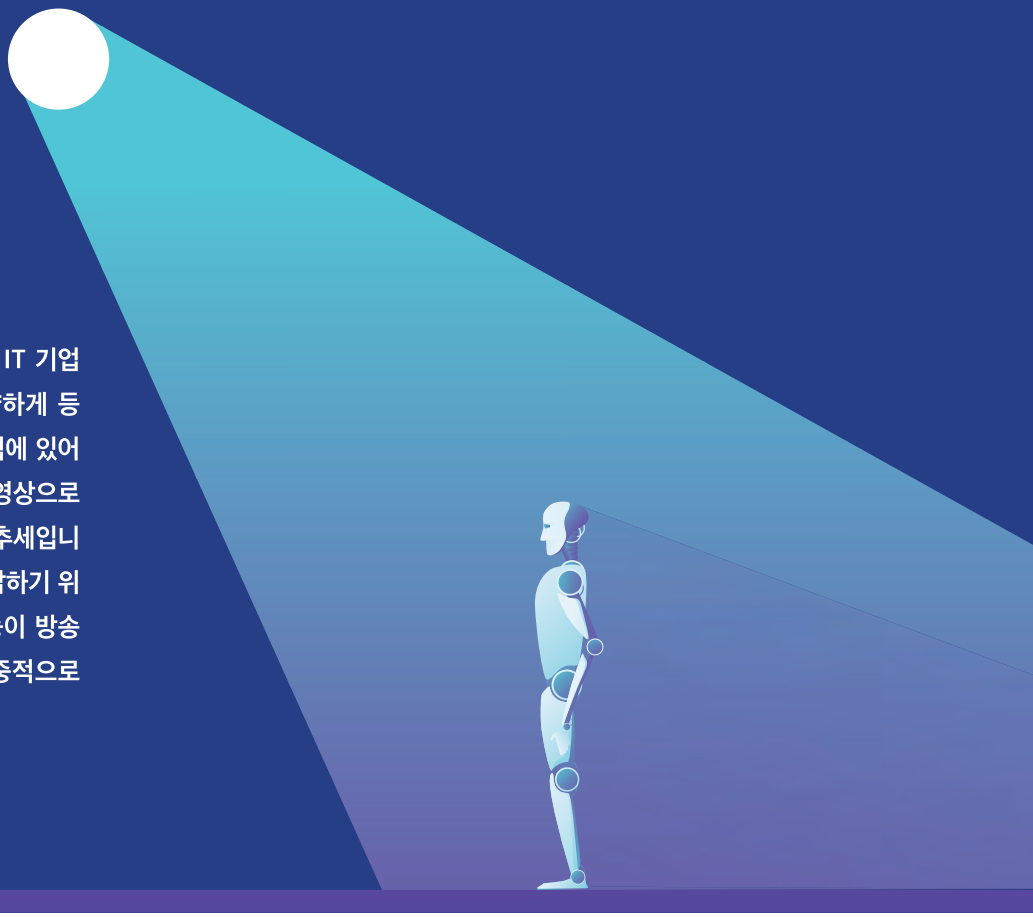
BBC, 인공지능을 제작현장으로 끌어들이다

글. 최인혜 前 EBS 미래교육연구소 연구원



최근 몇 년간 구글, MS, 페이스북 등 글로벌 IT 기업 주도의 인공지능 기반 기술 활용사례가 다양하게 등장했습니다. 최근까지 음성인식과 이미지 분석에 있어 괄목할만한 성장세를 보여 왔는데, 이제는 동영상으로까지 인공지능의 적용 범위를 확대하고 있는 추세입니다. 방송사의 움직임은 어떨까요? 이 질문에 답하기 위해 이번 글에서는 BBC 사례를 통해 인공지능이 방송 제작 환경에서 어떻게 적용될 수 있을지 집중적으로 살펴보겠습니다.

들어가며

BBC R&D는 200명이 넘는 전문가로 구성된 BBC의 연구 조직입니다. 연구진에는 엔지니어뿐 아니라 디자이너, 방송 제작자 및 혁신 전문가 등이 협업하는 구조로 알려져 있습니다. BBC R&D의 블로그를 살펴보면, IP, 클라우드, 인공지능 등 다양한 방송기술 연구를 꾸준히 지속해오고 있음을 알고 있습니다. 이 글에서 살펴보는 AI in Production은 논문명이면서도 동시에 BBC R&D의 프로젝트명이기도 합니다. 이 논문을 통해 연구진은 IBC 2018에서 베스트 논문상을 받기도 했습니다.

BBC wins best paper

BBC R&D took the prize for the best technical conference paper.

The publication described advances in the use of artificial intelligence (AI) and machine learning for live production, to the extent that it could automatically create finished programmes. The trophy was collected by project engineer Craig Wright who had presented the paper earlier in the day.

"The point of the work is to allow coverage of more events, to reach places we otherwise could not reach," BBC R&D project lead Mike Evans said.

"With conventional production we cover only about six of the nearly 100 places music is performed at the Glastonbury Festival, for example, or just a tiny fraction of the 50,000 performances in 300 venues at the Edinburgh Fringe.

"With this work we can reach many more of these, and do so with production techniques which are much less intrusive for the event itself," he explained.

"This technology will be suitable not just for the BBC, but for a whole range of use cases, like minor sports which need to increase visibility, and even vloggers who want to improve their online presence."

Executive Producer for Technology Sessions at the IBC Conference Dr Nick Lodge (left) and BBC R&D project engineer Craig Wright

2018

AI in Production 프로젝트란?

BBC 연구진의 인공지능 기반 제작 프로젝트 'AI in Production'은 AI를 활용해 기존 미디어 제작 사업을 어떻게 변화시킬 수 있는지를 파악하는 데에 목적을 두고 있으며, 이를 위해 제작부서와 협력해 머신러닝의 도움을 받을 수 있는 부분을 탐색하고, 제작 과정의 효율성을 높일 수 있는 알고리즘을 개발한 결과를 보여줍니다.

연구목적

BBC R&D의 이 연구는 스튜디오 촬영보다는 야외 촬영에서의 고민에서부터 출발합니다. 기본적으로 야외 촬영은 실시간 중계든 녹화 방송이든지 간에, 거대한 중계차량부터 여러 대의 카메라, 믹서, 촬영감독 등 충분한 장비와 인력이 필요합니다. 하지만 넓은 야외 이벤트에 비해 촬영으로 담아낼 수 있는 부분(시청자에게 제공되는 장면)은 매우 제한적입니다. 특히 BBC가 있는 영국은 세계 최대의 축제로 불리는 에든버러 페스티벌이 매년 개최되는데, 300개의 장소에서 5만 개의 공연이 열리지만 극히 일부만이 시청자에게 전달될 뿐입니다. 글래스톤 베리라는 유명 음악 축제에서도 BBC는 100곳 중 단 6곳만 촬영을 했다고 밝혔습니다. 기존 야외촬영 시의 한계-비용과 인력의 문제-때문입니다.

BBC R&D는 이러한 야외 라이브 이벤트 촬영 시에 발생하는 복잡성과 확장성의 한계를 극복하고, 인공지능 기반의 자동화된 방송 제작 환경을 만들기 위해 2015년부터 관련 연구를 수행했습니다. 먼저 IP 기반의 저비용 고품질의 비디오 캡처 장치를 개발하기 시작했고, 2017년부터는 인공지능과 머신러닝을 적용한 연구를 수행하기에 이르렀습니다. 페스티벌, 스포츠 경기 등 대형 이벤트를 실시간으로 시청자에게 전달해야 하는 방송사 입장에서 자동화와 효율성을 추구하는 것은 어찌 보면 당연한 고민이지만, 그 고민을 해결하려는 방법을 찾고 있다는 점에서 주목할 가치가 있다고 여겨집니다.

AI in Production 프로젝트는 인공지능 기반의 제작 자동화를 통해 라이브 이벤트의 적용 범위를 확대하는 것에 있습니다. 야외 촬영 시 넓은 지역에서 동시다발적으로 발생하는 이벤트를 놓치지 않기 위해, 저비용으로 고효율의 결과물을 만들어낼 수 있는 인공지능 기반의 자동화 말입니다. 따라서 BBC R&D가 목표로 하는 제작방식은 이벤트가 일어날 넓은 지역은 고정된 UHD 카메라를 설치해 촬영하고, 이후 샷 추출과 컷 편집은 'Ed(에드)'라는 자동화시스템이 대신하는 형태입니다. 이것이 가능한 이유는 이미 BBC 연구진은 UHD 해상도로 촬영한 와이드샷을 HD 화질로 잘라내 더 많은 가상의 샷을 만들어내는 실험에 성공했기 때문입니다. 따라서 인공지능 시스템인 'Ed'를 통해서는 분절된 가상의 샷들을 상황에 맞게 자동으로 편집하는 역할을 수행하도록 하는 것에 목표를 두고 있습니다.

Ed - 자동화를 위한 규칙 기반의 인공지능 시스템

보통 글로벌 기업에서 특정 서비스에 AI를 적용할 때는 다음의 3단계를 거친다고 합니다. 먼저, 현재 프로세스를 리뷰하고(Current State Assessment), AI를 통해 자동화하거나 개선할 수 있는 부분을 발견하고(Target Operating Model), 실제로 작은 규모의 실험을 통해 그를 증명하고(Proof of Concept), 목표에 도달하기 위한 계획을 세워 실행해가는 것입니다. BBC 연구진이 라이브 이벤트를 캡처하고 편집하기 위해 구축한 Ed는 위 단계 중 세 번째, PoC 단계에 해당합니다. 따라서 본 고에서 소개하는 내용은 BBC 연구진이 Ed로 AI 시스템을 테스트해본 작은 실험이자, 목표에 이르기 위한 수행 결과이기도 합니다.

간단히 Ed 시스템과 규칙에 대해 알아볼까요? Ed는 라이브 이벤트를 캡처하고 편집하기 위한 인공지능 시스템으로, 야외무대를 와이드샷으로 촬영하는 UHD 카메라에서 하나 또는 그 이상의 비디오를 인풋(input)으로 가져

오고, 샷 프레임 결정, 편집 순서 및 선택을 자동으로 수행합니다. Ed는 야외 공연 촬영 형식의 확대를 위해 개발된 만큼, PoC 단계에서의 실험 역시 페스티벌에서 흔히 볼 수 있는 '라이브 패널 쇼'와 비슷한 환경에서 제작되었습니다. 5명의 패널이 앉아있고 이야기를 주고받는 형식이며, 현장과는 다르게 실내 촬영으로 이루어졌습니다.
연구진은 Ed를 '규칙 기반 시스템'이라고 명명했습니다. 자동화를 위해서는 일련의 작업에 대한 순서와 규칙이 먼저 정의되어야 하는데, Ed의 규칙은 방송제작에 참여하는 이용자 경험 연구 및 실제 편집진의 권고사항에 기반을 두었습니다. 이렇게 도출된 Ed의 몇 가지 규칙 중 일부를 소개합니다.

샷 프레임 규칙

- 샷의 포커스를 중앙 또는 3번째 선에 위치시켜라(3분할 법칙).
- 사람이 바라보는 방향으로 루킹룸(looking room)이 있어야 한다.

샷 시퀀싱 및 샷 선택 규칙

- 일반적으로 화자는 샷을 유지
- 다양화를 위해 원샷과 투샷 전환
- 리액션을 위한 가끔의 화면 전환 등

형상 추출(Feature Extraction)

Ed는 영상에서 얼굴 감지 및 추적, 얼굴 표식 및 자세 추정 등을 사용해 여러 가지 특징을 추출합니다. 사람들이 각 프레임 중 어디에 있는지, 패널이 바라보고 있는 방향과 말할 때를 나타냅니다. [그림 1]의 왼쪽 사진은 Ed의 예시 프레임으로, 얼굴 탐지 경계 상자(녹색), 얼굴 랜드마크(파란색), 머리 자세 투영(head pose projection, 빨간색)을 인식하는 모습을 보여주고 있습니다.

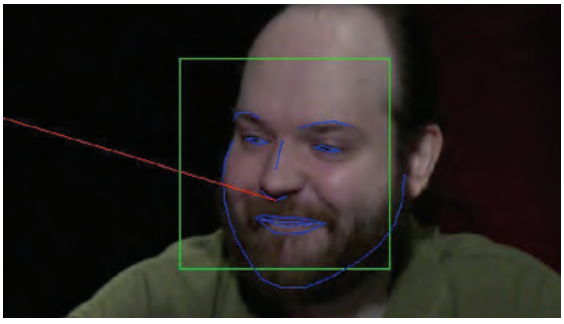


그림 1. 얼굴 탐지 경계 상자(녹색), 얼굴 랜드마크(파란색), 머리 자세 투영(head pose projection, 빨간색)

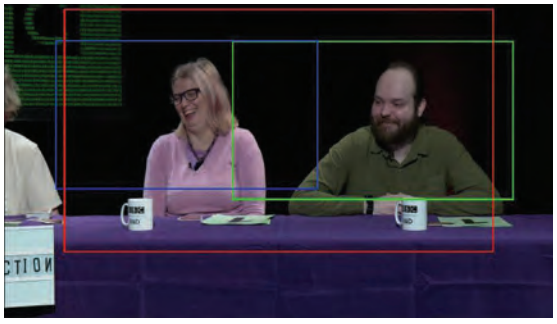


그림 2. 패널샷을 보여주는 3가지 후보 장면(crop)으로 표시된 카메라 뷰: mid-close shots(녹색 및 청색) 및 mid shot(빨간색)

연구진은 패널 샷을 결정할 때도 누구에게 초점을 맞출 것인가를 두고 고민했습니다. 패널의 얼굴과 자세를 추정해 각 조합에 맞는 와이드(WS), 미드(MS) 및 클로즈업(CU) 샷을 설정했고, [그림 2]와 같이 각 샷의 설정 범위에 맞게 패널이 위치하도록 했습니다.

샷 시퀀싱(Shot Sequencing)

시퀀스(Sequence)는 샷의 변화가 '언제' 일어날지를 정의하는 과정입니다. Ed에서의 시퀀스 방법은 샷 변경사항을 음성 이벤트(즉, 사람들이 말을 시작하거나 멈추는 경우)에 가깝게 스케줄하도록 설정했습니다.

샷 선택(Shot Selection)

샷 선택이란, 프레임의 장면 중 하나를 시퀀스의 각 샷 경계 사이의 구간에 '할당'하는 과정입니다. 연구진은 UX 인터뷰를 통해 장면 선택에 대한 전문가 의견을 다음과 같이 수렴했습니다.

- ☑ 일반적으로는 화자를 잡는다.
- ☑ 가끔 리액션 샷으로 장면을 전환한다.
- ☑ 가끔 배경샷(establishing shot)으로 장면을 전환한다.

라이브 패널 쇼에서 진행자와 패널은 일단 자리를 잡으면 움직이지 않습니다. 따라서 주어진 촬영 범위 내에서 편집을 결정하는 요소는 의외로 단순합니다. ▲장면 내에서의 연설량 ▲장면의 인원수 ▲장면 종류(클로즈업, 미드샷, 와이드샷) ▲장면 사용 시기까지입니다.

촬영된 수많은 영상을 두고 샷을 선택할 때, 위의 기준을 따르지만 여기서도 몇 가지 우선순위 또는 선호되는 상황이 있습니다. 우선 등장하는 사람은 적고, 더 많은 음성이 포함된 장면이 선호됩니다. 반대로, 음성이 감지되지 않을 때는 많은 사람들이 등장하는 와이드샷 등이 선호됩니다. 최근에 쓰이지 않았던 장면은 항상 선호되며, 생성된 샷 시퀀스의 각 샷은 시간 순서대로 선택됩니다. Ed는 해당 구간에서 동영상 콘텐츠로 이용할 수 있는 모든 프레임 장면과 가장 좋은 점수를 받는 장면을 고려합니다. 그 방법은 [그림 3]과 같습니다.

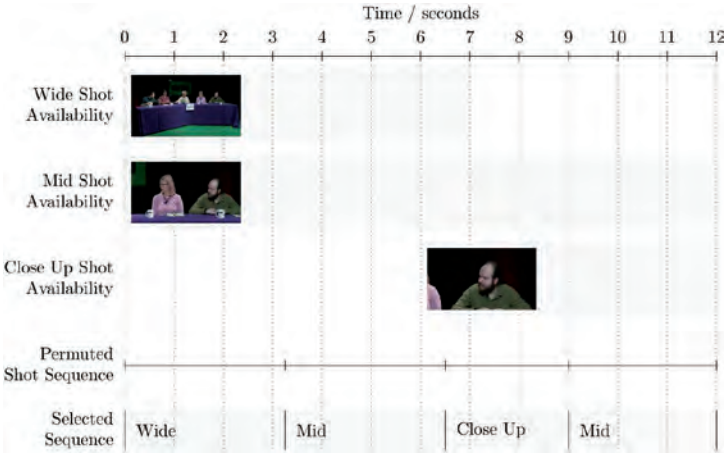


그림 3. 여러 장면들의 이용가능성과 샷 선택 예시

평가 및 개선을 위한 품질 인터뷰

BBC R&D는 앞서 Ed를 개발하기 위해 전문가들로부터 방송 촬영에서의 일련의 규칙들을 도출했고, 몇 가지 지침에 따라 Ed의 알고리즘을 개발했습니다. 공학적인 접근도 중요하지만, 방송은 사람이 보는 것이기에 연구진은 시청자와 제작 전문가의 '경험적인' 작업을 추가했습니다. '인간 중심 작업'이 중요하다는 연구진의 생각은 Ed의 성능과 아웃풋의 '품질'을 고려하는 다음의 두 가지 질문으로 정리됩니다. 연구진은 Ed의 규칙이 단순 컷 편집이 아닌, 시청자에게 얼마나 효과적이고 만족스러운 결과를 가져다주는지 검증하고 싶었습니다.

첫째, Ed의 샷 프레임, 시퀀싱, 샷 선택 및 결정은 인간이 동일한 자료 및 간략한 내용으로 만든 것과 어떻게 비교되는가?

둘째, 수용자의 시청 경험의 질은 어떠한가?

이 두 가지 질문에 답하기 위해, 연구진은 QoE(Quality of Experience) 접근법에 기초한 시청자 평가를 실시했습니다. 먼저 4명의 촬영전문가에게 의뢰해 주어진 세트 내에서 와이드, 미드, 클로즈업 샷 등을 촬영하도록 했습니다. Ed 역시 정확히 동일한 프레임 지시가 내려져, 유사하지만 뚜렷한 개별적인 장면을 산출했습니다. 그 결과, 총 수백 개의 프레임 클립이 확보되어 인간과 인간, 인간과 기계의 광범위한 쌍으로 비교가 가능해졌습니다.
연구진은 시청자 인터뷰를 통해 클립 선호도에 영향을 미치는 요소들을 조사하여, 이 정성적 데이터를 바탕으로 Ed의 프레임 가이드라인을 수정 및 개선하고자 했습니다.
따라서 연구의 두 번째 단계로 24명의 시청자에게 인간과 Ed가 캡처한 동일한 프레임의 클립 쌍을 나란히 제시해 보여주고, 왼쪽 또는 오른쪽의 동영상 중 어느 것이 더 호소력 있는지, 어떤 프레임에 선호가 있는지, 다른 샷을 선택하지 않은 이유는 무엇인지 등을 물었습니다.
일반적으로 시청자는 Ed가 알고리즘으로 캡처한 샷 프레임보다는, 인간의 프레임을 더 선호했습니다. 연구에 참여한 시청자는 자신이 해당 샷을 선택한 이유를 설명하기도 했는데, 인터뷰 결과를 바탕으로 연구진은 샷 프레임에 대한 몇 가지 추가 지침을 도출할 수 있었습니다. 본 고에서는 그중 일부를 소개합니다.

연구결과: 대조 실험을 통해 얻은 몇 가지 규칙들

가이드라인 #1 - 가장자리에는 물체가 없어야 한다.

그림 4. 미드샷 : Ed(좌)와 인간(우)의 프레임

시청자들은 클립에 있는 물체(식물, 간판, 머그잔)가 완전히 안으로 들어가거나 완전히 프레임 밖으로 빠져나가는 것에 대해 분명한 선호를 표시했습니다. 프레임 가장자리에 어중간하게 잘려있는 사물은 산만하고 전문적이지 못한 것으로 간주되었습니다.

가이드라인 #2 - 원샷의 과대한 줌은 피하라

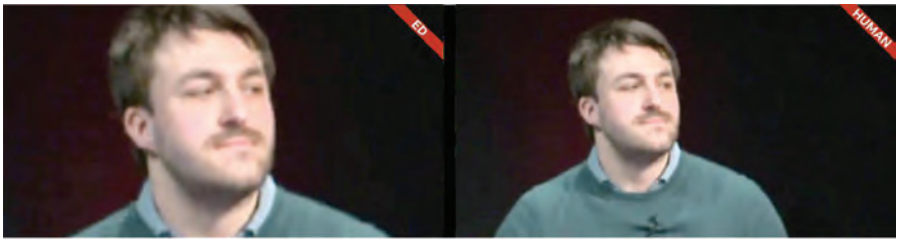


그림 5. 클로즈업 샷: Ed(좌)와 인간(우)의 프레임

원샷의 경우, 시청자들은 지나치게 확대된 얼굴을 보는 것은 기피했습니다. 연구진은 [그림 5]의 오른쪽 사진처럼, 시청자들은 머리 전체와 약간의 몸을 보여주는 원샷을 선호한다는 것을 발견했습니다. 영상 촬영 기법상 당연한 이야기 같지만, 이것은 아직 프로토타입인 Ed에게는 당연하지 않은 이야기일 수 있습니다. 전체적으로, 시청자들은 화면에 얼굴이 너무 부각되는 것은 거슬린다고 말했습니다.

가이드라인 #3 - 빈 공간은 피하거나 최소화하라

그림 6. 미드샷 : Ed(좌)와 인간(우)의 프레임

시청자들은 [그림 6]의 왼쪽과 같이, 빈 공간이 너무 많은 클립은 싫어했습니다. 시청자 중 한 명은 "죽은 공간과 블록 컬러 영역이 많아서 약간 허전한 느낌이다. 너무 많은 것이 없는 것 같다."라고 말했습니다. Ed가 이러한 공간을 최소화하기 위해서는 테이블 천의 보라색이나 배경의 검은색과 같은 블록 색상의 양을 최소화하는 프레임을 선택하도록 규칙을 추가할 수 있습니다.

나가며

지금까지 BBC R&D가 수행한 인공지능을 특정 제작 과정(야외 라이브 이벤트, 패널쇼)에 적용한 규칙기반 시스템 Ed의 개발 과정과 Ed의 자동화된 샷 프레임 추출과 샷 선택의 결과를 살펴봤습니다. 연구진은 Ed의 샷 순서 배열 및 선택을 개선하기 위해 이와 비슷한 인간 중심의 연구를 지속적으로 준비한다고 밝혔습니다. 반복적인 비교실험을 통해 Ed의 프로토타입의 품질을 개선하고, 언제쯤 알고리즘이 시청자에게 충분히 좋은 장면을 제공할지, 적합한 영상 콘텐츠 유형은 무엇일지에 대한 고민을 이어간다는 의미입니다.

최근 몇 년 동안 머신러닝은 이미지 분류, 얼굴 감지, 자세 평가와 같은 부분에서 큰 발전을 보여주었습니다. 구글, 트위터 등의 글로벌 기업을 주도로 이미지 프레임화 및 후처리 방법을 학습한 시스템이 등장하는 것, GPU의 기능 및 효과가 발전하는 것은 방송 영역에서 비디오 분석과 같은 대량의 데이터를 보다 쉽고 빠르게 처리하는 데 큰 도움이 되고 있습니다. 한편으로는 '인간 전문가'가 제작한 프로그램으로 가득한 TV 아카이브는 '무엇이 좋은 프레임인가'를 구성하는 매우 훌륭한 데이터 소스가 된다고 언급하고 있습니다.

국내 방송사에서도 인공지능을 적용한 제작 시스템을 구현하거나, 구축하려는 움직임이 조금씩 나타나고 있습니다. BBC의 사례를 살펴볼 때, 중요한 것은 알고리즘으로 구현할 '규칙'을 '무엇을 근거로 정의할 것인가'에 있다고 생각합니다. 어떤 형태의 작업이 되든 간에 기준을 마련하는 것이 우선시 되는데, 국내 방송사의 경우 우선 아카이브 작업을 고도화하거나 콘텐츠의 메타데이터를 고품질로 생성해두는 것이 필요하다고 생각됩니다. 제작 과정에 인공지능을 도입하는 것은 그 자체로도 고무적인 일이지만, 결과로 추출되는 영상 품질도 간과해서는 안 되는 영역이기 때문입니다.

궁극적으로, 인공지능과 머신러닝이 현재 예측하지 못한 작업 또한 자동화했을 때 인간의 '전문성'과 '창의성'에 대한 의미는 어디서 찾아야하는가에 대한 물음도 고민해야 할 부분입니다. 방송사에서 인공지능을 통해 잠재적인 이익을 취하려는 것은 매력적이고 중요한 활동임은 분명합니다.

BBC R&D의 개발 과정을 담은 AI in Production은 그런 의미에서 방송사의 혁신을 앞당기는 귀중한 결과물이라고 생각하며, 이런 의미에서 '방송과기술' 독자 여러분께 소개해드렸습니다. 방송기술인 여러분들도 함께 생각해보면 좋겠습니다.

참고문헌

- AI IN PRODUCTION: VIDEO ANALYSIS AND MACHINE LEARNING FOR EXPANDED LIVE EVENTS COVERAGE (BBC R&D, 2018)
- 기술주도형 미디어 환경변화에서의 BBC 대응전략 (한국방송통신전파진흥원, 2018)
- 인공지능 콘텐츠 혁명 (한빛미디어, 2018)