

데이터와 인공지능 알고리즘

글. 한영주 EBS 뉴미디어프로젝트팀 연구위원

SF(공상과학) 영화는 현실에서 경험할 수 없는 미래의 모습을 묘사해왔다. 아마도 SF 영화는 다른 장르 영화보다 인간의 상상력이 더 많이 집약된 콘텐츠일 것이다. 주요 소재로 과학기술을 다루며 허구의 이야기 속에서 진기한 장면들이 연출되기 때문이다. 그 안에서 나름대로 미래상을 제시하며 메시지를 전달해왔다. SF 영화로 묘사된 가상의 미래상 중에서 과연 현재와 가까운 모습은 어떻게 표현되었을까? 그 단서들은 1990년대~2000년대 초반 할리우드 SF 영화 속에서 간간히 찾아볼 수가 있다.

‘데폴리션맨(1993)’은 미래 범죄와 경찰에 대한 내용이다. 이 영화에서는 주인공이 체력단련을 위해 헤드마운트를 쓰고 가상공간에서 격투 대결을 벌이거나 감정 교류를 하는 장면을 볼 수 있다. 그 모습이 마치 요즘의 혼합현실(mixed reality)을 방불케 한다. ‘미션 임파서블(1996)’은 특수 요원 임무를 수행하기 위한 기발한 첨단 장비들이 등장한다. 그중에서 시리즈마다 단연 돋보였던 얼굴 복제는 3D 프린터와 흡사하다. ‘마이내리티 리포트(2002)’에서는 사전에 범죄를 예측할 수 있는 미래 도시 내용을 다룬다. 이 영화는



그림 1. 공상과학 영화 속 첨단 기술 / 출처: 인터넷 포털 영화 검색

흥미로운 기술들이 곳곳에 배치되어 시각적인 재미를 주었다. 홀로그램은 영상 속 인물을 사실처럼 입체화 시키고, 홍채 인식은 개인 정보를 관리하고 맞춤형 광고에 활용한다. 또한 모션 인식 인터페이스로 업무를 처리하고 무인 플라잉 카를 사용해 이동한다. 그 당시 영화를 관람했던 관객들은 SF 영화 속의 미래가 현실에서 빨리 재연될 것이라고 상상하지 못했을 것이다.

SF 영화 속에 미래상은 정보혁명이 시작된 이후 빠르게 현실로 다가왔다. 정보혁명은 인터넷 보급과 디지털 정보의 중요성이 대두되던 시기였다. 이전에 제조 산업이 높은 수준의 하드웨어 장치를 위주로 구성되었던 반면에 정보혁명 시대는 소프트웨어와 디지털 데이터의 가치가 높아지며 산업의 중심이 서비스로 변화했다. 즉 기업에서 고객에게 전달하는 가치 체계가 바뀐 것이다.

각종 정보가 디지털로 기록되고 관리되면서 인간의 모든 행동이 디지털 데이터로 누적되기 시작했다. 요즘은 스마트 기기 하나로 커뮤니케이션은 물론, 위치 정보를 활용한 지도나 간단한 결제까지 많은 일을 디지털로 처리한다. 이전보다 인간의 행동은 더 세밀하게 데이터로 기록되고 있다. 과거에 기업들은 개인정보처럼 특정 데이터를 수집하는데 주력했다. 그러나 지금은 거의 모든 데이터를 수집하는 양상으로 변화하고 있다. 데이터는 다른 자원들과 다르게 일단 확보

되면 쉽게 소멸되지 않고 지속적으로 활용하거나 다양한 형태로 가공할 수 있기 때문이다. 이제 데이터를 확보하지 못하면 경쟁할 수 없는 구조가 되었다. 그러나 기업의 규모와 자본에 따라 데이터 확보에 차이가 발생할 수밖에 없다. 거대 기업들은 우리가 생각하는 것보다 더욱 광범위한 데이터를 수집하기 위해 서로 경쟁한다. 그들의 방식은 가히 상상하기 어려울 정도인데 인터넷 브라우저, 운전자 시스템, 검색 시스템, CCTV, 웹사이트, SNS, 인공지능 비서, 번역기에서도 데이터를 수집해서 활용한다. 그러나 축적된 데이터

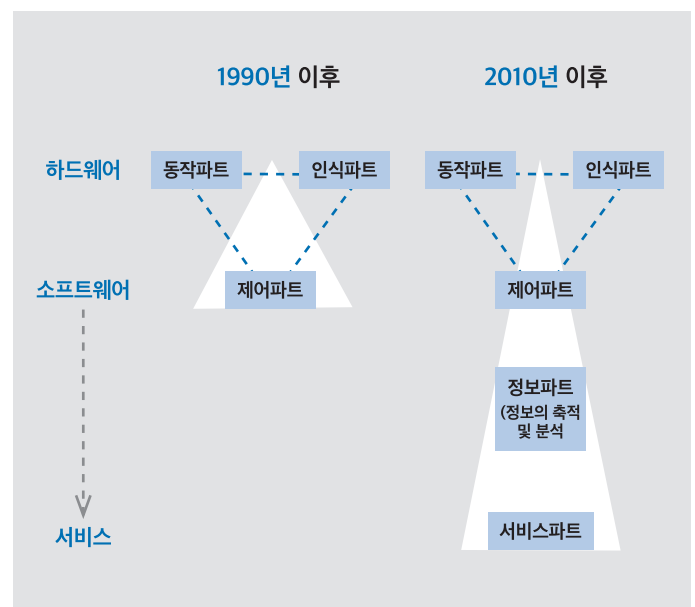


그림 2. 가치 변화 / 출처: EY 어드바이저리 (2016).

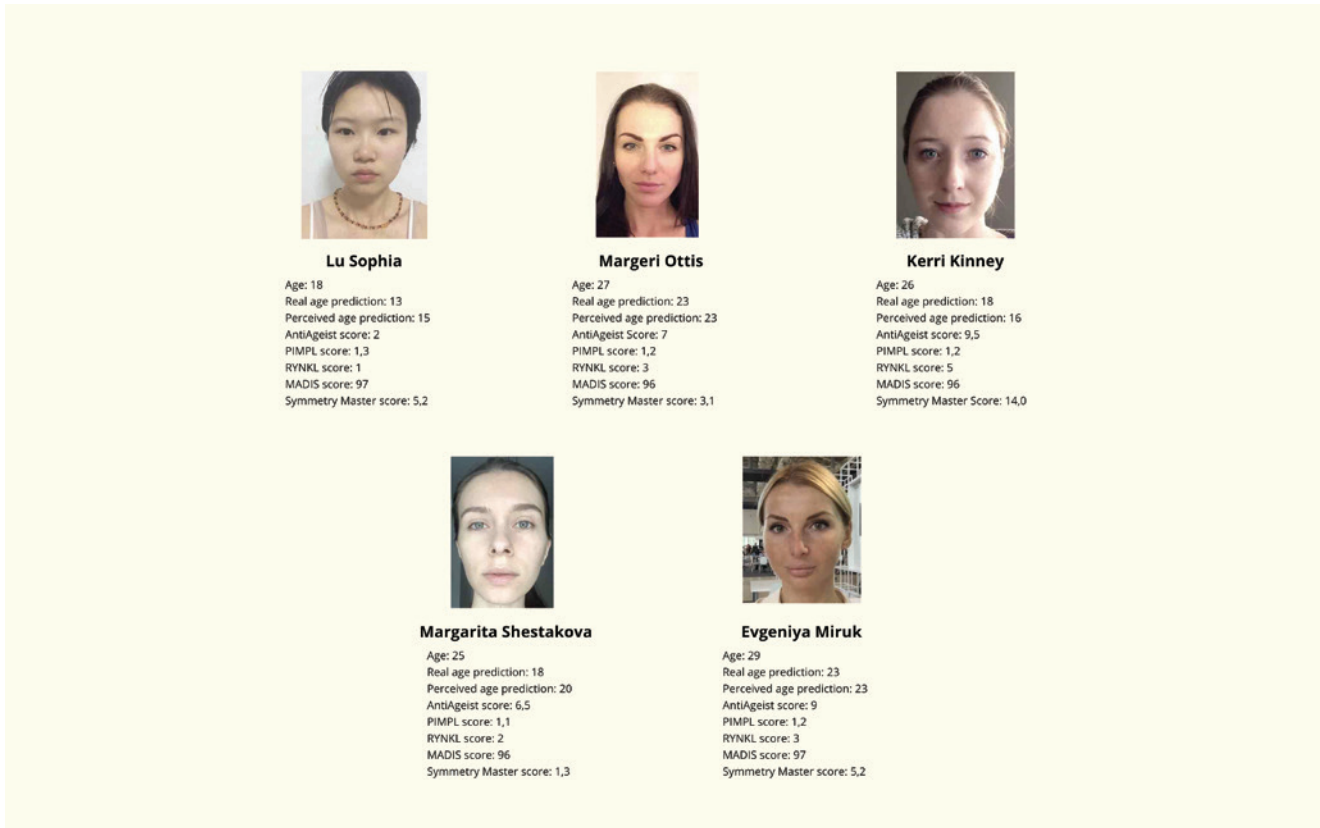


그림 3. 인공지능 미인 선발대회 우승자 / 출처 : Sam Levin (2016).

는 그 자체로는 별다른 의미를 갖지 못한다. 데이터는 사용 목적에 따라 분석 방법과 판단, 함의를 도출하고 나아가 예측이 가능해야 진정한 의미를 가질 수 있다. 수집된 데이터는 대규모 데이터 센터에 기록되어 분석된다. 보통은 빅데이터(big data) 형태와 스몰데이터(small data) 형태로 구분해서 분석한다. 빅데이터는 특정 집단의 변화나 트렌드를 분석해서 제품 개발이나 서비스에 반영한다. 스몰데이터는 개인 맞춤형 제품이나 서비스를 목적으로 분석된다.

미디어 산업에서는 주로 스몰데이터 형태로 자사 고객들의 취향과 관심에 맞는 서비스를 제공하는데 활용해왔다. 많은 테크(Tech) 기업들이 데이터 중심 콘텐츠 서비스를 전면에 내세우며 시장에서 혁신을 주도하고 있으며, 여기에 인공지능 알고리즘이 결합되며 더욱 정교해진 서비스로 시장을 선점하고 있다. 인공지능 알고리즘은 나와 유사한 사람들과 이전의 나의 행동 패턴을 분석해서 내가 알기도 전에 무엇을 좋아할지 미리 알고 적합한 콘텐츠를 찾아서 추천해준다.

표면적으로는 이용자의 시간과 노력을 절약하고 개인화된 콘텐츠에서 능동적인 선택권을 부여하는 것처럼 보인다. 그러나 실상은 많은 사람이 즐겨보는 인기 있고 영향력 있는 콘텐츠가 추천되어 노출될 확률이 더 높고 이것이 이용자의 선택 과정에 영향을 미칠 가능성이 크다.

우리가 일상에서 인터넷이나 정보 미디어를 사용할 때도 차별이 개입한다. 인공지능과 컴퓨터 프로그램은 우리가 선택한 내용을 수렴해서 자사 서비스를 위한 판단의 근거로 활용한다. 이런 점에서 인공지능의 알고리즘은 우리 사회에 존재하는 차별적인 선입견을 오히려 더 강화할 수도 있다. 인공지능의 판단에서 편향이나 편견이 그대로 반영되어 나타날 수가 있기 때문이다. 미인의 얼굴을 객관적으로 평가하기 위해 2016년 미인 선발대회에서 인공지능 알고리즘을 사용했던 일을 예로 들어보자. 미인의 평가 기준은 대칭적인 얼굴 모양, 주름 등의 다양한 요소가 포함되었다. 이 대회에는 100여 개의 다양한 나라에서 6천 명

이 사진을 보내 참가를 신청해서 평가되었다. 결과는 놀라웠다. 6천 명의 참가자 중에서 44명이 수상하였고 그중에서 소수의 동양인이 포함된 것 이외에 대부분이 백인이었다. 기타 유색 인종은 단 한 명도 없었다. 이 결과는 인공지능 알고리즘 학습에 사용된 훈련 데이터에서 백인의 사진은 많이 포함되었지만 유색 인종의 사진은 충분히 포함되지 않았던 것을 의미한다. 이 사례는 인공지능 알고리즘이 지식과 정보를 의도적으로 노출하거나 왜곡시킬 수 있다는 것을 보여준다. 또한 인공지능 알고리즘은 프로그램이지만 이것을 설계하고 입력하는 사람이나 세력에 의해 의사결정이 크게 좌우될 수 있다는 것을 보여준다.

인공지능 알고리즘은 사람들의 음성을 알아듣고 도움을 주는 응답을 하고 사람의 얼굴을 인식해 자동으로 일을 처리하기도 한다. 또한 어떤 인공지능은 낙서에서 예술을 창조하거나 사람들의 건강 상태를 진단하기도 한다. 인공지능 알고리즘은 기존 플랫폼이 제공하는 서비스를 더욱 풍부하게 만드는 보완적인 역할을 했지만 차츰 주도적인 비즈니스 영역을 만들어가고 있다. 주요 IT 기업에서는 무료로 인공지능 API를 공개해서 자사의 기술이나 상품에 대한 이용도를 높이려는 이른바 ‘쏟아림현상’을 유도하며 이용자의 거의 모든 데이터를 수집하고 활용하고 있다.

인공지능 알고리즘의 활용도에 따라 사람들의 인식도 변화하고 있다. 처음에는 그저 스마트폰에 있는 비서 정도로만 인식하고 편리성 측면에 관심을 두었다. 그러나 전 산업 분야에서 핵심 기반 기술로써 인공지능이 도입되면서 일자리와 같은 생존 문제, 개인정보 침해와 같은 윤리적 문제 등의 부정적인 견해들이 제기되고 있다. 과거 정보혁명 시절에 인터넷은 누구나가 참여해서 평등하게 누릴 수 있는 개방적인 공간이었지만 데이터가 사회와 경제의 핵심 동력이 되며, 특정 거대 기업들의 비즈니스 공간으로 변질되고 있다. 사회적 문제로 편향된 정보나 가짜 정보, 여론 왜곡 등이 도마 위에 오르며 책임론이 대두되고 있지만 실질적인 해결책은 미미한 상황이다. 정보 습득과 활용이 더 쉬운 환경이 되었지만 오히려 정보의 가치는 제대로 전달되고 있는지 의문이 든다.



대부분의 기업이 경쟁 요소로 데이터 분석과 활용에 집중하면서 본질적인 가치는 고려하지 못한 것 같다. 데이터 자체가 모든 현상을 해석할 수 있다는 잘못된 인식도 마찬가지이다.

다시 SF 영화 이야기로 돌아가서 생각해보자. SF 영화 내용을 보면 첨단 기술에 대한 장밋빛 전망만 제시하지 않는다. 첨단 기술이 가져온 긍정적인 측면 이면에 불안감, 폐해 등의 부정적인 측면도 함께 전달되고 있다. 현실에서도 마찬가지일 것이다. 현시대에서 데이터와 인공지능 알고리즘은 필연적인 기술이 되었다. 더욱이 인공지능 알고리즘이 탑재된 서비스가 우리의 소소한 일상 속에 들어오고 있다. 그동안 사업적 이익이나 기술적 유행에 치우쳤던 것에서 벗어나 데이터와 인공지능 알고리즘의 양면성을 잘 숙지하여 가치 전달에 좀 더 주력할 시기일 것이다. ☁

참고문헌

- 조병학 (2019). 2040 디바이디드. 서울 : 인사이드엔뷰.
- 서울대법과경제연구센터 (2017). 데이터이코노미. 서울 : 한스미디어.
- EY 어드바이저리 (2016). 세계 초일류 기업의 AI 전략. 서울 : 매일경제신문사.
- Sam Levin (2016). A beauty contest was judged by AI and the robots didn't like dark skin. The Guardian.