
11D

클라우드도 이런 것도 가능해? 유해 콘텐츠 필터 시스템 구축기

CloudEngineer Team (김학철)

요약

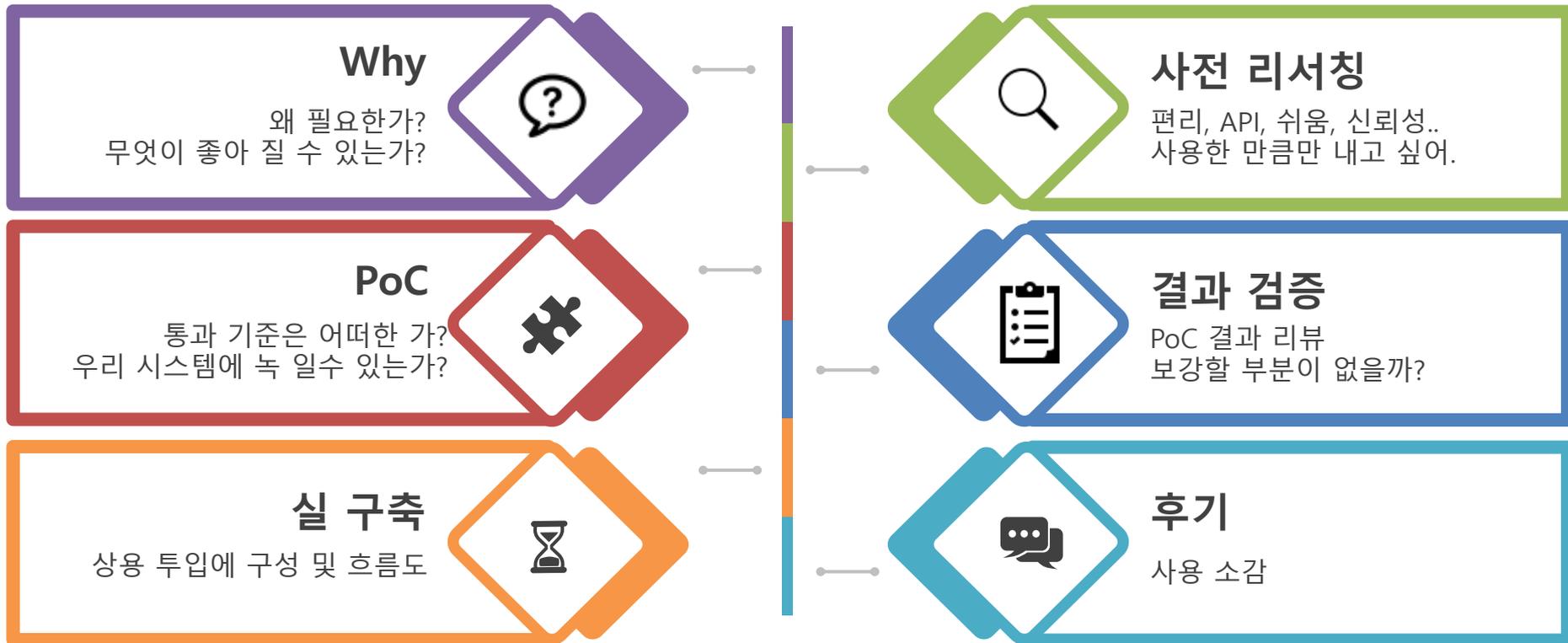
수작업 유해 리뷰 콘텐츠 분류 힘들지 않을까?

실수는 없을까?

얼마나 빠르게 차단할까?

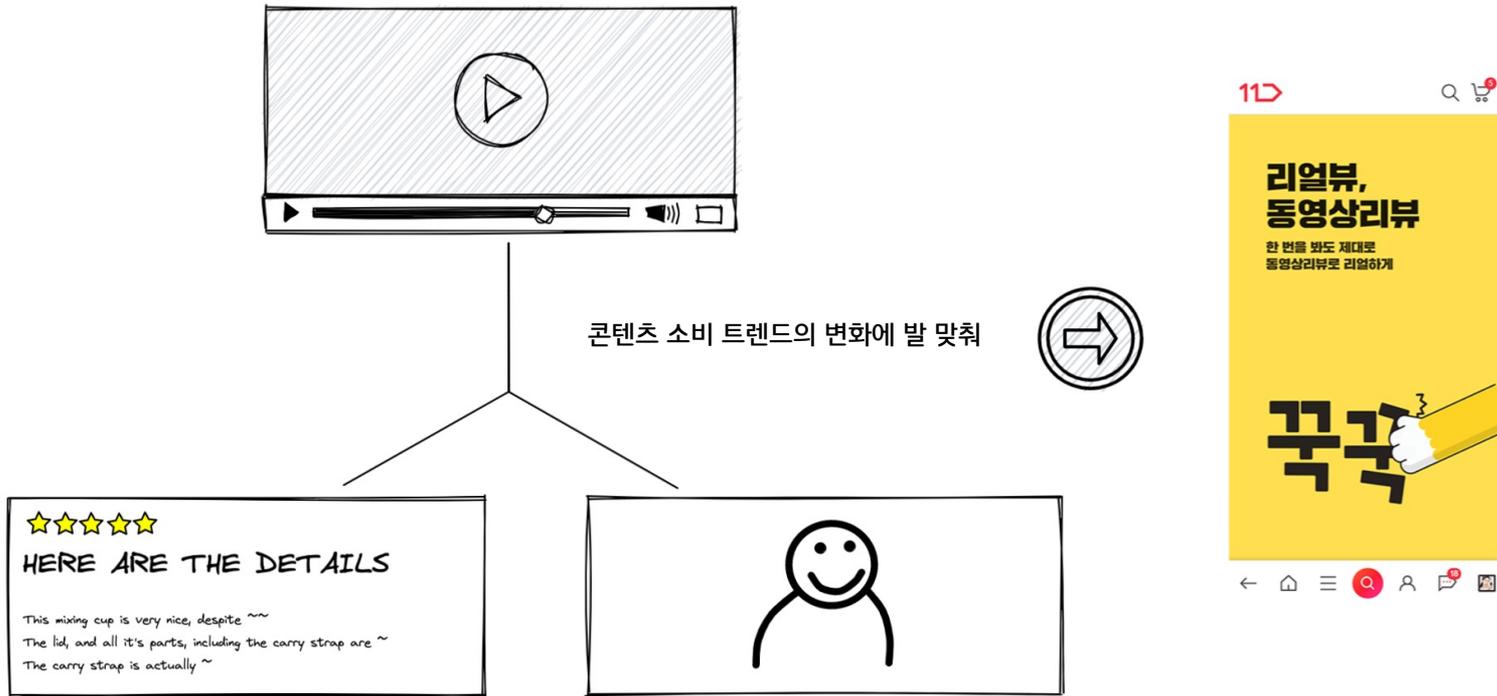
AWS 활용하여 쉽고 빠르게 적용한 유해콘텐츠 필터 시스템 구축기를 공유 하고자 합니다.

Contents



Why

리뷰 동영상 서비스 준비

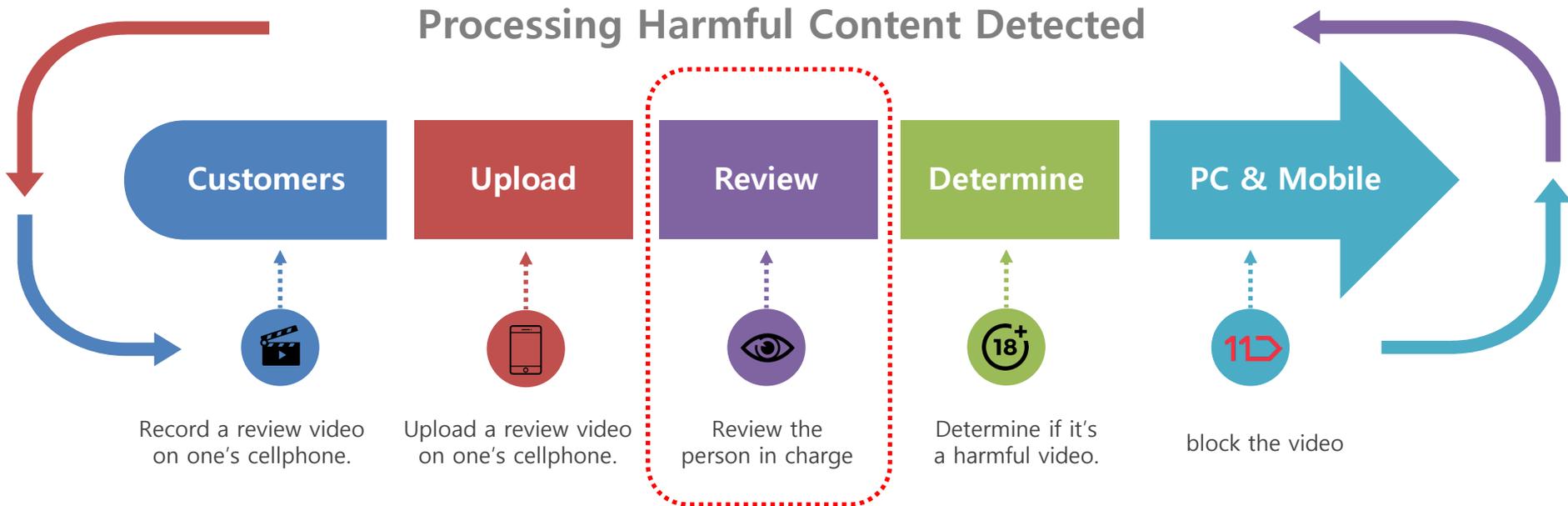


즐겁고 재미있는 경험을 제공

종합적인 트렌드에 발맞춰 사용자에게 좀 더 리얼하고 공감할 수 있는 리뷰 제공

실 구매한 사람이 직접 사용기를 촬영한 **영상**을 업로드하여 소비자에게 **신뢰**할 수 있는 정보를 제공

간략 업무 예측 흐름



영상 검수는 누가 하지?

이 부분을 누군가 대신 해주면 좋지 않을까? 최소한 1차 검증이라도 해주면 좋겠다.

대체제를 넣어, 얻을 것 측정

1 편리성

사람의 눈으로 보는게 최고 좋지만,
머신 러닝을 이용한 자동 프로세싱
이용하면 좀더 편리해 질것 같다.

2 안정성

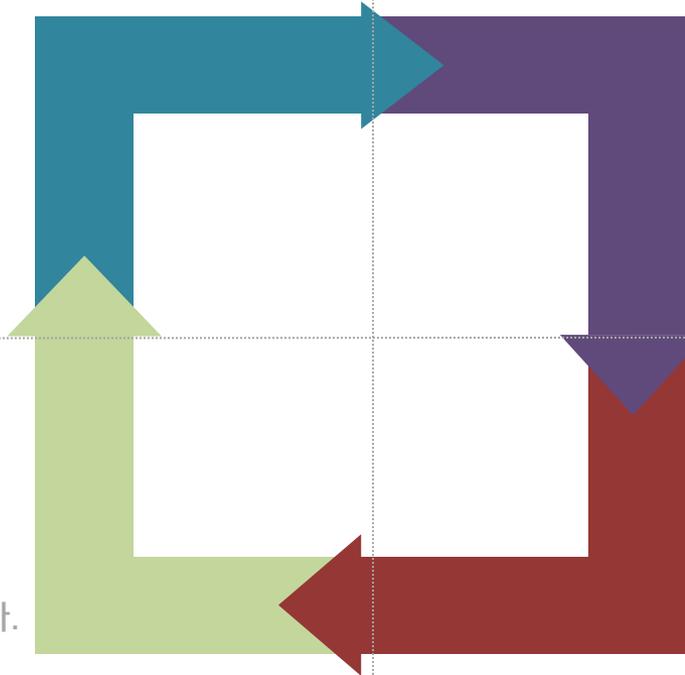
기계를 활용하여 필터링의 망을
좀 더 꼼꼼하게 만들어 수 있을 것
같다.

4 신뢰 받는 업체

결국, 신뢰성 있는 업체로 평가 된다.

3 법규 준수

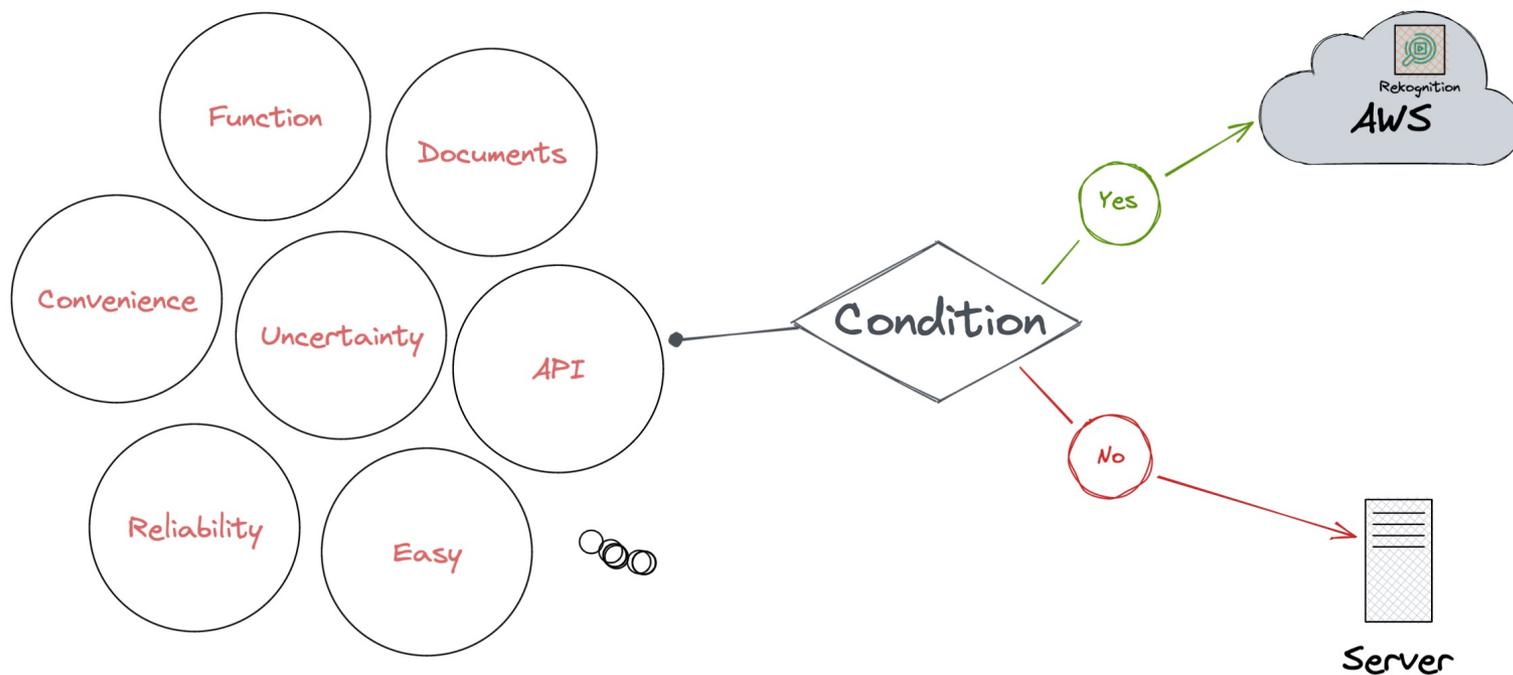
최소한의 대응 방안을 마련할 수 있다.



사전 리서칭

검색 조건

기능, 문서, 편함, 신뢰, 쉬움, API, 빠른 구축...
사용한 만큼만 내고 싶어...



PoC

PoC 방향

이 발표자료에서 다룰 건, 신뢰성 / 기능 전달

01

신뢰성

02

기능

03

비용

신뢰성 측정을 위한, 통과 기준 세우기

0 ~ 100%

- 선정성이 높을 때 = 100에 가까움
- 선정성이 낮을 때 = 0에 가까움

선정적인 동영상

$\geq 99\%$

애매한 동영상

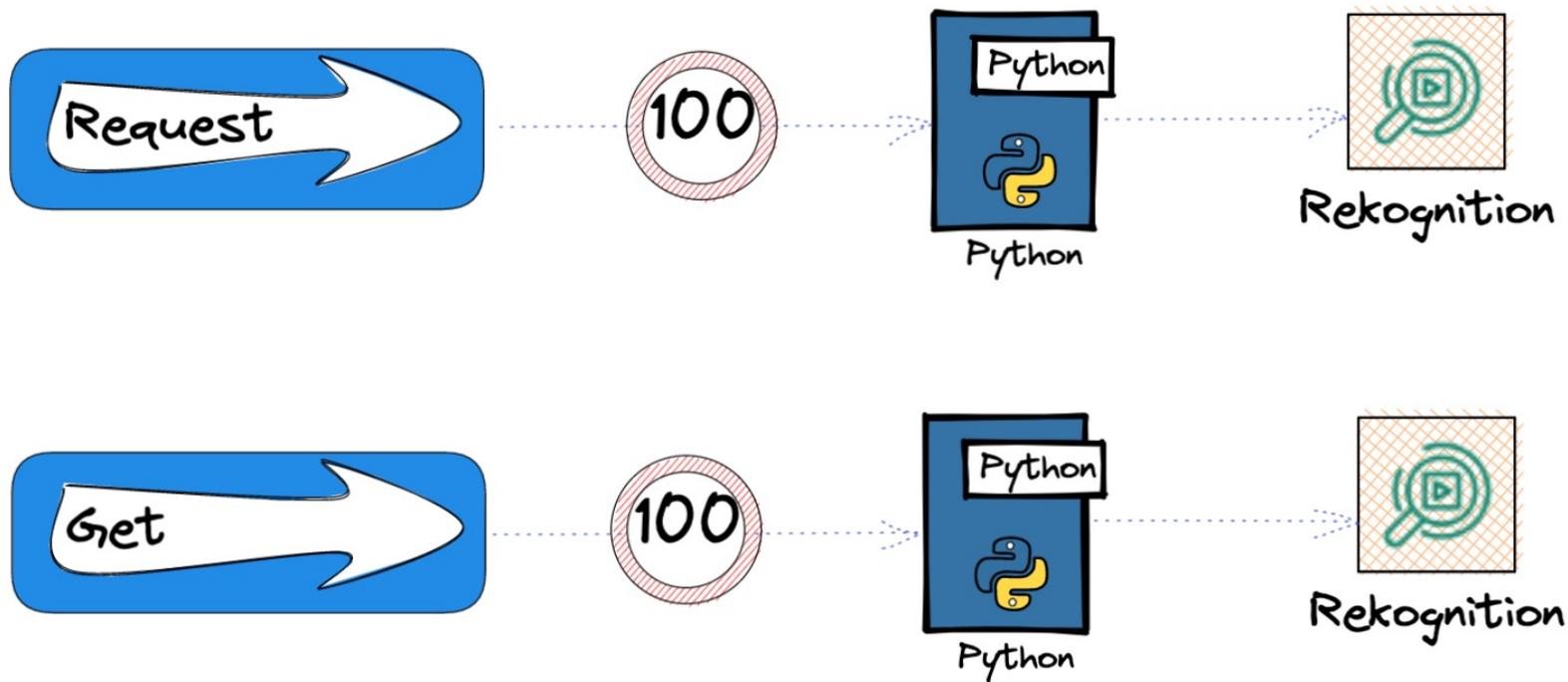
$\geq 80\%$

비 선정적인 동영상

$\leq 20\%$

분석 요청 및 결과 취합

각 기준에 맞춰 Python + Boto3 를 이용해 분석 요청 및 결과 취합



구축에 대한 기능 검토

11번가 시스템에 적용하기 위한 아키텍처링 검토 & 구성



결과 검증

필터 검증 결과

애매해진 결과

기준 수치를 낮추면, 선정적인 동영상을 높은 비율로 필터링 할 수 하지만, 비 선정적인 동영상의 필터 오탐도 높아지기 때문이다.

그래도, 높을 비율로 필터링 할 수 있다라는 것에 좀 더 무게를 두어 아키텍처에 대한 기능 검증도 시행

선정적인 동영상	
	
검출 기준	결과
>= 60%	100%
>= 80%	96.6%
>= 90%	93.3%

애매한 동영상	
	
검출 기준	결과
>= 70%	90%
>= 80%	80%
>= 90%	70%

비선정적인 동영상	
	
검출 기준	결과
>= 60%	20%
>= 70%	16%
>= 80%	5%

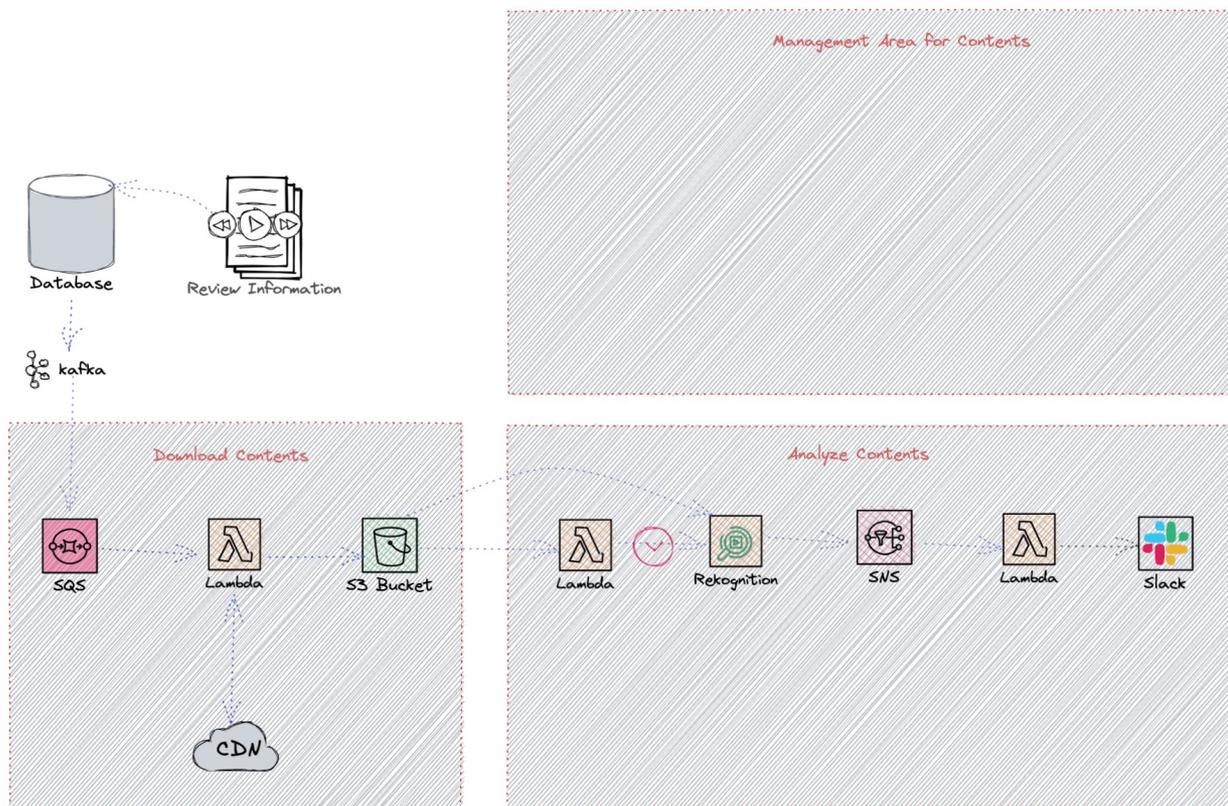
아키텍처에 대한 기능 검증

기본 기능에는 이슈 없음

기본 동시성보다 많은 건수가 들어 온 경우의 대안

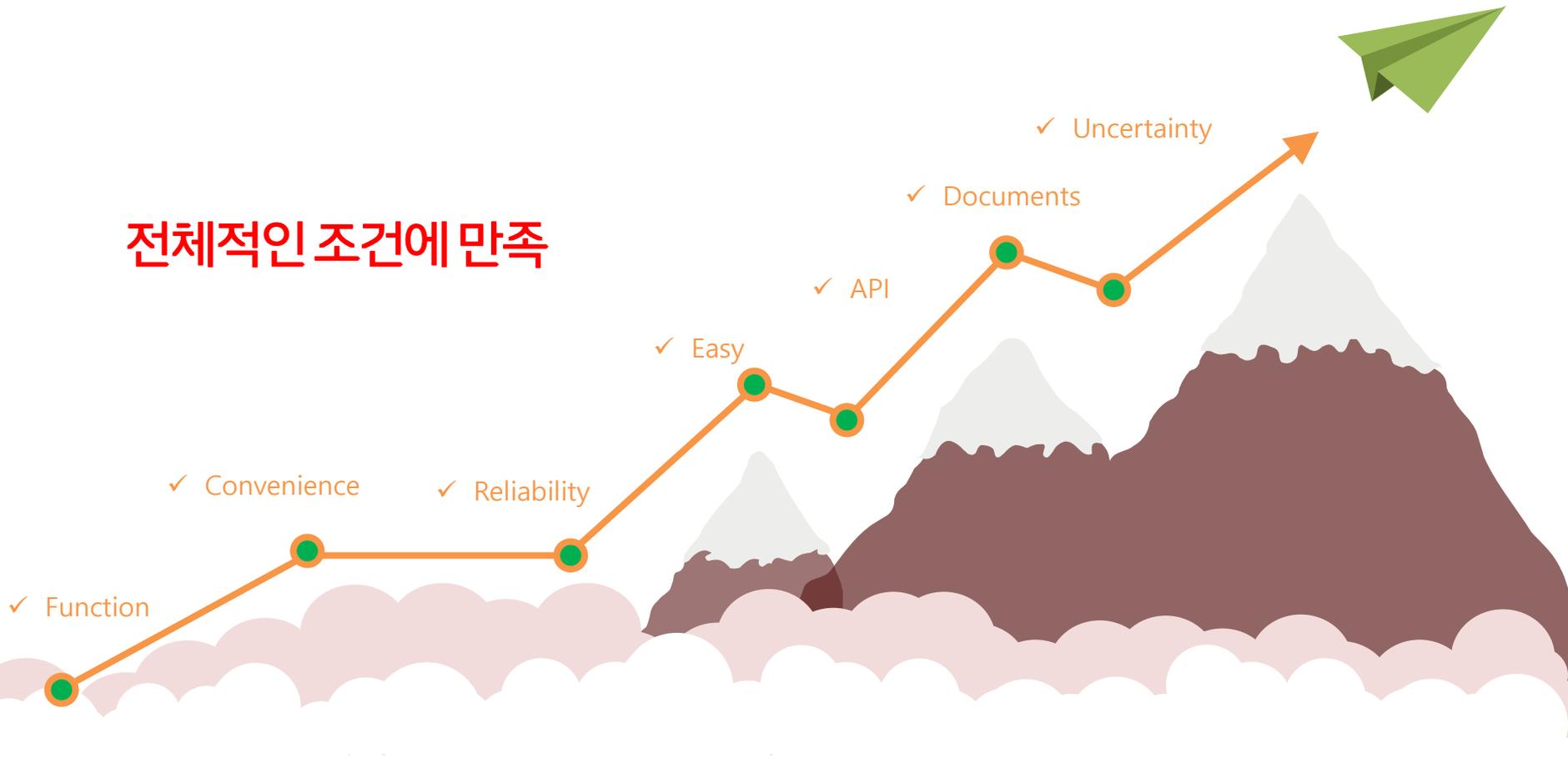
리뷰 동영상에 대한 비노출 제어 API 필요

분석 내용에 대한 히스토리 관리



종합 검토 결과

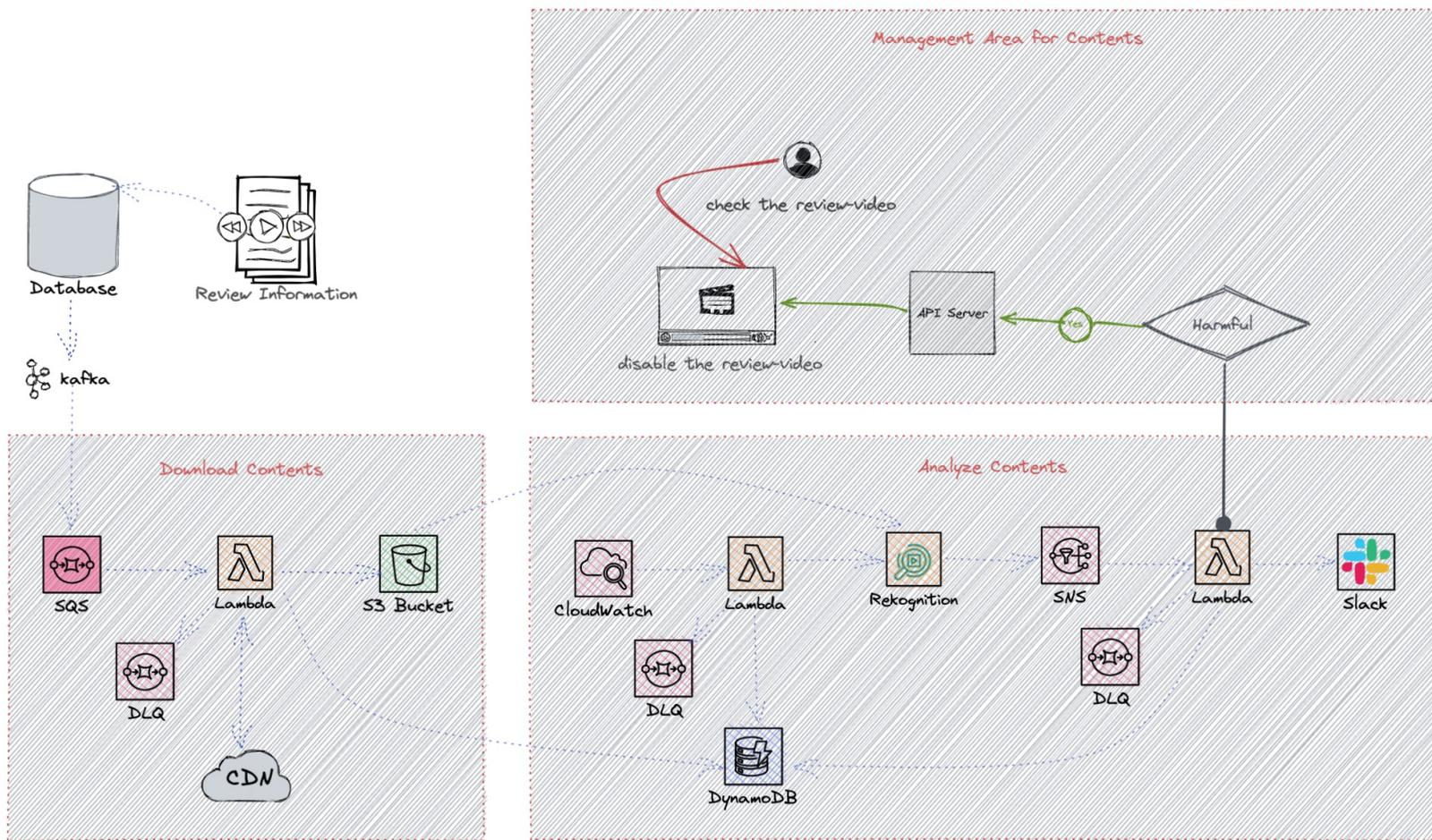
전체적인 조건에 만족



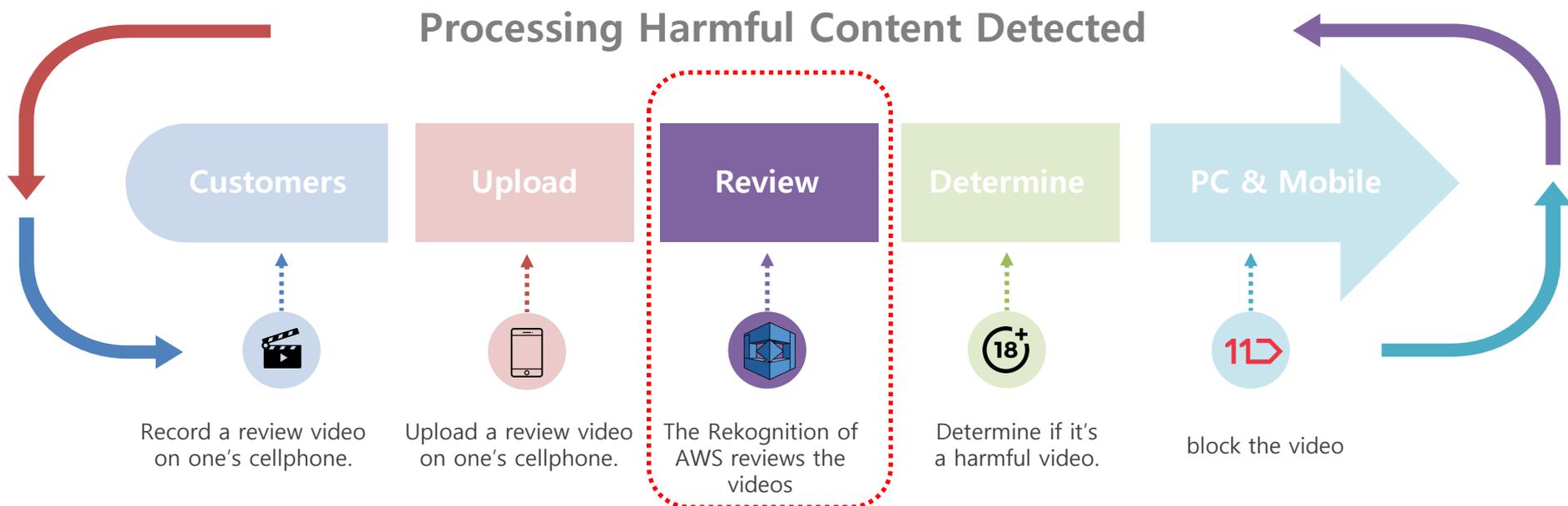
단, 실 상용에 투입하기 위해서는 **보안할 점**을 넣어야 할 필요가 있음.

실구축

상용 구축 아키텍처



최종 서비스 흐름도



Review 하는 부분에, AWS Rekognition 으로 처리

사용 소감

사용 소감

 좋았던 점		아쉬운 점 
1 실제 유해한 콘텐츠 발견 & 조치		트래픽이 높아짐에 따라 정상적인 건에 대한, 오탐 건수도 많아 짐 1
2 자동으로 선 조치		운영의 묘가 들어가야 한다는 점 2
3 SaaS 시스템의 낮은 난이도		트래픽이 많아 질수록 비용도 높아 짐 3

Q&A

E-mail) spakhc@sk.com

Thank you