# 2025-1학기 DU-도전학기 계획서

<table>
<tr><td>과제명</td><td colspan="4">Anti-성착취 딥페이크 기술 개발</td></tr>
<tr><td>신청 유형</td><td colspan="4">☐ 개인 ■ 팀(팀명: 딥트랙)</td></tr>
<tr><td>도전 영역</td><td colspan="4">■ 전공(주전공 또는 복수전공) ☐ 일반선택</td></tr>
<tr><td>신청 학점</td><td colspan="4">3학점</td></tr>
<tr><td rowspan="5">참여자</td><td>성명</td><td>소속</td><td>학번</td><td>비고</td></tr>
<tr><td>강동원</td><td>컴퓨터공학과</td><td>22124044</td><td>팀장</td></tr>
<tr><td>김다솔</td><td>컴퓨터공학과</td><td>22377260</td><td>팀원</td></tr>
<tr><td>김현송</td><td>컴퓨터공학과</td><td>22358386</td><td>팀원</td></tr>
<tr><td>박부휘</td><td>컴퓨터공학과</td><td>22029068</td><td>팀원</td></tr>
<tr><td>지도교수 의견</td><td colspan="4">상기 학생들이 선정한 주제는 최근 사회적 문제가 되고 있는 딥페이크 범죄를 탐지하고 예방하기 위한 기술로서 파이썬프로그래밍, 웹프로그래밍, 인공지능 교과목을 심화 학습할 수 있는 좋은 주제라고 생각합니다. 직접 딥페이크 성착취물 이미지 및 영상을 탐지하기 위한 기술을 개발하면서 IT 전공자로서 사회에 기여하는 경험도 해볼 수 있어서 교육적으로도 매우 긍정적일 것으로 기대합니다.<br>학생들이 수립한 계획 중, GNN(Graph Neural Network) 기반 딥페이크 탐지 기술은 기존의 CNN(Convolutional Neural Network) 기반의 딥페이크 탐지 기술과 비교하여 매우 도전적이고 선진적인 기술이라고 판단되며, 도전 결과를 학술대회에 발표함으로써 학계에도 큰 기여를 할 수 있을 것이라 생각합니다. 성공적인 도전학기가 되도록 성심껏 지도하겠습니다.<br>(소속) 컴퓨터정보공학부 (성명) 김지연 (서명 또는 날인)</td></tr>
</table>

## 1. 도전 배경

현재 온라인에 존재하는 딥페이크 영상의 약 96%가 불법 음란 동영상이며, 딥페이크 기술이 전문가가 아닌 일반인도 쉽게 사용할 수 있도록 퍼져있다. 딥페이크 음란물 사이트의 포르노 영상 중 25%가 한국의 K-POP 여가수를 대상으로 한 것으로 조사되었으며, 주위의 일반인을 대상으로 영상을 조작, 유포하는 범죄인 '지인 능욕'은 N번방 사건에서 보여지듯, 새로운 디지털 성범죄 유형으로 추가되었다.[1)]

최근 일반인 여성을 대상으로 제작된 불법 딥페이크 합성물이 텔레그램을 통해 무차별적으로 유포되면서 미성년자 등 수많은 피해자가 나오고 있다. 시큐리티 히어로의 '2023 딥페이크 현황 보고서'에 따르면 딥페이크 성착취물에 가장 취약한 국가 1위가 한국이다.[2)] 경찰청은 지난해 검거된 허위 영상 피의자는 120명, 그 가운데 10대가 무려 91명이나 포함됐다고 밝혔다. 이러한 문제점을 해결하기 위해서 딥페이크 탐지 기술은 개발되어져 왔다. 이러한 딥페이크 여부를 판별하는 소프트웨어가 개발되어져 왔지만 현재 80% 수준의 판독률로 낮은 정확도를 가진 소프트웨어가 개발되었고, 이마저도 주로 기업과 기관이 사용하고 있기에 개인이 이용하기는

1) https://koreascience.kr/article/JAKO202032254872919.pdf

2) https://m.boannews.com/html/detail.html?idx=133899

어려운 상황이다.[3] 최근 불법 딥페이크 성착취물 유포가 크게 주목받자 일부 업체에서는 탐지 프로그램을 무료로 지원하겠다고 나섰지만 기업, 관공서 등 B2B 서비스에 국한되어 개인이 이를 활용하기는 어려운 상황이다. 또한 딥페이크 생성 방법이 고도화되어감에 따라 더욱 실제 같은 가짜 동영상, 이미지 등이 만들어 지고 있다. 현재 개발된 딥페이크 탐지 방법은 분명 준수한 성능을 갖고 있으나, 새로운 딥페이크 생성 방법에 대하여 취약할 수 있다는 문제점이 있으며 이러한 문제점을 해결하고자 성착취물 딥페이크 탐지 기술 개발을 주제로 삼아 도전하고자 한다.

## 2. 도전 과제의 목표

### 가. 팀 목표

본 팀의 목표는 성 착취물의 딥페이크 여부를 탐지하는 모델을 개발한 후 딥페이크 탐지 성능을 분석하는 것이다. 먼저, 딥페이크 성 착취 이미지와 해당 성 착취 이미지의 원본을 수집하고, 해당 데이터셋을 전처리 수행 여부에 따라 분류하여 학습 데이터셋을 구축하려고 한다. 이후, 수집한 학습 데이터셋으로 학습을 진행하여 딥페이크 모델을 생성하고 이들의 성능을 평가한다. 이때, 딥페이크 탐지 모델은 '딥러닝 기반의 일반 이미지 딥페이크 탐지 모델'과 '딥러닝 기반의 전처리 이미지 딥페이크 탐지 모델', 'GNN(Graph Neural Networks) 기반의 딥페이크 탐지 모델'의 3가지 모델을 개발하고 성능 분석을 통해 가장 성능이 좋은 모델을 선정하고자 한다. 이를 통해 마지막으로, 성 착취물의 딥페이크 여부를 탐지하는 웹사이트를 개발하려고 한다.

### 나. 개인 목표

1) 딥페이크 성 착취물에 대한 이미지 데이터 수집과 전처리 및 딥페이크 탐지 웹사이트 구현(강동원)
   - 성 착취물 이미지들에 대한 딥페이크 수행
   - 딥페이크 이미지 데이터 전처리 진행
   - 성 착취물 이미지를 삽입하였을 경우 딥페이크 여부를 탐지하는 기능의 웹페이지를 구현하며 전공역량 강화
2) GNN 기반의 딥페이크 탐지 모델링 및 딥페이크 탐지 모델 성능 분석(김다솔)
   - 제작한 딥페이크 성 착취물 데이터셋을 기반으로 GNN 간선 예측 기반의 딥페이크 탐지 모델을 개발하며 전공역량 강화
   - GNN 기반의 딥페이크 탐지 모델 성능 평가
3) 딥러닝 기반의 딥페이크 탐지 인공지능 모델링 및 딥페이크 탐지 모델 성능 분석(김현송)
   - 제작한 딥페이크 성 착취물 데이터셋을 기반으로 딥러닝 기반의 딥페이크 탐지 모델을 개발하며 전공역량 강화
   - 딥러닝 기반의 딥페이크 탐지 모델 성능 평가
4) 딥페이크 성 착취물에 대한 이미지 데이터 수집과 전처리 및 딥페이크 탐지 웹사이트 구현(박부휘)
   - 성 착취물 이미지들에 대한 딥페이크 수행
   - 딥페이크 이미지 데이터 전처리 진행
   - 성 착취물 이미지를 삽입하였을 경우 딥페이크 여부를 탐지하는 기능의 웹페이지를 구현하며 전공역량 강화

3) https://www.donga.com/news/Society/article/all/20240905/126866797/1

## 3. 도전 과제 내용

### 가. 딥페이크 탐지 기술 조사

딥페이크(Deepfake)는 딥러닝(Deep-learning)과 가짜(fake)의 합성어로 인공지능을 기반으로 한 인간 이미지 합성 기술을 의미한다. 이 기술은 고도의 학습 알고리즘을 활용해 실제 사람의 얼굴, 목소리, 표정 등을 매우 정교하게 모방한다. 이러한 딥페이크 기술은 방송국에서 피해자의 모습을 모자이크 대신 딥페이크 기술을 활용하는 등 수많은 이점이 존재하지만 동시에, 기술에 대한 접근이 쉬워 딥페이크 기술은 성 착취물 제작, 가짜 뉴스, 금융 사기 등에 악용된다. 딥페이크 탐지 기술은 현재 눈의 움직임을 추적하여 딥페이크 유무를 판별하는 기술, 시간에 따라 변화하는 3차원 공간 내의 경로(3D Spatiotemporal Trajrctories)를 기반으로 딥페이크 유무를 판별하는 기술 등이 존재하며 이들은 모두 딥페이크로 인한 부작용을 감소하는데 중요한 역할을 하고 있다.

### 나. 딥러닝 조사 및 딥페이크 탐지 기술 제작

딥러닝(Deep Learning)이란 인간의 두뇌에서 영감을 얻은 방식으로 데이터를 처리하도록 가르치는 학습하는 기계 학습의 한 종류이다. 이러한 딥러닝은 데이터에서 특정한 패턴을 학습하여 복잡한 문제를 해결하여 인간의 지능이 필요한 작업을 자동화할 수 있다. 딥러닝의 알고리즘은 대표적으로 3가지가 있는데, 그중 CNN(Convolution Neural Network)은 이미지 처리에 강점을 보이는 알고리즘으로 컴퓨터가 이미지 속에서 색깔, 모양, 경계선과 같은 특징을 자동으로 찾아낸다. 본 팀은 CNN을 활용하여 딥페이크가 진행된 이미지와 딥페이크가 진행되지 않은 이미지 두 종류의 이미지 데이터를 기반으로 딥러닝을 진행하여 딥페이크 성착취물 탐지 모델을 제작할 것이다.

### 다. GNN 조사 및 딥페이크 탐지 기술 제작

GNN(Graph Neural Networks)이란 그래프 구조로 표현된 데이터를 분석하고 학습하는 딥러닝 기술의 일종이다. 그래프는 점(노드)과 점을 연결하는 선(엣지)로 이루어진 구조인데, GNN은 이러한 그래프 구조를 학습하여 노드의 클래스를 예측하거나 노드 사이의 연결 여부, 그래프의 클래스를 예측할 수 있다. 이러한 GNN은 현재 사용자와 사용자 간의 관계를 활용한 추천 시스템, 분자나 세포 등 기본 단위 간의 연관 관계로부터 새로운 현상을 분석하는 기술 등에 널리 활용되고 있다. 본 팀은 GNN을 활용하여 학습 데이터의 이미지를 노드로 삼아 해당 이미지의 성 착취 데이터에 대한 딥페이크 유무 탐지 모델을 제작할 것이다.

### 라. 업무 분장 내용

| 팀원 성명 | 소속 | 담당 업무 |
|---|---|---|
| 강동원 | 컴퓨터공학과 | - 성착취물 원본 이미지 학습 데이터 수집<br>- 웹페이지 기본 틀 제작 및 디자인 |
| 김다솔 | 컴퓨터공학과 | - GNN 기반의 딥페이크 탐지 모델 제작<br>- GNN 기반의 딥페이크 탐지 모델의 성능 분석 코드 작성 |
| 김현송 | 컴퓨터공학과 | - 딥러닝 기반의 딥페이크 탐지 모델 제작<br>- CNN 기반의 딥페이크 탐지 모델의 성능 분석 코드 작성 |
| 박부휘 | 컴퓨터공학과 | - 딥페이크 수행된 학습 데이터 수집<br>- 학습 데이터 이미지 전처리 및 분류 기능 코딩<br>- 웹페이지 내 딥페이크 탐지 시각화 코딩 |

## 4. 도전 과제 추진일정

| 주차 | 활동 목표 | 활동 내용 | 투입 시간 |
|---|---|---|---|
| 1주차 | 성 착취물 딥페이크 탐지를 위한 관련 기술 조사 | 강동원(팀장) : 딥페이크 제작 모델 조사 및 딥페이크 구동 준비 | 6 |
| | | 김다솔(팀원) : GNN 기술 조사 | 6 |
| | | 김현송(팀원) : 딥러닝 모델 중 CNN 알고리즘 조사 | 6 |
| | | 박부휘(팀원) : 이미지 전처리 모델 조사 | 6 |
| 2주차 | 시나리오 설계 | 강동원(팀장) : 딥페이크 제작 모델 선정 및 전체 시나리오 설계 | 6 |
| | | 김다솔(팀원) : GNN 모델 선정 | 6 |
| | | 김현송(팀원) : CNN 기반의 딥러닝 모델 선정 | 6 |
| | | 박부휘(팀원) : 이미지 전처리 방법론 설계 및 웹페이지 구상 | 6 |
| 3주차 | 성 착취물 학습 데이터 생성 | 강동원(팀장) : 성 착취물 원본 데이터 수집 | 6 |
| | | 김다솔(팀원) : GNN 학습 데이터 조사 | 6 |
| | | 김현송(팀원) : 딥러닝 학습 데이터 조사 | 6 |
| | | 박부휘(팀원) : 성 착취물 딥페이크 데이터 수집 | 6 |
| 4주차 | 성 착취물 학습 데이터 생성 | 강동원(팀장) : 학습 데이터 분류 | 6 |
| | | 김다솔(팀원) : GNN 학습 데이터 조사 | 6 |
| | | 김현송(팀원) : 딥러닝 학습 데이터 조사 | 6 |
| | | 박부휘(팀원) : 이미지 전처리 코드 작성 | 6 |
| 5주차 | 딥페이크 성 착취물 탐지 모델 개발 | 강동원(팀장) : 웹페이지 동적 모델 구현 조사 | 6 |
| | | 김다솔(팀원) : GNN 기반 딥페이크 성 착취물 탐지 모델 개발 | 6 |
| | | 김현송(팀원) : 딥러닝 기반 딥페이크 성 착취물 탐지 모델 개발 | 6 |
| | | 박부휘(팀원) : 웹페이지 동적 모델 구현 조사 | 6 |
| 6주차 | 딥페이크 성 착취물 탐지 모델 개발 | 강동원(팀장) : 웹페이지 동적 모델 구현 조사 | 6 |
| | | 김다솔(팀원) : GNN 기반 딥페이크 성 착취물 탐지 모델 개발 | 6 |
| | | 김현송(팀원) : 딥러닝 기반 딥페이크 성 착취물 탐지 모델 개발 | 6 |
| | | 박부휘(팀원) : 웹페이지 동적 모델 구현 조사 | 6 |
| 7주차 | 딥페이크 성 착취물 탐지 모델 개발 | 강동원(팀장) : 웹페이지 동적 모델 구현 조사 | 6 |
| | | 김다솔(팀원) : GNN 기반 딥페이크 성 착취물 탐지 모델 개발 | 6 |
| | | 김현송(팀원) : 딥러닝 기반 딥페이크 성 착취물 탐지 모델 개발 | 6 |
| | | 박부휘(팀원) : 웹페이지 동적 모델 구현 조사 | 6 |
| 8주차 | 중간 보고서 작성 | 강동원(팀장) : 중간 보고서 작성 | 6 |
| | | 김다솔(팀원) : 중간 보고서 작성 | 6 |
| | | 김현송(팀원) : 중간 보고서 작성 | 6 |
| | | 박부휘(팀원) : 중간 보고서 작성 | 6 |
| 9주차 | 성능 분석 | 강동원(팀장) : 딥페이크 성 착취물 탐지 모델 성능 평가 정리 및 선정 | 6 |
| | | 김다솔(팀원) : GNN 기반 딥페이크 성 착취물 탐지 모델 성능 평가 코드 작성 | 6 |
| | | 김현송(팀원) : 딥러닝 기반 딥페이크 성 착취물 탐지 모델 성능 평가 코드 작성 | 6 |
| | | 박부휘(팀원) : 딥페이크 성 착취물 탐지 모델 성능 평가 | 6 |
| 10주차 | 학술대회 논문 작성 | 강동원(팀장) : 학술대회 논문 작성 | 6 |
| | | 김다솔(팀원) : 학술대회 논문 작성 | 6 |
| | | 김현송(팀원) : 학술대회 논문 작성 | 6 |
| | | 박부휘(팀원) : 학술대회 논문 작성 | 6 |

| | | | |
|---|---|---|---|
| 11주차 | 학술대회 발표 준비 | 강동원(팀장) : 학술대회 대본 작성 | 6 |
| | | 김다솔(팀원) : 학술대회 발표자료 준비 | 6 |
| | | 김현송(팀원) : 학술대회 발표자료 준비 | 6 |
| | | 박부휘(팀원) : 학술대회 발표자료 준비 | 6 |
| 12주차 | 선정 모델 기반의 딥페이크 성 착취물 탐지 웹페이지 제작 | 강동원(팀장) : 웹페이지 전체 틀 작성 | 6 |
| | | 김다솔(팀원) : 선정 모델 웹페이지 내 구현 | 6 |
| | | 김현송(팀원) : 선정 모델 웹페이지 내 구현 | 6 |
| | | 박부휘(팀원) : 웹페이지 모델 동작 시각화 | 6 |
| 13주차 | 선정 모델 기반의 딥페이크 성 착취물 탐지 웹페이지 제작 | 강동원(팀장) : 웹페이지 전체 틀 작성 | 6 |
| | | 김다솔(팀원) : 선정 모델 웹페이지 내 구현 | 6 |
| | | 김현송(팀원) : 선정 모델 웹페이지 내 구현 | 6 |
| | | 박부휘(팀원) : 웹페이지 모델 동작 시각화 | 6 |
| 14주차 | 선정 모델 기반의 딥페이크 성 착취물 탐지 웹페이지 제작 | 강동원(팀장) : 웹페이지 전체 틀 작성 | 6 |
| | | 김다솔(팀원) : 선정 모델 웹페이지 내 구현 | 6 |
| | | 김현송(팀원) : 선정 모델 웹페이지 내 구현 | 6 |
| | | 박부휘(팀원) : 웹페이지 모델 동작 시각화 | 6 |
| 15주차 | 최종 보고서 작성 | 강동원(팀장) : 최종보고서 작성 | 6 |
| | | 김다솔(팀원) : 최종보고서 작성 | 6 |
| | | 김현송(팀원) : 최종보고서 작성 | 6 |
| | | 박부휘(팀원) : 최종보고서 작성 | 6 |

## 5. 활동 지원비 상세 내역

| 활동 지원비 신청내역 | | |
|---|---|---|
| 항 목 | 산출근거 | 금액(원) |
| 등록비 | - 학술대회 등록비 200,000원 | 200,000 |
| 항공료 | - 동대구 - 학술대회장 교통비(제주도)<br>- 200,000원 * 4명 = 800,000원 | 800,000 |
| 회의비 | - 회의비 10,000원 * 4명 * 15번 = 600,000원 | 600,000 |
| 재료비 | - 재료비[A4용지, 기타 품목(학술대회 포스터 발표)) 300,000원 | 300,000 |
| 자료구입비 | - 자료구입비 100,000원 | 100,000 |
| **합계(원)** | | **2,000,000** |

## 6. 과제 수행 후 제출할 수 있는 결과물

도전 학기 활동 수행 후 제출할 수 있는 결과물은 크게 팀 공통 결과물과 개인 결과물로 나뉜다. 먼저 팀 공동 결과물로는 성 착취물의 딥페이크 유무를 판별하는 모델과 해당 모델을 활용한 웹페이지가 있다. 또, 성 착취물의 딥페이크 유무를 판별하는 모델의 성능 분석을 기반으로 가장 유효한 모델 선정하는 것을 중심으로 학술대회 논문을 작성하여 구두 발표로 학술대회에 참여하고자 한다.

**가. 팀 공통 결과물** : 딥페이크 성착취물 탐지 모델 및 웹페이지, 학술대회 논문

**나. 개인 결과물**

| 팀원 성명 | 소속 | 개인 결과물 |
|---|---|---|
| 강동원 | 컴퓨터공학과 | - 원본 이미지 데이터셋<br>- 웹페이지 작성 코드 |
| 김다솔 | 컴퓨터공학과 | - GNN 기반 딥페이크 성착취물 탐지 모델의 작성 코드 파일 |
| 김현송 | 컴퓨터공학과 | - 딥러닝 기반의 딥페이크 성착취물 탐지 모델의 작성 코드 파일 |
| 박부휘 | 컴퓨터공학과 | - 딥페이크 이미지 데이터셋<br>- 이미지 전처리 코드<br>- 웹페이지 작성 코드 |

## 7. 결과물 활용 계획

**가. 일반 대중의 접근성과 편의성 제공**

Anti-딥페이크 탐지 기술을 통해 기관, 기업 이외에 개인이 쉽게 사용할 수 있는 웹사이트로 제작되어 쉽게 딥페이크 여부를 확인 가능하게 만들어 준다. 이를 통해 딥페이크로 인한 피해를 예방하고 정보의 신뢰성을 보장할 수 있게 된다.

**나. 교육 및 인식 개선**

Anti-딥페이크 탐지 기술은 개인을 대상으로 하는 소프트웨어이기 때문에 딥페이크의 위험성과 탐지 기술의 중요성을 알리기 위한 도구로서의 역할을 수행한다. 사용자는 딥페이크의 심각성, 탐지 가능성, 사회적으로 대응하는 방법에 대한 정보를 알아가는데 활용 가능하다.

**다. 법적 및 윤리적 지원 도구**

딥페이크 피해 사례가 발생한 경우, Anti-딥페이크 탐지 기술은 피해자들이 대응을 준비하는데 필요한 역할을 한다. 특히, 이 기술은 허위 정보 확산을 방지하고 디지털 콘텐츠의 진위 여부를 판단하는 데 중요한 역할을 한다.

**라. 지속적 기술 개선**

Anti-딥페이크 탐지 기술은 사용자 데이터를 익명화하여 수집, 지속적으로 딥페이크 탐지 모델을 개선하는데 활용된다. 이를 통해 기술의 정확도와 신뢰성을 높이고, 새로운 딥페이크 생성 기술에 대응하는데 큰 역할을 할 수 있게 된다.