현재의 선택과 선의의 기술로서 AI

Current Choice and AI as Good Faith

오래전에 보아온 AI 미래

역사를 공부하는 것은 과거로부터 현재를 살펴, 미래를 예측할 수 있기 때문이다. 컴퓨터에 대한 역사도 그러한 맥락이다. 1950년대, 인공지능이라는 표현을 처음 사용한 이래, 인공지능의 성장은 기술 투자가 있었기 때문에 가능했다. 그럼에도 불구하고, 인공지능은 범용 능력을 갖춘 인간과 같은 수준에 이르지는 못하고 있으며, 모라벡의 역설(Moravec's Paradox)처럼 인간이라면 쉽게 할 수 있는 인지, 운동, 표현 등이 기계 내지 인공지능에게는 어려운 것이 사실이다. 우려와 달리, 2020년 현재 터미네이터와 같은 인류를 위협하는 인공지능 또는 로봇의 실체는 찾아보기 어려운 이유다. 인공지능에 대한 막연한 두려움을 가질 필요는 없다. 대신 어떻게 인류에게 유익한 인공지능을 개발하고, 활용할 것인지, 그리고 막연한 두려움을 잠재울 수 있는 기술적, 제도적 방안을 강구할 필요가 있다. 선의(Good Faith)의 기술로서 AI의 미래는 우리가 만들어야 한다. 그것은 우리의 선택과 고민의 결과이기 때문이다.

김윤명
Kim, Yun-myung
한국디지털재산법학회 이사, 법학박사
Korea Association for Digital Property Law Studies, Director
digitallaw@naver.com

미지의 기술로서 AI

SW와 컴퓨팅 기술의 고도화에 따른 AI에 대해 '미지의 기술'이라고 한다. AI가 어떻게 진화할지, 어떠한 모습으로 인류와 함께할지 알 수 없기 때문이다. 여전히 AI에 대한 부정적인 얘기가 회자되고 있으며, 전문가들 사이에서도 AI의 역할에 대한 논란이 분분하다. AI가 인류를 지배할 것이며, 인류보다 뛰어난 능력을 갖기에 충분하다는 주장에서부터, AI는 절대로 인간과 같은 인지능력을 갖기 어렵다는 주장도 있다. 물론, AI가 가져올 수 있는 문제나 우려에 대해 AI를 셧다운 시킬 수 있는 기술적인 방법을 도입해야 한다는 주장도 마찬가지다.

이러한 다양한 주장은 AI의 궁극적인 모습을 알 수 없는 상황에서 당연한 반응이 아닐까 생각한다. 연혁적으로 기술발전에 따른 인간의 대응은 단순했다. 기술을 최대한 활용할 수 있는 방안을 강구하고, 그 과정에서 문제가 발생하거나 명확히 예측되는 경우에는 기술 자체보다는 기술을 응용한 서비스에 대한 규제를 도입해왔다. 기술 자체가 갖는 중립성을 해치지 않는다는 기술정책적 결과다. 기술이 갖는 폐해에 대해서는 어떻게 대응할 것인지, 산업적 논의와 경우에 따라서는 소송이라는 법적 논의과정을 통해 제도화시킬 것인지, 배척할 것인지 수렴되었다.

이와 같이 기술이 사회제도적으로 수용되어왔던 연혁적 고찰을 통해 보건데, AI도 그러한 과정을 거칠 것이다. 사회적으로 기술을 수용한다는 것은 기술이 갖는 효용성을 높이는 것이다. 해당 기술이 산업적으로, 인간의 실생활에서 어떠한 반응을 보일 것인지에 대한 고민을 하게 될 것이라는 점이다. 다만, AI는 직접적으로 실행가능하다는 점에서 기술의 패러다임을 바꾸어놓고 있다.

산업혁명이 갖는 명암

인류가 산업혁명이라는 '표현'을 사용하는 경우는 몇 차례 되지 않는다. 4차 산업혁명을 얘기하지만, 이도 혁명이 완성된 상태가 아닌 진행 중인 상황에서 붙여진 것이다. 역사적으로 혁명의 기준이나 구분은 명확하지는 않다. 3차 산업혁명을 정보혁명이라고 하지만, 여전히 정보혁명은 미완성의 진행형이기 때문이다.

인간의 관점에서 보면, 기술 발전은 질적, 양적 팽창과 함께하는 인류의 발전이었다. 그 과정에서 환경파괴, 질병의 창궐, 부의 집중화와 양극화 등 새로운 문제가 발생해왔다. 반면, 인간이 하기 어렵거나 단순한 일들을 자동화함으로써, 인간의 본질에 집중할 수 있는 시간과 비용을 절감할 수 있다. 그로 인하여, 태초부터 인간에게 부여된 인간만이 갖는 창조적인 결과물을 만들어 내거나, 또는 가치 있는 것을 연구하고 발견해낼 수 있음이다. 일례로, 1977년 8월에 우주 유형을 떠난 '보이저(Voyager) 2호'가 태양계를 넘어서 인류에게 우주의 메시지를 보내오고 있는 것도 다르지 않다. 보이저 2호를 통해 현재 진행 중인 에너지혁명과 함께 우주에서 지구로 보내오는 영상정보를 통해 정보혁명을 확인하고 있다. 실생활에서도 기술은 다양하게 응용되고 있다. 지금 작업하고 있는 PC는 산업분야, 의료분야를 포함하여 개인의 문서작업이나, 정보생활의 혁신을 이끌었다. 다양한 정보를 찾고, 커뮤니케이션할 수 있게 된 것도 이 덕분이다. 더 나아가, 음성인식 인공지능은 문자형 커뮤니케이션을 음성형 또는 감성형으로 바꿔놓고 있다.

AI에 대한 오해와 현실

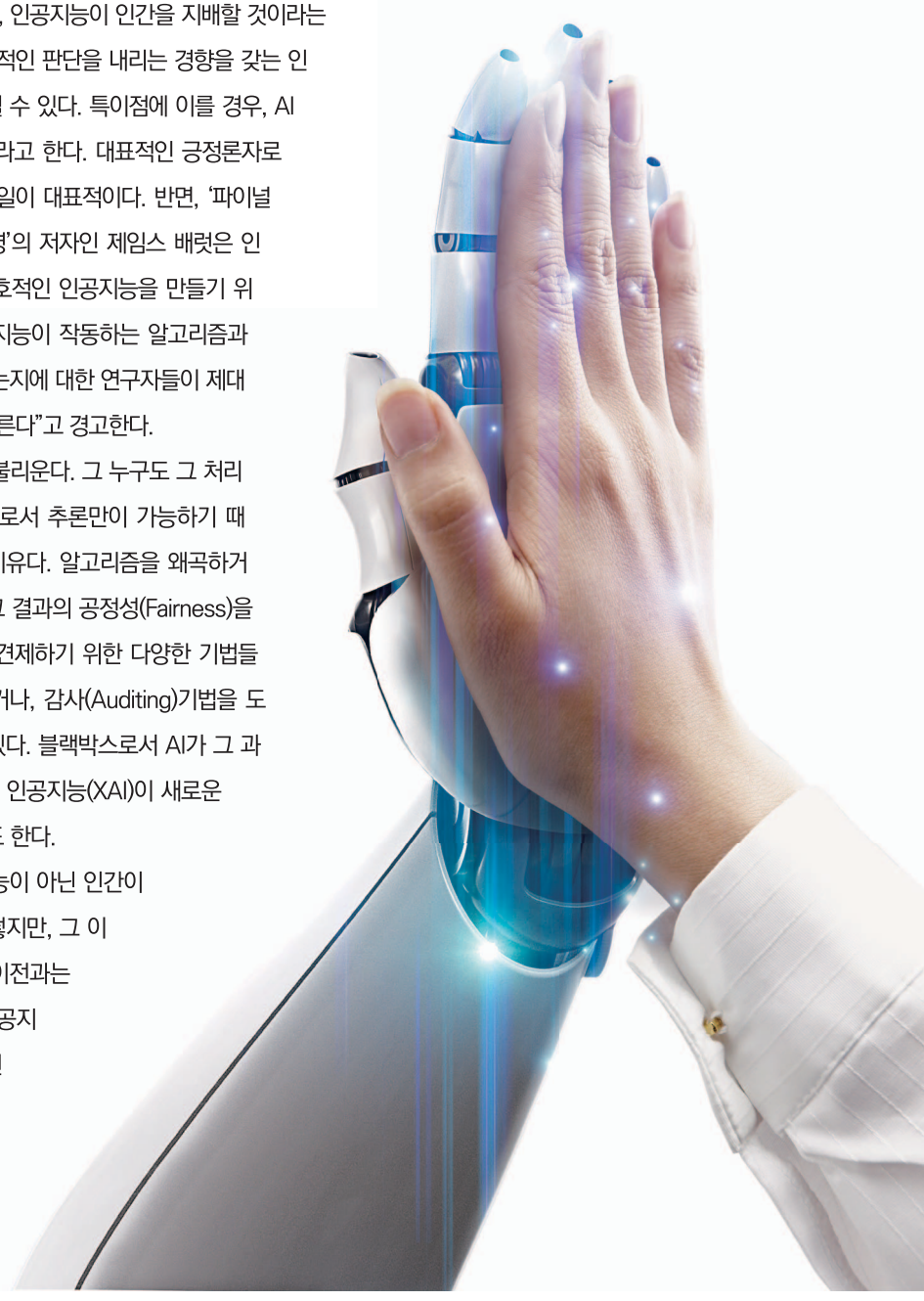
현재, 인류가 활용하고 있는 AI는 약한 인공지능(Narrow AI)이다. 특정 분야에서 인간의 능력을 보완할 수 있는 수준이다. 알파고는 바둑에, 왓슨(Watson)은 의료지원에 한정된다. AI스피커는 음성인식 기술을 활용하여 명령어 처리가 가능하다. 바벨탑을 쌓은 이후로 민족 간 커뮤니케이션이 단절된 언어분야도 AI를 통해 번역의 한계를 극복해가고 있다. 기계가 주식거래를 하는 경우도 있다. 인간이 갖는 한계를 기계적 도움으로 반복적이고 위험

한 일은 AI가 24시간 제한 없이 작동한다는 점에서 인간이 갖기 어려운, 효율적이고 지속적인 결과물을 만들어낼 수 있다는 장점을 갖는다. 물론, 이러한 것이 경제시스템에서 장점으로만 작용할 수 있을지는 다음 기회에 진행하고자 한다.

약한 AI와 달리, 인간과 다름없는 지능을 가진 AI를 부정적으로 인식하거나 또는 AI에 대한 오해도 있다. 대표적인 예가 강한 인공지능이 도래할 것이라는 점, 인공지능은 인간의 능력을 넘어설 것이라는 점, 인공지능이 인간을 지배할 것이라는 점이다. 또한, AI의 합리성에 따라, 비합리적인 판단을 내리는 경향을 갖는 인류는 제거대상이 될 것이라는 점도 포함될 수 있다. 특이점에 이를 경우, AI는 인간의 지능과 같은 능력을 얻을 것이라고 한다. 대표적인 긍정론자로는 '특이점이 온다'의 저자인 레이 커즈와일이 대표적이다. 반면, '파이널 인벤션(Final Invention), 인류 최후의 발명'의 저자인 제임스 배럿은 인공지능에 대한 우려를 제기하고 있다. 우호적인 인공지능을 만들기 위해 노력하고 있는 유드코프스키는 "인공지능이 작동하는 알고리즘과 복잡하게 연결된 시스템이 어떻게 움직이는지에 대한 연구자들이 제대로 이해하지 못하는 상황이 벌어질지도 모른다"고 경고한다.

인공지능은 블랙박스(Black Box)로 불리운다. 그 누구도 그 처리 과정을 이해할 수 없으며, 단순하게 결과로서 추론만이 가능하기 때문이다. 인공지능의 투명성이 강조되는 이유다. 알고리즘을 왜곡하거나, 또는 편견(Bias)이 들어갈 경우에는 그 결과의 공정성(Fairness)을 담보하기 어렵다. 이 때문에 인공지능을 견제하기 위한 다양한 기법들이 제시된다. 알고리즘의 소스를 공개하거나, 감사(Auditing)기법을 도입하는 것도 하나의 방안으로 논의되고 있다. 블랙박스로서 AI가 그 과정을 설명하지 못하기 때문에 설명가능한 인공지능(XAI)이 새로운 프로토타입(Prototype)으로서 제시되기도 한다.

여기까지는 자율적인 지능의 인공지능이 아닌 인간이 활용가능한 AI에 대한 대응방안이다. 그렇지만, 그 이상으로 진화하는 인공지능에 대해서는 이전과는 다른 방안이 제시될 필요가 있다. 실제 인공지능을 설계하는 과정에서 우호적이고, 인간의 통제가능하도록 해야 할 필요가 있다. 그렇지만, 이는 어디까지나 우려에 대한 대응이지, 원천적인 차단이어서는 안 된다.



기술 진보와 AI

'미지의 기술'로서 AI에 대한 부정적인 인식이나 긍정적인 주장이나 모습은 다양하다. 80년대, 90년대 암흑기를 지나온 AI는 이제야 가능성을 보여주고 있다. 이론적으로는 상당한 진전이 있었지만, 이를 뒷받침해줄 수 있는 HW능력이 그만큼 발전하지 못했던 것도 이유다.

AI는 스스로 학습하는 기계학습 알고리즘에 따라, 특정 업무에 대해서는 구체적이고 다양한 소스 코딩이 없이 스스로 학습해나가고 있다. 그만큼 정교하고, 시간과 비용을 절약할 수 있는 것이다. 지금도 몇 줄 되지 않는 공개된 소스코드를 이용하여, 누구라도 기계학습이 가능한 프로그래밍이 가능하다. AI 기반기술도 더욱 발전해가겠지만, AI를 활용한 응용도 다양한 분야로 확대될 것임은 자명하다. AI응용은 인간의 영역이라고 생각되어왔던, 창의적인 분야에서도 활용되고 있다. 작곡, 회화, 문학 등 다양한 분야에서 인간의 경쟁으로 묘사되고 있다. 아니, 인간의 것과 다름이 없다고도 한다. 물론, 이러한 결과를 가져오기 위해서는 다양한 데이터에 기반한 기계학습이 되어야만 가능하다. 데이터 확보가 경쟁력의 관건인 셈이다. 그렇지 않으면 인공지능은 인간과 경쟁할 수 있는 수준이 되기 어렵기 때문이다.

이러한 가치기반의 창작행위와 달리, 인간의 행동을 그대로 구현할 수 있는 프로파일링은 이미 다양한 분야에서 활용되고 있다. 특정 영역에 반응하거나, 추천서비스를 제공함으로써 관심을 가질만한 것을 제공하는 것이다. 그것은 서비스 사업자의 서비스 이용성을 높이거나, 또는 광고를 포함함으로써 수익을 높이는 방법이기도 하다. 물론 위험 지대나 위험도가 높은 산업영역에 AI가 활용되거나, 또는 AI가 탑재된 로봇을 활용하기도 한다. 방사능에 노출된 공간에서 작업은 인간에게 미치는 해가 크기 때문에 기계를 활용하기도 한다. 다만, 인명살상용 로봇은 아이러니하다. 터미네이터처럼 현실화가 될지 알 수 없으며, 어떠한 판단을 해야할지 쉽지 않다. 전쟁무기는 상대를 제압함으로써 아군의 인명피해를 줄이기 위해 개발된 것이다. AI가 탑재된 드론을 활용하거나, 킬러봇을 통한 방법은 AI의 부정적 인식을 높여준다. 이와 같이, AI는 긍정과 부정의 메시지를 다 같이 담고 있는 존재이다.

기술의 사회적 수용

우리는 어떤 경험을 하게 되면, 그러한 경험에 따라 변하고 부정적으로 반응하게 된다. 일종의 가용성 편향(Availability Bias)을 하게 된다. AI에 대해 직접적으로 비극적인 경험을 한 것은 아니지만, '알파고' 이래로 사람을 뛰어넘을 인공지능에 대해 간접적인 경험을 해왔다. 그렇기 때문에 AI에 대한 막연하지만, 경외감을 가졌는지도 모른다.

AI가 갖는 장점이나 그에 따른 긍정적 변화에 대해 널리 알릴 필요가 있다. 학교교육에서도 AI활용에 대한 내용도 포함될 필요가 있다. AI연구가 전문가의 손에 의해 이루어지긴 하지만, 보편적인 사용은 보통 사람들의 몫이기 때문이다. AI가 어떤 가치를 갖는지, 어떤 문제가 있는지에 알려줘야 한다. 그렇지 않을 경우, 막연한 우려에 따른 저항을 가질 수 있기 때문이다. 그러한 과정 없는 기술의 사회적 수용은 조심스러울 수밖에 없다.

AI를 포함한 나노기술, 바이오 기술 등 신기술에 대해 국가는 더욱 고민할 수밖에 없다. 시민단체, 연구자들도 이에 대해 우려하고, 그 파장을 고민한다. 만약 신기술이 시민들에게 미칠 수 있는 영향을 가늠할 수 없다면, 국가는 규제정책을 펴게 될 것이고, 시민단체는 해당 기술에 대한 우려를 과포장할 수도 있다. 우려나 긍정의 메시지만을 전달하는 것이 아닌, 우려에 대한 부분도 충분히 고려한 기술이나 정책이 수립되어야하는 이유이다. 기술에 대한 우려에 대해서 대응책이 없는 상황이 지속된다면 국가는 규제정책을 수립할 수밖에 없을 것이다. 다만, 이때의

규제정책도 원천기술에 대한 차단이 아닌 활용을 전제로 하는 것이어야 한다. AI 윤리가 하나의 대안이 되는 이유이고, 많은 나라나 국제기구를 중심으로 논의되는 이유이기도 하다. 산업계는 AI를 기획하는 초기부터 이러한 논란을 충분히 수용할 수 있어야 할 것이다.

AI윤리, 그리고 데이터 윤리

AI를 선의로 활용하기 위해서는 AI 자체, AI를 개발하는 개발자, AI 학습에 필요한 데이터셋(Data Set) 등 관련 분야에서 사회적 가치와 충돌할 수 있는 사항들을 배제시킬 필요가 있다. 쉽게 생각하는 것이 규제이나, 획일적인 규제는 기술 진보를 저해하기 마련이다. 기술을 살리고, 원래 의도대로 사용하려면, AI에 대한 윤리에 집중할 필요가 있다. 윤리는 최소한의 도덕으로서 법이 가져올 획일성을 유연성 내지 연성법(Soft Law)으로 대응할 수 있다는 긍정적인 의미도 있기 때문이다. 한계도 명확하다. 인공지능이 인식해야 할 윤리가 무엇인지 의문이기 때문이다.

인간에게도 어려운 윤리를 인공지능에게 학습시킬 수 있을 것인가? 자율주행차를 중심으로 논의되고 있는 인공지능 윤리는 어떠한 모습이어야 하는가? 한 단계 높은 철학적 논의에서 말하는 정의란 무엇인가? 이 질문을 포함하여, 인류가 최고선이라고 하는 윤리적 가치도 최고수준의 가치라고 단정하기 어렵다. 법적으로 살인은 금지되나 전쟁에서, 또는 생명을 위협하는 경우에는 정당방위로서 허용되기도 한다. 만약 로봇이나 AI가 가져야 할 윤리가 수립된다고 가정하더라도, AI만이 윤리적이어야 하는가? 이를 기획하고, 개발하고 때론 이용하는 인간은 어떠해야 하는가? 국가는 어떠해야 하는가? 정답은 없다. 다만, 지금 내리는 선택이 우리의 미래라는 점을 되새겨볼 일이다.

AI의 안전성, 공정성, 투명성 확보를 위해 윤리가 대안으로서 제시되지만, '어떻게' 라는 물음에 명확히 답하기는 쉽지 않다. 다만, 설계 단계에서 어떻게 윤리적이어야 하는지, 공정성을 담보할 수 있을지에 대한 가이드라인 수준의 논의가 이루어지고 있기는 하지만, 단순하게 '로봇3원칙'이 아닌 구체적이고 프로그래밍 가능한 또는 학습 가능한 가이드라인이 제시될 필요가 있다. 무엇보다, 인공지능이 가져야할 기본적인 지적 소양은 '인간을 위해야 한다'는 점이다.

인공지능 윤리에 더하여, 기계학습의 기반이 되는 데이터 윤리에 대한 관심도 필요하다. 아무리 인공지능이 윤리적이라고 하더라도, 학습에 필요한 데이터셋(Data Set)이 왜곡된 경우라면 알고리즘이나 소스코드를 공개하더라도, 그 원인을 찾기가 쉽지 않기 때문이다. 윤리적 적용이 어렵게 되면, 규제단계를 넘어 금지유형으로 형법 체계에서 인공지능이 다루어질 가능성이 크다. 인공지능 윤리는 인공지능과의 싸움 이전에, 규제기관과의 싸움을 피하기 위한 방안이라는 점을 인식할 필요가 있다.

인간과 AI의 공진화 - 사람중심에서 본 AI

인공지능이 어떻게 진화할지는 알 수 없으나, 분명한 것은 AI의 미래는 우리가 선택한 결과가 될 것이라는 점이다. 인공지능에 대해 막연한 두려움 또는 장밋빛 미래만을 기대해서는 곤란하다. 차별을 조장하거나 공정하지 못한 점은 배제하고, 다양한 사회적 기술로서 인공지능이 발전하고, 응용될 수 있도록 환경을 만들어야 한다.

이와 더불어, 인공지능이 작동하는 환경에서 빠지지 않을 가치는 '인간중심'이어야 한다는 점이다. 인간을 해치지 않고, 상호 공존할 수 있는 환경을 조성해야 한다. 그렇지 않는 로봇과 인류는 또다른 경쟁상대이자, 파괴할 객체가 될 수 있기 때문이다. 우호적인 인공지능을 위한 것은 최소한 파괴적이지 않은 인공지능이어야 한다는 것과 다르지 않다. AI에 탑재될 가치는 '우호적인 DNA'가 아닐까 생각된다. 우호적인 인공지능의 최소한의 기준으로서 윤리와 인간중심의 가치를 탑재할 수 있다면, 인공지

능이 인류를 지배하거나 파괴할 것이라는 우려는 줄일 수 있다고 보기 때문이다. 전 지구적으로 우호적인 DNA에 담겨야 할 기준이나 내용을 정하는 것도 방안이 될 수 있을 것이다.

인공지능이 중심이 되는 사회, '지능정보사회'에서 변치 않을 가치는 인간을 위한 AI여야 한다는 점이다. 그렇지 않은 인공지능은 인류와 경쟁하거나 그 과정에서 파괴하는 선택을 할 수도 있기 때문이다. 여전히 인공지능은 인류의 편리성을 위해서 역할을 하게 될 것이다. 검색, 번역, 추천, 거래, 창작 활동, 자율주행 등 다양한 분야에서 인간의 한계를 극복해가는 역할에 충실하기 때문이다. 사람중심의 인공지능에 대한 투명성, 공정성, 설명가능성을 담보할 수 있도록 신뢰를 주어야 한다. 인공지능의 가치는 '사람중심'을 넘어, '인권'이라는 인류 보편적 가치를 담는 것이어야 한다.

지금은 'AI의 민주화'가 도래했다고 볼만하다. 그렇지만 아무리 인공지능이 선하다고 하더라도, 인간의 욕심에 따라 악의적인 결과를 가져올 수 있다. 이를 막기 위해 우리가 AI에 대한 감시자가 되어야 한다. 제도적으로는 AI의 알고리즘에 대한 접근권의 보장과 그에 대한 절차적 정의가 수립되어야 한다.