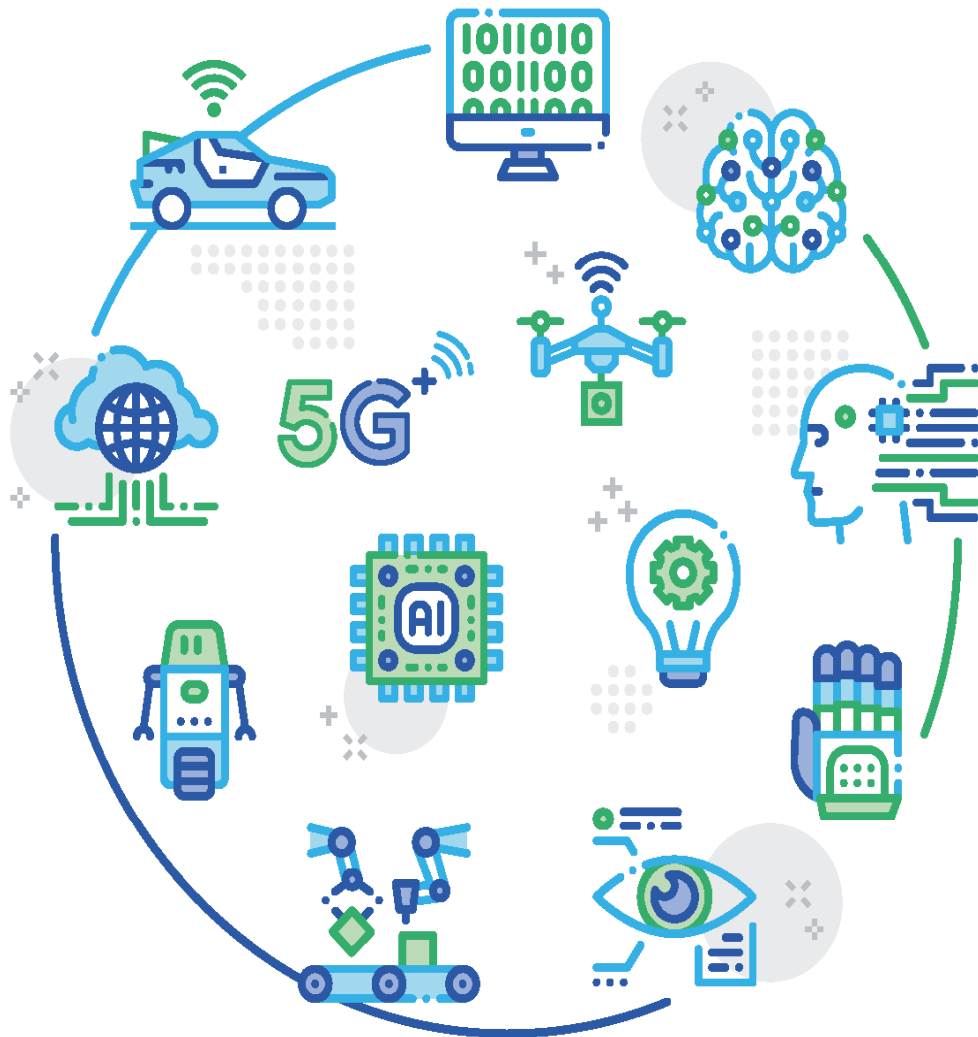


‘데이터 홍수’가 일으키는 새로운 경제

- 2/20 The Economist 특집 리포트



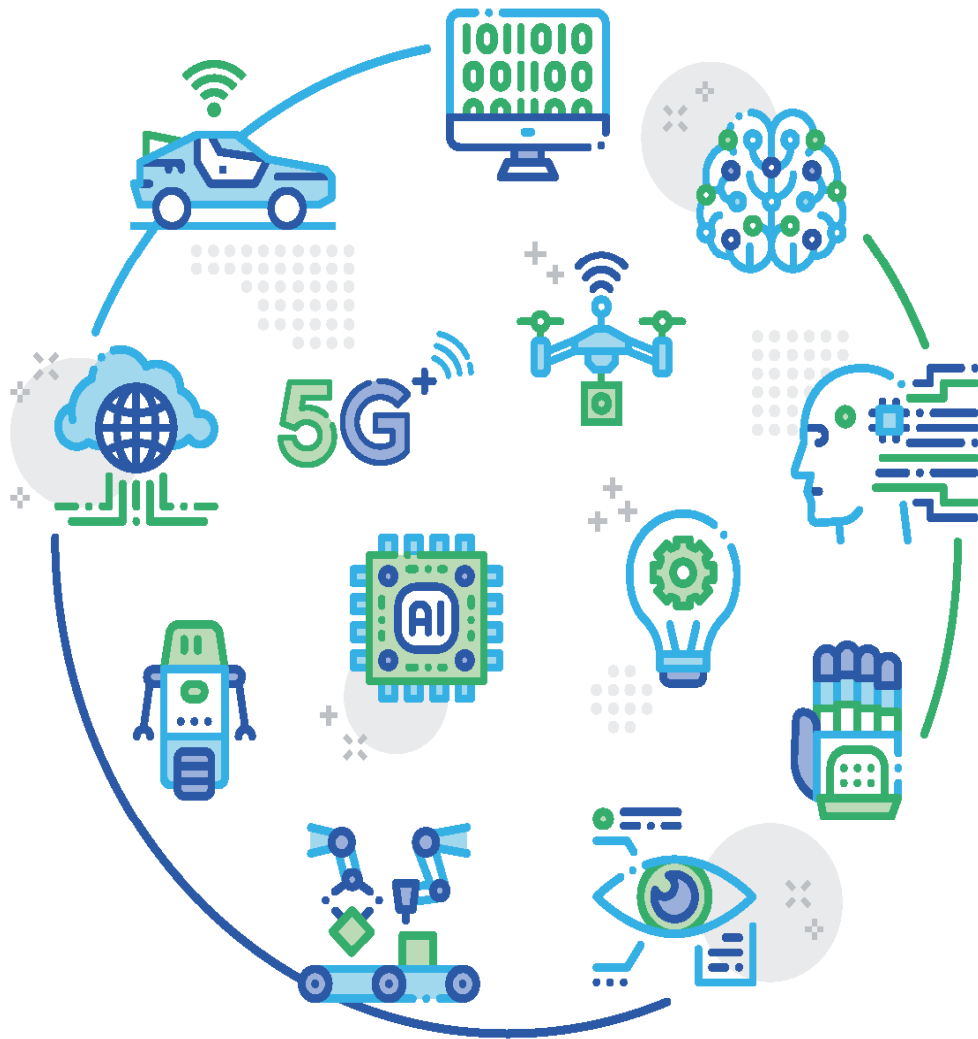
ICT POLICY PRIME

ICT POLICY PRIME은 새로운 디지털 대전환기에 대응한 정책 방향을 제안하기 위해 NIA에서 발간하는 보고서입니다.

-
- 이 보고서는 방송통신발전기금으로 수행한 과학기술정보통신부 정보통신·방송연구개발사업 (ICT진흥및혁신기반조성-지식정보사회의 국가발전전략연구사업)의 결과입니다.
 - 보고서 내용의 무단전재를 금하며, 가공·인용할 때는 반드시 출처를 밝혀 주시기 바랍니다.
 - 이 보고서의 내용은 한국정보화진흥원(NIA)의 공식 견해와 다를 수 있습니다.

‘데이터 홍수’가 일으키는 새로운 경제

- 2/20 The Economist 특집 리포트



목 차

I. 이코노미스트와 데이터 경제	1
II. 데이터경제(Mirror Worlds)	2
1. 데이터 홍수가 일으키는 새로운 경제	2
2. 데이터는 원유에 가까울까, 햇빛에 가까울까?	4
3. 데이터 저장·처리의 위치는 중앙? 혹은 엣지?	10
4. 갈수록 어려워지는 데이터 통합과 커지는 중요성	13
5. 각국의 데이터 국경 건설	17
6. 데이터 경제의 가장 많은 혜택을 얻는 이는 누구?	20

I. 이코노미스트(The Economist)와 데이터 경제

1 세계에서 가장 가치 있는 자원은 데이터('17년 5월)

□ '17년 5월 이코노미스트는 이미 데이터의 중요성을 강조

- 데이터를 21세기의 원유로 비유하며, 이번 세기의 성장과 변화의 원동력을 데이터가 담당 할 것으로 그 중요성을 역설
 - 데이터의 흐름은 새로운 인프라, 새로운 비즈니스, 새로운 독점, 새로운 정치 및 결정적으로 새로운 경제를 창출
 - 디지털 정보는 이전의 어떤 자원과도 다른 방식으로 추출, 정제, 가치 평가, 매매 및 판매된다 설명
- 그러므로 데이터는 시장의 법칙을 바꾸었고, 시장 규제에 대한 새로운 접근 방법에 대한 요구를 파생시켰음을 시사

2 데이터 홍수가 일으키는 새로운 경제('20년 2월)

□ 데이터 홍수가 일으키는 새로운 경제(p.2)

- 전세계 데이터 경제의 성장을 조망하고 관련 이슈 소개

□ 데이터는 원유에 가까울까, 햇빛에 가까울까?(p.4)

- 데이터의 유연한 경제적 특징에 따라 변모되는 개념을 소개

□ 데이터 저장·처리의 위치는 중앙? 혹은 엣지?(p.10)

- 엣지 컴퓨팅의 등장과 클라우드의 중요성, 데이터 세금에 대한 논의

□ 갈수록 어려워지는 데이터 통합과 커지는 중요성(p.13)

- 데이터 통합의 중요성 및 효과성을 소개하고 역량확보 방안 소개

□ 각국의 데이터 국경 건설(p.17)

- 데이터 주권 문제와 향후 우리가 선택할 수 있는 시나리오 제시

□ 데이터 경제의 가장 많은 혜택을 얻는 이는 누구?(p.20)

- 데이터로 인한 불평등 문제와 해결을 위한 아이디어 소개

II. 데이터경제(Mirror Worlds)¹⁾

1 데이터 홍수가 일으키는 새로운 경제

□ 실재하는 물리적 자아와 밀접한 연결을 통해 무수한 데이터 생성

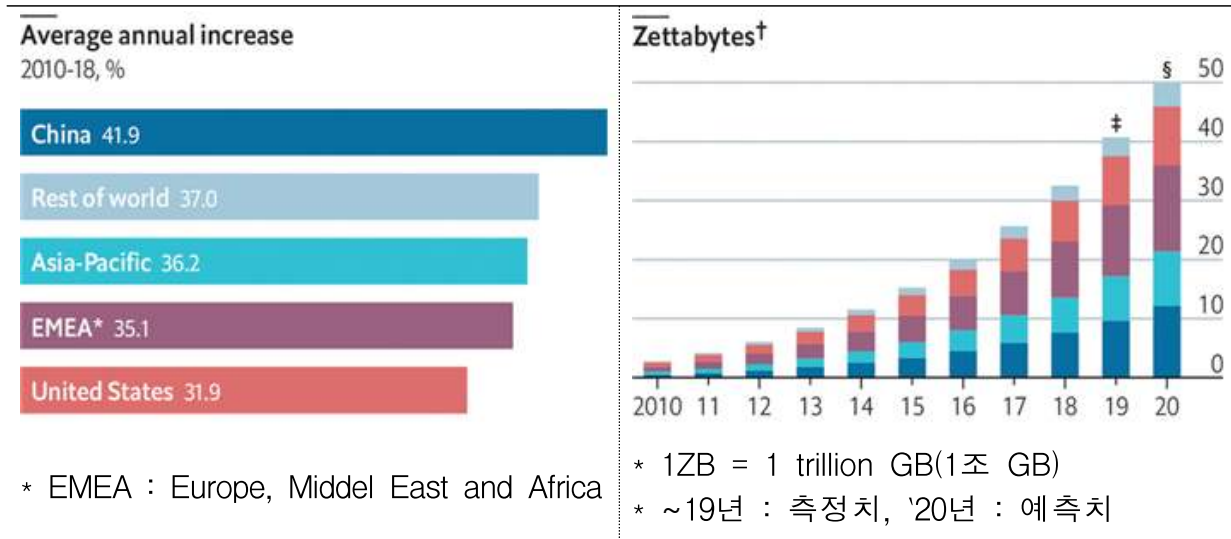
- 예일대 David Gelernter가 전망한 미러월드(Mirror Worlds), 즉 데이터에 기반하고 데이터로 돌아가는 새로운 인간 생활의 차원
 - 속칭 “디지털 트윈(digital twins)^{*}”은 다양한 센서와 네트워크를 통해 실재(true being)의 복제 이상의 의미를 가지며,
 - * 美 GE가 주창한 개념으로, 물리적 세계와 동일한 디지털 쌍둥이를 만들고, 현실에서 발생할 수 있는 상황을 시뮬레이션함으로써 결과를 예측하는 기술
 - 이러한 디지털 트윈은 세계 경제의 거대한 변화 중 일부에 불과하고, 미러월드의 등장은 다른 것과 뚜렷하게 구분되는 경제를 촉발할 것으로 예상
- 본 리포트에서는 ‘데이터 경제’의 전망과 위험을 집중적으로 조명

□ 데이터 경제의 급속한 성장보다 이윤배반성(Trade-off)에 초점

- 전 세계적으로 생성되는 데이터의 양만 본다면 데이터 경제는 빠르게 성장하고 있는 추세(표1. 참조)
- 시장조사 기관인 IDC는 올해와 내년에 전 세계적으로 약 90 ZB의(DVD 19조 장) 데이터가 만들어질 것으로 예측하며, 이는 컴퓨터 등장 이후 생산된 모든 데이터보다 더 많은 양
 - ※ 인간의 게놈 배열 순서는 17년전 최초로 밝혀졌으며(인간 게놈 데이터의 양은 3GB로 DVD 한 장을 거의 채울 만한 양), 지난 `19년 4월 유전자 검사 서비스를 제공하는 23andMe라는 회사는 고객 수가 천만명을 넘겼다고 밝힘
 - ※ 가장 최근의 자율주행 자동차는 여덟 시간의 운행마다 최대 30TB의 데이터를 생산(DVD 6,400장에 달하는 수준)
- 데이터는 어떤 면에서 원유와 같이 소유 및 거래가 가능한 자원이라 할 수 있으나*, 데이터를 통한 부의 창출을 극대화하기 위해서는 최대한 널리 사용되어야 하는 공공재의 성격도 존재
 - * `17년 이코노미스트에서는 데이터를 “세계에서 가장 가치있는 자원”으로 설명

1) The data economy(2020.2.20.), The Economist 재구성

< 표1. 세계에서 생성되는 데이터의 양 >



자료 : IDC, Seagate, The Economist

□ 데이터 인프라 또한 새로운 기회와 위험요인이 동시에 존재

- 데이터 수집이 이루어지는 곳에 더욱 가까운 “엣지(edge)”에서 더 많은 데이터가 처리되도록 데이터센터를 분산시키려는 추세
 - 현재는 주로 데이터 저장 및 처리를 위한 서버가 거대한 데이터센터에 몰려 있는 구조로, 에너지 소모가 심하고 프라이버시 위험이 발생할 수 있다는 단점을 보유
- 기업들도 디지털 전환을 마주하고 있으며, 많은 기업이 데이터를 활용하여 AI를 도입하기를 희망
 - 기업들은 “데이터 레이크(data lakes)” 와 같이 다양한 디지털 정보를 담고 있는 중앙 저장소를 구축하기 위한 시도도 추진
 - 그러나 ① 기업 및 직원이 데이터 활용에 필요한 역량(skills)을 갖추지 못하거나, ② 데이터를 신뢰하지 못하거나 ③ 심지어 내부적으로 공유하는 것을 거부한다면 이러한 시스템의 사용은 제한적일 것

□ 지정학적 측면에서 “디지털 주권(digital sovereignty)” 이슈 발생

- 디지털 자원은 기술 및 비용 조건이 최적인 곳에서의 가공을 위해 이동하기 때문에 온라인 기업들은 데이터 경제를 글로벌 차원의 문제로 고려
- 반면, 각국 정부는 점차 “디지털 주권”을 주장하며 데이터가 본국에서 이동하지 못하도록 국외이전을 제한하는 추세

□ 데이터에 대한 다양한 비유와 데이터의 유연한 경제학적 특징

• 데이터 흐름을 많은 비유들을 통해 다양한 의미로 묘사

- (원유 또는 태양광) 데이터는 미래의 연료라는 의미로 원유와 비교되다, 최근에는 태양광처럼 데이터가 어디에나 존재하고 모든 것이 바탕이 된다는 의미로 햇빛에 묘사
- (인프라) 도로나 철도 등을 디지털 트윈처럼 관리하기 위해 공공 투자와 새로운 기관이 필요하다는 의미로 인프라로도 표현

• 이는, 데이터의 유연한 경제학적 특징을 반영한 것

- (비경합성(non-rivalrous)) 무한대로 복제가 가능하여 다른 이의 사용을 제한하지 않고도 많은 사람들이 활용 가능
- (배제성(excludable)) 암호화(encryption)와 같은 기술은 데이터에 대한 접근을 통제 가능하며, 이것은 데이터 경제가 한 가지 유형만 존재하는 것이 아니라 서로 다른 유형(사적재, 공공재, 클럽재 등)으로 구분할 수 있음을 의미
 - ※ 데이터는 암호화 수준에 따라 원유와 같은 사적재(private goods)가 될 수도 있고, 햇빛과 같은 공공재(public goods)가 될 수도 있으며, 또는 클럽재(club goods)로 알려진 그 중간 어딘가에 해당될 수도 있음
- 주요 이슈는 이 중 한 가지 유형의 데이터 경제가 지배적이 될 것인가, 아니면 현실 세상만큼 미래월드에서도 여러 유형이 혼재적으로 존재할까 하는 부분

□ 데이터의 흐름에서 경제적 가치가 창출

• 원유와 마찬가지로 데이터도 사용하기 위해 정제되어야 하는데, 이는 부정확성을 걸러내고(정제, cleansed), 무슨 내용이 있는지 알려주는 표시를 하는 과정(태그, tagged)

- 이 과정은 주로 저임금 국가들에서 수십만명을 고용하는 글로벌 산업을 탄생시켰으며, 샌프란시스코의 한 스타트업인 Scale AI는 전 세계 3만명의 태거(taggers)를 고용하여 자율주행 차량이 전송하는 영상을 검토하고 SW가 사물을 정확하게 분류하는데 기여

- **일부 유형의 데이터와 이를 통해 추출한 분석 결과(insight)는 이미 널리 거래가 이루어지고 있음**

- AI 서비스를 실현하기에 앞서 데이터는 알고리즘을 거쳐 학습해야하고, 여러 데이터셋을 결합하여 통계적 패턴을 확인해야 함

- 또한 데이터 브로커들은 은행에서 통신사업자에 이르기까지 모든 곳에 개인정보를 판매하며 사업이 번창

- ※ '18년 컨설팅 기관 Strategy&은 온라인 광고 시장에서 각 뷰어의 디지털 프로필을 바탕으로 한 클릭 수의 판매 및 구입 규모는 전 세계적으로 1,780억 달러에 달하며, 데이터 브로커는 개인정보를 판매하여 연간 210억 달러가 넘는 수익을 창출

- **데이터 마이닝을 통해 분석을 수행하고 결과를 제공하여 수익을 창출하는 사례 또한 다수**

- 구글이 소유한 웹사이트로 머신러닝 대회를 주최하는 Kaggle에서는 데이터 과학자들로 이루어진 수천 개의 팀이 100만 달러가 넘는 상금을 위해 알고리즘 개발 경쟁 중

- ※ 건물의 에너지 소비량 예측 또는 “딥페이크(deepfake)” 비디오 탐지를 위한 최고의 알고리즘 개발 등

- 페이스북, 구글 등 역시 마찬가지로 데이터를 판매하는 것이 아니라 누가 최적의 광고 타겟인지에 대한 통찰을 판매

□ 데이터 가치책정에 대한 이슈는 여전히 존재

- **세계경제포럼(WEF)이* 2011년 예측한 바와 같이 데이터는 “새로운 자산군”으로의 편입에 실패**

- * WEF 저명한 기업인 · 경제학자 · 저널리스트 · 정치인 등이 모여 세계 경제에 대해 토론하고 연구하는 국제민간회의로서 싱크탱크의 역할을 수행

- 대부분의 데이터는 소유자가 변하는 일이 없으며 거래성을 높이기 위한 시도도 아직 시작되지 않은 실정

- 이를 바꾸기 위해, 특히 유럽의 경우 제조업체들이 자신의 제품을 통해 생성된 데이터에 대한 재산을 보장해 줄 것을 요구

- ※ 소비자들이 자신이 만들어내는 데이터를 소유하고 이를 판매할 수 있도록 하여 자신의 정보에 대해 우위를 차지하도록 요구하는 경우도 있음

- **주체에 따라서도 가치책정 및 귀속 여부에 대한 기준이 달라짐**

- 데이터는 주로 재화로 간주되지만 기업이 보유한 데이터셋의 경우 대체가 거의 불가능한 경향

- 각 기업의 수집 방법, 목적, 신뢰 수준이 모두 다르며, 각각의 가치를 비교하기가 힘들고 시간이 지나면서 가치도 변화하기 때문에 구매자와 판매자 간 가격 협의에 어려움

- 데이터를 누가 제어하는지에 따라 데이터셋의 가치가 달라지는데, 이는 어떤 기업에게는 단순한 데이터 기록이 다른 기업에게는 디지털 골드가 될 수 있음을 의미

- 개인정보의 경우 한 사람에게만 소유된다고 보기 힘든 경우가 많아 재산권을 정의하는 것이 어려움

- ※ (사례) 데이팅 사이트에서 커플을 연결시켜준 사실 정보는 누가 소유하는가? 커플인가? 서비스 회사인가?

- 데이터는 수많은 외부적 요인에 영향을 받으며, 이러한 외부 요인은 긍정적이기도 하고 부정적이기도 하여 시장의 실패로 이어지기도 함

- ※ (사례) 한 사람의 개인정보를 소셜 네트워크의 다른 이용자로부터 데이터를 수집·처리하는 방식으로 어느정도 정확하게 예측할 수 있다면, 굳이 왜 개인정보를 구매해야 할까?

- **데이터가 원유만큼 널리 거래될 확률은 낮지만 테크 기업들은 이를 쉽게 하려는 노력을 지속적으로 추진**

□ **오픈데이터 지지자들은 기관의 데이터 개방을 압박**

- **데이터를 원유에 비유하는 것의 문제점이 점차 드러남에 따라 햇빛 또는 공기, 물과 같은 자원에 비유**

- 이러한 비유를 선호하는 사람들의 주장은 데이터를 거래 가능한 재화로 만들 수 없다면, 차라리 데이터가 널리 사용할 수 있도록 보장하는 한편, 이를 통해 사회적 부를 극대화할 수 있을 것으로 전망

- 이미 이러한 논리는 “오픈데이터” 운동 이어졌고, 이를 지지하는 사람들은 조직 및 대학들이 보유한 데이터를 개방하여 스타트업과 같은 곳에서 널리 사용할 수 있도록 할 것을 주장

- **최근에는 기업들도 자체 보유한 데이터를 공개하기 시작**

- 구글의 모기업인 알파벳(Alphabet) 이 소유한 웨이모(Waymo) 등 자율주행차와 관련한 몇몇 기업들은 수집된 정보의 일부를 공개

☞ 웨이모 수석 과학자 Dragomir Anguelov,
“연구원들이 적절한 질문을 하기 위해서는 적절한 데이터가 필요하다”

- 데이터 공유를 더욱 쉽게 만드는 기술을 연구하는 곳도 있으며, 마이크로소프트를 비롯한 여러 소프트웨어 업체들은 머지않아 소위 “오픈데이터 이니셔티브” 를 도입하기 시작할 것으로 전망

☞ MS 최고법률책임자(CLO) Brad Smith 사장,
“우리는 시와 그 바탕이 되는 데이터를 민주화해야 한다”

- **원유로의 비유에서처럼 햇빛으로서의 데이터라는 비유도 한계**

- 오픈데이터도 어느 정도까지만 실현 가능하기 때문
 - 개인정보의 경우 EU의 개인정보보호규정(GDPR, General Data Protection Regulation)과 7월부터 시행될 캘리포니아주의 소비자 프라이버시법(CCPA, California Consumer Privacy Act) 등 프라이버시 관련 법령이 점차 엄격해지는 추세
 - 좋은 데이터를 생성하려면 비용이 많이 들고 기업의 경쟁력에 대해 지나치게 많은 정보를 노출 시킬 수도 있다는 점 등 경제성의 문제 대두

□ 데이터의 안전한 공유/유통에 대한 고민은 기술로 해결 가능

- **기술은 데이터 공유/유통에 대한 의사결정을 쉽게 해줄것으로 기대**

- “차별화된 프라이버시(differential privacy)” 는 의도적으로 데이터를 흐리게 근사치의 결과만을 나오게 만드는 것으로, 원 데이터셋의 통찰을 파괴하지 않음
- “동형암호(homomorphic encryption)” 기술은 알고리즘이 암호해독 없이 데이터를 처리하여 결과값을 도출 가능
- “블록체인(block chain)” 의 경우 누가 어떤 데이터에 접근할 수 있는지에 대한 정보를 상세하게 관리하고 접근 기록을 추적 가능

- 이러한 기술의 도입은 서서히 이루어지고 있는 추세

- EU에서 작년까지 자금을 지원했던 이니셔티브인 Decode의 경우 이러한 기술들을 모두 활용하여 도구를 개발

- ※ 현재 암스테르담과 바르셀로나에서 테스트가 진행되고 있으며, 사람들은 이 도구를 활용하여 소음, 공기질과 같은 주변 환경에 대해 자신이 만들어내고 수집하는 데이터를 관리 가능

- 샌프란시스코의 또 다른 스타트업인 오아시스랩스(Oasis Labs)는 건강 데이터와 관련하여 유사한 도구를 마련

- ※ 조만간 최초 서비스가 출시되면 이용자들은 자신의 유전 정보를 연구 프로젝트에 기증할 수 있게 될 것

□ 데이터를 인프라에 비유하는 사람은 데이터의 유형 구분을 선호

☞ 영국 Open Data Institute CEO Jeni Tennison ,

“어떤 도로는 요금을 내야 하는 민자도로일 것이고
어떤 도로는 다차선의 공공 고속도로이겠지만
많은 부분은 이용자들이 “공동체(Club)”로서 관리하는
공용 디지털 자원으로 운영되어야 한다.”

- 데이터 분석을 통해 새로운 통찰에 이르기 위해서는 많은 디지털 도로를 지나가야 하고 많은 데이터셋과의 결합이 필요

- “클럽재”를 창출하기 위해서는 기술만으로는 충분하지 않으며, 소위 “데이터 스튜어드십(data stewardship)*”을 제공하는 기관이 필요

- * 데이터 스튜어드십 : 데이터 자산을 관리하고 연구 분야나 조직의 요구에 맞춰 데이터를 활용할 수 있도록 운용하는 것

- 특정 데이터 유형을 생산하고 활용하는 사람들의 이해를 염두에 두는 방식으로 데이터 접근을 정리하는 거버넌스 구조를 제공

- ※ 데이터 신탁(trust), 데이터 협동조합(co-operatives), 개인정보 스토어 등 세부적으로 다르나 그 개념은 같음

- 초기에는 이러한 데이터 공동체(Club)가 여러 곳에서 등장

- (스위스) 마이데이터(Midata)는 회원들의 헬스케어 데이터를 수집·관리하는 협동조합

- (대만) 디지털 부처 장관인 Audrey Tang이 “대통령배 해커톤(Presidential Hackathon)”을 상시화하여 환경 데이터 관련 협업체 다수를 포함한 “데이터 협업기구(data collaboratives)”를 마련

- (핀란드) 정책 기관인 시트라(Sitra)가 유사한 대회를 마련하여 “공정한 데이터 거래”를 도모

□ 데이터 경제 유형 중 지배적인 것은 없으나 지역별 차이는 존재

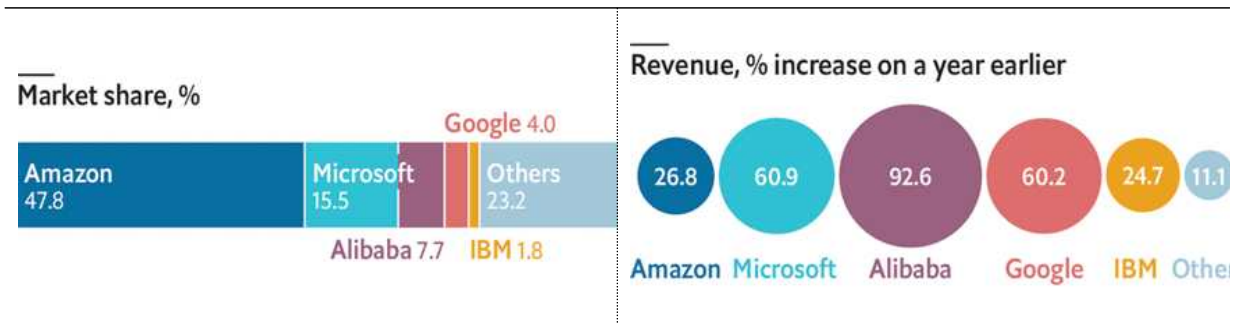
- (미국) 캐내는 사람이 소유하는 듯 데이터를 원유처럼 취급
- (중국) 알리바바와 텐센트처럼 데이터를 중시하는 온라인 플랫폼을 자체적으로 가지고 있으나, 데이터가 공공재에 해당하는 지역의 전형적인 사례로 꼽힘
 - 궁극적으로 정부가 데이터를 통제하고 기업들에게 건강 데이터와 같은 특정 데이터 종류를 수집하도록 압박
- (유럽) 많은 규제 기관들이 데이터를 인프라로 간주하기에 이르렀으며, 브뤼셀에서 새로이 출범한 유럽연합 집행위원회(EC)는 데이터 신탁 설립을 지원할 대규모 계획을 추진 중
 - EU가 기술에 뒤처져 보일 수 있으나, 이는 “공정한 데이터 경제”, 즉 미래 연료의 상당 부분을 생산해 낼 국민과 소비자들의 이해를 고려하는 경제로서 꽤 경쟁력이 있을 것으로 전망
 - 사람들과 기업들이 지역의 데이터 인프라를 신뢰한다면 질적으로 양적으로 더 좋고 많은 데이터를 기꺼이 공유할 것이며, 이는 모두를 위해 보다 나은 서비스로 이어질 것으로 예상
 - 이러한 “순환”이 이루어진다면 새로운 반전이 될 것으로 기대

□ “엣지 컴퓨팅(Edge computing)”의 등장

- 앞으로 데이터 경제의 인프라는 ① 데이터센터 등으로 일컬어지는 장소에 모든 데이터가 수집, 저장·처리되는 중앙집중형 모델과(예. AWS 등) ② 수집 장소에 최대한 가까운 곳에서 실시간으로 처리되는 엣지형 모델이(예. SWIM.AI*) 주도할 것

* SWIM.AI라는 스타트업의 SW는 도시 전체의 교통 흐름에 대한 일련의 계산을 수행하고, 각 교차로에 설치된 버튼을 누르는 보행자, 도로의 차량 등의 정보를 수집하여 교통신호의 2분 후 변화를 지속적으로 예측

< 표2. 2018 클라우드 서비스 경쟁 >



자료 : Gartner, The Economist

- 오늘날 비트(bit)는 웹사이트에서 클릭을 하거나 스마트폰을 터치함으로써 생성되고, 앞으로는 소위 “사물인터넷(IoT, Internet of Things)” 이라 불리는 기기*에 의해 생성될 것

- 네트워크 기기 제조사인 에릭슨(Ericsson)은 IoT 기기의 숫자가 2019년 110억 개에서 2025년 250억 개에 이를 것이라 예측
- 또한 에릭슨은 2025년까지 전 세계 모바일 망*에서 매월 160 엑사바이트(EB**)를 처리할 것으로 전망하며, 이는 현재의 4배에 해당

* 차세대 모바일 기술인 5G는 1평방 킬로미터당 100만 개의 연결을 지원하기 위해 개발된 것으로, 이는 맨하탄에서만 6천만 개의 연결을 의미

** 1EB = 10¹⁸ bytes = 100경 바이트

- 생성된 비트의 목적지 또한 최초 생성된 기기에 머무르는 것에서 클라우드 흘러가는 것으로 바뀌고 있는 추세

□ 대규모 클라우드 공급자는 중앙집중형 모델을 강화하기 위해 노력

- 이러한 접근의 주된 이유 중 하나는 데이터 중력(data gravity)*

- * 데이터가 더 많은 데이터를 끌어들이는 현상을 표현

- 기업의 주요 데이터가 클라우드에 자리잡게 되면, 더 많은 애플리케이션이 클라우드 컴퓨팅 기반으로 이동하게 되고,
 - 이는 클라우드 공급자들이 더욱 풍부한 서비스를 제공할 수 있도록 하여 고객들이 자신의 데이터를 마이닝하여 통찰을 얻을 수 있도록 선순환

- 그러나 중앙집중형 모델을 구축하기 위해서는 비용이 수반

- 한가지는 기업들이 데이터를 다른 클라우드로 이동할 때 지불해야 하는 높은 비용이 발생
 - 또한 데이터를 중앙으로 보내기 위해 발생하는 에너지가 소모로 데이터를 대규모 센터에 집중하는 것은 환경에 많은 부담될 우려
 - 중앙에 모은 후에도 지속적으로 데이터를 가공·처리하려는 유혹이 큼

- ※ Open AI에 따르면 최첨단 AI 프로젝트에 사용되는 컴퓨팅 역량은 2012년 폭증

- ※ 스탠퍼드대의 보고서²⁾에 따르면 '12년 이전에는 컴퓨터 칩의 연산 능력이 2년마다 2배씩 증가한다는 무어의 법칙대로 증가해 왔으나 그 이후는 3.4개월에 2배씩 증가하고 있다고 소개

□ 이에 대응하여 데이터가 생성되는 엣지에 대해서도 변화가 시작

- 대규모 데이터센터의 서버만이 아니라, 보다 규모가 작은 로컬 센터를 비롯하여 강력해진 연결된 기기를 바탕으로 데이터 소스와 더욱 더가까운 위치에서 데이터 분석 가능

- 뿐만아니라, 최적의 성능을 발휘할 수 있도록 IoT 기기 또는 해당 기기 근처로 컴퓨팅 역량을 이동하기 위한 소프트웨어도 마련

- ※ (예시) 자율주행차와 같은 애플리케이션은 반응속도가 매우 빠른 연결을 필요로 하며 연결이 끊어지는 위험을 감수할 수 없기 때문에, 가까운 데이터센터나 자동차 내부에서 컴퓨팅이 이루어져야 함

2) 미국 스탠퍼드대 인간중심인공지능연구소(HAI)가 국제컨설팅그룹 맥킨지 등과 공동으로 작성한 '인공지능 인덱스 2019 연례 보고서(The AI Index 2019 Annual Report)

□ 머지않아 데이터 생산에 대한 세금 부과 논쟁이 치열해질 것

- 데이터가 가져다주는 인센티브를 고려하면 대규모 클라우드 공급자들은 계획해서 많은 데이터를 수집하고 저장·처리하려 할 것
- 이러한 논쟁에 대한 답은 입장에 따라 각양각색일 것으로 예상

- 데이터의 집중과 분산에 관한 인터뷰 사례

- (AWS, AI 담당 Matt Wood) “엣지가 중요하다는 점을 인정하나, 모형의 훈련을 위해 언젠가는 가지고 있는 모든 데이터를 취합해야 한다”
- (MS, Azure IoT 사업부문 Sam George) “컴퓨팅이 클라우드와 엣지에 고르게 분포되어 있기를 기대한다”
- (SWIM.AI, 수석 기술자 Simon Crosby) “엣지형 모델 접근법이 ‘모두에 적용되는 것은 아님’ 을 인정하는 한편, 클라우드로 전송되기에는 엣지에서 생산되는 데이터가 많으며 모든 모형을 중앙에서 훈련하도록 도울 데이터 과학자들이 절대 부족할 것이다”

□ “단일 버전의 진실(a single version of the truth)”을 찾는 과정

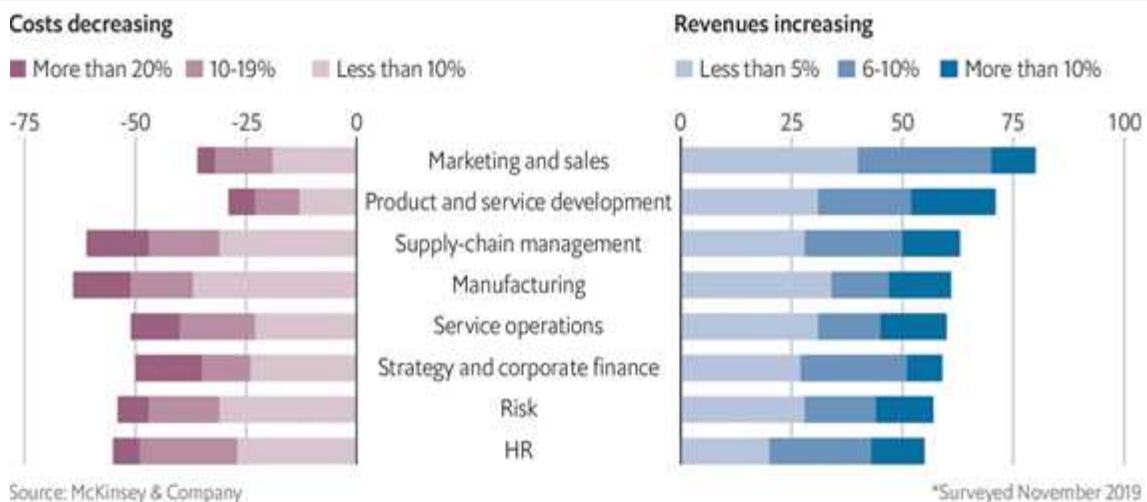
• 데이터 경제 속에서 기업의 주요 불안을 압축하고 있는 표현

- 디지털 진실을 찾는 것, 즉 현실을 정확하게 나타내는 데이터를 찾아내고 결합하는 것은 더욱 어려워지고 있지만, 그 중요성은 커지고 있어 경쟁력 유지를 위해 기업들은 수많은 노력을 경주

- (어려워지는 이유) 데이터와 그 소스가 크게 증가
- (중요해지는 이유) 기업들이 AI의 혜택을 얻기 위해서 데이터 하우스를 마련하고, 이를 통해 경쟁력을 유지*

* 맥킨지의 최근 조사에 따르면 AI가 수익을 높여주는 것으로 나타남

< 표3. AI 도입 후 평균 비용 감소 및 수익 증가와 응답자 비율(%) >



* 2019년 11월 조사 기준

자료 : McKinsey & Company, The Economist

• 데이터 처리 소프트웨어와 클라우드 컴퓨팅은 “AI-조립라인*”의 구축 가능성을 더욱 높일 것으로 기대

* IT 산업 부문 투자자이자 관찰자인 George Gilbert가 100년 전 공장의 주요 동력원이 증기에서 전기로 대체된 때에 빗대어 언급

- 디지털 시대의 기계는 “기업의 비즈니스 애플리케이션과 이를 구축하기 위한 소프트웨어” 로,

- 이는 새로운 동력원인 디지털 저장소(repositories), 즉 “데이터 웨어하우스” 또는 “데이터 레이크” 를 중심으로 사실상 재편성

□ 디지털 진실을 찾는 과정의 어려움과 개선방안

• **나날이 데이터의 소스와 그 형식이 믿기 어려울 정도로 다양해지고 통합하기가 어려워지는 실정**

- 디지털 저장소들을 통합하고 연계하려는 시도가 실패한 사례 다수
 - ※ 고객의 이름과 같이 간단한 데이터조차도 여러 가지 다른 방식으로 정의 및 저장될 수 있고, 기업들이 사용하는 SW 및 애플리케이션은 각기 다른 자체 데이터베이스를 가지고 있을 수 있는 등
- 오늘날 기업의 데이터는 여러 로컬 데이터베이스에 분산되어 있을 뿐만 아니라 여러 클라우드 서비스에 올라가 있으며, 제3자 및 연결된 기기들로부터도 흘러들어와 통합이 더욱 어려워진 실정

• **디지털을 사용하기 쉽게 만들어 주고 있는 것이 “데이터 웨어하우스”와 “데이터 레이크”**

- 이들은 정보를 구조화하는 방식에 있어 차이가 있는데, 데이터 웨어하우스가 데이터 레이크보다 더욱 엄격한 접근을 취하나, 두 가지 모두 클라우드와 연계할 수 있어 차이는 점차 줄고 있음
- 이를 통해 디지털의 관리가 더욱 저렴해질 뿐 아니라 다양한 소스로부터의 데이터를 쉽게 피드할 수 있고 사용자들의 이용이 쉬워짐
 - ※ (사례) 스노우플레이크(Snowflake)라는 스타트업은 데이터 웨어하우스를 여러 클라우드에 걸쳐 적용 가능한 “데이터 플랫폼”으로 전환

• **또다른 개선방안은 특정 유형의 데이터를 다루는 특수 데이터베이스**

- 실시간 디지털 스트림으로 들어오는 데이터가 증가함에 따라 그 취급 방식도 달라져야 함
 - ※ (사례) Confluent 라는 스타트업은 오픈소스 프로그램인 Apache Kafka를 기반으로 하여 데이터 스트림을 분석하고 이를 데이터 레이크에 덤핑하는 클라우드 서비스를 판매하며, 독일의 대기업인 보쉬(Bosch) 등이 이를 활용)
- “AI-조립라인” 으로 전환시키는 SW와 서비스는 데이터의 수집·저장을 비롯하여, AI 알고리즘의 훈련 설계, 의사결정 자동화 등을 지원
 - ※ 전기회사 에넬(Enel), 정유 회사 셸(Shell), 비영리 대출기관인 키바(Kiva) 등이 이러한 기능을 활용하여 서비스를 개발하고 업무 프로세스를 개선

□ 데이터 정제와 통합은 기업의 데이터 생산성 증가의 열쇠

• 기업의 데이터 활용 역량 강화를 위한 애플리케이션 공급 기업 증가

- (IBM) 기업들이 AI 애플리케이션을 개발하기 위해 사용하는 프로그램의 집합체인 “데이터 플레인(data plane)” 을 구축할 수 있도록 지원하고, 시각 인식과 번역 같은 다양한 AI 서비스를 제공하여 다른 기업이 자사의 제품에 프러그인 할 수 있도록 허용
- (Oracle) “자율운영 데이터베이스(autonomous database)” 라는 것을 제공하며, 이러한 서비스는 모든 종류의 디지털 저장소와 AI를 결합하고 자동화하여 고객들이 직접 취합하지 않아도 되도록 지원
- (Salesforce*) 기업이 고객의 행동을 예측하고 매장에서든 온라인에서든 맞춤형 서비스 제공할 수 있도록 지원

* 신생 IT 회사로 지난 2년간 아인슈타인(Einstein)이라고 불리는 자체 AI 기술을 개발하고 MuleSoft와 Tableau라는 빅데이터 회사 인수에 수십억 달러를 투자

- (Databricks) 데이터 정제, 알고리즘 개발 및 적용을 위한 도구로 완비된 AI 플랫폼을 마련
- (C3.AI) Databricks와 유사한 서비스를 제공하나, 그 목적은 대기업들의 디지털 전환을 도와주기 위한 것
- (Qlik) 데이터 분석 및 시각화로 알려진 회사로 최근 AI분야로 진출

• 가트너(Gartner)의 Debra Logan는 이러한 도구들이 있음에도 불구하고, 많은 AI 프로젝트에 여전히 한계가 있다 지적

- 기업의 내부 경계선이 반영된 데이터 사일로(silos) 존재하여 부서간 데이터 공유를 꺼리는 현상 발생
- 또한 데이터 수집 종류 및 방식을 변화에 거부반응도 다수 존재
- 이로 인해 많은 기업들이 사업 목표를 달성하는데 필요한 데이터를 수집하고 분석 하도록 하는 일관된 “데이터 전략” 을 개발하지 못하고 있는 실정이라고 주장

□ 데이터 정제와 통합은 기업의 데이터 생산성 증가의 열쇠

☞ “우리는 데이터가 객관적이라 생각하지만
사실은 셰익스피어만큼 해석이 가능하다”

- **디지털 격차를 극복하기 위해 일부 기업들은 조직적 변화를 단행**
 - “최고 데이터 책임관(CDO, Chief Data Officer)” 을 임명하여 부서 및 업무 단위들이 협력하도록 중재하고, 이를 통해 AI 조립라인과 비슷한 기능을 수행할 수 있도록 지원
- **그러나 문화 조성, 직원역량 강화 등 기업의 나머지 부분에서도 준비가 되어 있지 않다면 최고위층 인사, 그리고 기술 방면에서 이루어지는 변화는 그 가치를 발휘하지 못할 것**
 - ※ 최근 가트너의 조사에 따르면 기업의 데이터 프로젝트에 걸림돌이 되는 가장 큰 문제는 “변화를 수용하려는 문화적 도전”이고, 그 다음으로 “심각한 데이터 지연”으로 나타남
 - 이를 변화시키기 위해서 모든 직원들이 데이터 과학자가 되어야 한다는 것은 아니나,
 - 데이터가 무엇을 위해 사용될 수 있는지 등에 대한 기본적인 이해는 갖추어야 할 필요

□ 데이터 주권이 큰 이슈로 급속히 부상

- **각국 정부는 자국의 국민, 주권, 경제를 보호하기 위해 데이터의 자유로운 흐름을 점차 막고 있는 추세**
 - 영국 Songkick 회장 Ian Hogarth에 따르면, 이와 같은 디지털 보호주의 움직임은 각국이 단순히 자신의 데이터 자산을 방어하고 스스로의 데이터 경제를 구축하는 선을 넘어 전면적인 “AI 국가주의” 로 돌아설 수 있다 우려
 - 아리조나 대학의 Andrew Woods는 “범세계적 이상(cosmopolitan ideal)” 은 데이터의 자유로운 흐름이 세상을 더 나은 곳으로 만들지는 못하더라도 최소한 더 효율적인 곳으로 만드는 것이라 표현
- **미국 첩보기관의 엠탐행위, 글로벌 정보유출 등의 문제가 드러나게 됨에 따라 각국 정부들은 데이터 인프라의 중요성을 넘어 데이터 경제 자체에 대한 중요성을 이해하기 시작**
 - 이러한 문제들에 대한 디지털 증거가 해외에 저장되어 있다면 법적 집행에도 제약이 발생

□ 그 결과, 최근에는 가상 국경이 세워지는 추세

- **각국은 데이터의 국외이전에 대응하기 위한 조치를 마련**
 - (EU) GDPR은 개인정보가 EU를 벗어날 수 있는 때는 기업이 적절한 안전장치를 갖추고 있거나 목적지 국가가 “적절한 보호 수준” 을 마련하고 있을 때에만 가능하다고 명시
 - ※ EU는 역내 상품과 마찬가지로 데이터에 대한 단일 시장 구축 방안에 대해 논의
 - (인도) 결제 정보가 국외로 나가지 못하게 막고 있으며 조만간 특정 유형의 개인정보도 국가를 벗어나지 못하게 될 것으로 예상
 - (러시아) 영토 내에 있는 서버에서만 데이터 처리 및 저장이 가능
 - (중국) 대부분의 국제 데이터 흐름을 차단
- **데이터 주권을 찾기 위한 노력은 이미 국제 외교의 최상위 무대에서도 논의를 촉발**
 - `19년 7월에는 20개의 개발도상국과 선진국들로 구성된 G20에서 “오사카 트랙(Osaka Track)” 이 출범
 - ※ “신뢰를 기반한 데이터의 자유로운 흐름(free flow of data with trust)”을 바탕으로 “데이터 거버넌스”를 위한 글로벌 규칙을 마련하자는데 합의

□ 데이터 주권 확보 방안과 시나리오

• (시나리오1) 모든 데이터 이동의 원천적 차단

- 모든 데이터의 이동을 방지하기 위해 국가들은 인터넷 연결을 필수적으로 차단해야 하며, 이것이 데이터가 정말로 빠져나가지 않게 하는 유일한 방법일 것
- 과도하게 보호조치를 취하는 국가의 경우, 클라우드 컴퓨팅 공급자들이 시장이 작다는 이유로 서비스를 거부하는 결과를 보게 될 것이며, 내수를 위한 클라우드를 구축은 어렵고 많은 비용이 발생

• (시나리오2) 여러가지 데이터 유형을 위한 연합 형성

- 이는 이미 일어나고 있는 일이며, 실제로 GDPR의 적정성에 대한 의무요건이 이를 효과적으로 만드는데 일조
 - EU에서 개인정보를 수출해야 함에 따라 미국과 일본을 포함한 많은 국가들은 엄격한 데이터 보호 규정에 동의
 - 미국은 2018년 통과된 클라우드법(Cloud Act)에서 이와 유사한 공동체(club)을 마련하여 정부간의 호혜협정을 타결하도록 조치

※ 영국은 이미 이러한 협정에 서명했고, EU도 곧 동참할 것으로 예상

- 오사카 트랙이 글로벌 규칙을 마련하기 위한 것이긴 하지만 결국 또 다른 데이터 연합을 만들어내는 결과를 가져올 것
 - 일본 정부의 제안으로 시작된 이 계획은 미국, EU와 함께 동맹을 구축하여 가입국 간의 자유로운 데이터 흐름을 촉진하는 한편 중국과 같이 데이터 보호주의에 빠져있는 국가들의 접근을 제한하기 위한 것

※ 이에 대한 우려로 인도와 몇몇 개발도상국들은 오사카 트랙에 서명을 거부

• (시나리오3) 글로벌 데이터 스피어(data sphere)의 미래

- 독일에서 시작된 가이아-X(GAIA-X)가 대표적인 사례
 - 이는 독일 기업들과 정부기관을 해외 거대 클라우드의 힘으로부터 보호할 합법적인 SW 레이어인 “위버클라우드(über-cloud)” 구축을 준비하기 위한 것
 - 이 프로젝트는 기업간, 국경간 데이터 공유를 보다 쉽게 하기 위한 “국제 데이터 공간(International Data Spaces)” 이라는 계획도 포함

- 가이아-X는 저속한 디지털 보호주의에 안주하는 대신 세밀한 국가 데이터 정책을 도입하게 해주는 도구가 될 것으로 예상
 - 또한, 글로벌 데이터 인프라를 지배하는 미국 또는 중국 기업들에 대응하는데 도움이 될 것으로 기대
- ※ 독일 정부는 유럽이사회 의장국을 맡게 되면 이를 추진할 계획이며, 프랑스는 이미 지지를 표명했으며 다른 국가들도 동참할 것으로 기대
- ※ 또한 대규모 클라우드 공급자들을 포함한 100여개 기업들과 조직들이 이미 동참하고 있으나, 놀라운 점은 마이크로소프트가 아직까지 불참하고 있다는 점
- 가이아-X의 비전이 실현된다면 데이터 경제가 “블록”으로 형성되기보다는 새로운 다양성을 나타낼 수 있을 것으로 기대

□ 데이터 경제의 최종문제는 창출된 부의 분배 문제

☞ 산타페연구소 Brian Arthur,

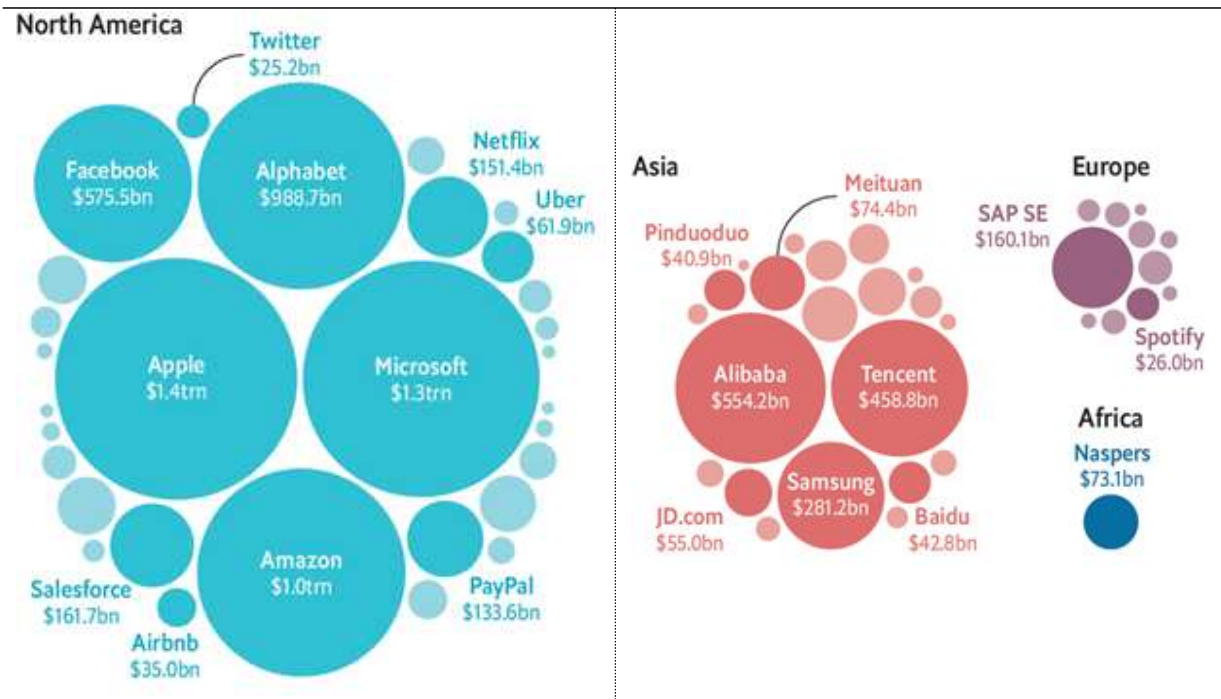
“제2의 경제(second economy)라고도 할 수 있는 데이터 경제가 ‘더 많이, 더 효율적으로 생산하는 것이 주요 관건이었던 경제에서 생산된 부의 분배가 가장 큰 이슈가 되는 경제로 이동할 것이다’

- 데이터 경제는 이미 네트워크 효과로 인해 거대 데이터 플랫폼 기업에 의한 시장 지배력 강화 등 불평등이 심화

- 데이터 경제가 확대됨에 따라 불평등은 비기술 기업들을 비롯해, 심지어 국가 간에도 점차 적용될 것이라는 전망이 강화되는 추세

※ 국가 단위로는 미국과 중국이 전 세계 70대 플랫폼 시가총액의 90%를 차지하고 있으며(표4. 참고), 아프리카와 남미는 1%만 보유

< 표4. 거대 글로벌 플랫폼 기업 위주의 경쟁 >



* 주요 글로벌 플랫폼 시가총액(30억 달러 이상)

* 2020년 2월 1일 기준 (단위: 10억 달러)

자료 : Bloomberg, CB Insights, The Economist

- 또한 주거용 부동산 분야의 Compass, 자율주행차 분야의 테슬라 등 많은 부문에서 데이터 플랫폼 지배를 위한 경쟁이 심화되는 중

□ 자본과 노동사이의 불균형한 소득분배는 심각한 문제로 대두

- 많은 노동이 미래월드로 이동하게 될 것이고, 사람들이 단지 디지털을 더 많이 사용하게 될 뿐만 아니라 실제로 “데이터 업무”를 수행하여 AI 서비스를 훈련하고 개선하는데 필요한 디지털 정보를 생산하게 될 것
 - 이미 온라인을 돌아다니며 피드백을 제공하는 단순한 것에서부터 사진에 제목을 달고, 데이터를 수집하는 차량을 운전하며, 디지털 트윈의 역량을 시험해보는 등 점점 더 많은 활동이 포함될 것
 - 이러한 이유로 일각에서는 AI를 사실 “집단 지성(collective intelligence)” 라고 불러야 한다고 말하기도 하며, 이러한 활동에 사람의 투입이 상당량 필요한 것이 사실
- 문제는 이러한 데이터 근로자들의 임금을 책정하는 것
 - MS의 Glen Weyl은 현 상황에서 이들의 업무는 시스템적으로 가치가 과소평가될 수 있다고 주장
 - 한 가지 이유는 온라인 시장의 구조인데, 거대한 플랫폼이 판매에 있어서 뿐만 아니라 구매에 있어서도 독점
 - 즉, 거대 플랫폼이 데이터 노동의 임금을 결정할 권력을 쥐고 있다는 것을 의미
 - 또한, 데이터의 경제학 또한 데이터 노동의 가격을 높이지 못하도록 압박
 - ※ (예시) 주변 다른 사람의 정보를 가지고 저렴한 비용으로 한 개인의 정보를 추론할 수 있다면 개인의 데이터에 대해 기업이 굳이 높은 가격을 지불할 이유가 있는가?
 - 주요 투입의 상당 부분을 생산하는 주체가 지속적으로 제값을 못 받는 구조의 데이터 경제는 건강하다고 볼 수 없음
 - ※ 고급 전문가들은 자신의 지식을 비밀로 지키고 데이터 제공에 참여를 거부할 수도 있을 것이며, 또한 자신이 생성하는 정보에 대해 발언권이 없으며 저임금을 받는 데이터 근로자들은 점차 소외감을 느끼게 될 것

□ 이의 해결을 위한 다양한 방안이 지속적으로 논의되는 중

- (방안1) 캘리포니아의 민주당 출신 주지사 Gavin Newsom은 “디지털 배당금(digital dividend)”의 일환으로 거대 테크 기업들로부터 세금을 거두어 캘리포니아주 시민들에게 나누려 시도

- 그러나 재분배를 통해 이 문제를 해결하는 것은 데이터 경제에 부담이 될 수 있으며 거래의 갈등으로 이어질 우려가 있으며, 정권 변화에 따라 정책의 연속성 보장이 어려움

- (방안2) 개인정보에 재산권 제도를 도입하여 사람들이 더 큰 협상력을 가질 수 있게 하는 방안

- 그러나 재산권 제도는 문제를 더 심각하게 만들 우려가 존재

- 대부분의 개인정보는 근본적으로 한 사람 이상이 권리를 가진 사회적 산물이기 때문에 개개인의 권리 주장이 불분명

※ (예시) 한 가족의 구성원 각각이 자신의 유전자 정보를 판매한다고 해도 이 데이터를 통해 친척들의 DNA 정보까지 노출

- (방안3) 협동조합 형태로 데이터에 대한 통제권을 부여하는 방안

- 이는 기존 경제의 노동조합과 같은 개념으로, 이 협동조합들은 데이터 업무에 대한 비용을 협상하고 조합원의 디지털 결과물의 품질을 보장하며 이러한 결과물로 혜택을 얻는 기업들에게 비용을 청구하고 수익금을 분배

※ (사례) MS의 Glen Weyl비롯해 그가 설립에 도움을 준 활동가 조직인 래디컬익스체인지(RadicalxChange)는 개개인 모두가 “데이터 협동조합”이라는 곳에 가입하기를 희망

- 데이터 신탁과 마찬가지로 견고한 데이터 협동조합은 모든 관련 영역에서의 도움이 필요

- 일부 서구권 국가들은 래디컬익스체인지가 마련한 초안을 바탕으로 “데이터 자유법(Data Freedom Act)”의 논의를 시작할 수도 있으며, 이를 통해 해당 목적에 맞는 새로운 규제대상의 주체가 생겨날 수 있음

□ 앞으로의 데이터 경제를 맞이하기 위한 철저한 준비가 필요

- 거대 테크 기업의 수장들 중에는 MS 최고경영자 Satya Nadella가 1월 다보스에서 개최된 세계경제포럼에서 “데이터 존엄성(data dignity)”에 대해 산업계가 더 많은 존중을 표시해야 한다고 주장
 - 이는 사람들에게 자신의 데이터에 대한 권한을 더 많이 부여하고 이러한 데이터가 창출하는 가치에 더 많은 비중을 두어야 한다는 의미
 - 대중은 자신의 데이터로 어떤 일이 벌어지는지에 대해 점점 더 우려를 나타내고 실정
 - ※ 예를 들어 이제 미국 국민 10명 중 약 8명은 기업들이 내게서 수집해가는 데이터에 대해 내가 어떻게 할 수 있는 부분은 미미하거나 아예 없다고 생각
- 데이터 경제의 적절한 메커니즘과 제도를 수립하는 데에 많은 시간이 걸릴 것으로 예상되나, 우리는 이에 대비하여 많은 준비와 논의를 앞으로도 진행해 나가야 할 것
 - 데이터 경제가 성장할수록 이와 같은 논의는 더욱 뜨거워질 것으로 예상되며, 이에대한 기대와 우려가 혼재
 - 1850년대 산업혁명은 생산의 큰 증가를 가져왔으나, 사회가 적응하는 데에는 100년이 걸렸으며 적응하지 못한 곳도 있음

ICT POLICY PRIME

▶ 작 성

- 한국정보화진흥원 정책본부 정책기획팀
우창완 선임

▶ 기 획

- 한국정보화진흥원 정책본부
박원재 본부장
- 한국정보화진흥원 정책본부 정책기획팀
이정아 팀장

- 이 보고서는 방송통신발전기금으로 수행한 과학기술정보통신부 정보통신·방송연구개발사업(ICT진흥 및 혁신기반 조성-지식정보사회의 국가발전전략 연구 사업)의 결과입니다.
- 보고서 내용의 무단전재를 금하며, 가공·인용할 때는 반드시 출처를 밝혀 주시기 바랍니다.
- 이 보고서의 내용은 한국정보화진흥원(NIA)의 공식 견해와 다를 수 있습니다.