

1장-1

데이터 사이언스와 AI

이경전(경희대학교 경영대학 & 빅데이터응용학과 교수)

1. 데이터 사이언스와 AI 분야에서 융합의 의미, 필요성과, 중요성

데이터 사이언스(Data Science: 데이터 과학)는 통계학, 컴퓨터 과학 및 데이터 분석 기법을 이용하여 데이터에서 인사이트와 지식을 추출하는 학제적 분야라고 정의할 수 있다. 인공지능(AI: Artificial Intelligence)은 인간으로부터 주어진 목표를 달성할 수 있도록 적절히 행동하는(Albus, 1991) 시스템을 만드는 컴퓨터 과학의 한 분야다(Nilsson, 2010). 각 학문 분야의 이름이 지어진 역사를 살펴보면, ‘인공지능’은 1950년대에, ‘데이터 사이언스’는 1970년에 처음 사용되었으므로, 인공지능이 더 오래된 학문이라고 볼 수 있다. 하지만, 데이터 사이언스는 사실상 통계학의 새로운 이름이라는 점에서 데이터 사이언스가 인공지능보다 더 오래된 학문이라고 주장할 수도 있다.

통계학은 수학의 한 분야로서 데이터의 수집, 분석, 해석, 표현 및 조직화를 다루는데, 추론을 도출하고 데이터로부터 결정을 내리기 위해 수학적 방법을 사용하는 데 중점을 둔다. 데이터 사이언스는 통계학 방법론에 컴퓨터 과학 분야에서 발전한 데이터 처리 방법론을 결합한 학문 분야다.

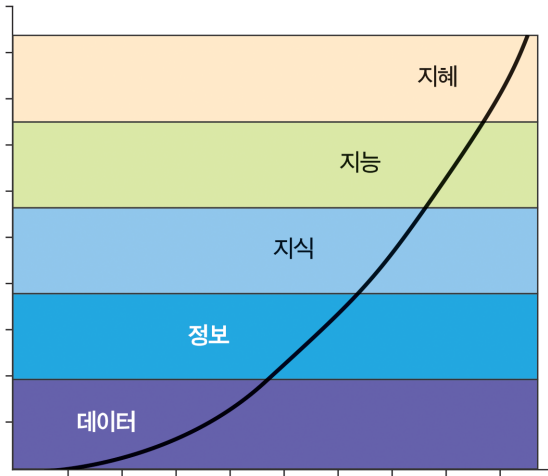
데이터 사이언스는 AI의 방법론을 차용하여 데이터를 처리하는 데에 활용하고, AI는 지능 시스템을 엄밀하게 구현하는 방법론을 탐색하는 과정에서 데이터 사이언스의 발전 결과를 활용한다. 학문의 관점으로 구분하면, 데이

터 사이언스는 과학에, 인공지능은 공학에 가깝다.

최근 인공지능 분야에서는 기계 학습(Machine Learning)이 각광받고 있다. 기계 학습의 정의는 다음과 같다. 어떤 기계의 성과(Performance)가 그 기계의 경험(Experience)에 따라서 향상되면 그 기계는 학습한다고 정의한다 (Mitchell, 1997). 여기서 그 기계의 경험이 많으면 데이터로 남게 되므로 많은 기계 학습 기법은 늘 그런 것은 아니지만 데이터를 다루게 된다. 인공지능을 기계 학습으로 이해하는 사람들은 인공지능이 데이터 사이언스와 거의 같은 분야라고 인식하게 되는 것이다. 그러나 기계 학습 외에도 다양한 인공지능 방법론이 있다. 예를 들어 대표적인 인공지능 교과서로 꼽히는 스투어트 러셀(Stuart Russell)과 피터 노빅(Peter Norvig)의 공저 『인공지능: 현대적 접근 방법(Artificial Intelligence: A Modern Approach)』에서는 28개 장 중 오직 4개 장에서만 기계 학습을 다룬다(Russell and Norvig, 2020). 이런 관점에서 보면 데이터 사이언스와 인공지능은 크게 다른 분야라고 할 수 있다. 탐색(Search), 게임(Game), 컴퓨터 논리(Computer Logic), 계획(Planning), 멀티 에이전트(Multi Agent) 등의 인공지능 방법론에서는 데이터의 처리나 분석을 다루지 않기 때문이다. 예를 들어 2016년에 이세돌 9단을 이겨 세상을 깜짝 놀라게 했던 알파고(AlphaGo)는 인간 바둑 기사들의 기보를 많이 학습해, 즉 데이터를 활용해 인공지능을 개발한 사례이지만, 2017년에 발표된 알파고 제로(AlphaGo Zero)나 알파 제로(Alpha Zero)는 기보 데이터를 전혀 사용하지 않았다는 점에서 제로(Zero)라는 단어가 붙었다. 인공지능은 데이터를 전혀 사용하지 않고도 개발할 수 있는 것이다. 다만 최근의 인공지능의 성과는 대부분 데이터를 기반으로 학습하는 딥러닝 등 기계 학습의 발전에 기반을 두고 있어, 데이터 사이언스와 인공지능의 유사성이 더 부각되었을 뿐이다.

결국 데이터 사이언스와 인공지능의 교차점은 데이터이므로 데이터를 대하는 관점을 명확하게 할 필요가 있다. 일반적으로 데이터는 세계와 사물을 낱것 그대로 표현하는 것이라고 정의하고, 그 데이터를 분석하고 정제해 정보

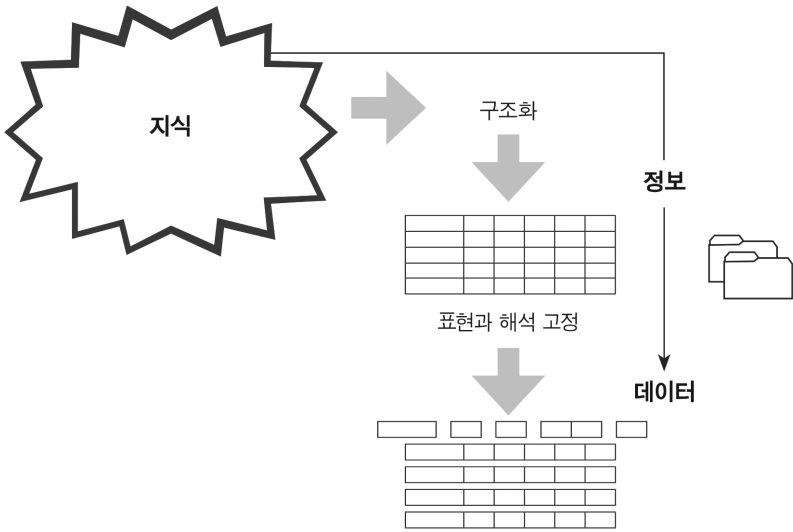
그림 1.1-1 데이터, 정보, 지식에 관한 전통적인 관점



자료: Tuomi(1999).

를 추출하며, 정보를 더 정제하고 분석하여 지식을 도출한다고 설명한다. 이러한 관점에서 데이터 사이언스는 보통 데이터에서 지식을 도출하는 과학이라고 이해한다. 전통 통계학이 데이터에서 평균, 표준 편차 등의 지식을 도출하고, 가설을 과학적으로 검증하는 학문인 것처럼 데이터 사이언스도 데이터, 정보, 지식으로 이어지는 과정을 기본으로 한다(〈그림 1.1-1〉).

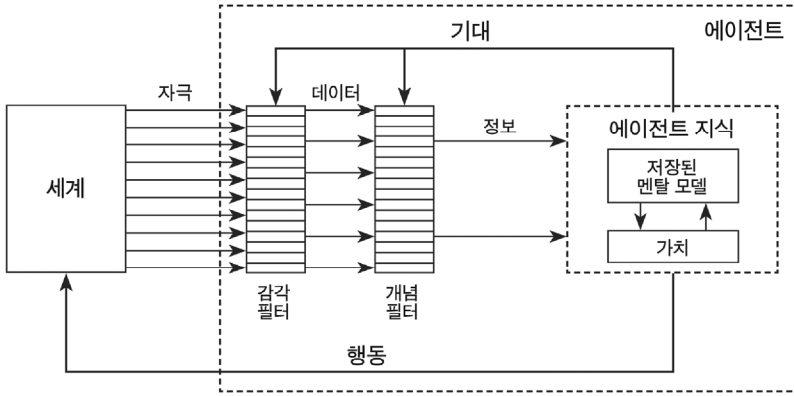
정반대로 생각할 필요도 있다. 어떤 지식도 없이 데이터가 존재할 수 있을지를 생각해 보아야 한다. 우리가 날씨를 “덥다”, “춥다”라고 하지 않고 “섭씨 30도다”, “영하 10도다”라고 데이터로 이야기 할 때, 그 30, -10이라는 데이터는 섭씨라는 어떤 지식에 바탕이 있다는 것이다(Tuomi, 1999: 103~117). 따라서 세상에 존재하는 데이터들은 사실 어떤 지식을 객관화하는 과정에서 정보로 만들고, 이를 더욱 구조화하는 과정에서 데이터가 도출된다고 할 수 있다(〈그림 1.1-2〉).



자료: Tuomi(1999).

이러한 두 가지 관점은 사실 데이터 사이언스와 인공지능을 구분하는 관건이 될 수 있다. 데이터 사이언스가 데이터에서 정보를, 정보에서 지식을 뽑아내는 과학적 방법론에 관심이 있다면 인공지능은 어떤 지능 시스템을 구축해서 그 시스템이 세계 곳곳의 어떤 이벤트나 개체를 감지하고 관찰해, 그러한 것들을 감각 필터를 통해 지능 시스템으로 입력하고 처리함으로써 적절한 행동을 산출하는 방법론에 관심이 있다. 즉, 데이터의 해석에 관심을 가지고, 이를 인간에게 제공하겠다는 것이 데이터 사이언스의 주된 관심사라면, 인공지능은 주된 관심사는 데이터 해석보다는 데이터를 입력하고 그 시스템의 지능으로 해석해 적절한 행동을 산출하는 것이다(그림 1.1-3).

한편 과학적 관점에서 인공지능은 데이터사이언스보다 엄밀하지 못한 점



자료: Boisot and Canals(2004: 43-67).

이 있다. 데이터사이언스는 통계학의 연장이자 확장^{이므로} 수학의 엄밀성을 매우 중요시하지만 인공지능은 수학의 엄밀성보다는 그동안 풀지 못했던 문제나 과제를 정복하는 공학 가치를 더 중요시한다. 정복하는 과정이 얼마나 엄밀하고 얼마나 재현 가능한가에는 크게 관심을 두지 않는다. 특히 최근 인공지능 분야에서 가장 각광받는 딥러닝 방법론은 아직 연금술의 단계에 있다고 비판받는다. 오픈AI가 개발한 Dall·E 2나 챗GPT는 새로운 이미지나 텍스트를 정말 잘 생성하는데, 그 성과를 수학적으로 엄밀하게 설명하거나 예측할 수 없다. 통계학자나 데이터 사이언스 학자의 관점에서는 과학이 아닌 것이다. 그러나 통계학자나 데이터 사이언스 학자가 하지 못한 것을 오픈AI는 딥러닝으로 성취해 낸다. 아직 수학적으로 엄밀하지 못하고 왜 그렇게 잘 작동하는지 설명하지는 못하나, 그동안 인류에게 없었던 새로운 기계를 만들어낸 것이다. 바로 이 대목에서 인공지능과 데이터 사이언스의 융합이 필요하다. 인공지능은 이 분야의 지능형 기계를 더 엄밀하게 만들고 그 성과를 수학적으로 검증하기 위해 데이터 사이언스의 방법론을 활용할 필요가 있다. 데이터

사이언스는 데이터의 종류와 데이터 처리 기법을 확장하기 위해 눈부시게 발전하는 인공지능 분야의 기법을 신속하게 도입할 필요가 있다.

2. 데이터 사이언스와 AI 분야의 발전 현황

데이터 사이언스의 발전상은 일반 대중에게 잘 보이지 않는다. 반면 AI의 발전상은 일반 대중에게 훨씬 더 쉽게 전달된다. AI 분야는 눈부시게 발전하고 있으나 여전히 갈 길이 멀다는 말로 요약할 수 있다. 2016년 3월 구글의 자회사 딥마인드의 알파고 대 이세돌 9단 간의 바둑 대결은 인공지능의 발전 상황을 세계인이 충격으로 받아들이는 계기가 된 일대 사건이다. 물론 바둑 인공지능이 사람 바둑 기사를 넘어서게 되리라는 것은 이전에 이미 예측되었다. 체스에서 IBM의 딥 블루가 세계 챔피언 카스파로프(Garry Kasparov)를 이긴 것이 1997년이기 때문이다. 2022년 말 오픈AI가 발표한 초거대 언어 모델 챗GPT는 2016년의 알파고만큼이나 충격을 주었다. 한편 테슬라나 웨이모로 대표되는 인공지능 기반 자율주행자동차는 그 발전 속도가 기대에 미치지 못한 채 지지부진하다. 애플의 시리, 아마존의 알렉사, 구글의 어시스턴트 등 인공지능 비서 또는 인공지능 스피커, 대화형 AI 역시 그 발전이 정체되고 있는 상황이다.

인공지능의 발전 현황이 이렇게 엇갈리는 것은 현재 인공지능을 대표하는 딥러닝 기술의 특징 때문이다. 딥러닝은 불가능했던 작업(Task)을 가능하게 한다는 점에서 놀라운 기술이다. 딥러닝은 문자 인식을 정복했고, 이미지 인식을 정복했다. 여기서 정복했다는 의미는 인간의 능력을 넘어섰다는 의미다. 체스를 정복한 딥블루(Deep Blue)에는 딥러닝 기술이 사용되지 않았지만 알파고에는 딥러닝 기술이 들어갔다. 물론 딥러닝 기술만 사용된 것이 아니라 탐색 기술이 결합되어 있는데 CNN(Convolution Neural Network: **컨볼루션**

신경망의 발전 덕분에 결국 바둑을 정복했다. 이렇게 딥러닝 기술은 불가능했던 작업을 가능하게 하면서 세상을 깜짝 놀라게 한다. 딥러닝 기술은 데이터 사이언스의 관점에서 보면 결국 비선형 회귀분석(Nonlinear Regression) 기법이라고 할 수 있는데 비선형 회귀분석은 완전히 최적화될 수가 없다. 완전히 최적화될 수 없다는 것은 그 모델이 국지적 최적해(Local Optima)에 빠질 수 있다는 의미가 되는데 이것은 현실 작업에서 오류(error), 실수(mistake)를 의미한다. 또 세상의 많은 문제가 조합최적화(Combinatorial Optimization)인데, 이 문제들 역시 어떤 기법을 사용하더라도 상당수가 주어진 시간 안에 주어진 컴퓨터로 최적해를 찾아내는 것이 불가능하다. 이를 NP-Complete, NP-Hard 문제라고 한다. 결국 제아무리 딥러닝 기법을 포함한 인공지능 기법을 사용한다고 해도, 현실의 문제 자체가 이미 풀기 어려운 NP-Complete, NP-Hard에 해당하면, 그 최적해를 구하지 못하므로 현실 작업에서 오류, 실수를 저지르게 된다.

인공지능은 이러한 기법의 한계와 현실 문제 자체의 복잡성 때문에 오류 및 실수가 빚어질 수 있다. 인공지능 기술의 발전 현황을 엿갈리게 설명할 수밖에 없는 이유다. 오류나 실수가 크게 문제 되지 않는 분야에서는 인공지능이 많이 사용되고, 그 수용도가 높아지게 된다. 체스, 바둑 등 게임 분야에서 인공지능이 발전해 사람들을 깜짝 놀라게 한 것도, 게임은 인공지능이 실수하더라도 문제가 되지 않는 응용 분야이기 때문이다. 반면 자율주행자동차는 인공지능이 실수하면 인명 피해가 발생할 수 있으므로 인공지능 기반의 자율주행자동차의 실현은 많은 사람들의 예상보다 늦어지고 있다. 인공지능 스피커도 마찬가지다. 여러 인공지능 스피커 제품이 사용자의 질문이나 요구에 완전 자동화된 대답을 하거나 행동을 할 수 있는 것으로 광고하고 마케팅했으나, 실제로는 인공지능 시스템의 오류가 잦아서 소비자들이 인공지능 스피커에 걸었던 기대를 접고 있는 실정이다. 그러나 오픈AI가 2020년에 GPT-3을 발표하고, 2022년 말에 챗GPT를 발표하면서 대화형 인공지능의 가능성이 급

속하게 높아진 것도 사실이다. GPT-3 사용자가 전문가에 한정되었던 것과는 달리 챗GPT는 비전문가에게도 개방되었고, 그 성능이 GPT-3보다 크게 향상되어 2016년 알파고가 불러일으켰던 AI 열풍이 또 다시 불게 될 개연성이 커졌다.¹ 한편 한국에서는 2020년 GPT-3의 등장을 보고 네이버, 카카오, LG, SK 등이 초거대 언어 모델을 개발하겠다고 발표했다. 네이버는 2021년 하이퍼 클로바를 개발 완료하고 2022년 일반에 공개했지만, 다른 국내 기업들은 아직 초거대 언어 모델을 공개하지 못하고 있다. 그러나 네이버는 오픈AI의 챗GPT가 하이퍼 클로바보다 월등하다는 점에서 기회와 위기를 동시에 마주하게 되었다. 초거대 언어 모델의 성능이 더 좋아질 수 있음을 챗GPT에서 확인한 네이버로서는 이를 큰 기회로 인식할 수 있다. 하지만 GPT-3의 수준으로 빨리 따라왔다고 생각한 네이버로서는 다시 챗GPT 수준으로 추격해야 한다는 점에서 위기에 직면했다. 초거대 언어 모델은 한 번 개발하고 운영하는 데, 연간 수백억에서 수천억 원의 비용을 써야 하므로 네이버로서는 세계 최고 수준의 언어 모델을 개발해 운영하는 의사 결정을 하기가 만만치 않을 수 있다. 이러한 의사 결정은 네이버뿐만 아니라 국내의 관련 기업들 모두에 닥친 중요한 문제다.

챗GPT의 발전은 앞서 인공지능 스피커 서비스나 콜센터 자동화 서비스, 인공지능 챗봇 서비스 등 대화형 AI를 개발해 온 인공지능 기업들에게도, 똑같은 관점에서 큰 기회이자 위협이 된다. 또 하나 명심해야 할 것은 이러한 초거대 언어 모델은 인공지능 기술 전체에서 차지하는 비중이 100분의 1 정도에 불과할 것이라는 점이다. 콘블루션신경망 기술의 발전에 따라 문자·인식, 영상·인식 기술은 보편적 기술로 자리 잡았다. 그리고 초거대 언어 모델은 크게

1 챗GPT는 2022년 11월 30일 출시한 후 다양한 활용 가능성을 보여주며 출시된 지 5일 만에 100만 명, 2주 만에 200만 명의 사용자를 확보했다. 이는 넷플릭스(3.5년), 에어비앤비(2.5년), 페이스북(10개월) 등 주요 서비스가 100만 명을 확보하는 데 걸린 기간에 비해 매우 짧으며, AI 서비스에서도 전례가 없는 기록이다(김태원, 2023).

생성AI(Generative AI)의 영역에 속한다. 챗GPT를 개발했던 오픈AI에서 발표한 DALL·E 2와 같이 이미지를 생성하는 AI 기술 및 기업도 많이 나타나고 있다. 예를 들어 BTS로 유명한 하이브는 2022년 인공지능 오디오 기술 개발 업체 슈퍼톤을 인수했다. 슈퍼톤은 합성 AI 기술을 표방한다. 한국에서는 생성 AI 기술을 AI 휴먼, 가상 연예인들에게 적용하는 시도가 활발히 이루어진다. 문제는 AI 휴먼, 버추얼 인플루언서, 가상 연예인 비즈니스모델이 과연 정착할 수 있겠느냐는 점인데, 이러한 점에서 인공지능도 여타 다른 기술과 마찬가지로 기술뿐만 아니라 이를 활용하는 비즈니스모델의 개발과 정착이 더 중요한 상황에 처해 있다.

흥미로운 것은 인공지능 기술이 기존의 데이터사이언스, 다시 말해 통계학이 다루기 힘들었던 언어의 영역, 이미지의 영역에서는 큰 발전을 보이고 있지만 숫자만을 다루는 영역에서는 이렇다 할 성과를 보이지 못하고 있다는 점이다. 금융을 예로 들자면 오래전부터 쿼트 분야에서 컴퓨터 기술과 금융 모델을 결합한 시스템 투자 기술이 발전해 왔는데, 인공지능의 딥러닝 기술이나 강화 학습 기술 등이 적용되어 큰 성과를 내는 부분은 크게 소개되고 있지 않다. 이는 학술적으로 매우 흥미로운 점을 시사한다. 우리 인간이 사용하는 언어나 인간이 그린 그림, 또는 자연과 사회에 존재하는 여러 이미지, 사회에 존재하는 각종 연결 및 네트워크에는 어떤 구조(Structure)가 있고, 이를 잘 분석하고 학습하는 도구가 딥러닝 분야에서 발전했다는 점을 인식할 필요가 있다는 점이다. 인간이 사용하는 언어에 있는 어떤 구조를 잘 파악해 언어를 생성하는 기술이 딥러닝 분야에서 결국 어텐션(Attention)에(ahdanau et al., 2014) 기반을 둔 트랜스포머(Transformer) 기술로 고안되어(Vaswani et al., 2017) 앞서 설명한 GPT-3, 챗GPT가 탄생하게 되었다. 인간이 생성한 이미지나 자연과 사회에 존재하는 여러 이미지 역시 어떤 구조가 있으며, 그 구조를 잘 파악해 이미지를 인식하고 생성하는 기술이 딥러닝 분야에서 컨볼루션 신경망 기술과 GAN(Generative Adversarial Network) 기술로 발전해 각 분야에서 사용

된다. 사회에 존재하는 각종 연결 및 네트워크에 있는 어떤 구조를 파악하고 분석하는 데 기존에는 SNA(Social Network Analysis) 또는 네트워크 과학이라는 기법을 주로 썼다면 최근에는 그래프 신경망(Graph Neural Network)을 많이 활용한다.

결국 세상의 모든 문제를 해결하는 통일된 방법론은 없으며 특정 문제를 해결하는 여러 방법론, 즉 모델을 구비하는 것이 중요하다는 의미로 해석할 수 있다. 이것을 인공지능 분야에서는 앙상블 모델이라고 하는데, 일반적으로는 매니 모델 어프로치(Many Model Approach)라고 하기도 한다. 이러한 면에서 데이터사이언스와 AI의 융합은 한층 더 요구된다.

한편 과학기술 연구에도 인공지능이 많이 활용된다. 딥마인드의 알파폴드(AlphaFold)와 제너레이트 바이오메디신의 크로마(Chroma), 그리고 미국 워싱턴 **대학교**가 공개한 로제타 폴드 디퓨전(RoseTTA Folded Diffusion), 페이스북의 모회사 메타가 개발한 AI ‘ESM 폴드’ 등이 대표 사례로, 단백질 구조를 예측할 뿐만 아니라 새로운 단백질을 디자인한다. 특히 ESM 폴드는 초거대 언어 모델을 응용한 것이다. 흥미로운 점은 AI 전문 회사와 기존의 과학 분야 전문가 그룹이 서로 경쟁적으로 연구개발에 나서고 있다는 점이다.

인공지능 발전의 또 다른 중요한 방향은 연합 학습(Federated Learning)의 활용이다. 딥러닝 등 기계 학습 시스템은 데이터가 많이 필요한데 중요한 데이터일수록 공개하기가 어렵다. 이는 개인이든 조직이든 마찬가지다. 공개하고 싶어도 개인정보**보호법**, GDPR 등 법적으로 불가능하기도 하고, 법적으로 가능하더라도 개인과 조직의 이해관계상 공개, 공유, 전송, 통합을 좀처럼 하지 않으려 하는 것이 현실이다. 연합 학습은 각 주체의 데이터를 통합하지 않고도 마치 통합한 것과 유사한 수준의 성능을 보일 수 있는 기법이다. 주체 간에 민감한 데이터를 서로 교환하고 공유하는 것이 아니라 각 주체가 자신의 데이터로 학습한 인공지능 모델을 서로 교환하고 공유해 더 강력한 성능의 인공지능 모델을 만들어내는 방법론이다. 이 연합 학습 기법은 의료, 금융, 제

조, 교통 등 다양한 분야에서 활발히 응용되어 이 기술에 기반을 둔 플랫폼 회사, 기술 회사 등이 설립되면서 산업 생태계가 형성되고 있다.

3. 데이터 사이언스와 AI 분야의 발전 방향 및 중장기적 전망

AI는 단기적으로는 약속을 못 지키는 것으로 보일 수 있으나 긴 호흡으로 전망해 보면, 계속 발전 중인 것이 분명하다. 인공지능으로 움직이는 완전 자율 주행 자동차는 2020년대에는 어렵겠지만, 2030년대에는 가능할 것이라는 전망이 지배적이다. 여기서 중요한 것은 2030년대에는 된다는 것이 아니라, 2020년대에는 안 된다는 점이다. 이 부분은 많은 사람에게 실망스러운 소식 이겠지만 여기서 배울 점이 있다. 앞서 밝힌 바와 같이 인공지능은 실수할 수 밖에 없는 존재이기 때문이다. 자동차 운전처럼 실수의 결과가 인간의 사망으로 이어질 수 있는, 즉 실수에 따르는 위험이 큰 응용 분야는 그 실현 속도가 더디다는 점을 이해해야 한다. 그래서 위험이 큰 분야에서 인공지능을 잘 활용하려면 사람과 인공지능이 잘 협업할 수 있도록 해야 한다.

현재 인공지능 분야에 가장 큰 영향을 미치는 것이 초거대 언어 모델이다. 초거대 언어 모델은 생성 AI의 한 분야로 자리매김하면서 다양한 분야에서 응용될 것이다. 생성 AI는 초기에는 텍스트 또는 이미지 등 단일 모드의 콘텐츠를 생성하는 방식으로 응용되기 시작했지만 멀티 모달(Multi-Modal) 생성 AI로 발전하고 있다. 처음엔 AGI(Artificial General Intelligence, 일반 인공 지능) 형태로 시작되었지만 결국 상업화 과정에서 전문화될 것이다. 이를 비유적으로 생각하는 방법으로 월드와이드웹의 탄생 및 성장 과정과 비교할 수 있다. 월드와이드웹 초기에 야후와 아미존닷컴이 등장했다. 이후 구글 검색 엔진이 등장했는데, 야후닷컴(Yahoo.com)과 구글닷컴(Google.com)은 이렇다 할 수익모델이 없었다. 현재 많은 사람들이 챗GPT와 같은 초거대 언어 모델의 등

장을 구글 검색 엔진의 초기에 비유하는데, 이는 매우 적절하다. 챗GPT는 그 성능에 비해 수익모델이 아직 분명하지 않다. 그러나 구글 검색 엔진 초기보다 상황은 좋다고 할 수도 있다. 마이크로소프트가 십수조 달러를 투자하며 후원하고 있기 때문이다. 구글과 야후의 수익모델을 만들어준 것은 고토닷컴(Goto.com)으로 출발한 오버추어였다. 키워드 검색 광고 특허 비즈니스 기법을 개발하고 보유한 기업으로 훗날 야후에 인수되었다. 챗GPT는 아직 수익 모델을 발견하지 못한 초기의 구글 서비스와 유사하지만, 구글이 오버추어를 만난 것처럼 챗GPT도 곧 수익모델을 발견할 것이다. 특히, 스마트폰 앱스토어, 플레이 마켓의 산업 생태계를 경험했으므로, 챗GPT 같은 초거대 언어 모델도 일종의 앱 생태계를 구축하게 될 것이다. 이러한 과정에서 전문화의 길을 걷게 될 것이다. 초거대언어 모델을 의료, 과학기술연구, 금융, 창작, 대고객 서비스에 활용해 수익을 꾀하는 사업자들이 나타나고 경쟁하면서 초거대언어 모델 기반 서비스가 발전해 나갈 것이다.

챗GPT 같은 B2C 초거대 언어 모델의 대중화는 기업이나 정부의 B2B 초거대 언어 모델의 활용 의도를 키울 것이다. 그런데 기존의 초거대 언어 모델이 인터넷에 공개된 데이터를 대규모로 활용하여 일반 인공지능 같은 생성 AI를 만들어내는 것이라면, 기업이나 정부의 인공지능 시스템은 서로 공유하고 싶지 않은 자체 데이터만으로 개발하는 것이다. 이렇게 되면 조직 안에서만 존재하는 데이터로는 챗GPT와 같은 강력한 인공지능 시스템을 만들 수 없다는 패러독스에 빠지게 된다. 이 문제를 해결할 수 있는 방법이 연합 학습이다. 따라서 빅테크 중심의 강력한 AI 시스템이 계속 나올수록 이에 고무되고 확신하게 되면서도 이에 대항하고 경쟁하기 위한 방법으로서 기업 간 연합, 정부 간 연합, 기업 정부 간 연합, 시민 간 연합 등 다양한 연합에 의한 인공지능 시스템의 개발과 이를 공유하려는 시도가 늘어날 것이며, 이와 관련된 산업 생태계가 전 지구적으로 성장할 것으로 보인다.

이러한 연합 학습 기법의 확산은 B2C 서비스 패러다임에도 큰 영향을 미칠

것으로 예상된다. 지금까지 플랫폼 서비스의 개념은 개인 고객의 데이터를 거의 무한정으로 플랫폼이 소유하고, 이에 기반을 두고 서비스를 제공하는 형태인데, 데이터를 플랫폼에 주지 않아도 지능 서비스가 가능하다는 것을 자각함에 따라 자신의 프라이버시를 보호받으면서도 지능 서비스를 받을 수 있게끔 사용자 중심의 서비스를 제공하는 기업이나 플랫폼을 선호하게 될 것이다. 이러한 구조가 확산되면 헬스케어, 뷰티, 교육, 상담, 자산 관리, 마음 관리 등 거의 대부분의 대인 서비스 구조가 혁신될 수 있다. 사업자는 저마다 고객 정보 제로를 외치면서 고객을 유치하려고 할 것이다. 이러한 움직임은 플랫폼 산업 생태계를 크게 변혁하는 계기로 작용할 수 있다.

4. 데이터 사이언스와 AI 분야가 미래 사회에 미칠 가능한 영향

인공지능이 미치는 가장 중요한 영향은 사회 각 부문에서의 생산성을 제고해 새로운 산업을 창출하는 것이다. 인공지능이 생산성을 높여서 많은 직업이 사라질 것이라고 전망하는데 이는 그동안 자동화 기술이 인류에 미친 영향의 양상에 관한 인식이 부족한 결과다. 그림을 그리는 것을 자동화하는 카메라 기술은 수많은 직업과 산업을 창출했고, 마차보다 더 자동화 정도가 높은 자동차 기술 역시 수많은 직업과 산업을 창출했다. 인공지능도 마찬가지다. 인공지능 기술이 적용되는 분야는 생산성이 높아지면서 가격이 급격히 낮아지고, 이를 통해 새로운 수요가 생기면서 그 분야의 새로운 고객을 맞이해 번영하게 될 것이다. 많은 제품과 서비스의 단가가 낮아지면서 새로운 고객과 산업이 창출되고, 기존의 제품과 서비스는 심미성과 품질, 개인화와 지능화의 정도에 따라 차별화될 것이다.

그렇다면 어떤 산업이 새롭게 창출될 것인가? 유발 하라리(Yuval Harari)가 2015년에 저술한 『호모데우스(Homo deus)』를 참조하면 도움이 되지 않을까

한다. 하라리는 전쟁, 기근, 역병을 우리 인류가 해결했다고 주장하면서 새로운 인류인 호모데우스는 불멸, 행복, 신성을 추구할 것이라고 주장했다. 물론 하라리는 틀렸다. 인류는 여전히 우크라이나 전쟁 등 각종 전쟁의 위협으로 고통받고, 코로나19의 창궐처럼 여전히 역병에서 자유롭지 못하다는 것을 확인했다. 그러나 인류가 점점 불멸(Eternity), 행복(Bliss), 신성(Divinity)을 추구할 것이라는 주장은 인공지능이 창출할 증장기의 미래 가치와 상당히 맞닿아 있다.

인공지능에 의한 과학기술 연구의 급속한 발전은 인류의 신체 건강, 정신 건강 증진에 큰 도움을 주어 인류의 수명을 연장하며 궁극적으로는 불멸의 수준으로까지 나아가는 데 계속 기여하면서 새로운 산업과 서비스를 창출할 것이다. 또 인공지능에 의한 새로운 지식의 발견과 뷰티, 미용 등 자기 꾸미기와 서비스의 창출, 새로운 엔터테인먼트 산업의 창출은 인간의 행복도 높이기에도 계속 기여하면서 새로운 제품, 서비스, 산업, 직업을 창출할 것이다. 예를 들어 버추얼 인플루언서, AI 연예인 등 새로운 엔터테인먼트 산업이 정착할 수 있을지는 불확실하지만 인공지능에 의해 새로운 콘텐츠의 형식이 개발되면 지금으로서는 상상할 수 없는 새로운 예술, 새로운 미디어가 생겨날 것이다. 인간이 신성(Divinity)을 추구한다는 것은, 인간이 궁극적으로 신과 같은 전지성(Omniscience, 全知性), 전능성(Omnipotence, 全能性), 편재성(Omnipresence, 遍在性)으로 나아간다는 것인데 인공지능 기술에 기반을 둔 로봇은 인간을 위협하기보다는 인간의 능력을 강화하는 형태로 발전해 인간이 전능성으로 나아가는 데 기여할 것이며, 앞서 설명한 챗GPT 같은 인공지능 서비스는 인간이 전지성으로 나아가는 데 기여할 것이다.

이러한 인공지능의 발전이 사회 구성원들의 격차를 키울 것이라는 우려도 있다. 일반적으로 새로운 기술은 재화와 서비스의 생산과 유통의 비용을 줄여서, 상대적으로 가난한 사람이 더 많은 재화와 서비스를 향유하는 데 기여해 왔다. 이처럼 인공지능 기술 발전 역시 인류의 생활을 더욱 윤택하게 하는

데 기여할 것이다. 물론 FANG(페이스북, 애플, 넷플릭스, 구글)로 대표되는 빅테크의 독점 현상이 인공지능 시대에도 지속될 것이라는 우려가 있다. 이 부분은 적절한 기술과 비즈니스모델을 통해 미리 방지할 수 있다. 특히, 연합 학습 기술은 데이터의 소유 주체와 인공지능 시스템 소유 주체를 분리할 수 있는 방법으로, 데이터를 독점하는 현상과 데이터가 과편화되는 현상을 예방하면서 동시에, 데이터를 개인과 조직, 조직과 조직 간에 통합하지 않아도 협업이 가능하게 한다는 점에서 앞으로 널리 사용될 것으로 예상된다.

어쩌면 사용자들이 자신의 데이터를 빅테크 플랫폼에 모두 뺏긴 시대는 긴 역사로 볼 때 과도기로 평가될 것이다. 기술은 개발 초기에는 소수의 전유물이 되지만 발전할수록 모든 대중이 향유하는 문명의 이기가 된다. 자동차도 처음에는 귀족의 레저용품이었으나 헨리 포드(Henry Ford)가 모델T를 내놓자 두 달 치 월급이면 자동차를 소유할 수 있는 시대가 열렸다. 컴퓨터도 처음에는 일부 대학, 일부 대기업에만 설치되는 기계였으나 스티브 잡스(Steve Jobs)와 빌 게이츠(Bill Gates) 같은 사람들의 기여로 개인용 컴퓨터가 나타나고, 이제 세상 사람 모두가 스마트폰이라는 엄청난 성능의 컴퓨터를 사용하는 시대가 되었다. 미디어도 마찬가지다. 예전엔 오직 왕과 정부 기관만 미디어를 소유했다. 종교가 힘이 셀 때는 교회와 성당이 찬송가와 예배라는 미디어를 가졌다. 그러다가 TV, 라디오, 신문, 잡지라는 4대 매체, 즉 대중매체가 나왔다. 이때까지만 해도 미디어는 개인의 소유가 아니었으나 이제는 유튜브, 페이스북, 인스타그램, 틱톡, 트위터 등의 미디어를 개인이 가지는 시대가 되었다.

인공지능도 예외가 아닐 것이다. 지금의 인공지능 개발은 구글, 네이버, 아마존, 애플 등 거대 기업이 주도하고, 기존 대기업에서나 인공지능을 자체 개발하는 상황이지만 중장기적으로는 개인이 인공지능을 소유하고, 개발하고, 운영하는 시대가 도래할 것이다. 개인이 인공지능을 소유하려면 스스로 데이터를 소유해야 한다. 개인이 자신의 디바이스나 개인 클라우드에 데이터를

소유하고, 이 데이터를 활용해 인공지능을 만들어내는 시대가 올 것이다. 그런데 개인의 데이터만으로 만든 인공지능은 성능이 좋지 못할 수 있으므로 개인은 자신의 의지로 자신의 인공지능을 다른 개인과 상호 공유하고 통합함으로써 성능을 높이고, 더욱 일반적이면서도 자신을 잘 아는 인공지능을 소유하고 운영할 수 있게 될 것이다.² 이렇게 개인과 소상공인, 중소기업이 인공지능을 가지게 되면 디지털 디바이드, 인공지능 격차 등의 문제가 다소 해결될 수 있을 것이다. 데이터의 교환도 데이터를 물리적으로 교환하는 데이터 교환 시장이 나타나기보다는, 각 주체가 자신의 데이터를 그대로 가진 상태에서, 자신의 데이터로 학습하려는 외부 AI 모델의 접근을 허용하며 대가를 받는 형태의 데이터 접속 시장이 나타날 것이다.

참고문헌

- 김태원. 2023. 「ChatGPT는 혁신의 도구가 될 수 있을까?: ChatGPT 활용 사례 및 전망」. 《THE AI REPORT》, 2023-1.
- Albus, James. 1991. "Outline for a Theory of Intelligence." *IEEE Transactions on Systems, Man, and Cybernetics*, 21(3).
- Bahdanau, Dzmitry, Cho Kyunghyun and Yoshua Bengio. 2014. "Neural machine translation by jointly learning to align and translate." arXiv preprint(cornell university), arXiv:1409.0473.
- Boisot, Max and Agusti Canals. 2004. "Data, Information and knowledge: Have we got it right?." *Journal of Evolutionary Economics*, 14(1).
- Mitchell, Tom, 1997. *Machine Learning*. NY: McGraw Hill.
- Nilsson, Nils. 2010. *The Quest for Artificial Intelligence: A history of ideas and achievements*. GB: Cambridge University Press.

2 이러한 것을 가능하게 하는 기술이 개인화 연합 학습(Personalized Federated Learning)이다.

- Russell, Stuart and Peter Norvig. 2020. *Artificial Intelligence: A modern approach*(UK: Pearson Education), 4th Edition.
- Tuomi, Ilkka. 1999. "Data is more than knowledge: Implications of the reversed knowledge hierarchy for knowledge management and organizational memory." *Journal of Management Information Systems*, 16(3).
- Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan Gomez, Lukasz Kaiser and Illia Polosukhin, 2017. "Attention is all you need." *Advances in neural information processing systems*, 30.