

I. 인공지능과 예술. 어디까지가 예술인가?

1. AI가 그린 그림, 1위에 오르다

현지시간 3일 CNN·뉴욕타임스(NYT) 등은 지난달 미국에서 열린 한 미술전에서 인공지능(AI)이 생성한 그림이 우승을 차지하면서 논란이 일고 있다고 보도했습니다. 보도에 따르면 지난달 ‘콜로라도 주립 박람회 미술대회’의 디지털아트 부문에서 게임 기획자인 제이슨 M. 앨런(39)이 AI로 제작한 작품 ‘스페이스 오페라 극장’(Theatre D’opera Spatial)이 1위에 올랐습니다.

우승한 작품은 환하고 비현실적인 분위기에 마치 르네상스 작품과 SF영화제작 장르 중 하나인 스팀펍크(steampunk) 페인팅의 혼종인 것처럼 보입니다. 앨런이 우승을 차지한 이 작품은 텍스트로 된 설명문을 입력하면 몇 초 만에 이미지로 변환시켜주는 ‘미드저니’(Midjourney)라는 AI 프로그램으로 만들어낸 작품입니다. 앨런은 이런 방식으로 얻은 작품 중 3개를 골라 대회에 제출했고, 이 중 하나가 1위를 한 것입니다. 대회에 출전한 계기에 대해 그는 “미드저니를 시험해보다가 AI가 생성한 사실적인 이미지가 매력적으로 느껴졌고, 사람들에게 이런 예술이 얼마나 대단한지 보여주고 싶었다”고 설명했습니다.

미술대회 디지털아트 부문의 규정을 보면 창작과정에서 디지털 기술을 활용하거나 색깔을 조정하는 등 디지털 방식으로 이미지를 편집하는 행위를 인정합니다. 그렇기에 디지털 프로그램 사용 자체는 문제가 되지 않습니다. 그러나 앨런이 3개 작품과 함께 우승 소식을 자신의 SNS에 올리고, 이것이 다시 트위터로 옮겨가 급속히 퍼지면서 논란에 불이 붙기 시작했습니다.

“예술이다” vs “부정행위다” 갑론을박

AI 그림이 우승을 차지했다는 소식을 접한 누리꾼 사이에서는 앨런의 작품을 두고 “예술가가 단 한 번의 붓질조차 하지 않은 작품이 우승을 차지하는 게 정당인가”부터 “사람이 아닌 예술작품으로 볼 수 있는 지” 등 다양한 논의를 펼치기 시작했습니다. 일부는 앨런이 AI가 그린 그림으로 우승한 것은 ‘부정행위’라고 비난하기도 했습니다.

이에 앨런은 뉴욕타임스를 통해 “애초에 자신은 대회에 작품을 제출할 때 ‘미드저니를 거친 제이슨 M. 앨런’이라고 명시해 AI로 작품을 생성했다는 점을 밝혔고 작품의 출처를 속인 사실이 없다”고 항변했습니다. 또, 그는 CNN과 인터뷰를 통해 “대회에 제출한 작품 3개를 얻기 위해 80시간이 넘게 걸렸다”고 설명했습니다. 그러면서 “나는 어떤 규칙도 어기지 않았다”며 “AI 기술을 증오히기보다는 강력한 도구로 인정하고 이를 잘 활용해야 앞으로 나아갈 수 있을 것”이라고 덧붙였습니다. 박람회를 감독하는 콜로라도 농업부 측도 앨런과 같은 이유로 문제가 없다는 입장을 밝혔습니다. 앨런이 작품을 제출할 때 AI 프로그램을 활용했다는 사실을 밝혔고, 해당 부문 규정에서도 창작 과정에서 디지털 기술을 활용한 그 어떤 예술 행위도 용인한다고 설명했습니다.

AI 그림, 예술일까

AI가 예술작품을 만드는 것은 새로운 현상은 아닙니다. 올해 미드저니뿐 아니라 오픈 AI가 제공하는 ‘달리-2’(DALL-E 2)와 구글의 이메진(Imagen) 등 AI 프로그램이 잇따라 출시되면서 누구나 텍스트를 입력하기만 하면 손쉽게 복잡하고 사실적인 그림을 ‘그릴’ 수 있게 됐습니다. 이에 대해 AI가 생성한 예술품은 본질적으로는 첨단기술로 둔갑한 표절의 한 형태일 뿐이라는 주장이 나옵니다. AI가 그리는 그림은 결국 기존에 존재하는 이미지를 활용하기 때문입니다. 이에 반해 일부 예술가는 AI로 작품을 만드는 것은 포토샵 같은 이미지 편집 프로그램을 사용하는 것과 큰 차이가 없고, 이미지 변환을 위한 적절한 문구를 생각해내는 것에도 인간의 창의성이 요구되기에 예술작품으로 볼 수 있다고 반박합니다.

AI가 인간 고유의 영역으로 여겨지던 예술까지 학습하기 시작하면서 최근 AI로 그려낸 그림을 예술을 둘러싼 논의가 활발해졌습니다. 핵심은 인공지능이 만든 작품을 인간이 만든 작품과 같은 의미의 예술작품으로 바라볼 수 있을지에 대한 부분입니다.

2. 조영남 대작사건

가수 조영남이 다른 무명 화가를 고용해 대리 제작 하게 하고 적은 돈만 준 사건으로 미술계와 대중, 고전미술계와 현대미술계의 인식 차이를 드러내는 계기가 된 사건.

조영남 측 주장

미술계의 관행을 따랐을 뿐이다.

또한 12월 21일 공판에서 조영남은 “제가 직접 마지막 (붓)터치를 했고 사인을 했기 때문에 제 작품”이라며 무죄를 주장했다. 그는 “일부 그림을 그리는데 초안에 도움을 받았다고 해서 사는 사람에게 일일이 고지할 의무는 없다”며 “예술의 자유를 국가 형벌권으로 제한할 것인지 면밀히 판단해달라”며 무죄를 주장했다. 피고인 신문에서 조영남은 “팝아트에서는 아이디어나 개념을 중시한다. 화투를 그리는 것을 팝아트로 보고 저 스스로 팝아티스트라고 한 것”이라며 “제가 (송씨에게) 화투를 그리라고 한 것이니까 당연히 제 작품이 맞다”고 밝혔다.

관점 1 : 수작업 없이 아이디어만 제시하는 예술은 사기다.

그러나 이와 별개로 대중과 고전미술계는 여전히 미술가가 손수 수작업을 하는 것이 중요하다고 생각했다. 장인이 한땀한땀 작업하듯, 예술가도 한끝씩 돌조각을 떼어내고 한붓씩 물감을 칠하는 과정을 반드시 거쳐서 작품을 완성해야 정직한 예술가라는 것이 이들의 생각이다. 이에 대해서는 신제남 한국전업미술가협회 이사장이 제출한 고소장 내용 중 일부를 보는 것이 좋을 것이다.

관점 2 : 기계적인 수작업보다 더 중요한 것은 예술가의 구상작업이다.

하지만 현대미술계에서는 이런 생각을 하지 않는 작가나 전문가도 많았다. 마르셀 뒤샹 이후 현대미술은 더이상 그런 수작업이나 기교를 중요시하지 않는다는게 이들의 생각이다. 돌을 깎아내고 붓질을 하는 기계적인 반복작업을 열심히 했다고 그게 수준높은 예술작품이 아니란 것이다. 그리고 구상만 했다고 성의가 없는 게 아니다. 오히려 다른 사람을 도구로 삼아서 자신의 작품을 만들었다고 볼 수 있다는 것이다. 이 입장을 대변하는 대표인물이 진중권이었다. 그는 한 기고문에서 수작업보다 창의적인 발상이 더 중요하다고 주장하며 ‘대작 관행’에 대해 설명했다.

II. 생명윤리, 유전자 복제와 낙태

1. 인간 복제의 윤리적 쟁점

(1) 배아 복제 : 질병 치료를 주목적으로 줄기세포를 활용하는 복제

① 찬성 입장

복제 과정에서 이용하는 배아는 아직 완전한 인간이 아님.

생식 초기에 관한 연구, 인체 조직과 장기 복구, 질병 치료 등에 활용할 수 있음.

② 반대 입장

복제 배아는 인간의 지위를 지닌 생명이므로 보호되어야 함.

배아 복제를 위한 난자 확보 과정에서 여성의 건강권과 인권을 훼손할 수 있음.

(2) 개체 복제 : 새로운 생명체를 탄생시키는 복제

① 찬성

불임 부부가 유전적 인권이 있는 자녀를 가질 수 있음

복제인간도 서로 다른 선택과 경험, 환경에서 독자적인 삶을 살아갈 것임

② 반대

인간의 존엄성 훼손 : 복제를 위한 사람의 의도에 따라 복제 인간을 도구로 이용할 수 있음.

자연스러운 출신 과정에 위배 : 한 사람의 체세포로부터 인간이 복제된다면 인간의 상호 의존심이 파괴될 수 있음

인간의 고유성 위협 : 복제된 인간은 체세포를 제공한 사람과 유전 형질이 같으므로 자신의 고유성을 갖기 어려움.

가족 관계에 혼란 초래 : 체세포와 난자를 제공한 사람과 복제인간이 부모 자녀 관계인지 형제자매 관계인지 불분명함

2. “배아도 생명..폐기되는 잔여배아 줄이는 노력 필요”

‘난임의학과 잔여배아의 생명윤리’를 주제로 한 제2회 샘병원 생명사랑포럼이 지난 17일 안양 샘병원에서 개최됐다.

생명윤리법 제2조 제4호에 따르면, 잔여배아란 체외수정으로 생성된 배아중 임신의 목적으로 이용하고 남은 배아를 말한다. 입법 취지에 따르면 연구목적 사용을 전제한 용어다. 시험관 기술이 보편화된 최근에는 잔여배아 폐기가 새로운 이슈로 부각되는 중이다. 배아에 대해 과학자들은 14일 이전의 배아는 생명체라기보다는 세포일 뿐이라고 본다. 배아를 생명체로 보는 종교계와 시민단체는 배아 연구를 반대한다. ‘생명윤리 및 안전에 관한 법률’에 따르면 배아는 임신을 목적으로 할 때만 형성을 허가받는다. 다만 인공수정 시도로 추출 후 남은 배아 중 권리권자가 기증을 허용할 경우 연구에 이용될 수 있다.

발표에 나선 백수진 국가생명윤리정책원 부장은 “배아와 이월배아 말고 더 이상 임신 목적으로 이용 계획이 없는 폐기에정인 연구용 잔여배아의 보존기관은 147개소 중 139개소이며, 8개소는 연구용 잔여배아 보관량이 없다. 2021년 12월 말 기준, 누적 잔여배아 보관량은 29만 1344개”라고 설명했다.

김종한 안양 샘병원 가임센터장은 “배아는 형성 중에 있는 생명이란 개념에서 배아, 생식세포에 관한 투명한 데이터 통합관리 및 투명성, 공공성 제고 등 아직은 더 논의되고 공론화 되어야 하는 문제 등이 있다”고 말했다.

이명진 성산생명윤리연구소 소장은 태아의 생명을 존중한다는 프로라이프 운동가로서 잔여배아 문제 해결의 의견을 전했다. 이 소장은 “보조생식술로 알려진 시험관 기술은 현재 성공률이 15% 이하이기에 그동안 경제적 비용과 난임치료에 참여하는 여성의 건강을 고려해 한번에 많은 배아를 생산해오고 있다”며

“현재 35세 미만에서는 한 번에 2개, 35세 이상 부부에서는 3개까지 배아를 사용하도록 기준을 정하고 있다”고 설명했다.

2021년 보건복지부 자료에 따르면, 2016년 생성된 배아는 33만 4687개였으나, 2020년에는 49만 7133개로 약 50% 가까이 증가했으나, 생성배아 대비 이용률은 점점 떨어져 2016년에는 32.6%였지만 매년 감소해 지난해에는 27.7%로 떨어졌다. 생성된 배아 대비 폐기되는 배아 역시 증가해 2016년에는 15만 6713개로 전체 배아 대비 생성량의 46.8%였지만, 지난해에는 25만 2930개가 폐기돼 전체 대비 50.9%로 절반이상 폐기되는 것으로 나타났다.

이명진 소장은 “복지부는 2021년 9월 17일 ‘생명윤리 및 안전에 관한 법률 시행령 및 시행규칙’ 개정안을 입법예고하면서, 폐기 기한을 30일에서 6개월로 연장하고 기관위원회가 승인하면 배아 보존기간을 5년 이상으로 할 수 있도록 했으나 누적된 잔여배아와 새로 증가하는 잔여배아를 처리하기는 역부족”이라며 “가능한 생성배아의 수를 줄이고, 잔여배아의 윤리적 처리방법이 필요하다”고 밝혔다. 이 소장은 잔여배아 누적문제를 해결하기 위해 배아 생산수를 줄일 것을 제안했다. 그는 “잔여배아 누적 문제를 해결하기 위해서는 생산을 가능한 적게 해야 한다. 의학 기술에 맞게 배아 생산을 해야 한다”며 “생명을 죽이는 잔여배아 폐기문제를 해결하기 위해서는 한 번에 사용할 수정란만 생성해야 한다”고 주장했다. 또한 이 소장은 “우리 프로라이프 입장에서 볼 때, 배아 입양과 영구보존의 두 가지 방법에 동의할 수 있다”며 “반대로 냉동 해제를 통한 배아의 자연 폐기는 생명연장을 위한 예외적 수단 사용의 철회라는 맥락에서 냉동을 해제함으로써 잔여배아의 폐기를 받아들이는 방법으로 해석될 수도 있으나, 배아의 자기 결정권을 담보할 수 없는 방법이기엔 동의하기 어렵다”고 말했다. 이 소장은 “배아 입양은 다양한 윤리적, 법적, 도덕적 문제를 안고 있다. 입양한 배아가 출산해 성장한 후 자신의 출생비밀에 대해 알려줄 것인지, 알려준다면 어느시기에 어떤 방법으로 알려줄 것인지 논의가 필요하다”며 “배아 입양이 살아있는 생명인 잔여 배아에게 출생의 기회를 주는 방법이기엔 하지만, 입양된 배아가 건강하게 성장하고 보호받을 수 있도록 법적 보호장치를 마련해야 한다”고 강조했다. 이어 그는 “잔여배아로 두는 것은 인위적으로 파괴해 죽이는 일이 아니기에 윤리적으로 맞는 방법이기엔 하나, 인간으로 출생하지 못한 상태로 지속적으로 머물게 하는 것은 인간의 욕망의 댓가를 잔여 배아가 치르고 있는 것 같아 마음이 편하지 않다”며 “생물학적 친모의 출산이 없고 배아 입양의 길이 보이지 않는 상황에서는 무기한 냉동을 통한 과도한 기계적 개입에 의존해 생명을 유지시키는 것이 일종의 무의미한 연명이 아닌지도 검토해봐야 하는 의견도 있다”고 밝혔다.

III. 인공지능면접과 편향의 문제점

1. 공공기관의 AI 면접, 무엇이 문제인가 (KOICA, 한전KDN 사례)

정보인권단체가 공공기관의 AI 면접 정보공개 소송에서 ‘일부 승소’했습니다. 이제 AI는 공상과학이 아니라 현실입니다. AI 면접도 그 중 하나죠. 면접관의 주관적인 요소가 배제되니 공정하겠네요! 인공지능이 인간을 면접한다고요? 감히! AI 면접에 대한 막연한 기대감이나 거부감보다는 AI를 어떻게 적용하고 활용하는지, 그 과정과 절차는 올바른지 능동적으로 참여해 살펴보는 일이 더 중요합니다. 그런 의미에서 이번 AI 면접 정보공개 소송 ‘일부 승소’는 값진 결과로 생각합니다. 공공기관의 AI 면접에 관한 더 많은 정보를 마땅히 국민에게 공유해야 합니다. 소송에 참여한 정보인권단체의 공동 논평을 여러분께 공유합니다. 우리 단체들은 2022년 5월 26일과 6월16일, AI면접 등 공공기관의 채용 과정에 도입된 인공지능 시스템에 대하여 제기했던 정보공개 거부처분 취소소송에서 일부 승소하였다. 우리 단체들은 이번 소송에서 인공지능 채용 도구를 무분별하게 도입해 온 공공기관들이 법적으로 요구되는 자료들조차 구비하지 않는 등 심각한 책무성 부족을 확인하였다.

이제 인공지능(AI)으로 면접을 보는 시대입니다. 하지만 인공지능을 만드는 것도 운용하는 것도 사람입니다. AI 접속오류로 불합격 처리한 뒤 재응시 기회도 주지 않고(한국국제협력단), 용역업체에 모든 업무를 일방적으로 맡겨버리는 행태(한전KDN)은 무책임합니다.

법원, AI 면접 정보공개청구 정당성 직접 언급

우리 단체들은 지난 2020년, 공공기관에서 사용하는 인공지능 채용도구의 △공공기관의 공정한 채용절차 준수 유무 △개인정보 침해 여부 검토 △AI면접의 차별성과 편향적 결과 검토 등의 실태를 파악하고자 인공지능 채용을 진행한 13곳의 공공기관에 정보공개를 청구하였다. 그 중 단순한 계약 자료 등 기초적인 정보에 대해서도 대부분 부존재한다고 답변했던 기관인 한국국제협력단, 그리고 AI면접과 AI서류평가 등의 채용도구를 적극적으로 사용하고 있는 기관인 한전KDN 두 기관을 대상으로 2020년 10월 정보공개 거부처분 취소소송을 제기하였다. 그리고 약 1년 6개월 이후인 2022년 5월(수원지방법원 한국국제협력단 사건)과 6월(광주지방법원 한전KDN 사건)에 모두 원고 일부 승소 판결을 받았다. 각기 공개가 인용된 정보들은 다르지만, 우리가 청구한 정보 중 부존재하거나 이미 공개되었다고 볼 수 있는 정보를 제외한 나머지 기관이 보유하고 있는 모든 정보는 공개하라는 것이 법원의 일관된 판단이었다. 법원은 특히 정보공개 청구의 정당성을 판결문에서 직접 언급하였다. “이 사건 정보를 공개함으로써 공공기관의 계약 관련 정보에 관한 국민의 알 권리를 충족하고, 용역업체를 통해 인공지능 프로그램을 면접에 활용하는 과정에서 개인정보가 안전하게 잘 관리되고 있는지 등을 확인할 수 있는 계기가 될 수 있다.”

AI 면접 결과... 합리적 설명 자료 없음 확인

이번 소송을 통해 우리 단체들은 공공기관이 채용 과정에 인공지능 도구를 도입해 판단을 자동화하면서도, 그 결과에 대해 적절하고 합리적인 설명을 보장하기 위한 자료를 갖추어 두지 않았음을 확인하였다. 피고 기관들은 인공지능 도구를 제공하는 사기업에 채용절차를 일임한 채 의사결정에 필요한 자료는 물론이고 공정한 채용절차, 개인정보보호법, 나아가 공공계약에 있어 책임을 다하기 위해 필요한 자료들 또한 구비해두거나 점검하지 않고 있었다. 특히 피고 기관 중 하나인 한국국제협력단의 경우, 2019년 AI 면접 프로그램 접속 오류로 면접이 중단된 구직자에게 재응시 기회를 주지 않고 불합격 처리하였고 이로 인해 감사원에서 주의 조치까지 내린 사실이 있다. 그러나 해당 기관은 ‘인공지능 프로그램 구동 과정에서 발생하는 돌발상황 및 응시자 민원에 대한 해결 내역을 알 수 있는 문서’, ‘인공지능 프로그램이 채용에 실제 투입될 당시 인공지능의 기능별 오류/오차율’ 등에 대한 정보공개 청구에 대하여 해당 정보가 존재하지 않는다고 답변하였으며, 소송 과정에서도 기관의 AI면접이 문제없이 진행되었고, 그렇기에 그것을 기록한 정보도 없다고 답변하는 무책임한 모습을 보였다. 한편 피고 기관 중 하나인 한전KDN의 경우 2019년부터 AI면접을 사용하며 이를 점수에 반영하는 등 적극적으로 활용해왔으며 최근에는 AI면접을 넘어 자기소개서까지 자동화된 평가에 맡기는 AI서류평가 소프트웨어를 사용하고 있다. 해당 기관은

채용 과정 중 1차 서류전형을 오로지 AI서류평가로 진행하여 구직자를 탈락시키거나, 최종 합격을 위한 전형을 AI면접+AI자기소개서 평가로 구성하는 등 완전히 자동화된 방식을 통해 채용을 진행하며 사기업의 인공지능 프로그램에 의존해왔다. 그럼에도 한진KDN은 AI면접시 주어지는 질문 사항에 관한 어떠한 검토도 없이 그 내용을 용역업체에 포괄적으로 위임한 채 채용의 의사결정자에게는 AI면접과 관련한 아무런 자료도 제공하지 않았다. 채용 당사자나 감독기관이 채용 당락에 대한 이유를 물어도 제대로 설명할 수 없는 상태인 것이다. 인공지능 면접을 수행하는 공공기관이라면 더 책임 있고 꼼꼼한 모습을 보여줘야 할 텐데... 뻔뻔하고 무책임한 모습은 실망스럽습니다.

무분별한 인공지능 도입... 제도적 조율 관리 필요

이러한 공공기관의 무분별한 인공지능 도입과 그에 비해 한참 부족한 책무성, 설명불가능성은 국내외 인공지능과 관련된 인권 규범에 모두 역행하는 것이다. 지난 5월 국가인권위원회가 발표한 ‘인공지능 개발과 활용에 관한 인권 가이드라인’(2022. 5.)은 공공기관이 인공지능의 개발과 활용 계획 등을 사전 공개하고 관련 당사자들의 의견을 공청회 등으로 수렴할 것을 권장하고 있으며, 나아가 인공지능을 통한 의사결정을 할 때 설명할 수 없는 인공지능을 활용해서는 안 된다는 것을 분명히 하고 있다. 또한, 인공지능에 의한 자동화된 의사결정에 있어 영향을 받는 당사자가 그 결정의 이유에 대한 설명을 듣고 당사자 진술을 하며 이의를 제기할 수 있어야 한다는 점, 완전히 자동화된 의사결정으로만 개인에게 법적 효력 또는 생명·신체·정신·재산에 중대한 영향을 미치는 일은 제한되어야 하며 이 경우 당사자에게 해당 방식을 거부하거나 인적 개입을 요구할 수 있는 권리가 보장받아야 한다는 점도 강조되어 있다. 국외에서는 일반적인 가이드라인을 넘어 고용 영역에서의 인공지능 기술 사용에 대한 보다 강력한 규제가 마련되었거나 준비되고 있다. 유럽연합은 2021년 4월 ‘인공지능법’(AI ACT)을 발의하고 입법 절차를 밟고 있는데, 해당 법안은 고용 영역에서의 인공지능을 고위험 인공지능으로 분류하고 위험 평가 및 완화, 데이터셋의 고품질 보장, 기록 및 문서화, 정보 공개, 인적 통제 등의 의무를 부과하고 있다.

미국 정부와 의회는 고위험 자동화된 의사결정 시스템에 영향 평가 의무를 부여하는 연방 입법을 추진하고 있으며, 일리노이주와 뉴욕시 같은 지방정부는 채용과정의 인공지능에 특별히 주목해 ‘인공지능 영상 면접법(Artificial Intelligence Video Interview Act)’, ‘자동화된 채용 결정 도구에 관한 법(Automated employment decision tools)’ 등을 통과시켜 인공지능 면접에 대한 정보 제공, 편향성 감사의 의무화 등의 규제를 마련하고 시행 중에 있다.

무책임 도입 중단, 차별 방지 제도적 대책 시급

현재도 다수 공공기관은 채용 과정에 인공지능 기술을 도입하여 사용하고 있다. 그러나 채용은 당사자에게 현재와 미래의 생계로 이어지는 경제적 기회를 제공하는 중대한 의사결정이다. 공공기관이 채용에 사용되는 기술의 의사결정에 대하여 설명할 준비를 갖추지 않고, 검증도 감독도 없이 민간회사에게 일임하는 방식의 무책임한 인공지능 도구의 도입은 중단되어야만 한다. 또한, 국가와 감독기관은 공공 영역의 인공지능 기술 사용에 대한 인권영향평가 등 인권에 중대한 영향을 미칠 수 있는 영역에서의 인권침해와 차별을 방지하기 위한 제도적 대책을 서둘러 마련해야 할 것이다.

2. 불균형 데이터가 초래하는 불균형 AI

“사람들이 인공지능(AI)에 대한 환상을 갖고 있는 것 같습니다. AI를 맹신하면 안 돼요. 이미 AI는 알게 모르게 편향된 결단을 내리고 있고, 그 피해를 보는 건 취약층과 소수층입니다.”

AI 전문가들의 한결같은 말이다. 성별이나 인종, 연령대에 대한 AI의 차별은 이미 여러 차례 논란이 됐다. 마이크로소프트에서 개발한 대화형 AI ‘테이’의 인종 차별적 발언, 아마존 채용 AI의 성차별적 판단, 국내 대화형 AI ‘이루다’의 소수자 혐오 발언이 그 예다. 단, 기술적 의미에서 AI의 편향(bias)은 조금 더 넓은 범위를 포괄한다. 최재식 KAIST AI대학원 교수는 “사회적·윤리적 이슈에 대해서만이 아니라, 사실관계를 예측하는 AI가 한쪽으로 치우치는 문제도 AI 편향성에 포함된다”고 말했다. 가령 공장의 폐쇄 회로(CC)TV를 보고 사고 위험성을 예측하는 AI가 있다고 가정해보자. 이 AI는 수많은 데이터를 학습하

지만, 그 데이터 중 공장 폭발이란 결과에 도달한 것은 극히 일부다. AI가 이런 데이터로 계속 학습할수록 공장 폭발이란 예측할 가능성은 점점 줄어든다. 이 때문에 현장에서 공장 폭발의 징후가 나타나도, AI는 자신이 학습한 데이터에 따라 공장 폭발이란 경우를 배제해버릴 수 있다. 이처럼 개발자가 의도하지 않았음에도 AI가 편향된 예측을 내놓는 문제 역시 AI 편향 문제에 해당한다. 2019년 미국 델러스에서 열린 IT기업 아마존의 채용박람회 현장. 아마존은 2014년부터 채용에 도입할 AI를 개발했으나 여성 지원자를 차별하는 문제를 해결하지 못하고 2017년 폐기에 이르렀다.

AI 편향성을 일으키는 주요 원인 중 하나는 데이터의 불균형이다. KAIST 인공지능공정성연구센터장을 맡고 있는 유창동 KAIST 전기 및 전자공학부 교수는 “데이터가 많은 집단에 대해서는 AI가 학습을 통해 그 집단의 비교적 보편적인 결과를 도출할 수 있지만, 데이터가 적은 집단에 대해서는 AI의 결론이 그 집단을 대표한다고 볼 수 없는 경우가 발생한다”고 말했다. 한 예로 직원 대다수가 한국인이고, 소수만이 외국인 직원으로 구성된 회사에서 채용 AI에 이전 합격자의 이력서 데이터를 입력하는 경우를 들 수 있다. 이 경우 다수의 한국인 직원 이력서를 학습한 AI는 한국인 지원자를 판단할 때는 합격 기준선이 비교적 합당하게 설정되지만, 외국인 지원자에 대해서는 몇몇 외국인 직원의 이력서 데이터로만 학습한 탓에 그 기준선이 비합리적으로 설정될 수 있다. 이로 인해 동등한 능력을 지닌 지원자임에도 한국인은 합격하고, 외국인은 탈락하는 결과가 나타날 수 있다. 물론 반대 경우도 나타날 수 있다. 이처럼 집단 간에 표본 수가 달라 발생하는 편향은 지금까지 사회적으로 차별 받은 계층에게 흔히 일어난다. 데이터 축적 역시 기득권에게 더 유리한 일이기 때문이다. 그래서 데이터 불균형은 곧 사회적 차별 문제로 직결되는 경우가 잦다.

편향되지 않아 보이는 데이터도 문제가 될 수 있다. 아마존의 채용 AI가 대표적 예다. 아마존은 2014년부터 채용 과정에 도입할 AI를 개발했다. 개발팀은 이전 10년간의 합격자 이력서를 학습시켰고, 이를 통해 AI는 지원자를 1~5점으로 평가했다. 당시 전 세계 아마존 소속 직원 가운데 남성의 비율은 약 60%로 남성이 약간 많기는 했지만 심하게 차이가 나보이지 않았다. 하지만 1년 뒤, AI가 여성 지원자를 차별한다는 사실이 발견됐다. AI는 ‘여자 체스 클럽 주장’ ‘여자 대학 졸업’ 등 ‘여자’라는 단어가 포함된 이력서에 대해 점수를 낮게 조정했다. 이후 개발팀은 데이터에서 성별에 대한 정보를 제거했지만 문제가 해결되진 않았다. AI가 자기소개서 등 지원자가 작성한 문서의 문체에 따라 성별을 암묵적으로 구분하는 현상이 발견됐다. 성별 데이터가 제거되도, AI가 스스로 성별을 유추할 수 있는 특성을 끄집어낸 것이다. 개발팀은 이 편향 문제를 끝내해결하지 못했고, 결국 2017년 팀 해체와 함께 채용 AI 개발 프로젝트는 폐기됐다.

열 수 없는 블랙박스, 딥러닝

AI 편향성 문제 해결에 가장 앞장선 건 세계적 정보기술(IT) 기업인 IBM, 구글, 페이스북이다. 유 교수는 “유럽연합(EU)에서는 편향된 AI를 법적으로 판매하지 못하도록 하고 있으며, 한국을 포함해 이외 지역에서도 정책적·법적 조치를 마련 중”이라며 “기업에서 공정한 AI를 개발하는 건 차별에 대한 사회적 책임을 넘어 자사의 AI 상품을 판매하기 위해서 반드시 선행해야 하는 과제”라고 말했다. AI의 편향성을 해결하기 위해서는 먼저 편향의 원인이 무엇인지 파악하고, 그 부분을 고쳐야 한다. 현재는 알고리즘보다는 데이터상의 문제를 파악하고 제거하는 연구가 더 활발히 이뤄지고 있다. 요즘 각광받는 딥러닝 AI는 그 속을 들여다보는 것이 불가능에 가깝기 때문이다.

다른 기계학습이 한두 단계(층, layer)에 걸쳐 데이터를 학습하는 것과 달리, 딥러닝은 3개 이상의 여러 단계에 걸쳐 데이터를 학습하는 걸 뜻한다. 용량이 크고, 형태도 다양한 빅데이터를 정교하게 학습하는데 탁월하기에 ‘AI=딥러닝’이라고 인식될 만큼 최근 AI 대부분은 딥러닝 방식으로 학습된다. 문제는 딥러닝 AI가 학습하는 여러 단계 중에는 사람이 알아낼 수 없는 숨겨진 단계(hidden layer)가 많다는 것이다. 그러다 보니 개발자조차 AI가 왜 그런 방식으로 작동했는지 알 수 없다. 이 탓에 ‘답은 맞는데 풀이과정은 알 수 없는 AI’로 비유되곤 한다. 이런 AI를 뜯어보기 위해 최근 들어 사람이 이해하는 형태로 작동방식을 설명하고 제시할 수 있는 설명가능 AI(XAI=eXplain AI)란 기술이 주목받고 있다. 하지만 당장 편향성 문제를 해결할 수준은 아니다. KAIST 설명가능인공지능연구센터장이기도 한 최재식 교수는 “궁극적으로 딥러닝 AI의 오류를 제거하기 한 XAI 연구가 급격히 증가하는 추세인 하나 현 상황에서 AI의 편향성을 해결하는 데는 데이터의 편향을 제거하는 것이 더 적합하다”고 말했다.

이에 따라 데이터상의 문제를 파악하고 제거하는 연구가 주로 이뤄지고 있다. 이 때 ‘불공정하다’는 기준을 가를 수학적 정의가 필요하지만, 수식에 따라 공정에 대한 판단이 변해 일률적으로 적용할 수가 없다. 마치 똑같은 사안에 대해서 누구는 공정하다고 하고 누구는 불공정하다고 하는 것 같이, 수학적으로도 어떤 수식을 쓰느냐에 따라 불공정을 달리 판단한다.

그렇다 보니 데이터 편향을 줄이기 위한 기술은 대부분 편향성과 관련이 있는 변수를 찾아 조정하는 데 초점을 맞추고 있다. 가령 채용 AI에 입력된 한 지원자의 정보에서 성별을 남성에서 여성으로 바꿔보거나, 나이를 20세에서 50세로 바꾸면 성별과 나이가 합격과 불합격을 가르는 데 불공정한 영향을 얼마나 끼치는지 도표로 표시된다. 이렇게 데이터의 불공정성을 측정하다 보면, 개발자가 의도한 것과 달리 차별적 결과를 내놓는 특정 변수를 발견할 수 있다. 그 변수를 ‘보호변수(Protective Attribute 또는 Sensitive Attribute)’로 지정해 불공정성을 최소화할 수 있다. 유 교수는 “보호변수는 AI의 판단에 영향을 주지 않도록 설정한 변수를 의미한다”며 “공정한 AI를 만들기 위해서는 성별, 인종, 지역 등 편향된 결과를 유발할 수 있는 요소를 보호변수로 지정하고 결과에 미치는 영향력을 조절하는 것이 매우 중요하다”고 말했다.

보호변수를 지정했다면, 다양한 알고리즘을 통해 보호변수의 영향력을 조절할 수 있다. 대표적 예로 IBM이 2018년에 내놓은 ‘AI 공정성(Fairness) 360’이라는 오픈소스 툴킷(소프트웨어 개발 도구 모음)가 있다. 이 툴킷은 편향을 완화할 수 있는 10가지 알고리즘을 포함하고 있다. 이 가운데에는 AI가 학습할 데이터에서 보호변수의 가중치를 수정하는 알고리즘, 학습 과정 중 보호변수의 영향을 감소시키는 알고리즘도 있다. 보호변수를 조절하는 것 외에도 AI의 학습 전후에 편향성을 감지하고 완화시킬 수 있는 알고리즘도 포함돼 있다.

이처럼 공정한 AI를 만들기 위한 다양한 방안과 도구가 마련되고 있지만 안타깝게도 완전히, 절대적으로 공정한 AI를 개발하는 일은 사실상 불가능하다. 유 교수는 “공정성이라는 개념 자체가 누구도 하나로 정의할 수 없는 개념”이라며 “병원, 법원, 기업에서 바라보는 공정성도 제각기 다르기 때문에 영역별로 공정성에 대한 개념을 확립하는 게 우선 과제”라고 말했다. 더군다나 공정성은 계속해서 진화하고 있는 개념이다. 예를 들어 과거에는 국내에 남성중심적 사고가 지금보다 더 강했다. 때문에 지금이라면 논란이 될만한 장면이 불과 10년 전만 해도 대중매체에서 거리낌 없이 등장했다. 유 교수는 “과거에 공정하다고 생각했지만 지금은 공정하지 않다고 여겨지는 것들이 많다”며 “하나의 공정한 AI를 만들었다고 끝나는 일이 아니라, 컴퓨터 바이러스 백신 프로그램을 계속 업그레이드하는 것처럼 공정한 AI도 현실의 공정성을 반영하는 수정을 끊임없이 지속해야 한다”고 말했다.

2019년 10월 국제학술지 ‘사이언스’에 의료 서비스가 필요한 환자를 예측하는 AI가 인종 차별을 한다는 연구결과를 낸 브라이언 파워스 브리검여성병원 부원장도 “알고리즘 편향을 측정하는 툴은 존재하지만 이를 활용한다 해도 편향의 모든 원인을 제거할 수 없다”며 “모든 알고리즘에 적용 가능한 하나의 해결 방법은 존재하지 않는다”고 말했다. 결국 AI의 공정성은 그 AI를 만드는 사람들 손에 달렸다. 불공정을 판가름할 수학적 기준을 정하는 일도, 그런 불공정을 조절할 알고리즘을 설계하는 일도 각 개발자들의 판단에 따르기 때문이다.

사람 손으로 만드는 공정한 AI

그래서 개발자들이 공정함에 대해 올바른 판단을 하기 위한 가이드라인이 필요하다. 법적·정책적 제도다. 이상욱 한양대 철학과 교수는 “한국에서 ‘윤리’는 일종의 바른생활과 같은 좁은 의미로 사용되지만, 해외에서 ‘ethics’는 바른생활을 뒷받침할 법적·정책적 제도까지 포함하는 개념”이라고 말했다. 각국에서는 공정성을 포함해 AI 윤리에 대한 법적·정책적 제도를 마련하기 위한 논의가 활발하다.

또 하나 필요한 것은 교육이다. AI를 배우는 단계에서부터 올바른 윤리관을 정립할 필요가 있다. 미국 매사추세츠공대(MIT), 하버드대 등 우수 대학은 AI 윤리 교육 프로그램을 별도로 만들었다. 뉴욕대에서는 데이터사이언스 학위를 취득하기 위해 ‘책임감 있는 데이터사이언스’란 AI 윤리 과목을 이수해야만 한다. 국내에서는 2022년 초·중·고교에서 AI 윤리를 포함한 ‘AI 교육’이 정식 도입될 예정이다. 최 교수는 “알고리즘을 만드는 개발자뿐만 아니라, AI 전체를 설계하는 기획자까지 모두 교육이 이뤄질 때 효과가 나타날 것”이라고 말했다.

IV. 윤리적인 문제를 AI를 통해 해결할 수 있나?

안락사 윤리 판단하는 AI ‘델파이’ 등장

철학 분야 베스트셀러 ‘정의는 무엇인가’에 등장하는 딜레마에 대해 윤리적 대안을 제시하는 인공지능(AI)이 등장했다. 미국 시애틀 앨런(Allen) 연구소 AI 연구팀이 개발한 ‘델파이(Delphi)’ 이야기다.

연구진은 델파이에게 사람 4명과 1명이 각기 묶여있는 2가지 철로를 전차가 지나가야 한다면 어떤 선택을 할 것인지 물었다. 이에 대한 델파이의 대답은 “4명을 살리기 위해 1명을 치는 선택을 하는 것이 좀 더 윤리적으로 허용 가능할 것”였다.

30일 서울대 인공지능 정책 이니셔티브(SAPI) ‘인공지능정책 국제 컨퍼런스 2021’을 열고 델파이 프로젝트를 이끄는 최예진 워싱턴대 교수를 발표자로 초대했다.

앨런 연구소는 지난 10월 윤리적 판단을 내리는 AI 프로그램 델파이의 데모인 ‘Ask Delphi’를 누구나 사용할 수 있도록 공개했다. 입력 창에 궁금한 질문을 입력하면 AI가 바로 답변을 제시한다.

델파이 활용 예시에 대해 최예진 교수는 “어린이를 구하기 위해 꿈을 죽이는 것은 괜찮냐는 질문에는 동의하면서, 어린이를 구하기 위해 핵폭탄을 터뜨리는 것에 대해서는 옳지 않다고 대답한다”고 말했다. 이어 “질병으로 극심한 고통을 느끼는 고령의 환자에게 의사가 인공호흡기를 계속 쓰게 해야 할지 아니면 치료를 중단해야 할지와 같은 어려운 윤리 문제에 대해서도 대답할 수 있다”고 강조했다.

윤리 문제 이외 일반적인 지식이나 문화, 통념에 기반한 판단도 가능하다는 설명이다. 유당불내증이 있는 사람이 우유를 마시는 것, 장례식에 밝은 주황색 셔츠를 입고 가는 것에 대해 델파이는 각기 바람직하지 않다고 대답한다.

특히 AI가 다양한 요소가 관여되는 상황을 파악할 수 있도록 개발하는 것이 관건이었다는 설명이다. 최 교수는 “잔디를 깎는 것은 괜찮고, 한밤중에 잔디를 깎는 것은 괜찮지 않으며, 외딴 곳에 떨어져 살면서 한밤중에 잔디를 깎는 것은 괜찮다고 답하는 식이다. AI가 여러 요소가 개입하는 상황에 맞춰 판단하는 것은 인간과 달리 쉽지 않다”고 말했다.

GPT-3로 상식·윤리 판단 어려워...상식 개념 뱅크 활용성 기대

GPT-3와 같은 기존 초거대 AI 언어모델이 규모의 경쟁을 하는 방식으로는 상식·윤리를 반영해 판단하는 AI를 개발하기 어려울 것으로 보인다.

최예진 교수 연구팀은 GPT-3 계열 모델들과 델파이의 윤리적 판단 능력을 비교하는 연구를 진행했다. 연구 결과, 사례 예시를 주지 않은 GPT-3 모델의 성과가 크게 떨어졌다는 설명이다. 최 교수는 “이 연구에 따르면 GPT-3와 같은 기성품(off-the-shelf) 모델은 인간 가치와 규범을 스스로 학습하는 일에는 완전히 희망이 없는 정도다. 예제를 많이 줄수록 성과는 나아지지만 매번 많은 예제를 보여주고 그 답을 얻으려 하면 컴퓨팅 자원을 많이 소모하게 된다”고 말했다. 이어 “더 큰 네트워크와 더 깊은 레이어를 사용한다고 극복할 수 있는 문제가 아니다. 고층 빌딩을 아무리 지어도 달에 닿을 수는 없는 것과 같다. 사람 표현(representation)과 기계 표현의 차이(gap)를 줄여야 한다”고 강조했다.

동시에 “반면 기성품 모델을 상식 개념 뱅크(Commonsense Norm Bank)를 통해 학습시켰을 때, 이들의 도덕 판단이 크게 개선됐다. 이는 뉴럴 모델을 서술적 윤리 관점을 통해 학습시키는 방법에 대한 가능성을 보여준다. 언어 기반 상식 도덕 모델의 가능성과 한계를 연구하는 것이 델파이 연구 목표”라고 말했다.

V. 4차 산업혁명과 불평등

4차 산업혁명의 기회와 위협

4차 산업혁명(4IR)은 기회일까, 위협일까? 기회와 위협은 전 세계 모든 국가, 국민에게 균등하게 나타날 것인가? 우리나라의 기술/산업 경쟁력에는 어떤 영향을 끼칠 것인가? 또, 사회적 과제 해결에 도움이 될 것인가, 아니면 더 큰 문제를 만들 것인가? 이는 많은 국민들이 4IR에 대해 갖고 있는 공통 관심사일 것이다. 이번 글에서는 국내/외 여러 전문가들의 소견에 필자 나름의 주관적 판단을 덧붙여서 이 문제를 정리해 보려 한다.

기술-경제-사회 상호작용

먼저 인류문명 발전에 대한 종래의 2가지 상반된 이론 즉, 기술결정론과 사회구조론(또는 기술의 사회결정론)의 의미를 살펴보자. 앨빈 토플러, 다니엘 벨, 마셜 맥루한 등이 주장한 기술결정론은 기술이 독립 변수로서 문명 발전을 주도한다고 설명한다. 반면, 사회구조론은 기술은 종속변수(즉, 도구)로서 사회적 요구/기대에 따라 선택, 발전된다고 한다. 기술결정론은 기술의 영향력을 너무 크게 인정하고 기술과 사회 간의 관계를 지나치게 단선적(單線的)으로 설명한다는 비판을 받고 있다. 반면, 사회구조론은 기술의 사회적 영향 즉, 신기술이 개인과 사회를 변화시키는 것을 너무 가볍게 본다는 비판을 받고 있다. 결국, 두 이론 중 어느 하나가 늘 옳다고 할 수는 없으며, 기술과 경제-사회는 서로 영향을 끼치면서 지금에 이르렀다고 보아야 할 것이다.

슈밥(2018)은 기술에 대한 종래의 2가지 견해가 바람직하지 않을 뿐만 아니라 위험하며 세 번째 견해 즉, “모든 기술은 정치적이다(All technologies are political)”라는 전제가 필요하다고 하였다. 특히, 4IR 기술은 발전/확산 속도가 매우 빠르고 (따라서, 사회가 개입할 시간적 여유가 별로 없음), 사용자에게 많은 권한을 부여하며 (예: 개인이 3D 프린터로 총기를 만들 수 있음), 인간을 감싸고 인체에 내재되기 (예: 사물인터넷, 인공지능) 때문에 기술과 사회는 분리할 것이 아니라 밀접하게 연결된 것으로 보아야 한다고 하였다. 따라서, 신기술을 활용해서 구현되는 시스템 내부에 포함되어야 할 가정, 가치, 원칙 등이 사회에 끼칠 영향력과 사회구조 및 신분 변화에 대해 이해관계자들이 함께 논의/심의하는 과정이 필요하다고 하였다. 기술이 사회를 바꾸는 것은 물론, 거꾸로 사회가 기술 발전을 촉진/억제하는 것이 모두 가능하고 그 속도, 범위, 영향력이 매우 크다는 점을 인정한다면, 기술-사회 간 상호작용을 효과적으로 조정/통제하는 일은 이제 모든 국가의 최상위 어젠다 중 하나가 되어야 할 것이다. 그러기 위해서는 최소한, 기술(개발)자는 소비자들이 기술을 어떻게 활용하게 될 것인지에 대해, 또 소비자는 신기술이 만들어내는 기회와 위협에 대해 정확하게 알아야 한다.

인류의 도전과제와 기술의 기회

UN은 2010년에 MDG(Millennium Development Goal, 2015년까지 달성하고자 하는 8대 목표)를 발표한 이래, 2016년 1월에 SDG(Sustainable Development Goal, 2030년까지 달성할 17대 목표)를 발표하였다. 17대 목표에는 빈곤 탈피, 굶주림 해방, 건강/웰빙, 고품질 교육, 양성 평등, 식수와 위생, 경제적/청정 에너지, 노동의 질 향상과 경제성장, 산업의 지속가능 인프라와 혁신, 불평등 감소, 지속가능 도시/커뮤니티, 지속가능 소비/생산, 기후제어, 지속가능 해양생태계와 육상생태계, 평화/공정/포용 사회, 목표 달성을 위한 협력 등이 포함되어 있다. UN의 SDG는 국가, 지역, 인종, 이념 등과 관계없이 전 인류가 추구하는 공통가치라 해도 무방할 것이다.

AI, IoT, 클라우드, 빅데이터, 모바일 기기, 5G 통신 등 고도화된 디지털 플랫폼은 개인생활, 기업활동, 정부서비스 등 인류의 모든 경제/사회생활에 긍정적 영향을 끼친다. 개인/기업은 실시간으로 수집-분석되는 데이터에 입각해서 합리적 대안을 선택-실행하고, 불필요한 시간적 지체나 비용/자원의 낭비를 없애며, 현실공간과 가상공간이 일체화된 환경에서 활동할 수 있게 된다. AR/VR은 미지의 세계나 시/공간을 경험하게 해 줌으로써 인간의 통찰력이나 영감을 키워 준다. BT, IT, NT 등 융합기술은 농업/환경/에너지 등을 혁신시켜서 기후 변화와 무관하게 식량의 생산/증산, 영양가는 높으면서 저비용인 대체식

품 개발, 해수의 담수화(淡水化), 오염된 물/공기/토양의 정화(淨化) 등을 가능하게 해 준다. NT는 강도, 내구성이 좋으면서 초소형인 소재와 장치를 경제적으로 생산-소비할 수 있게 해 준다. 3D 프린팅의 보급에 따라 누구나 쉽게 집이나 거주지역에서 각종 생활용품을 직접 만들어 쓰고 저비용으로 가구나 집도 지으며 심지어 음식도 만들어 먹을 수 있게 된다. 각종 질병, 의료기록, 나아가 개인의 일상생활 기록이 빅데이터로 축적되고 AI로 분석되며, 유전자 편집 기술을 제약 없이 쓸 수 있는 단계에 이르면 '4P 의료'(Predictive, Preventive, Participative, Personalized: 예측, 예방, 참여, 개인화/맞춤 의료)가 실현된다. BT, NT, 3D 프린팅 등의 융합기술을 통해 손상된 장기/신체를 대체할 수 있게 되면 건강은 물론, 장수를 누릴 수 있게 된다. 블록체인은 중앙집중화된 권력의 분산, 중개자/관리자가 필요 없는 경제체제, 각종 지식/콘텐츠의 창작과 소비의 확대를 가능하게 할 것이다. AI와 로봇/드론의 발달에 따라 인간은 단순, 반복적이거나 힘들고 위험한 작업에서 벗어나 보다 창의적, 감성적 활동과 지적/영적 계발에 집중할 수 있을 것이다.

신기술이 만드는 경제/사회 위협

기술이 인류문명 발전에 커다란 기여를 한 것은 부인할 수 없는 일이지만, 여러 가지 폐해를 만든 것도 사실이다. 예를 들면, 2차 산업혁명은 환경오염, 자본가-노동자 계급구조, 제국주의 확산등의 폐해를 만들었다. 정보혁명도 디지털 디바이드(divide, 양극화), 사생활 침해, 사이버 범죄, 불건전 콘텐츠 유통 같은 문제를 야기하였다. 4IR 경우, 이전 혁명에 비해 훨씬 더 큰 기회가 있는 것만큼, 더 큰 위협을 예상할 수 있다. 인류가 이를 지혜롭게 준비, 대응한다면 유토피아(Utopia)가 되겠지만, 그렇지 못한다면 디스토피아(Dystopia)에 이를 것이다. WEF의 Global Risk Report는 향후 등장하게 될 다음과 같은 위험요인들을 예시하였다(슈밥, 2018). 파괴적 무기(예: 바이오무기)의 보편화, 신물질이 인간의 건강이나 환경에 부정적 영향을 끼치게 됨, 청정에너지 등장에 따라 석유생산국가의 지정학적 영향력이 약화됨, 기후 변화 문제 해결 시도가 생태계에 예기치 않았던 부정적 영향을 끼치게 됨, 양자컴퓨팅 등장에 따라 기존 암호체계가 무용지물이 됨, (주어진 문제를 어떻게 풀었는지 유추할 수 없는) '블랙박스 AI'로 인해 경제 시스템이 위협받게 됨, 신경 기술을 이용해서 사람의 생각/행동을 읽을 수 있을 때 나타날 부작용 등이 그것이다.

또한, 인간이 AI, 로봇, 드론 등을 통제하지 못하는 상황이 될 경우, 일자리는 물론, 프라이버시, 나아가 생명이나 존엄성까지 위협받을 수 있다. 강인공지능 또는 일반인공지능(AGI), 나아가 초인공지능(ASI)의 도래 가능성이나 도래 시기에 대해서는 전문가들조차 조금씩 다른 견해를 갖고 있지만, 가능성이 낮다고 해서, 또 먼 훗날의 문제라고 해서 준비/대응조차 안 할 수는 없는 일이다. IoT 확산에 따라 실현될 초연결사회는 감시사회가 될 수도 있고 네트워크에 연결된 작은 개체 하나가 전체 사회(예: 교통, 통신, 에너지, 교육, 생산, 유통 등)를 마비시키거나 파괴하는 고위험사회가 될 수도 있다. 블록체인 기술의 첫 번째 응용인 비트코인을 통해 경험했던 것처럼, 신기술이 개발자의 의도와는 다른 부정적 결과를 야기할 경우 그 폐해는 고스란히 이용자의 몫이 된다. 유전자 편집이나 줄기세포 기술은 생태계의 질서를 위협하고 정신과 육체가 획일화된 인간을 양산함으로써 인류의 자멸을 초래할 수도 있다. NT는 육안으로는 확인할 수 없는 초미세 세계를 다루기에 그 과정이나 결과물을 통제하기 어렵다는 것이 고민거리이다. 게다가 개별 기술이 갖고 있는 불안/위험요인이 기술 융합을 통해 증폭된다면 이는 더욱더 심각한 상황이 될 것이다. 모든 연구자/기술자, 기업, 국가가 선의(善意)만 갖고 새로운 기술과 제품/서비스를 개발할 거라고 믿는다면 이는 너무 순진한(naive) 생각이다. 설사, 그것을 선의로 개발했다고 치더라도 예상치 않았던 부작용(side effects)이 나타날 가능성은 충분히 있다는 점을 감안해야 한다.

우리나라의 기회와 위협- 일자리

위에서 4IR이 인류에게 엄청난 기회와 그에 못지않은 위협이 되리라는 전문가들의 예상을 축약해서 소개했는데.. 그렇다면, 4IR은 우리나라에 어떤 기회와 위협으로 다가오게 될까? 국내에서 가장 많이 회자된 것은 AI와 로봇에 의한 일자리 감소, 'IT 강국'으로서의 기대 또는 자신감, 반면, 4IR 준비도가 세계 25위라는 UBS(스위스 은행)의 평가 등일 것이다. 일자리의 미래에 대한 WEF/슈밥, Techcast사 윌리엄 할랄 교수, 미래학자 토마스 프레이 등의 예측은 낙관론일 수도, 또 비관론일 수도 있다. 분명한 것은 ① 단순, 반복적 작업이나 일정한 규칙을 정의할 수 있는 고임금 직업 경우, 머지않아 AI나 로봇으로 대체될 수 있다는 점, 그러나 ② (지난 산업혁명 때도 그랬지만) 개인이든 기업이든 국가든 새로운 일자리를

많이 만들어 낸다면 기회가 될 것이고 그러지 못한다면 위협이 될 것이라는 점이다. 새로운 일자리는 당연히 신기술/신산업 영역에서 많이 생길 것이기에 디지털 기술은 물론, BT, NT, 그리고 융합기술의 개발과 활용에 집중 투자해야 한다. 할랄 교수는 “AI가 커버하지 못하는 거대한 미개척 영역(예: 창의성, 기업가 정신, 비전, 협업, 외교, 마케팅, 감독 등)이 존재한다”라고 했고, 프레이는 “(다가올) 물리적 경제는 (지금의) 디지털 경제의 5~6배 규모에 이를 것이다”라고 했다. 필자는 4IR이 진전됨에 따라 나타나게 될 일자리의 증감 규모는 지금의 국가별 경제규모나 수준에 비례하지 않을 것이라고 본다. 지난 수십 년간 축적된 기술, 경제, 사회의 내공 위에서 향후 5~10년 동안 정부와 기업, 전체 국민들이 어떻게 준비하는가에 따라 얼마든지 달라질 것이기에 지금 예단(豫斷)할 수는 없다는 말이다.

4IR 시대, 우리나라의 위상

4IR은 우리나라에 기회가 될 것인가 아니면 위협이 될 것인가? 필자는 기회보다는 위협이 훨씬 더 크기에 치밀한 준비와 실행이 없다면 지금 누리고 있는 기술/경제적 위상은 추락할 수밖에 없을 것이라 생각한다. 그 이유를 몇 가지만 꼽는다면, 첫째, 기술/산업 측면에서 우리나라는 기초과학, 원천기술, 기반산업 등이 선진국들에 비해 여전히 크게 뒤지고 있기 때문이다. 산업 시대 내지 정보화 시대는 ‘빠른 추종(fast follower)’ 방식의 공정혁신을 통해 세계 10위권의 경제대국이 될 수 있었는지 모르지만, 4IR 시대는 장기간의 실험과 시행착오 속에 축적된 내실과 기초 없이 ‘시장창조(first mover)’ 수준의 제품혁신을 이룩할 수는 없을 것이다. 반도체를 예로 든다면, 미국은 1968~2012년 중에 정부-민간이 약 1조불(약 1,100조원)을 투자한(출처: Lundstrom & Wong, 2012) 반면, 우리나라는 지난 10여 년간 삼성전자가 110조원을 투자했다고 한다(물론, 타 반도체 기업과 연간 최대 1천억원 수준의 정부 투자도 있었음). 이런 기적(?)이 4IR 시대에도 계속될 수 있을까? 우리나라에 비해 5배의 시간, 10배의 자금을 투자하고도 시장에서 2~3등 하고 있는 미국은 바보일까? 4IR의 기반 산업이 될 센서/부품, SW/AI, 데이터, 생산기계 등은 단기간에 집중적으로 투자한다고 해서 성과를 거둘 수 있는 기술/산업이 아니다. 대학/연구소-기업-정부 간 역할분담과 긴밀한 협력, 지속적 투자 등이 전제되어야 한다. 센서는 적어도 10~20년 전부터 그 중요성을 인지했지만, 글로벌 센서 시장 점유율은 현재 2% 수준이다. ‘IT 강국’이라는 표현은 적어도 4IR 시대에는 엄청난 착각이다. 여기에서 IT는 과거 전자산업으로 분류되던 휴대폰, 반도체, 디스플레이 등 HW와 내수용인 초고속통신망을 가리킬 뿐, 정작 중요한 SW는 글로벌 SW 시장 점유율이 1% 정도인 약소국이기 때문이다. 둘째, 경제 측면에서도 여러 가지 이유를 꼽을 수 있지만, 그중에서도 가장 심각한 것은 이미 시작된 플랫폼 경제에서 생태계 내의 주도권은커녕 적정 수준의 협상력조차 확보하지 못하고 있기 때문이다. 지금까지는 개별기업이 가진 차별화된 제품/서비스/브랜드만으로도 먹고살 수 있었는지 모르지만, 향후에는 미국의 FANG, 중국의 BAT 같은 데이터/콘텐츠/IT서비스 회사나 미국의 GE, 독일의 지멘스 같은 제조서비스 기업이 플랫폼의 오픈 및 관리자로서 시장가치의 대부분을 가져가게 될 것이다. 셋째, 사회 측면에서 우리나라는 4IR 시대가 필요로 하는 상생, 협력, 조화 등과는 거리가 먼 갈등과 대립이 심화되고 있기 때문이다. 빠르게 이룩한 경제성장의 대가로 여러 가지 사회적 해결 과제(예: 경제/사회 불평등, 전통적 가치관 붕괴)가 누적되어 있는 상황에서 점점 더 큰 과제(예: 고령화, 저출산, 일자리와 삶의 질, 다문화 확산)들이 불거지고 있는데 이를 해결할 수 있는 역량은 오히려 줄어들고 있다.