 답변 확률 중 가장 높은 것이 정답일 경우에 대한 정확도이며, Top5-accuracy는 모델의 예상 답변 확률 중 높은 5개에 정답이 포함될 경우에 대한 정확도이다.

그림 2는 여성과 남성의 각 모델별 Top1-accuracy를 나타낸 것이다. 여성과 남성 모두 ResNet34, ResNet18, VGGNet, MobileNetV1, MobileNetV2 순으로 Accuracy가 높았으며 가장 Accuracy가 높았던 ResNet34에서는 여성과 남성이 91%, 89.9%로 1.1% 차이가 났으며, Accuracy가 가장 낮은 MobileNetV2는 여성과 남성이 각각 66.6%, 70.3%로 3.7% 차이가 나는 결과를 나타냈다. 또한, ResNet34와 MobileNetV2 간의 Accuracy는 여성과 남성 각각 33.5%, 19.6% 차이가 났다.



(그림 2) 모델별 성별에 대한 Top1-accuracy

또한, 인공지능이 학습한 데이터셋이 남성의 비율이 여성에 비해 10% 더 높을 때, 이를 가장 잘 나타낸 모델은 MobileNetV2이다. MobileNetV2는 여성 대비 남성에서 정확도가 3.7% 크게 나타났다. 이는 데이터의 편향이 정확도에 영향을 줄 수 있음을 가장 잘 나타낸 것으로 해석할 수 있다. 반면, 여성의 학습 데이터 수가 남성에 비해 적었지만, ResNet18, ResNet34에서는 여성의 정확도가 각각 1.2%, 0.9% 높게 나왔다. 이 두 모델을 사용함으로써 학습 데이터량의 차이가 학습의 결과인 정확도에 미치는 영향을 조절할 수 있다는 것을 확인하였다.



(그림 3) 모델별 성별에 대한 Top5-accuracy

그림 3은 여성과 남성의 각 모델별 Top5-accuracy를 나타낸 것으로, 가로축은 각 모델의 명칭이며, 세로축은 Accuracy에 대한 수치이다. 그림 1과 동일하게 Accuracy가 높은 모델은 ResNet34, ResNet18, VGGNet, MobileNetV1, MobileNetV2 순으로 나타났다. 가장 Accuracy가 높았던 ResNet34에서는 여성과 남

성이 98%, 97.1%로 1.1% 차이가 났으며, Accuracy가 가장 낮은 MobileNetV2의 Accuracy인 85.6%, 87.1%에서는 1.5% 차이가 났다. 또한, 두 모델에서 여성과 남성의 정확도를 비교했을 때, 각각 14.4%, 10% 차이가 나는 결과를 나타냈다.

또한, Top1-accuracy와 마찬가지로 인공지능이 학습한 데이터셋이 남성의 비율이 여성에 비해 10% 더 높을 때, 이를 가장 잘 나타낸 모델은 MobileNetV2이다. MobileNetV2는 여성 대비 남성에서 정확도가 1.5% 크게 나타났다. Top1-accuracy와 동일하게 Top5-accuracy에서도 MobileNetV2를 통해 데이터의 편향이 정확도에 영향을 줄 수 있음을 확인하였다. 반면, 여성의 학습 데이터 수가 남성에 비해 적었지만, ResNet18, ResNet34에서는 여성의 정확도가 각각 1.2%, 0.9% 높게 나왔다. 이 두 모델을 사용함으로써 학습 데이터량의 차이가 학습의 결과인 정확도에 미치는 영향을 조절할 수 있다는 것을 확인하였다.

<표 2> 모델별 성별의

Top1-accuracy & Top5-accuracy

Model	Top1-accuracy (%)		Top5-accuracy (%)	
	여성	남성	여성	남성
VGGNet	76.4	78.4	91.3	92.2
ResNet18	89.1	88.6	97.5	96.3
ResNet34	91	89.9	98	97.1
MobilNetV1	72.9	75.6	89.8	90.1
MobilNetV2	66.6	70.3	85.6	87.1

Top1-accuracy, Top5-accuracy를 나타낸 그림 2와 그림 3을 요약하여 정리한 표 2를 살펴보면, Top5-accuracy가 Top1-accuracy 대비 평균적으로 여성과 남성 각각 13.24%, 12% 더 높은 정확도를 가졌다. 이는 Top1-accuracy는 1개 데이터를 이용해 정확도를 계산하지만, Top5-accuracy는 5개의 데이터 중 하나만 해당되는 것이 있으면 되기 때문에 정확도가 높아진다.

Top1-accuracy 측면에서 ResNet34가 MobileNetV2에 비해 성별 간의 정확도 차이가 1.1%로 가장 작은 반면, MobileNetV2는 1.5%로 가장 컸다. Top5-accu

acy 측면에서는 MobileNetV1이 성별 간의 정확도 차이가 0.3%로 작았으며 반면, ResNet18은 1.8%로 가장 컸다.

이처럼 남성은 55%이고, 여성의 경우 45%으로 남성의 데이터가 10% 많은 편향적 데이터셋이 모든 모델에 있어서 남녀 간의 정확도와 비례하는 것은 아니었음을 확인하였다. 또한, 각 모델의 정확도 차이는 모델을 활용한 서비스 이용 시 성별 간의 차등적인 서비스 품질로 이어질 수 있다. 목표 변수에 편향이 존재하는 경우에는 모델의 공정성을 정량화하고 편향을 줄이는 프레임워크가 필요하다[14].

#### 4. 결론

본 연구에서는 VoxCeleb1을 이용하여 AI 공정성을 살펴보았다. 모델별로 데이터셋에 대한 정확도는 차이가 있었으며, 남성과 여성의 데이터는 각각 55%, 45%를 차지하여 남성의 데이터가 10% 비중이 높았다. 이러한 실험을 통해 불균형한 데이터셋을 학습한 인공지능 모델 간 남녀별 화자 인식 정확도가 얼마나 상이한지 확인하였다. Top1-accuracy 측면에서 ResNet34가 MobileNetV2에 비해 성별 간의 정확도 차이가 가장 작은 반면, MobileNetV2는 남녀 간의 정확도 차이가 컸음을 확인하였다. 그리고 Top5-accuracy 측면에서는 MobileNetV1이 성별 간의 정확도 차이가 가장 작은 반면, ResNet18은 가장 정확도 차이가 컸다. 즉, 서비스 이용자들은 성별에 따라 공정하지 않은 서비스를 받을 수 있으며, 이는 성별에 따라 저하된 서비스의 품질을 받을 수 있는 성차별로 이어진다. 이 문제를 해결하여 정확하고 공정한 서비스를 제공하려면, 모델의 공정성을 정량적으로 측정하고 편향을 줄이는 관리 프레임워크가 필요하다. 또한, 데이터의 불균형으로 인한 AI 불공정성을 개선하기 위해 데이터를 인공지능에 학습시키기 전에 데이터가 성별, 인종, 국가 등 다양한 기준으로 분류된 비율에 대하여 데이터 불균형의 문제가 없는지 확인한 후 비율을 조정하는 기술에 관한 연구가 요구된다. 향후 연구에서는 성별에 관한 학습 데이터셋의 비율을 다르게 조정하여, 데이터셋의 비율이 정

확도에 미치는 영향을 평가하고, 결과에서 나타나는 편향을 줄이는 방안을 연구하고자 한다. 또한, 성별 간의 불공정성에 대한 정량화를 통해 편향을 줄이는 프레임워크를 연구할 계획이다.

#### 참고문헌

- [1] Mahata, A., Saini, N., Saharawat, S., & Tiwari., "Intelligent movie recommender system using machine learning," *Intelligent Human Computer Interaction*, Vol. 10127, pp. 94-110, 2016.
- [2] Qiu, J., Wu, Q., Ding, G., Xu, Y., & Feng, S., "A survey of machine learning for big data processing," *EURASIP Journal on Advances in Signal Processing*, Vol. 1, No. 67, pp. 1-16, 2016.
- [3] Markets and Markets, "Voice biometrics market by component, type(active and passive), application (authentication and customer verification, transaction processing), authentication process, organization size, deployment mode, vertical, and region - global forecast to 2026," <https://www.marketsandmarkets.com/Market-Reports/voicebiometrics-market-104503105.html>, 2021.
- [4] Canbek, N. G., & Mutlu, M. E., "On the track of artificial intelligence: Learning with intelligent personal assistants," *Journal of Human Sciences*, Vol. 13, No. 1, pp. 592-601, 2016.
- [5] 오원준, 연명흠. (2019). 스마트홈 환경에서 발생하는 다중사용자 간 인터랙션 이슈 발굴 및 디자인 솔루션 제안 - 제어 권한의 조절을 중심으로. *Journal of Integrated Design Research*, 18(3), 77-90.
- [6] Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A., "A Survey on Bias and Fairness in Machine Learning." *ACM Comput. Surv.* 54, 6, Article 115 (July 2022), 35 pages. 2021.
- [7] Toussaint, W., and Ding, A. Y., "Sveva fair: A framework for evaluating fairness in speaker verification," *arXiv preprint*, 2021.
- [8] 정원섭, "인공지능 알고리즘의 편향성과 공정성," *인간.환경.미래*, No. 25, pp. 55-73. 2020.
- [9] An, N. N., Thanh, N. Q., & Liu, Y., "Deep CNNs with self-attention for speaker identification," *IEEE access*, Vol. 7, pp. 85327-85337, 2019.
- [10] Chung, J. S., Huh, J., Mun, S., Lee, M., Heo, H. S.,

Choe, S., ... & Han, I., "In defence of metric learning for speaker recognition," arXiv preprint, 2020.

[11] Shen, H., Yang, Y., Sun, G., Langman, R., Han, E., Droppo, J., & Stolcke, A., "Improving fairness in speaker verification via Group-adapted Fusion Network," ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 7077-7081, 2022.

[12] Nagrani, A., Chung, J. S., & Zisserman, A., "VoxCeleb: a large-scale speaker identification dataset," arXiv preprint, 2017.

[13] Derpimort (2020) VGGVox <https://github.com/Derpimort/VGGVox-PyTorch>.

[14] Fogliato, R., Chouldechova, A., & G'Sell, M., "Fairness evaluation in presence of biased noisy labels," International Conference on Artificial Intelligence and Statistics, pp. 2325-2336, 2020.

**[ 저자 소개 ]**



신 나 연 (Na Yeon Shin)  
 2021년 08월 성신여자대학교 융합보안공학과 학사  
 2021년 09월~현재: 성신여자대학교 미래융합기술공학과 석사과정  
 email : 220216135@sungshin.ac.kr



이 진 민 (Jin-Min Lee)  
 2021년 08월 성신여자대학교 융합보안공학과 학사  
 2023년 02월 성신여자대학교 미래융합기술공학과 석사  
 2023년 03월~현재: 성신여자대학교 미래융합기술공학과 박사과정  
 email : 220237019@sungshin.ac.kr



노 현 (Hyeon No)  
 2022년 08월 성신여자대학교 융합보안공학과 학사  
 2022년 09월~현재: 성신여자대학교 미래융합기술공학과 석사과정  
 email : hyunn915@gmail.com



이 일 구 (Il-Gu Lee)  
 2003년 2월 서강대학교 전자공학과 학사  
 2005년 2월 KAIST 정보통신대학원 석사  
 2012년 2월 KAIST 지식재산대학원 석사  
 2016년 2월 KAIST 전산학부 박사  
 2017년 2월~현재: 성신여자대학교 융합보안공학과/미래융합기술공학과 부교수  
 email : iglee@sungshin.ac.kr