



국민대학교  
전자정보통신대학  
컴퓨터공학부


# 캡스톤 디자인 I

## 종합설계 프로젝트

프로젝트 명	캡스톤 계획서 안내
팀 명	캡스톤 팀
문서 제목	계획서

Version	1.2
Date	20XX-MAR-29

팀원	이태훈 (조장)
	이인평
	이주형
	김성수
	김민재

 <b>국민대학교</b> <b>컴퓨터공학부</b> <b>캡스톤 디자인 I</b>	<b>계획서</b>		
	<b>프로젝트 명</b>	인공지능을 이용한 영상 및 음성 필터링	
	<b>팀 명</b>	유하	
	Confidential Restricted	Version 1.2	20xx-MAR-25


#### CONFIDENTIALITY/SECURITY WARNING

이 문서에 포함되어 있는 정보는 국민대학교 전자정보통신대학 컴퓨터공학부 및 컴퓨터공학부 개설 교과목 캡스톤 디자인I 수강 학생 중 프로젝트 "Youtube Hi"를 수행하는 팀 "YouHi"의 팀원들의 자산입니다. 국민대학교 컴퓨터공학부 및 팀 "YouHi"의 팀원들의 서면 허락없이 사용되거나, 재가공 될 수 없습니다.

## 문서 정보 / 수정 내역


<b>Filename</b>	계획서-유하.doc
<b>원안작성자</b>	이태훈
<b>수정작업자</b>	이인평, 이주형

수정날짜	대표수정자	Revision	추가/수정 항목	내 용
2020-03-09	이태훈	1.0	최초 작성	
2020-03-11	이인평	1..1	내용 수정	개발 목표 수정 및 AWS 내용 추가
2020-03-20	이주형	1.2	내용 수정	시스템 구조 변경사항 적용 및 내용 수정

 <b>국민대학교</b> <b>컴퓨터공학부</b> <b>캡스톤 디자인 I</b>	<b>계획서</b>		
	<b>프로젝트 명</b>	인공지능을 이용한 영상 및 음성 필터링	
	<b>팀 명</b>	유하	
	Confidential Restricted	Version 1.2	20xx-MAR-25

## 목 차

<b>1</b>	<b>개요</b>	<b>4</b>
1.1	프로젝트 개요	4
1.2	추진 배경 및 필요성	4
<b>2</b>	<b>개발 목표 및 내용</b>	<b>8</b>
2.1	목표	8
2.2	연구/개발 내용	11
2.3	개발 결과	11
2.3.1	결과물 목록 및 상세 사양	11
2.3.2	시스템 기능 및 구조	12
2.4	기대효과 및 활용방안	13
<b>3</b>	<b>배경 기술</b>	<b>14</b>
3.1	기술적 요구사항	14
3.2	현실적 제한 요소 및 그 해결 방안	15
3.2.1	하드웨어	15
3.2.2	소프트웨어	15
3.2.3	기타	16
<b>4</b>	<b>프로젝트 팀 구성 및 역할 분담</b>	<b>17</b>
<b>5</b>	<b>프로젝트 비용</b>	<b>17</b>
<b>6</b>	<b>개발 일정 및 자원 관리</b>	<b>18</b>
6.1	개발 일정	18
6.2	일정별 주요 산출물	19
6.3	인력자원 투입계획	20
6.4	비 인적자원 투입계획	20
<b>7</b>	<b>참고 문헌</b>	<b>21</b>

 <b>국민대학교</b> <b>컴퓨터공학부</b> <b>캡스톤 디자인 I</b>	<b>계획서</b>		
	<b>프로젝트 명</b>	인공지능을 이용한 영상 및 음성 필터링	
	<b>팀 명</b>	유하	
	Confidential Restricted	Version 1.2	20xx-MAR-25

# 1 개요

## 1.1 프로젝트 개요

본 프로젝트는 동영상 파일에서 영상과 음성을 분리하여 각각 사전 학습된 딥러닝 모델을 통해 태깅을 실시한 후, 만약 미성년자가 시청하기에 부적절한 영상 및 음성이 포함되어있으면 해당 구간의 위치와 위배된 내용을 이용자에게 알려준다.

## 1.2 추진 배경 및 필요성


영상 시청 시 사용할 수 있는 동영상 플랫폼에는 다양한 종류들이 있다. 그 중에서도 YouTube 플랫폼은 2019년 7월 기준 타 플랫폼 대비 가장 많은 순이용자를 보유하고 있다. 리서치 전문업체 닐슨코리아클릭의 통계에 따르면 YouTube의 2019년 7월 기준 이용자 수는 2,631만명에 달했고 이는 전년 대비 6.7% 증가한 수치이다.

주요 OTT 월간 순이용자			(단위=명)
구분	2018년 7월	2019년 7월	증감률 (%)
유튜브	2,466만	2,631만	6.7
넷플릭스	41만	185만	342
틱톡	87만	180만	107
옥수수	313만	336만	7.1
네이버TV	323만	193만	-40
폭	76만	139만	83

※ 자료=닐슨코리아클릭

### <사진 1. 동영상 플랫폼 이용자 현황>

YouTube의 시스템은 접근성이 뛰어나 누구나 쉽게 자신이 원하는 동영상을 시청하거나 업로드할 수 있다는 장점이 있다. 하지만 동영상 업로드에 관한 제한이 없어서 일부 사용자가 올린 부적절한 내용의 동영상은 제재없이 그대로 미성년자에게 노출되기도 한다. 이

 <b>국민대학교</b> <b>컴퓨터공학부</b> <b>캡스톤 디자인 I</b>	<b>계획서</b>		
	<b>프로젝트 명</b>	인공지능을 이용한 영상 및 음성 필터링	
	<b>팀 명</b>	유하	
	Confidential Restricted	Version 1.2	20xx-MAR-25

는 청소년들에게 부정적인 영향을 주며 청소년 범죄와 같은 사회적 문제로 이어질 수 있다. 현재 YouTube는 이를 예방하기 위해 다양한 가이드 라인을 사용자에게 제시하고 있고, 가이드 라인에 적합하지 않은 영상들은 삭제하거나, 연령 제한을 통해 미성년자들이 시청하지 못하도록 하였다.

## 정책 및 안전


YouTube를 사용한다는 것은 전 세계 사람들과 함께하는 커뮤니티의 일원이 된다는 것입니다. YouTube의 새롭고 멋진 커뮤니티 기능을 사용하려면 일정 수준 이상의 신뢰가 필요합니다. 수많은 YouTube 사용자들이 이러한 신뢰를 존중하고 있으며 여러 분도 책임감 있는 모습을 보여주시리라 믿습니다. 아래 가이드라인을 준수하시면 YouTube를 모두에게 즐겁고 신나는 공간으로 유지할 수 있습니다.

YouTube에 표시되는 콘텐츠가 모두 마음에 들 수는 없을 것입니다. 콘텐츠가 부적절하다고 생각하시면 YouTube팀에서 검토할 수 있도록 **신고 기능**을 통해 신고해 주세요. YouTube팀에서는 신고된 콘텐츠를 연중무휴 검토하여 커뮤니티 가이드의 위반 여부를 판단합니다.

### <사진 2. 현재 Youtube 영상 검열 시스템>

현재 YouTube에서는 가이드라인을 준수하지 않는 영상에 대하여 YouTube 운영자들의 모니터링과 사용자들의 신고를 통해서 필터링하고 있으며 일부 영상에 대해서는 "노란딱지"라는 유튜브 자체 AI시스템이 검열을 진행한다. 하지만 운영자의 모니터링과 사용자의 신고를 통한 필터링 방법은 처리 과정이 길다는 단점이 존재하며 자체 AI시스템은 사용자가 자신의 영상 중 어떤 부분이 가이드라인을 위반하였는지 직접 확인이 불가능하여 오검열 발생시 확인 및 대처가 어렵다는 한계가 있다.

이러한 문제들이 제대로 해결되지 않아 몇몇 영상들은 미성년자에게 부적합한 영상임에도 불구하고 미성년자에 대한 제약이 걸려있지 않은 채로 청소년들에게 노출되고 있다. 영화 <신세계>는 청소년 관람 불가 등급의 영화이다 YouTube에서도 또한 이에 대한 내용을 확인할 수 있다.




**신세계**  
YouTube 영화

드라마 · 2013 · 청소년 관람불가 · 한국어 오디오

세 남자가 가고 싶었던 서로 다른 신세계 "너, 나하고 일 하나 같이 하자" 경찰청 수사 기획과 감과장(최민식)은 국내 최대 범죄 조직인 ...

출연: 이정재, 최민식, 황정민  
감독: 박훈정

구매 또는 대여

 <b>국민대학교</b> <b>컴퓨터공학부</b> <b>캡스톤 디자인 I</b>	<b>계획서</b>		
	<b>프로젝트 명</b>	인공지능을 이용한 영상 및 음성 필터링	
	<b>팀 명</b>	유하	
	Confidential Restricted	Version 1.2	20xx-MAR-25

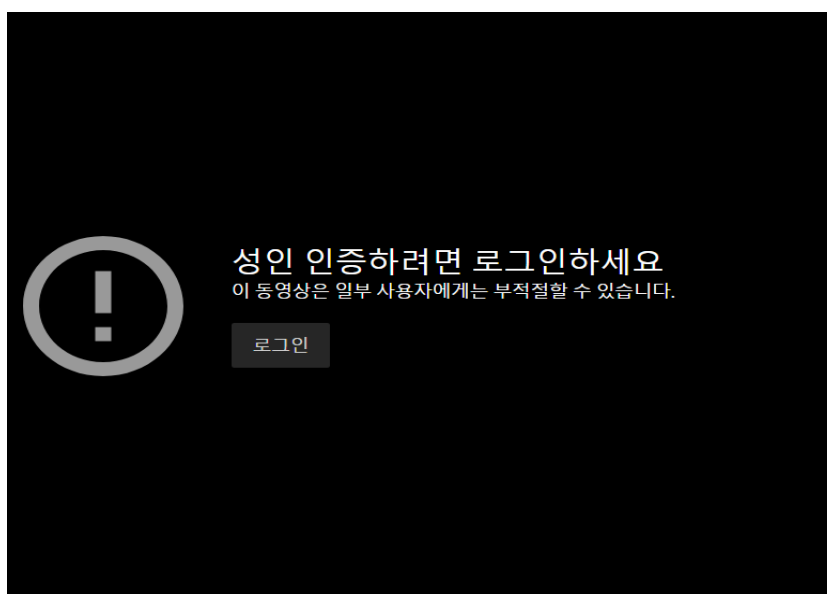
### <사진 3. 신세계 영화 - Youtube 제공>


그러나 해당 영화에서 선정적이거나 폭력적인 장면이 고스란히 담겨있는 장면들을 편집해 놓은 영상들은 누구나 시청 가능하게 설정되어 있고, Youtube 정책에 의해 검열되지 않았다. 이러한 영상들 중, 400만이 넘는 조회수를 기록한 영상도 존재했다.



### <사진 4. 신세계 명장면 - Youtube>

미성년자에게 부적합한 영상은 다음과 같이 개제되어야 한다.

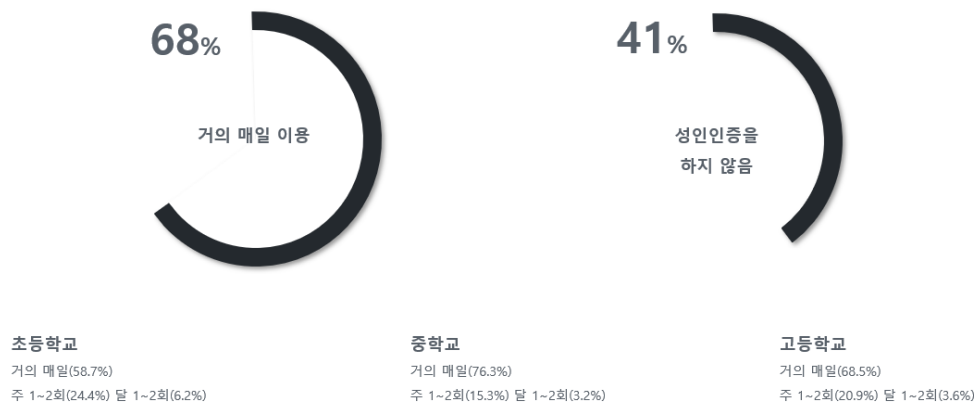


 <b>국민대학교</b> <b>컴퓨터공학부</b> <b>캡스톤 디자인 I</b>	<b>계획서</b>		
	<b>프로젝트 명</b>	인공지능을 이용한 영상 및 음성 필터링	
	<b>팀 명</b>	유하	
	Confidential Restricted	Version 1.2	20xx-MAR-25


### <사진 5. 미성년자 부적합 영상 - Youtube>

<사진 4>과 같은 예로 많은 영상들이 미성년자에게 부적합한 내용을 담고 있음에도 불구하고, <사진 5>와 같은 화면이 나오지 않기 때문에 미성년자들은 성인 인증과 같은 절차없이 손쉽게 영상을 볼 수 있다. 아래 <사진 6>의 내용은 여성가족부가 실시한 청소년 매체 이용 및 유해 환경 실태에 대한 설문조사 결과이다. 설문조사에 따르면, 조사에 응한 청소년 중 반이 넘는 인원들이 거의 매일 YouTube 를 포함한 동영상 시청 플랫폼을 이용한다고 응답했고 41%에 해당하는 청소년들은 청소년 유해 영상에 성인 인증 절차를 밟지 않고도 접근할 수 있었다고 답했다. 이를 통해 많은 청소년들이 별다른 제한없이 손쉽게 청소년 유해물에 접근하고 있음을 알 수 있다. 따라서 본 프로젝트에서는 미성년자들에게 초점을 맞추어 해당 연령대가 시청하기에 부적합한 영상에 대해 '19 세이상 시청가능' 조건을 걸지 않으면 YouTube 에 업로드 하지 못하도록 하는 자동 필터링 시스템을 개발하기로 했다. 필터링의 기준은 YouTube 이용정책 - 연령별 등급 제한을 참고하였다.

청소년의 YOUTUBE 사용 실태(여성가족부, 『청소년매체이용및유해환경실태조사』)



### <사진 6. 청소년들의 Youtube 이용 실태>

 <b>국민대학교</b> <b>컴퓨터공학부</b> <b>캡스톤 디자인 I</b>	<b>계획서</b>		
	<b>프로젝트 명</b>	인공지능을 이용한 영상 및 음성 필터링	
	<b>팀 명</b>	유하	
	Confidential Restricted	Version 1.2	20xx-MAR-25

## 2 개발 목표 및 내용

### 2.1 목표

- 현재 우리나라에서 가장 많이 이용되고 있는 영상 플랫폼인 YouTube에서 영상 제한 콘텐츠 가이드라인에 맞게 동영상의 업로드될 수 있도록 도움을 주며 미성년자들이 건전하고 건강하게 YouTube를 이용할 수 있도록 장려한다.
- 또한 현재 YouTube "노란 딱지" 정책 기준의 모호성과 불공정성에 대한 문제를 해결하기 위하여 사용자가 직접 자신의 영상 중 어떤 부분이 부적합한지 확인이 가능하게 하여 크리에이터들의 영상 제작에 있어서 효율성을 높인다.
- YouTube에 업로드 하기 전 필터링 과정을 거치며 각 영상과 음성에 대한 필터링이 실시되어 YouTube 가이드라인에 부적합한 구간이 발견되면 업로드하고자 하는 이용자에게 해당 구간을 알려주고, 가이드라인의 어떤 사항에 의해 부적합한지 알려준다.
- 현재 YouTube에 존재하는 다양한 가이드라인 중 선정성, 폭력성, 모방성(흡연, 욕설 등)에 대한 가이드라인을 충족시키는 지 중점적으로 확인하는 필터링을 실시한다.


### 2.2 연구/개발 내용

#### 2.2.1 영상 필터링

딥러닝 모델을 이용하여 영상 필터링을 진행한다. 필터링 되는 내용은 폭력적인 장면(만 19세 이상 영화에서 추출한 흥기에 찢리거나 베이는 장면, 만 18세 이상 게임에서 추출한 총, 흥기에 의해 피가 튀는 장면), 선정적인 장면(신체의 과도한 노출), 흡연 장면에 해당된다. 영상 필터링 시스템의 세부 모델은 다음과 같다.

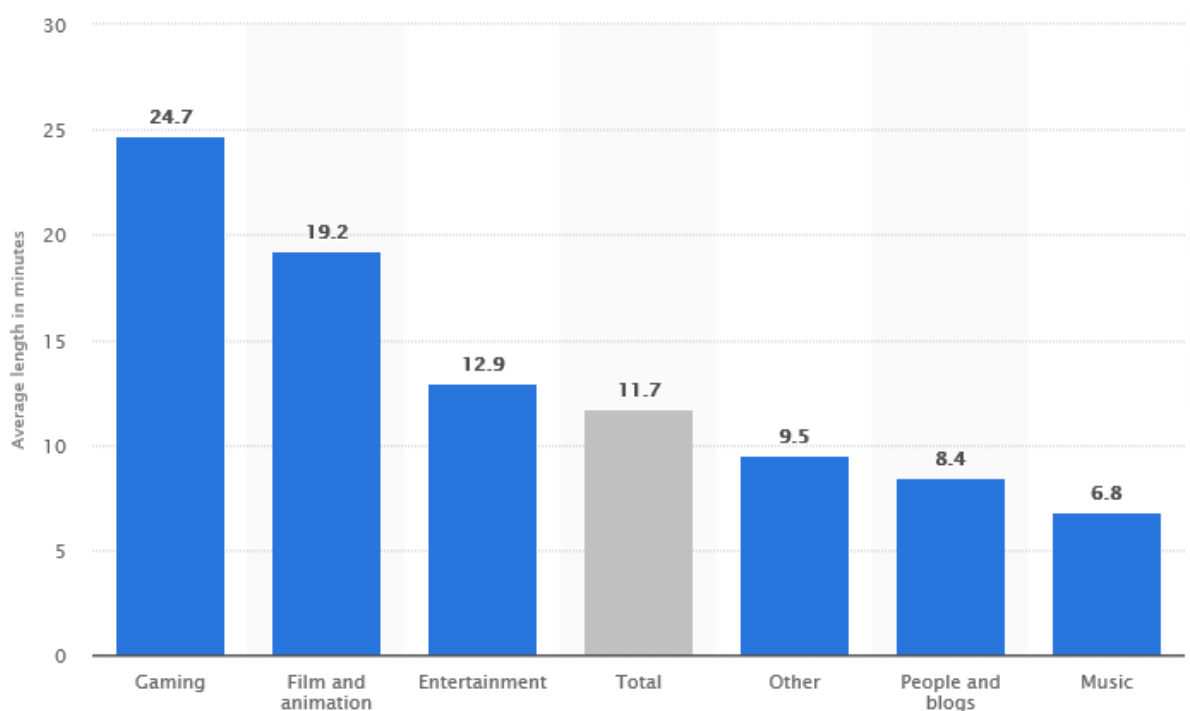
- 해당 데이터셋으로 학습된 딥러닝 모델을 준비한다. 딥러닝 모델은 CRNN을 이용해 설계된 ResNetCRNN을 사용하여 학습을 진행한다. 학습 데이터는 YouTube에 게시되어 있는 영상을 컴퓨터 화면 동영상 녹화 프로그램인 bandicam로 녹화하여 직접 구축한다.
- 영상 필터링의 작동원리는 영상의 길이에 따라 다른 방법을 사용한다. 영상의 전수조사를 기본 원칙으로 하지만 영상 필터링에 소요되는 최대 시간을 제한하기 위하여 일정 길



 <b>국민대학교</b> <b>컴퓨터공학부</b> <b>캡스톤 디자인 I</b>	<b>계획서</b>		
	<b>프로젝트 명</b>	인공지능을 이용한 영상 및 음성 필터링	
	<b>팀 명</b>	유하	
	Confidential Restricted	Version 1.2	20xx-MAR-25

이 이상의 영상은 구간을 나눈 후 랜덤샘플링을 진행한다. 이러한 경우에는 샘플링 된 구간에 한해서만 필터링을 실시한다. 이 후, 준비된 딥러닝 모델로 테스트하여 부적절한지에 대한 여부를 판단한다. 구간을 나누는 길이에 대해서는 서버환경에 따라 유동적으로 적용한다.

<사진 7>에 따르면 2018년 기준 Youtube에 업로드 되어있는 영상의 총 평균 시간은 11.7분이다.



<사진 7. Average Youtube video length as of December 2018 – statista >

## 2.2.2 음성 필터링

딥러닝 모델을 이용하여 음성 필터링을 진행한다. 필터링 되는 내용은 한국어로 표현된 욕설에 해당되며, 다양한 비속어를 포함한다. 사용되는 기술로 STT, FastText, Khaiii가 있다..

- STT(Speech To Text)기술을 이용하여 음성을 텍스트로 변환한다. STT는 현재 구글 API를 이용해서 진행하고 있으며, Kaldi 기반의 한국어 음성 인식기인 Zeroth와의 성능비교를 통해 높은 성능의 솔루션을 사용한다.

- 사전 훈련된 FastText 모델을 이용하여 텍스트와 욕설의 cosine similarity를 구한다.

 <b>국민대학교</b> <b>컴퓨터공학부</b> <b>캡스톤 디자인 I</b>	<b>계획서</b>		
	<b>프로젝트 명</b>	인공지능을 이용한 영상 및 음성 필터링	
	<b>팀 명</b>	유하	
	Confidential Restricted	Version 1.2	20xx-MAR-25

FastText는 Word2Vec을 기본으로 하지만 부분 단어들을 임베딩하는 기법으로, 주변에 있는 단어들을 가지고 중심에 있는 단어를 맞추는 cbow 모델 또는 중심에 있는 단어로 주변 단어를 예측하는 skip-gram 모델로 구현될 수 있다. 만약 cosine similarity가 70% 이상이라면 입력 텍스트는 Khaiii를 이용하는 다음 단계에서 필터링된다.

- 카카오에서 개발한 khaiii(Kakao Hangul Analyzer III) 형태소 분석기 모델을 통해 입력 텍스트를 형태소 단위로 분석한 후 필터하려는 욕설과 매칭이 되는 형태소를 찾는다. khaiii는 CNN 알고리즘을 이용하여 형태소 분석을 구현했고 신경망이 오분류할 경우를 대비하여 알고리즘 앞 단과 뒤 단에 사용자 사전 장치를 마련하여 사용자가 설정한 입력 어절은 설정한 값 자체로 분석하도록 한다.


### 2.2.3 AWS 및 웹서버

딥러닝 모델 학습과 모델의 영상과 음성 필터링에 있어 필요한 Amazon EC2 instance, Amazon S3, Amazon Lambda, Amazon Gateway API를 생성한다. 웹 페이지는 동영상의 업로드와 필터링 결과를 출력하고 필터링 결과에 대한 사용자의 의견의 수렴과 이에 따른 사용자의 구글 로그인 기능을 구현한다. 그리고 웹 페이지의 배포를 위해 웹서버를 구축한다.

- 동영상이 업로드 되는 공간, 업로드 된 동영상의 프레임 추출, 딥러닝 모델에 넣는 작업을 위한 Amzon S3, Amazon Lambda, Amazon Gateway API를 이용한다.

- AWS EC2 instance는 AWS Deep Learning AMI를 채택하여 딥러닝 모델의 학습을 진행한다. 모델 학습에 있어 필요한 데이터셋은 Amazon S3에 저장한다.

- 사용자 인터페이스를 만들기 위한 JavaScript 라이브러리인 React를 이용하여 웹페이지를 제작한다. 웹 구현을 위해 이용되는 React의 component 중 우리 프로젝트에 있어 많은 비중을 차지하는 것은 동영상 업로드 box, 영상 필터링 결과 box, 음성 필터링 결과 box이다. component마다 필요한 기능에 따라 적합한 라이브러리를 채택하고 CSS를 적용하여 레이아웃과 스타일을 정의한다.

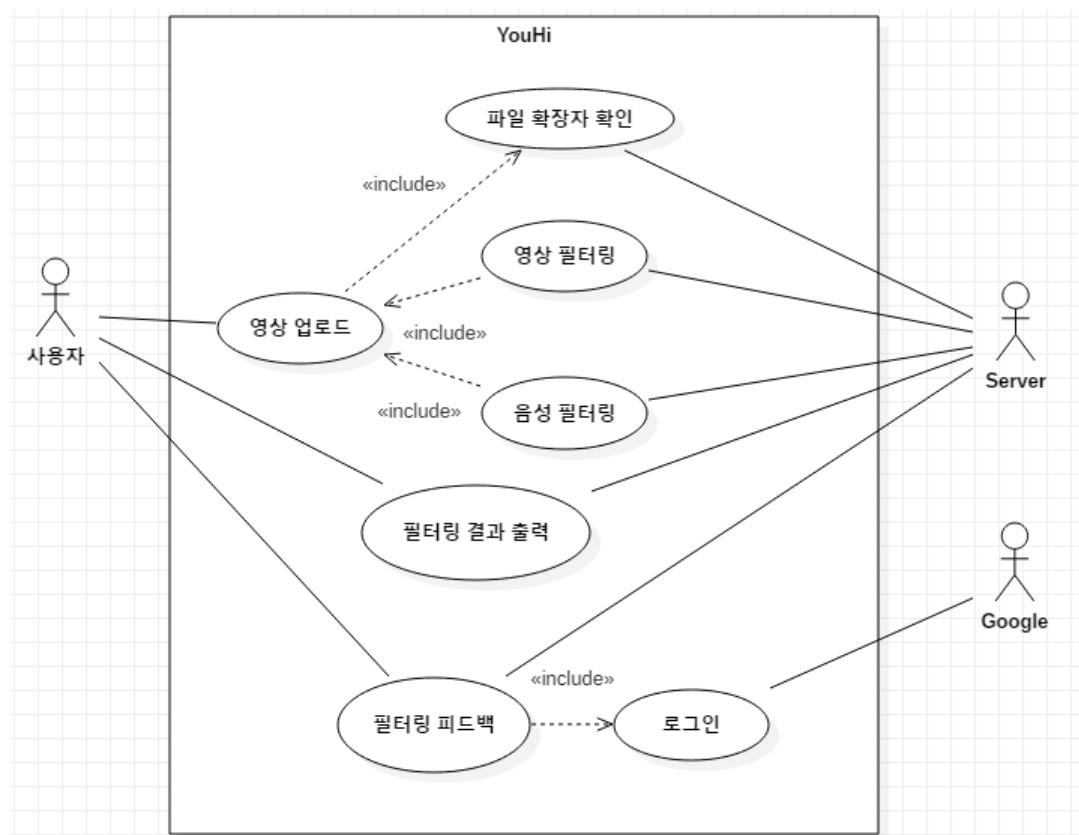
 <b>국민대학교</b> <b>컴퓨터공학부</b> <b>캡스톤 디자인 I</b>	<b>계획서</b>		
	<b>프로젝트 명</b>	인공지능을 이용한 영상 및 음성 필터링	
	<b>팀 명</b>	유하	
	Confidential Restricted	Version 1.2	20xx-MAR-25

- React로 구현한 웹 페이지 배포를 위해 apache server를 웹 서버로 선택한다.

## 2.3 개발 결과

### 2.3.1. 시스템 기능 요구사항

- 웹페이지는 동영상을 서버에 업로드한다.
- 웹페이지는 영상 및 음성에 대한 필터링 결과를 출력한다.
- 웹페이지에서 사용자의 의견을 수신할 수 있게 사용자가 구글 계정으로 로그인한다.
- 웹페이지는 로그인 된 사용자의 의견을 수신한다.
- 서버는 업로드 된 동영상의 확장자를 확인하고 저장한다.
- 서버는 업로드 된 동영상에서 음성 및 프레임을 추출한다.
- 서버는 추출된 음성 및 프레임을 딥러닝 모델에 넣어 그 결과를 저장한다.
- 서버는 로그인 된 사용자의 의견을 송신한다.



 <b>국민대학교</b> <b>컴퓨터공학부</b> <b>캡스톤 디자인 I</b>	<b>계획서</b>		
	<b>프로젝트 명</b>	인공지능을 이용한 영상 및 음성 필터링	
	<b>팀 명</b>	유하	
	Confidential Restricted	Version 1.2	20xx-MAR-25

### 2.3.2. 시스템 비기능(품질) 요구사항

사용성 : 사용자가 한 눈에 사용방법을 알 수 있도록 UI 를 제작한다

필터링 결과를 한 눈에 알아볼 수 있도록 가시화한다

신뢰성 : 이용자들이 필터링 결과를 납득할 수 있게 정확하게 필터링을 한다.

기능성 : 사용자들의 편의성을 위해 다양한 파일 확장자의 업로드를 가능하게 설계한다.

효율성 : 필터링에 소요되는 시간은 최대 10 분이 넘어가지 않도록 설계한다.

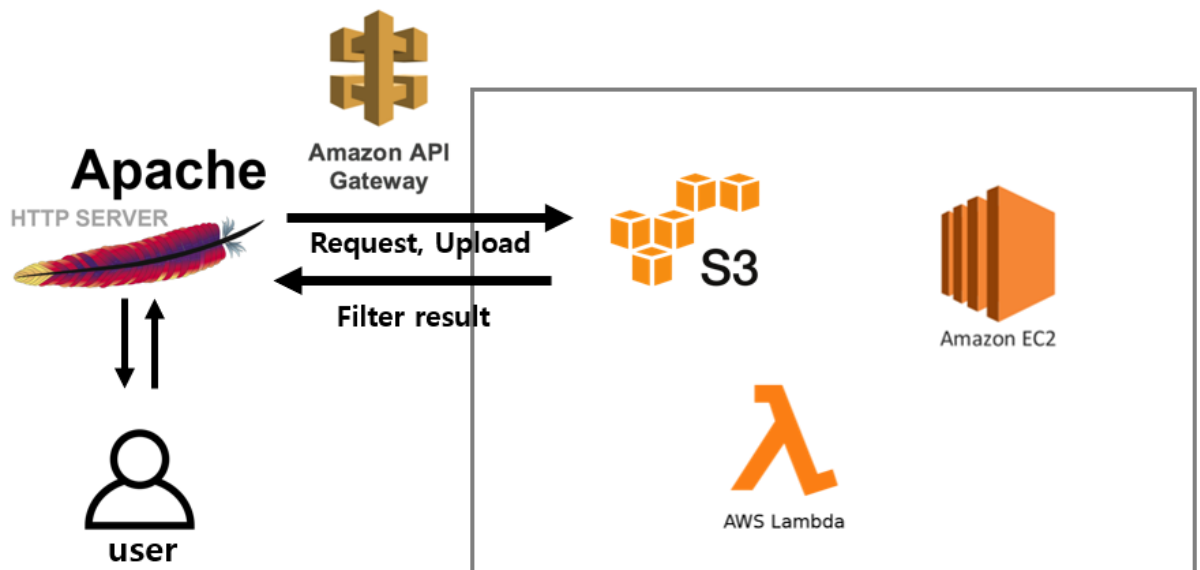
이식성 : 동영상 업로드 기능 및 필터링 결과 출력을 웹페이지에서 구현함으로써 다양한 플랫폼(스마트폰, 태블릿 PC 등)에서 사용가능 하게 한다.

서버를 AWS 를 이용해서 구현함으로써 다른 컴퓨터나 환경에서도 쉽게 구축할 수 있게 한다.

보안성 : 보안성을 위해 필터링이 끝난 데이터는 즉시 클라우드에서 삭제한다.

서버기능의 보안성을 위해 mp4,avi 형태의 파일만 업로드 가능하도록 제한한다.

### 2.3.3. 시스템 구조



 <b>국민대학교</b> <b>컴퓨터공학부</b> <b>캡스톤 디자인 I</b>	<b>계획서</b>		
	<b>프로젝트 명</b>	인공지능을 이용한 영상 및 음성 필터링	
	<b>팀 명</b>	유하	
	Confidential Restricted	Version 1.2	20xx-MAR-25


### 2.3.4. 결과물 목록 및 상세 사양

대분류	소분류	기능	형식	비고
프로그램	<i>프레임 분할</i>	영상 프레임을 추출한다	python	
	<i>음성 추출</i>	영상 파일에서 음성을 추출한다	python	
API	<i>GOOGLE STT API</i>	음성을 텍스트로 변환한다	API 모듈	
딥러닝 모델	<i>Kaldi - Zeroth</i>	음성을 텍스트로 변환한다	python	
	<i>ResNetCRNN</i>	연속된 이미지(프레임)을 학습시킬 수 있는 모델	python	
	<i>FastText 모델</i>	단어 임베딩(단어들간의 유사도를 확인하는 것)을 하는 모델	python / C++	
웹페이지	<i>영상 업로드</i>	Drag and Drop 과 Choose 로 동영상 업로드 할 수 있다.	Javascript	
	<i>결과 출력</i>	필터링된 결과를 출력한다.	Javascript	

## 2.4 기대효과 및 활용방안

필터링 시스템은 Youtube 의 가이드 라인에 맞춰서 영상과 음성을 검열한다. 이는 현재 신고 기반으로 이루어지는 수작업 검열 과정보다 효율적으로 작동된다. 업로드 이전에 일어나는 자동 검열 시스템이므로 다양한 기대효과와 활용방안이 있다.

1. 필터링 시스템을 적용함으로써 청소년들에 있어서 부적합한 영상들을 1 차적으로 검열할 수 있어 동영상 업로드 플랫폼에서 실시하는 신고 기반 검열 시스템보다 더 많은 영상을 검열할 수 있고, 작업량과 비용 측면에서 효율적이다.
2. Youtube 이외의 다양한 실시간 스트리밍 서비스(Twitch, Affreca TV 등)에서도 효과적으로 사용할 수 있다. 각 스트리밍 서비스들은 운영진이 직접 실시간 모니터링과 시청자들의 신고를 통해 제재가 가해지는데, 그 숫자가 매우 많아 효과적으로 이루어지지 않는다. 따라서 방송되고 있는 장면들을 실시간으로 딥러닝 모델에 넣어 검열할 수 있다.
3. 이러한 검열 작업이 원활하게 이루어지면, 현재 미성년자들이 무분별하게 시청하고 있는 영상들이 자동으로 검열됨으로써 청소년에게 유해한 매체물과 약물 등이 청소년에게 유통되는 것과 청소년이 유해한 업소에 출입하는 것 등을 규제하고 청소년을 유해한 환경으로부터 보호·구제함으로써 청소년이 건전한 인격체로 성장할 수 있도록 함을 목적으로 한다.

 <b>국민대학교</b> <b>컴퓨터공학부</b> <b>캡스톤 디자인 I</b>	<b>계획서</b>		
	<b>프로젝트 명</b>	인공지능을 이용한 영상 및 음성 필터링	
	<b>팀 명</b>	유하	
	Confidential Restricted	Version 1.2	20xx-MAR-25


4. 현재는 미성년자들을 대상으로 부적합한 영상들을 검열하는 시스템을 만들었지만, 점차 데이터셋이 많아지고 다양한 라벨에 대해서 학습을 시킨다면, 교통사고, 자연 재해, 길거리 싸움, 집단 구타 및 학대 등 검열 대상을 확대시켜 활용이 가능하다.
5. 또한 음성 필터링은 단순 욕설에 대해서만 진행되지만, 데이터셋 양을 확대시켜 학습시킨다면 본질적 속성을 토대로 한 악의적인 모욕(인종차별적 비방), 미성년자에게 수치심을 주거나 모욕감을 주는 등 검열 대상을 확대시킬 수 있다. 여기서 본질적 속성이란, 보호 대상 집단 신분, 신체적 특징 또는 성폭행, 가정폭력, 아동 학대 등의 피해자 신분이다.
6. 이 시스템이 적극적으로 활용된다면, 해당 동영상 업로드 플랫폼에 대한 사람들의 신뢰도와 인식의 향상에 도움이 된다. 또한 미성년자 뿐만 아니라 일반인에 대해서도 각종 범죄에 쉽게 노출되지 않게 함으로써, 많은 사회적 범죄를 예방할 수 있다.

## 3 배경 기술

### 3.1 기술적 요구사항

#### 3.1.1 프로젝트 개발

1. 웹페이지 개발 언어
  - Javascript(react 포함), Css
2. 웹페이지 개발 환경
  - Window 10, Visual Studio Code, Chrome
3. 서버 개발 언어
  - Javascript(웹 서버 및 AWS Lambda 구현), Python(영상 프레임 추출)
4. 서버 개발 환경
  - AWS EC2 인스턴스(Apache 웹서버), AWS S3, Linux(Ubuntu 16.04), Apache Web Server, AWS Gateway API, AWS Lambda
5. 영상, 음성 필터링 개발 언어
  - Python

 <b>국민대학교</b> <b>컴퓨터공학부</b> <b>캡스톤 디자인 I</b>	<b>계획서</b>		
	<b>프로젝트 명</b>	인공지능을 이용한 영상 및 음성 필터링	
	<b>팀 명</b>	유하	
	Confidential Restricted	Version 1.2	20xx-MAR-25

#### 6. 영상, 음성 필터링 라이브러리

- Pytorch(Video Classification 딥러닝 모델), Kaldi - Zeroth(STT 딥러닝 모델), FastText(욕설 판별 딥러닝 모델)

#### 7. 영상, 음성 필터링 개발 환경

- AWS EC2 인스턴스(Deep Learning AMI), Linux(Ubuntu 16.04), Jupyter Notebook

### 3.1.2 프로젝트 결과

#### 1. 웹페이지(Chrome)

- 동영상을 업로드하고, 필터링된 출력 결과를 확인한다. 업로드 파일의 확장자는 .avi, .mp4로 제한하며, 그에 대한 주의사항이 표기되어 있는 UI를 제공한다.

## 3.2 현실적 제한 요소 및 그 해결 방안


### 3.2.1 하드웨어

1. 모든 딥러닝 모델(Video Classification, Kaldi - Zeroth, FastText)을 학습시키기 위해서는 일반적인 로컬(데스크탑 PC, 노트북 등)에서는 RAM 메모리와 GPU 메모리의 양이 절대적으로 부족해 원활하게 진행되지 않는다. 따라서 우리는 AWS 의 Deep Learning AMI 를 이용해 학습 및 테스트를 진행할 것이다.

2. Internet Explorer 에서는 우리가 만든 웹페이지 기능이 정상적으로 작동되지 않을 수 있다. 각종 라이브러리가 제대로 적용되지 않기 때문이다. 이를 크리에이터들이 Chrome 에서만 접속하여 이용할 수 있도록, 주의사항을 작성해 해결한다.

### 3.2.2 소프트웨어

1. 단순 AWS EC2 인스턴스 도메인으로 여러 이용자가 대용량 File Upload 를 진행할 시 서버의 속도가 느려질 뿐만 아니라, 용량의 제약에도 큰 영향을 받아 문제가 생긴다. 따라서 AWS S3 를 통해 이를 해결한다. AWS S3 는 무제한 클라우드로서 파일을 업로드하거나 다운로드할 때 S3 을 이용하면 여러 제약사항이 사라진다.

 <b>국민대학교</b> <b>컴퓨터공학부</b> <b>캡스톤 디자인 I</b>	<b>계획서</b>		
	<b>프로젝트 명</b>	인공지능을 이용한 영상 및 음성 필터링	
	<b>팀 명</b>	유하	
	Confidential Restricted	Version 1.2	20xx-MAR-25


2. 음성 필터링을 진행할 때 Google API 만 사용해서는 Youtube 환경에서 음성을 텍스트로 바꾸는 STT 기술이 정확하게 작동하지 않을 수 있다. 따라서 Kaldi - Zeroth 딥러닝 모델을 통해 직접 학습시킴으로써 해결한다.

3. '빨간 원숭이의 해'(예: 2016 년)를 뜻하는 '병신년'과 시작점을 뜻하는 '시발점'등 욕설이 아님에도 불구하고, 욕설과 같은 글자들을 오분류하는 문제점이 발생한다. 따라서 FastText 딥러닝 모델을 이용하여, 문맥을 고려함으로써 해결한다.

### 3.2.3 기타

1. 모든 Youtube 가이드라인을 맞추어서 필터링을 진행하기에는 그 데이터셋 구축하는 시간과 노력이 상당하고, 단순 Video Classification 으로 필터링할 수 없어서, 가이드라인 중 4 개에 충족되게끔 목표를 설정했다.



 <b>국민대학교</b> <b>컴퓨터공학부</b> <b>캡스톤 디자인 I</b>	<b>계획서</b>		
	<b>프로젝트 명</b>	인공지능을 이용한 영상 및 음성 필터링	
	<b>팀 명</b>	유하	
	Confidential Restricted	Version 1.2	20xx-MAR-25

## 4 프로젝트 팀 구성 및 역할 분담

이름	역할
이태훈	<ul style="list-style-type: none"> <li>Software Project Leader</li> <li>Video Classification 모델 학습</li> <li>AWS 서버를 이용한 소프트웨어 설계</li> </ul>
이인평	<ul style="list-style-type: none"> <li>웹서버 구축 및 영상 데이터셋 구축</li> <li>FastText 모델링과 학습</li> </ul>
이주형	<ul style="list-style-type: none"> <li>STT(Speech To Text) API 적용</li> <li>웹페이지 구축</li> </ul>
김성수	<ul style="list-style-type: none"> <li>Kaldi - Zeroth 모델링과 학습</li> </ul>
김민재	<ul style="list-style-type: none"> <li>FastText 학습을 위한 데이터셋 구축</li> <li>웹페이지 디자인 UI 제작</li> </ul>

## 5 프로젝트 비용


항목	예산치 (MD)
AWS EC2 인스턴스 제작 및 웹서버 구축	5MD
웹페이지 제작 및 UI 설정	20MD
Video Classification 데이터셋 구축	10MD
FastText 데이터셋 구축	5MD
Video Classifcation 모델 학습	15MD
Kaldi - Zeroth 모델 학습	15MD
각 모듈 연동	10MD
합	80MD

 <b>국민대학교</b> <b>컴퓨터공학부</b> <b>캡스톤 디자인 I</b>	<b>계획서</b>		
	<b>프로젝트 명</b>	인공지능을 이용한 영상 및 음성 필터링	
	<b>팀 명</b>	유하	
	Confidential Restricted	Version 1.2	20xx-MAR-25

## 6 개발 일정 및 자원 관리


### 6.1 개발 일정

항목	세부내용	1 월	2 월	3 월	4 월	5 월	6 월	비고
요구사항분석	요구 분석 - Youtube 등 동영상 업로드 플랫폼들의 가이드라인 조사							
	데이터셋 조사							
관련분야연구	주요 기술 연구 - RNN + CNN 을 결합한 딥러닝 모델 - STT 기술을 위한 딥러닝 모델							
	관련 시스템 분석 - AWS 서비스들 중 필요한 서비스 분석							
	데이터셋 구축 - 영상 데이터셋 - FastText 를 위한 텍스트 데이터셋							
설계	시스템 설계 - AWS EC2 인스턴스 설계, AWS S3 설계 - 웹서버와 웹페이지 설계							
구현	영상 필터링 구현							
	음성 필터링 구현							
	웹페이지 구현							
테스트	시스템 테스트							

 <b>국민대학교</b> <b>컴퓨터공학부</b> <b>캡스톤 디자인 I</b>	<b>계획서</b>		
	<b>프로젝트 명</b>	인공지능을 이용한 영상 및 음성 필터링	
	<b>팀 명</b>	유하	
	Confidential Restricted	Version 1.2	20xx-MAR-25

## 6.2 일정별 주요 산출물

마일스톤	개요	시작일	종료일
계획서 발표	<p>프로젝트 구체화 및 개발 환경 완성 (GCC 설치, 기본 응용 작성 및 테스트 완료), 데이터셋 구축</p> <p><b>산출물 :</b></p> <ol style="list-style-type: none"> <li>1. 프로젝트 수행 계획서</li> <li>2. AWS EC2 인스턴스</li> <li>3. 계획서 발표용 PPT 및 영상</li> <li>4. 영상 및 음성 학습용 데이터셋</li> </ol>	2020-01-02	2020-03-20
설계 완료	<p>시스템 설계 완료</p> <p><b>산출물 :</b></p> <ol style="list-style-type: none"> <li>1. 시스템 설계 사양서</li> </ol>	2020-03-20	2012-03-27
1 차 중간 보고	<p>기능 영상 필터링 구현 완료</p> <ul style="list-style-type: none"> <li>- 영상 프레임 추출 및 샘플링 완료</li> <li>- ResNet CRNN 딥러닝 모델 학습 완료</li> </ul> <p>웹서버 구축 완료</p> <ul style="list-style-type: none"> <li>- AWS EC2 인스턴스에서 Apache 웹서버 구축 완료</li> <li>- AWS S3 생성 및 EC2 인스턴스와 연동</li> <li>- Javascript 와 Css 로 웹페이지 구축 및 배포 완료</li> </ul> <p><b>산출물 :</b></p> <ol style="list-style-type: none"> <li>1. 프로젝트 1 차 중간 보고서</li> <li>2. 프로젝트 진도 점검표</li> <li>3. 1 차분 구현 소스 코드</li> </ol>	2020-03-27	2020-04-04
2 차 중간 보고	<p>기능 음성 필터링 구현완료</p> <ul style="list-style-type: none"> <li>- Google STT API 로 텍스트 변환 완료</li> <li>- Kaldi - Zeroth STT 딥러닝 모델 학습 완료</li> <li>- FastText 로 문맥 파악을 통한 욕설 필터링 완료</li> </ul> <p><b>산출물 :</b></p>	2020-04-10	2020-04-24

 <b>국민대학교</b> <b>컴퓨터공학부</b> <b>캡스톤 디자인 I</b>	<b>계획서</b>		
	<b>프로젝트 명</b>	인공지능을 이용한 영상 및 음성 필터링	
	<b>팀 명</b>	유하	
	Confidential Restricted	Version 1.2	20xx-MAR-25

	1. 프로젝트 2 차 중간 보고서 2. 2 차분 구현 소스코드		
구현 완료	시스템 구현 완료 <b>산출물:</b> 1. 영상 및 음성 필터링 모델이 들어있는 AWS EC2 인스턴스 2. 동영상 파일을 업로드 할 수 있는 웹페이지	2020-04-24	2020-05-20
테스트	시스템 통합 테스트 <b>산출물:</b>	2020-05-20	2020-06-03
최종 보고서	최종 보고 <b>산출물:</b> 최종보고서	2020-06-03	2020-06-12

### 6.3 인력자원 투입계획

이름	개발항목	시작일	종료일	총개발일(MD)
이태훈	AWS 서버 및 영상 업러닝 구축	2020-03-16	2020-05-20	80
이인평	웹서버 구축 및 텍스트 업러닝 학습	2020-03-16	2020-05-20	80
이주형	STT API 적용 및 웹페이지 제작	2020-03-16	2020-05-20	80
김성수	Kaldi - Zeroth 모델링 및 학습	2020-03-16	2020-05-20	80
김민재	웹페이지 디자인 UI 제작	2020-03-16	2020-05-20	80

### 6.4 비 인적자원 투입계획

항목	Provider	시작일	종료일	Required Options
AWS	국민대학교	2020-03-16	2020-06-12	
노트북 (5 대)	개인 소유	2020-03-16	2020-06-12	
데스크탑 (1 대)	개인 소유	2020-03-16	2020-06-12	

 <b>국민대학교</b> <b>컴퓨터공학부</b> <b>캡스톤 디자인 I</b>	<b>계획서</b>		
	<b>프로젝트 명</b>	인공지능을 이용한 영상 및 음성 필터링	
	<b>팀 명</b>	유하	
	Confidential Restricted	Version 1.2	20xx-MAR-25

## 7 참고 문헌

번호	종류	제목	출처	발행년도	저자	기타
1	기사	[NDC2018] "스 111 발" 도 잡아내는 욕설 탐지기, 딥러닝으로 만들기	INVEN	2018	정필권	
1	논문	청소년 매체이용 및 유해환경 실태조사	여성가족 부	2018	여성가 족부	
2	논문	청소년의 스마트폰 중독과 인터넷 유해매체 노출이 성폭력 가해행동에 미치는 영향	청소년정 책연구원	2017	청소년 정책연 구원	
1	규정집	방송언어 가이드라인	방송통신 심의위원 회	2019	방송통 신심의 위원회	
1	법령	청소년보호법	국가법령 정보센터	2018	여성가 족부	