



방송 영상 인물 검색 시스템 개발

글. 신지업 SBS 기술연구소 매니저

AI 기술은 어느새 방송 현장 곳곳에서 자연스럽게 사용되고 있습니다. SBS 역시 AI 기술을 다양한 분야에 적용해보며 그 가능성을 탐색하고 있는데, 그중에서도 최근 성과를 낸 사례가 바로 AI 기반 '방송 영상 인물 검색 시스템'입니다. SBS는 개국 후 지금까지 30년 넘는 시간 동안 약 수십만 시간 분량의 방송 영상을 축적해왔습니다. 이 방대한 영상 아카이브는 매우 가치 있는 자산이지만, 실제로 필요할 때 원하는 장면을 찾기란 쉽지 않았습니다. 기존에는 CMS(Contents Management System)를 통해 키워드 기반으로 영상을 검색했지만, 메타데이터가 누락된 경우 원하는 영상을 찾지 못하는 경우가 많았습니다.

최근 들어 유튜브나 SNS를 통해 구작 영상을 활용한 2차 콘텐츠 제작이 활발해지면서, '어떤 인물이 언제 어떤 영상에 출연했는지'를 빠르게 찾아야 하는 수요가 커졌습니다. 특히 인물 중심의 클립 편집 콘텐츠가 높은 반응을 얻으면서, 영상 검색의 효율화가 절실한 과제로 떠올랐습니다. 이러한 배경에서, 저희는 AI 기반 '인물 검색 서비스'를 개발하게 되었습니다. 출연자 이름 정보가 아닌 얼굴 이미지를 기반으로 검색할 수 있는 시스템을 구현하여 과거 데이터의 활용도를 대폭 높일 수 있었습니다.

본 글에서 인물 검색 서비스의 핵심 요소인 얼굴 인식 기술과 벡터 검색의 원리에 대해 소개하고, 서비스를 개발하는 과정에서 겪었던 주요 이슈와 해결 경험을 공유하고자 합니다.

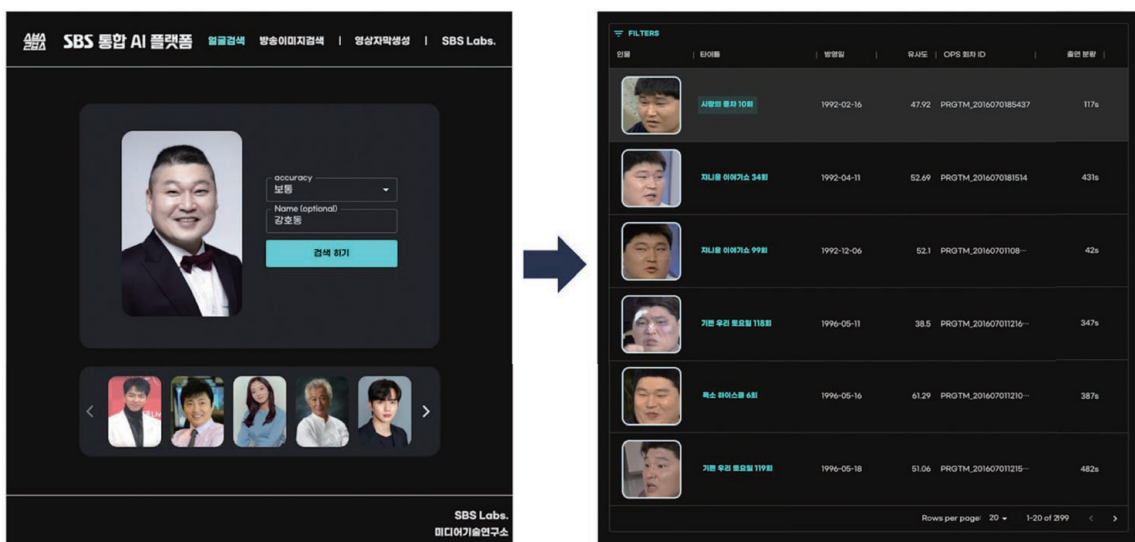


그림 1. 인물 검색 서비스 화면

얼굴 인식 기술

2010년대 중반까지 대부분의 딥러닝 기술은 정해진 데이터셋 내에서 카테고리를 분류하는 데 집중되었습니다. 예를 들어 이미지를 입력하면 개와 고양이를 구분하거나, 치와와와 머핀을 분별하는 객체 인식 모델들이 딥러닝 연구의 주류를 이루었습니다.

얼굴 인식 기술도 마찬가지로 특정 데이터셋 내에서 인물을 분류하는 방식이었습니다. A, B, C라는 인물의 얼굴 이미지를 학습하여, 이 셋 중 누구지를 분류하는 식입니다. 문제는 새로운 인물(D, E)이 등장하면, 기존 학습 범위 밖이므로 인식에 실패하거나 오인식하게 된다는 점이었습니다. 이는, 이른바 Closed-set Classification 방식의 한계입니다. 이러한 한계로 인해 딥러닝 기술이 인간 성능을 뛰어넘었다는 환호가 나오던 시기에도, 실제 응용 가능성에는 회의적인 시각이 적지 않았습니다.

2010년대 후반부터는 딥러닝 기술의 패러다임이 전환되었습니다. 정해진 카테고리를 분류하는 것이 아니라, 사물의 고유한 ‘특징 벡터(Representation)’를 학습하는 방식이 주류가 된 것입니다. 이전에는 정해진 A, B, C라는 인물에 대해 분류하는 법을 학습했다면, 새로운 방식에서는 많은 인물에 대한 데이터(수천 ~ 수만 명의 얼굴 데이터)를 바탕으로 각 사람을 분별할 수 있는 고유한 특징점들을 학습하게 됩니다. 사람의 얼굴에는 사람마다 지니는 고유한 특징들이 존재한다고 알려져 있습니다. 인종과 미간 사이의 길이 비율, 귀의 세로 길이, 귀의 곡률, 눈동자의 색깔, 점의 위치 등 다양한 특징이 얼굴에 존재하는데, 이러한 특징을 수치화하여 조합하면, 전 세계 모든 사람의 얼굴을 분별할 수 있는 고유한 특징 수치를 추출할 수 있게 됩니다. 이러한 특징 수치를 인공신경망이 스스로 설계하고 추출하는 방식이 바로 특징 벡터 방식입니다.

이런 특징 벡터를 학습하는 방식의 가장 큰 장점은 데이터셋 안에 존재하지 않는 사람도 분별할 수 있는 능력이 생긴다는 점입니다. 학습한 데이터셋을 넘어서는 능력을 갖추게 되는 이러한 성질을 일반화라고 하는데, 이러한 일반화 성능의 급격한 향상이 2010년대 후반에 일어났고, 당시 유행했던 Metric Learning이나 Few-Shot Learning, One-Shot Learning 등이 바로 이런 맥락에서 나온 개념들입니다.

벡터 검색

학습된 벡터에 대해 조금 더 살펴보겠습니다. 특징 벡터는 결국 수치의 나열이라고 볼 수 있습니다. 벡터라는 용어가 낯설게 느껴질 수 있겠습니다만 저희가 고등학교 수학과 물리 시간에 배웠던 바로 그 벡터입니다. 당시 2~3차원의 벡터를 다뤄서 (x, y)나 (x, y, z)로 표현했었는데, 여기서는 차원이 약 1,000차원으로 나열되는 수치가 1,000개로 많아진 점만 다릅니다. 이 특징 벡터로 하려는 작업은 굉장히 단순합니다. A라는 사람의 이미지로부터 추출된 벡터와 B라는 사람의 벡터 사이의 거리를 측정하는 것입니다. 이 벡터 간 거리가 특정값보다 크면 얼굴의 고유한 특징이 크게 다른 것이므로 A와 B는 다른 사람으로 판단하고, 특정값보다 작으면 고유 특징이 유사하니 A와 B를 같은 사람으로 인식하는 방식으로 동작하게 됩니다.

2010년대 후반 이뤄진 패러다임 전환 이후에 많이 사용되는 임베딩(벡터), 특징 벡터 등의 개념은 모두 이렇게 거리를 측정하는 방식으로 활용되고 있습니다. “성인 남자가 신나게 춤추는 영상을 찾아줘”라는 검색을 해주는 영상 검색 서비스나 특정 음악의 일부를 들려주고 원곡을 찾아주는 ‘샤잠’ 같은 서비스 모두 이러한 특징 벡터를 기반으로 동작하는 것으로 알려져 있습니다.

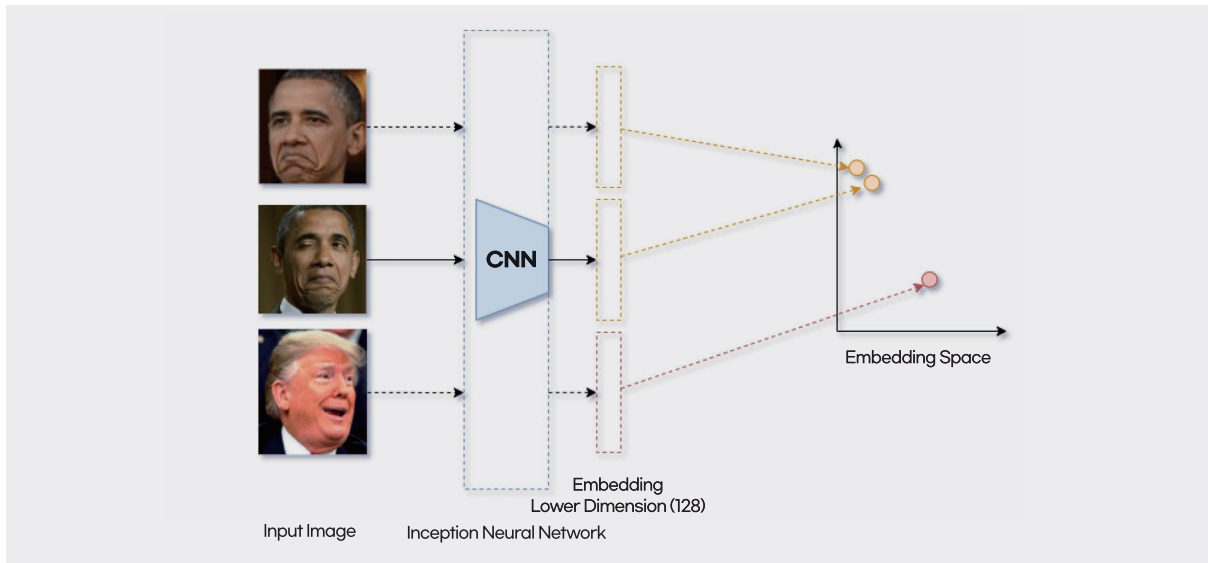


그림 2. 얼굴 특징 벡터 개념도 / 출처 : medium.com/analytics-vidhya/building-a-dog-search-engine-with-facenet-65d1ae79dd8a

보유한 영상이나 음원 데이터에 대한 특징 벡터를 미리 추출해두어 비교할 수 있게 준비해 둔 후, 검색하려는 텍스트 혹은 음원의 특징 벡터를 추출해서 가장 가까운 거리에 있는 벡터를 찾는 방식으로 동작하는 것입니다. 머신러닝 분야에서는 이러한 방식을 k-NN(k-Nearest Neighbor)라고도 부르는데, 보유한 모든 벡터와의 거리를 측정/비교하여 검색하려는 벡터와 거리상 가까운 k개의 이웃(벡터)을 찾는다는 개념입니다.

이러한 벡터 기반의 k-NN 알고리즘은 검색뿐만 아니라 유튜브, 틱톡 등의 추천 알고리즘, RAG(Retrieval Augmented Generation) 기반 LLM(Large Language Model) 등 AI 기반의 애플리케이션에는 핵심적인 기능이 되었습니다. 이와 같은 NN 방식의 활용 수요가 많아짐에 따라 기존 텍스트 검색엔진이나 데이터베이스에서도 벡터 검색 기능을 지원하기 시작했고, Pinecone, Weaviate, chroma, Milvus 등의 전문화된 벡터 검색엔진도 출시됐습니다. 공개된 벡터 검색엔진은 사용하는 검색 알고리즘이나 타겟으로하는 애플리케이션이 조금씩 다르지만, 대규모 데이터에 대해서 빠르고 정확하게 이웃 벡터를 찾아주는 기능을 제공합니다.

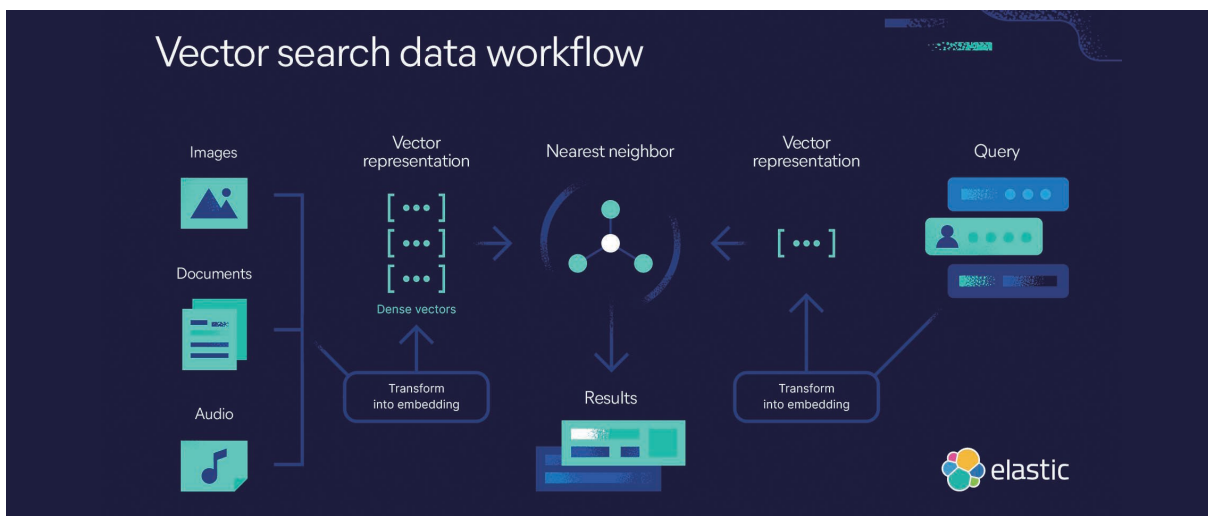


그림 3. 벡터 검색엔진 개념도 / 출처 : www.elastic.co/kr/what-is/vector-embedding

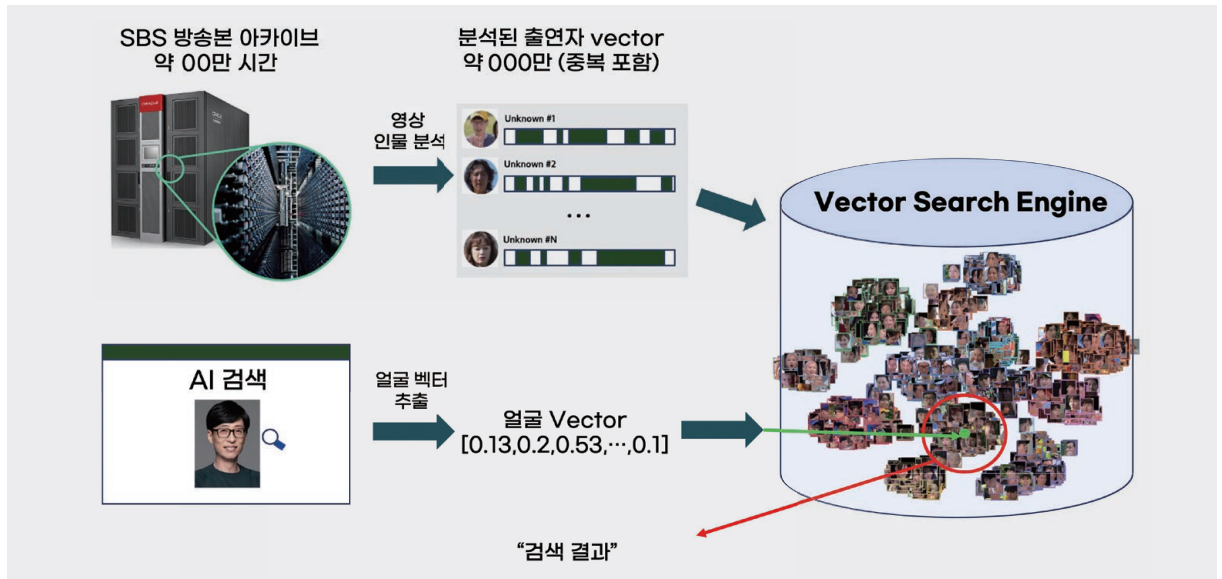


그림 4. 인물 검색 서비스 개념도

AI 인물 서비스 구현

프로젝트 초기에는 외부 AI 서비스를 활용하여 인물 검색 기능을 구현하는 방안을 먼저 검토했습니다. 당시 글로벌 클라우드 서비스들이 제공하는 영상 분석 기능을 활용하면 빠르게 서비스를 구축할 수 있을 것으로 기대했기 때문입니다. 여러 유명 서비스들에 대해 자체 PoC를 진행했지만, 곧바로 두 가지 문제에 직면했습니다.

유명인 인식 기반 서비스

당시 AI 얼굴 인식 기술을 활용한 서비스들은 유명인 인식 기술을 기반으로 했습니다. 영상 내에서 인물을 인식하고, 해당 인물이 유명인이면 이름을 텍스트 형식의 메타데이터로 저장 및 검색하는 구조였습니다. 이러한 구조로는 정해진 인물 DB 내의 인물만을 검색할 수 있습니다. 인물 DB에 누락된 인물이나 예전엔 무명이었으나 갑자기 인기가 많아진 인물은 검색할 수 없는 한계가 있습니다. 이를 보완하기 위해서는 충분히 많은 수의 인물 DB를 구축/관리해야 하는 이슈가 추가로 따라오게 됩니다.

더 심각한 문제는 인물 DB가 업데이트되면 영상 분석을 다시 해주어야 한다는 점입니다. 기존에 인식하지 못했던 인물을 새로 인식하기 위해서는 업데이트된 인물 정보를 바탕으로 새로 분석해주어야 하는데, 여기에 소모되는 리소스가 상당합니다. 따라서 방송국처럼 대규모 아카이브를 보유한 기업의 경우 인물 DB 방식은 시간과 비용의 문제로 적용하기 어렵습니다. 이러한 인물 DB 기반의 유명인 인식 방식의 구조적 문제를 해결하기 위해서 검색할 인물의 이미지를 업로드하여 검색하는 방식으로 서비스를 구성하였습니다. 인물 검색 방식으로는 비유명인도 모두 검색할 수 있는데, 이는 앞서 언급한 얼굴 특징 벡터와 벡터 검색엔진을 활용하면 구현할 수 있습니다.

영상 분석 비용

클라우드 서비스의 경우 영상 1시간당 1~3만 원의 비용이 발생합니다. SBS가 보유한 아카이브 영상이 수십만 시간 영상에 달하는데, 단순 계산으로 10만 시간에 10~30억 정도 됩니다. 기존에 서비스하는 아카이브 검색 기능에 ‘AI 인물 검색 기능’ 하나를 추가하는 데에 수십억이 필요한 셈인데, 단일 프로젝트로 실행 가능한 수준을 상당히 넘어서는 수준이었습니다.

이러한 시행착오를 거친 끝에, 결국 자체 개발로 방향을 전환하게 되었습니다. 얼굴 영역 탐지(Face Detection), 특징 벡터 추출(Face Recognition), 벡터 클러스터링 등의 기술들을 조합하여 영상 내의 출연자 정보와 출연자별 출연 구간 정보를 자동으로 분석하는 인물 분석기를 만들었습니다.

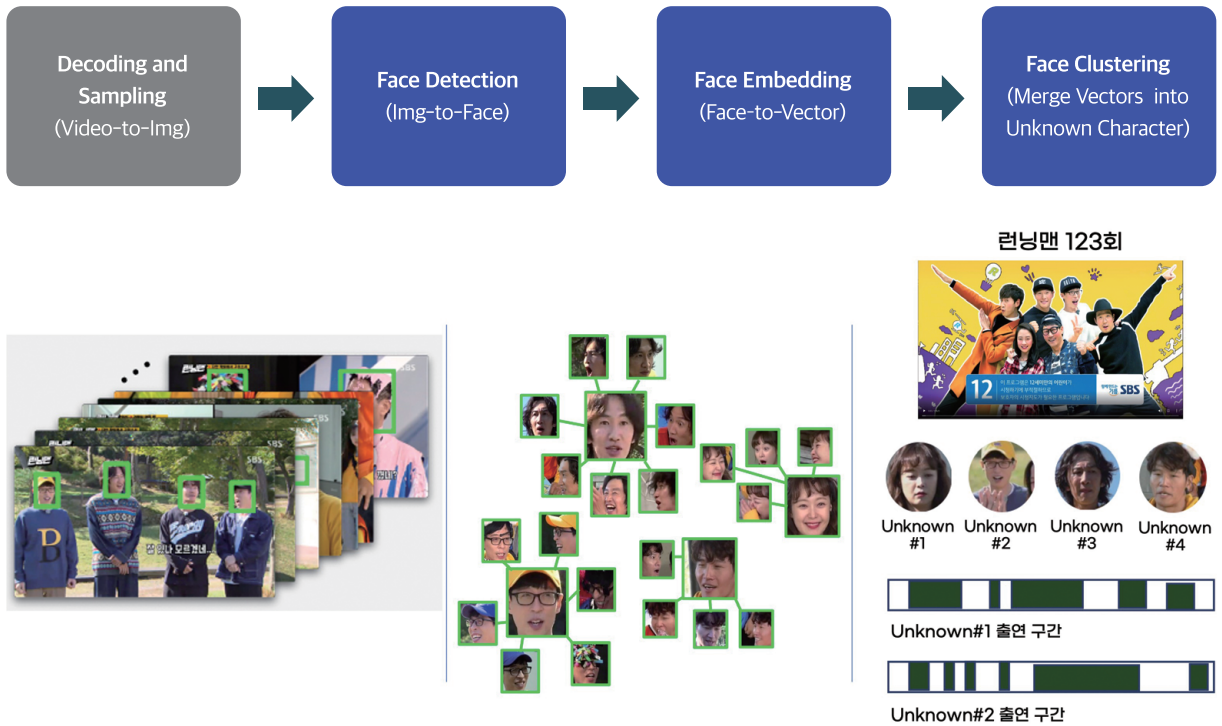


그림 5. SBS 자체 영상 인물 분석기 개념도

이 분석기를 만드는 과정의 핵심은 영상을 빠르게 분석할 수 있도록 분석기를 최적화하는 작업입니다. 영상 분석은 GPU가 탑재된 서버에서 동작하게 되는데 영상을 빠르게 분석할 수 있다는 것은 더 적은 GPU 자원을 사용한다는 것이고, 이는 결국 분석 비용이 줄어든다는 것을 의미합니다. GPU 가속 디코딩, 영역 트래커 기반의 분석 구조 설계, 신경망 모델 최적화 등 효율적인 영상 분석을 위한 최적화를 수행하여 단일 GPU에서 영상의 길이 대비 약 10배속 정도의 속도로 분석할 수 있었습니다. 분석 속도와 도입 GPU 서버 수 등을 바탕으로 실제 분석에 투입되는 비용을 산정할 수 있는데, 대략 10만 시간의 영상 분석에 2천만 원 수준으로 분석할 수 있게 되었고, 현재는 SBS 제작/보도 아카이브 영상 모두에 대해 인물 검색이 가능합니다.

활용 사례

AI 인물 검색 시스템은 현재 SBS의 다양한 제작 현장에서 사용되고 있습니다.

먼저, 방송 제작 분야에서는 과거 예능 프로그램에 출연했던 인물의 자료 화면을 찾을 때 큰 도움이 되고 있습니다. 이전에는 특정 출연자의 데뷔 초 영상이나 짧게 등장했던 장면을 찾으려면, 담당자가 수십 편의 프로그램 영상을 처음부터 끝까지 돌려봐야 했습니다. 하지만 이제는 찾고자 하는 인물의 사진 한 장만 업로드하면, 아카이브에 저장된 수십만 편의 영상 중 해당 인물이 등장한 장면을 자동으로 추출할 수 있게 되었습니다. 덕분에 클립 제작이나 하이라이트 영상 편집이 훨씬 빠르고 효율적으로 이뤄지고 있습니다.

또한, 뉴스 제작 현장에서도 AI 인물 검색은 중요한 역할을 하고 있습니다. 최근 사회적 이슈나 사건 보도를 준비할 때, 과거 아카이브 영상에서 메타데이터만으로는 찾기 어려운 인물이 필요한 경우가 종종 발생합니다. 예컨대, 대통령 경호처 직원과 같이 사전에 이름이나 정보가 등록되기 어려운 인물이 화면에 잠깐 등장하는 경우가 그렇습니다. 과거라면 역대 대통령 출연 영상을 하나하나 재생하며, 화면에 스치듯 잡힌 얼굴을 살살이 확인해야 겨우 찾을 수 있었습니다. 이제는 보도하고자 하는 대상의 이미지 하나만으로 AI 인물 검색을 하면, 몇 초 만에 해당 인물이 등장한 장면 수십 개를 자동으로 찾아낼 수 있습니다. 덕분에 뉴스 제작 속도가 한층 빨라지고, 더 풍부한 자료 화면을 신속하게 확보할 수 있게 되었습니다.

마치며

AI 기반 인물 검색 시스템을 개발하며 가장 크게 느낀 점은, 기술적 성능뿐 아니라 비용과 활용성 간의 균형을 어떻게 맞출 것인가가 중요하다는 점이었습니다. 아무리 정확도가 높은 AI 기술이라도 실제 방송 제작 환경에서 사용할 수 있는 수준의 속도와 비용 구조가 아니면, 현업 도입은 어렵다는 교훈을 얻었습니다. 이번 프로젝트를 통해 저희는 단순히 AI 기술을 도입하는 것을 넘어, 현업 환경에서 실질적인 가치가 있는 서비스를 만드는 방법을 고민하게 되었습니다.

앞으로 멀티모달리티 영상 분석, 편집툴과의 자연스러운 연계, 방송 제작용 생성 AI 등으로 서비스를 더욱 확장해나갈 계획입니다. 아직 AI 서비스에는 '완성'이라는 단어가 없는 만큼, 꾸준한 기술 고도화와 현업 피드백 반응을 통해 더욱 유용한 시스템으로 발전시키고자 합니다. 이번 사례가 방송 영상 분야에서 AI 기술을 효과적으로 적용하는 데 작은 참고가 되기를 바랍니다. 

