# 기계가 만든 소프트웨어가 가져오는 불평등과 데이터 과학의 역할

이광춘

KPMG Korea 상무

1. 인간과 기계의 경쟁관계와 불평등의 발생

기계(Machine)라 하면 기계장치를 떠올릴 수 있지만, 영어로 머신(machine)은 인공지능을 탑재한 컴퓨터를 의미한다. 사람들이 하는 작업의 자동화 수준을 여러 단계로 나눌 수 있는데, 기계의 도움 없이 모든 결정과 행동을 사람이 취하는 수준부터 인간을 배제하고 기계가 모든 의사결정을 내리고 자율적으로 운전, 판결, 세금계산 등 완전한 자동화 수준이 있다 [1].  이렇게 작업 자동화 수준이 나눌 수 있는 원인이 인간과 기계가 서로 잘하는 영역에 근본적인 차이가 있다는 믿음이 있었지만 급격한 자동화 기술의 발전으로 이러한 구분도 의미가 없어지고 있다.

단어를 찾거나 글자 수를 세는 단순하고 반복적인 작업은 컴퓨터에게는 쉬운 작업이다. 반면, 논문이나 책을 읽고 그 속에 숨겨진 맥락을 파악하는 것은 현재 기술로도 한계가 있다. 인간은 지루하고 반복된 문제를 해결하는 데 적합하지 않고, 컴퓨터도 추상적이고 일반화하는 작업에 적합하지 않았다.

생산성과 임금격차, 보울리의 법칙, 노동인력 참여율을 통해 확인되는 공통된 사항은 1980년 이후 일자리에 구조적인 변동이 생겼다는 점이다[2].  과거 일자리와 관련하여 국가내에 정규직과 비정규직 프레임 혹은 외국인 노동자로 대표되는 국외노동자와 국내 일자리 프레임에 추가하여 사람과 기계 프레임이 추가되었다. 예를 들어, 국가의 근간을 이루는 세무업무를 살펴보자. 과거 숫자를 다룰 수 있는 소수의 사람만이 숫자의 계산을 암산에서 벗어나 주판의 도움으로 생산성을 주판을 사용하지 못한 사람과 비교하여 수십배에서 수천배의 정확도와 함께 빠른 계산을 달성하게 되었다. 이러한 주판은 중간에 기계장치 계산기(찰스 배비지)도 있었지만, 일제 전자계산기로 자리를 내어주지만 사칙연산만 이해하면 기존 주판과 비교하여 어마어마한 생산성을 향상과 정확도를 높인 것은 분명하다. 이후, 개인용 컴퓨터의 보급으로 비지칼크와 로터스 1-2-3가 그 가능성을 열었다면 마이크로소프트 엑셀 스프레드시트 프로그램이 세무사 업무의 생산성을 또한 엄청나게 올린 것도 사실이다. 아마도 여기 까지가 세무사가 기계와 경쟁을 하지 않고 기계가 세무사의 생산성 향상에 도움을 준 것으로 볼 수 있다. 이제부터 문제가 되는 것은 PC 매거진, “The Best Tax Software for 2019”에 소개된 세금관련 프로그램이 $39 달러에 불과하다는 점이다. 1년 세무업무가 개인의 경우 4만원에 불과한데 세무사가 이런 자동화된 기계와 경쟁에서 승리할 수 있는가? 결과는 명확하다. 이제 기계와 세무사간의 일자리 경쟁이 시작된 것이다.

2. 데이터 과학의 시대

과거에는 사람이 직접 설계한 소프트웨어를 연구실에서 꺼내 실생활에 적용함으로써 큰 변화를 이끌어 냈다면, 현 시대는 빅데이터 기술을 통해 축적한 데이터를 활용하여 기계로 하여금 소프트웨어를 만들게 하는 시대이다. 이런 변화는 학문분야 뿐 아니라 실생활에까지 폭넓게 영향을 미치고 있으면 그 영향의 정도는 혁명적이라고 할 수 있다. 이러한 혁명적 변화를 가능하게 한 기술의 중심에는 요즘 자주 언급되는 기계학습(Machine Learning)과 딥러닝(Deep Learning)이 자리하고 있다. 하지만, 기계로 하여금 소프트웨어를 작성하게 하고, 기계가 작성한 소프트웨어를 검증하고, 실생활에 적용시키는 것 또한 아직까지는 인간이 큰 부분을 담당하고 있으며 이러한 추세는 한동안 유지될 것으로 예측되고 있으며, 이와 관련된 업무를 과거 빅데이터라고 통칭했다면 현재는 **데이터 과학(Data Science)**이라고 부른다.

데이터 과학은, 석유화학 산업이 석유를 자원으로 여러 단계의 공정을 거쳐 플라스틱, 섬유, 고무와 같은 상품으로 재탄생시켜 경제적 가치를 창출하듯이, 데이터 자원을 여러 단계의 과정을 거치면서 기계가 기존에 없던 가치를 창출한다[5]. 기존의 통계학도 데이터를 원자재로 사용하고 데이터에서 유용한 가치를 찾아낸다는 점에서 유사하지만, 통계학이 통계이론과 GUI 통계 팩키지를 결합하여 통계기반 의사결정을 지원하는 보고서 작성에 집중한 반면, 데이터 과학은 CLI 프로그래밍 언어(R/파이썬)를 사용하여 데이터를 원재료로 삼아 데이터 속 패턴을 찾는 알고리즘을 만들고, 대쉬보드를 제작하여 시각화를 통해 인간과 커뮤니케이션은 물론 의사결정도 지원하고, 예측모형을 RESTful API로 제작하여 기계간의 자동화 업무도 포괄한다. 데이터 과학은 전통적인 소프트웨어 개발과 비교해서 보면 모든 출발이 데이터로부터 시작된다는 점에서 확연한 차이가 난다. 하지만 데이터 과학 제품의 복잡성이 증가함에 따라 애자일 소프트웨어 개발 방법론, DevOps, 아키텍처, 소프트웨어 품질, SPLE 등 Best Practice와 이론이 속속 도입되고 있다.

데이터 과학 소프트웨어는 데이터를 원재료로 삼아 기계가 소프트웨어를 작성하게 되면서 기존 인간이 작성한 소프트웨어가 자연어 처리(NLP)와 시각분야에서 가졌던 한계를 뛰어 넘는데 큰 기여를 했다.

3. 인간과 기계의 경쟁관계에서 데이터 과학이 할 수 있는 역할

언제나 소수의 과학자와 기술자들이 제품과 서비스를 만들고 기업가가 확대시키고 시간의 흐름에 따라 선순환을 타고 확대발전하면 많은 사람들에게 과학기술의 혜택을 주는 과정을 거치게 되지만 그 이면에는 독과점과 불평등이라는 어두운 면이 동전의 양면처럼 함께한다.

단기적으로 데이터 과학은 기계가 데이터로부터 소프트웨어를 제작하는 과정에 있어 블랙박스를 투명하게 만들어 투명성을 증진시키고, 소프트웨어 성숙도(Software Maturity)를 가속화하여 Commodity화 시킴으로써 독과점으로부터 발생되는 불평등을 방지하는데 기여할 것이다. 중장기적으로는 데이터 과학을 통해 기계가 만드는 소프트웨어를 경험이 축적되고 데이터 과학 민주화를 통해 더 많은 분들의 참여가 이뤄짐에 따라, 기계가 만든 소프트웨어와 사람이 만든 소프트웨어를 아우르는 통합된 소프트웨어를 제작함으로써 그 이전 아무도 가지 않은 새로운 세상을 열어가는데 기여할 것이다.

참고문헌

1. Cummings MM, Man versus machine or man+machine, IEEE Intelligent Systems, 2014;29:62–9.
2. Mishel L., Gould E., Bivens J., Wage stagnation in nine charts. Economic Policy Institute, 2015;6:2–13.
3. Kaplan J., Humans need not apply: A guide to wealth and work in the age of artificial intelligence, Yale University Press 2015.
4. Wing JM., Computational thinking, Communications of the ACM, 2006;49:33.
5. 김기범, 이효정, 박도휘, 기업 운영 혁신을 위한 데이터 과학: 기업의 활용 방안, 삼정KPMG 경제연구원, 2020;제121호.