

데이터 경제 시대, 데이터 공유와 활용을 위한 과제

정도범 책임연구원
한국과학기술정보연구원 조사분석평가팀
dbchung@kisti.re.kr

DATA

들어가며

불과 작년까지만 해도 거의 언급되지 않았던 용어인 챗GPT가 최근 가장 큰 화두로 등장하고 있다. 챗GPT는 OpenAI가 2022년 11월 30일에 출시한 대화형 인공지능(AI)으로, 대화(Chat) 기능과 언어 모델(GPT-3.5 또는 2023년 3월 출시된 GPT-4)을 결합해 다양한 분야의 질문에 대해 정교한 답변과 수준 높은 결과물을 생성하고 있다. 2016년에 AI 바둑 프로그램인 알파고(AlphaGo)가 이세돌 9단과의 대국에서 승리해 인공지능에 대한 인식을 확산하는 계기가 됐다면, 2023년에 챗GPT는 우리의 일상생활에서 인공지능을 누구나 쉽게 체험할 수 있는 기점이 됐다고 볼 수 있다. 실제로, 인스타그램(30개월)이나 틱톡(9개월) 등과 비교했을 때 챗GPT는 월간 사용자 수 1억 명을 달성하는데 고작 2개월이 소요되는데 불과해 향후 인공지능이 일상 속으로 더욱 스며들어 전 세계에 사회·경제적으로 매우 큰 영향을 미칠 것으로 예측된다.

그리고 챗GPT 열풍으로 인해 데이터의 중요성이 다시 주목받고 있다. 초거대 AI 또는 대규모 언어 모델로 알려진 챗GPT는 방대한 규모의 데이터를 학습하는 것이 필수적이며, 이에 따라 AI 모델의 성능도 학습하는 데이터의 양과 질에 의해 크게 좌우될 것이다. 오늘날, 국내외 우수 기업들은 전략적으로 데이터를 수집·구축하는데 초점을 맞추고 있으며, 데이터와 AI를 접목해 챗GPT 외에도 혁신적인 비즈니스 모델을 발굴해 나가고 있다. 향후 데이터의 활용을 통해 과거에 존재하지 않던 새로운 가치가 더욱 많이 창출될 것이고, 기존 산업의 혁신이나 신산업 발굴 등이 가능해질 것이다. 우리는 이러한 변화를 의미하는 데이터 경제 시대에 대해 올바르게 이해하고, 세계적인 변화의 흐름에 능동적으로 준비하고 대응할 필요가 있다.

데이터 경제 시대

요즘 ‘디지털 경제’ 또는 ‘데이터 경제’라는 표현이 거의 흔재돼 사용되고 있다. 보통 ‘디지털 경제’는 1990년대 후반에 인터넷을 기반으로 이루어진 경제 활동, 즉 전자상거래(e-Business)를 의미하는 용어로 사용됐다. 2020년 이후 코로나19가 경제·사회 전반에 걸쳐 비대면화·온라인화·디지털화를 촉발함에 따라, 문재인 정부 그리고 2022년에 출범한 윤석열 정부에서도 디지털 기술을 다양한 분야에 적용해 전통적인 운영 방식이나 구조 등의 혁신을 추구하는 디지털전환(혁신)을 추진 중에 있다. 특히, AI, Cloud, IoT, AR/VR/XR 등 디지털 기술의 급속한 발전으로 인해 디지털 경제의 개념이 과거에 비해 폭넓은 의미로 사용되고 있으며, 전체 경제에서 차지하는 비중도 점점 증가하고 있다.

이러한 디지털 경제 하에서 전산화(Digitization), 디지털화(Digitalization)를 통해

일상생활을 포함한 모든 활동이 데이터로 축적될 수 있는 환경이 마련되고 있다. 이에 따라 삼일회계법인(2020)은 디지털 경제를 ‘디지털화된 데이터를 활용한 모든 활동’과 같이 광의의 의미로 설명하기도 한다. OECD(2021)도 데이터에 대해 관찰 가능한 현상에 대해 디지털 형식으로 측정하고 기록한 정보 콘텐츠로 정의함으로써 (아날로그 자료를 제외한) 전자적으로 접근·처리할 수 있는 자료임을 강조하고 있다. 즉 디지털 경제에서는 디지털 형식으로 저장된 데이터를 활용해 다양한 가치를 창출하는 것이 가능해지고 있다. David Newman(2011) 등이 데이터 경제에 대해 다양하게 정의하고 있지만, ‘데이터 경제’라는 용어도 결국 디지털 경제 환경에서 (디지털) 데이터의 효과적인 관리와 활용에 보다 중점을 둔 것으로 설명할 수 있을 것이다. 오늘날 우리는 데이터의 중요성이 강조되는 데이터 경제 시대에 살고 있으며, 앞으로 (디지털) 데이터 산업 및 시장은 더욱 성장할 것으로 전망된다.

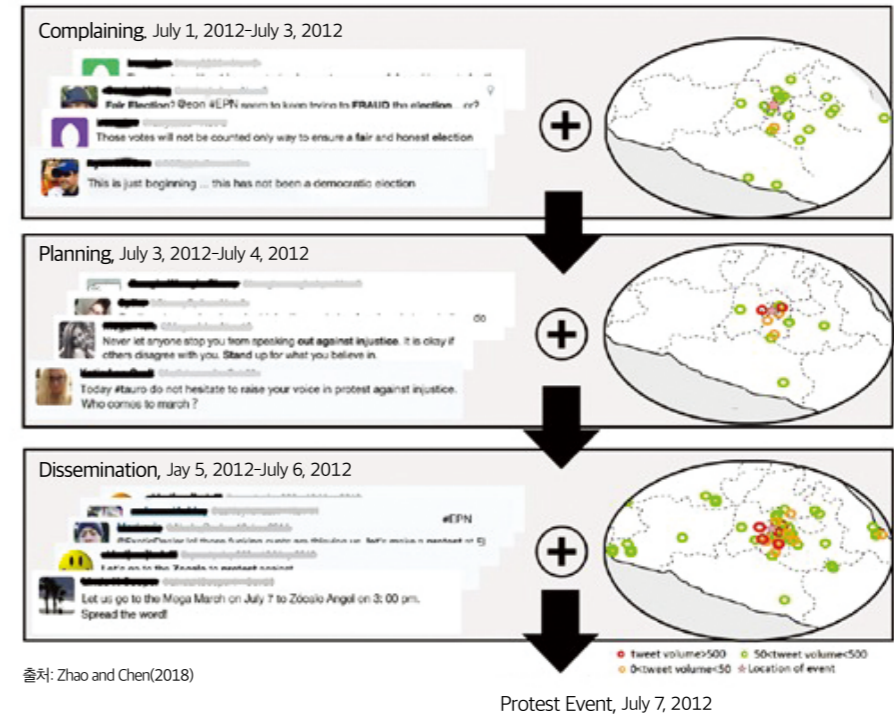
데이터 활용 사례

영국의 경제 전문지인 이코노미스트는 데이터를 ‘세계에서 가장 가치 있는 자원(The world’s most valuable resource)’으로 언급했다(The Economist, 2017). 데이터는 실제로 산업 발전의 촉매 역할을 담당할 뿐만 아니라, 개인적·사회적 측면에서도 새로운 기회와 가치를 창출하고 있다. 데이터의 활용을 통해 기존의 제품이 스마트 제품으로 전환되기도 하고, 데이터 기반의 제품/서비스를 개발해 새로운 혁신을 추구하기도 한다(WEF, 2021). 서울시 심야버스 노선 개발 사례는 버스와는 전혀 관련이 없는 휴대전화 통화이력 데이터에서 추출한 개인위치 정보를 활용해 최적의 심야버스 노선을 구축한 것으로 우리에게 잘 알려져 있다. 또한 특정 지역에서 관광 정책을 수립 및 개선하기 위해 현지인과 구분된 관광객의 신용카드 결제 데이터를 분석한 관광객 소비 지도를 바탕으로 새로운 관광 권역을 설정한 사례도 존재한다. 미세먼지, 침수·지진 피해 등 현실 세계의 각종 사회 현안과 관련해서도 한국과학기술정보연구원 등의 기관들이 여러 분야의 데이터를 연계·융합해 해결 방안을 모색하고 있다. 예를 들어 캐나다의 AI 스타트업인 ‘블루닷’은 항공, 동식물 질병, 뉴스 등의 다양한 데이터를 활용해 WHO보다 먼저 코로나19 사태를 예측하기도 했다.

최근에는 SNS 데이터가 특정 사안에 대한 대중의 생각이나 행동을 실시간으로 파악할 수 있다는 점에서 많이 활용되고 있다. 기존의 많은 연구들은 대중이 무인 자동차 도입에 대해 반대하는 이유로 안전에 대한 불안, 직접 운전하지 않는다는 데에서 오는 거부감 등을 주로 제시하고 있었다. 반면, Buch et al.(2018)은 트위터 등의 데이터를 분석함으로써

과거에 드러나지 않았던 무인 자동차에 대한 대중의 부정적인 심리(자신 혹은 타인의 고용 불안정)를 밝히기도 했다(문희진 외, 2019). 이 외에도 SNS 데이터는 폭동이나 테러, 교통 혼잡, 선거 예측 등과 같은 사회적 사건을 신속히 예측하는데 많이 활용되기도 한다(Zhao and Chen, 2018).

<그림 1> SNS 데이터를 활용한 폭동 예측



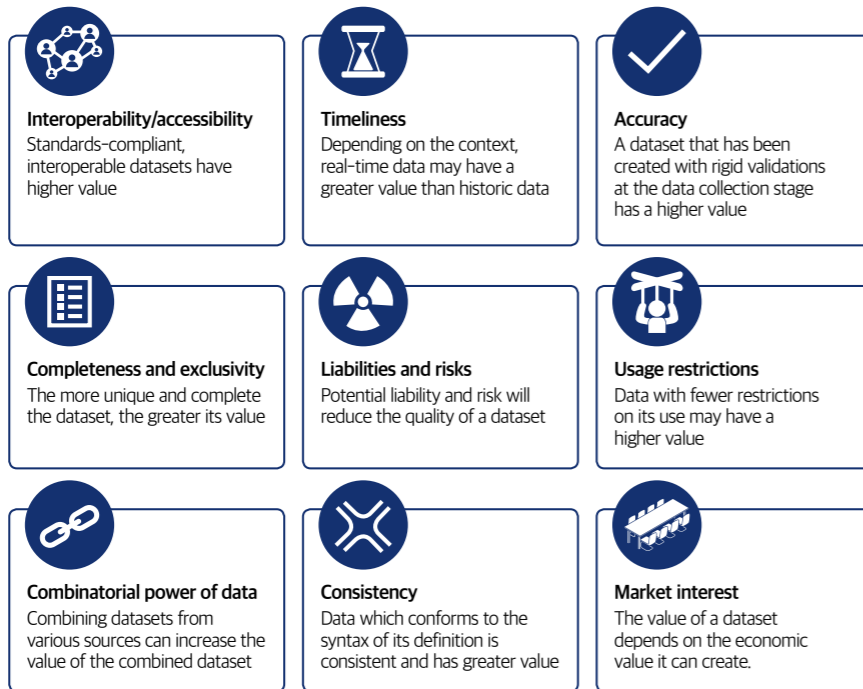
이처럼 데이터는 다양한 분야에서 분석 및 활용되고 있으며, 구글, 아마존, 메타(페이스북), 애플 등 글로벌 대형 플랫폼 기업들은 축적된 데이터를 바탕으로 금융 산업 등으로 범위를 확장하고 있다. 우리나라에서 2021년 유니콘 기업으로 선정된 직방이나 당근마켓 등도 공급자와 수요자 간에 필요한 데이터를 공유해 가치를 창출하고 있다. ‘한계비용 제로 사회’의 저자인 제러미 리프킨(2014)은 데이터(정보 상품)의 경우 유통의 한계비용이 제로에 가까워, ‘협력적 공유사회(Collaborative Commons)’로 패러다임이 전환될 것이라고 설명했다. 따라서 앞으로 데이터가 자유롭게 공유·연계·활용된다면 지금보다 더욱 큰 가치를 창출할 것이며, 다학제적(Multidisciplinary) 문제를 해결하는 데에도 기여할 수 있을 것이다.

데이터 공유·활용의 한계

디지털전환의 가속화로 인해 디지털 기술 또는 데이터 기술(데이터 기반의 가치 창출을 위해 데이터 수집부터 데이터 분석·활용에 이르는 과정에서 밀접하게 적용되는 기술)이 빠르게 발전하고 있음에도 불구하고, 데이터 공유는 여전히 잘 이뤄지지 않는 실정이다. 이는 사람들이 관심을 가지는 데이터가 개방됐는지, 개방된 데이터를 쉽게 활용할 수 있는지, 데이터의 품질이 보장되는지 등과 같은 다양한 이유가 존재한다. 대규모의 데이터가 공개·개방됐다고 하더라도 필요하지 않거나 활용하기 어렵고 정확하지 않은 데이터라면, 단지 데이터의 볼륨(Volume)은 데이터 공유를 촉진하는데 있어 중요한 요인이 되지 않을 것이다.

WEF(2021)는 데이터의 가치를 결정하는 요인으로 상호 운용성/접근성, 적시성, 정확성, 안전성/독점성, 책임과 위험, 사용 제한, 데이터 결합력, 일관성, 시장 관심 등을 제시했다. 상호 운용 가능한 최신(실시간) 데이터, 정확하고 완전한 데이터, 규정을 준수하고 시장의

<그림 2> 데이터의 가치를 결정하는 요인



출처: WEF(2021)

수요가 높은 데이터 등은 그 가치가 상승한다. 다른 무엇보다 시장의 수요, 즉 사람들이 필요로 하는 고품질의 데이터가 많이 개방돼야 데이터 공유를 촉진할 수 있을 것이며, OECD(2021)에서도 데이터가 포함하고 있는 내용(Content)의 중요성을 강조했다. 반면, 잠재적으로 책임과 위험이 따르거나 사용 제한이 많은 데이터는 가치가 낮아지게 된다. 데이터의 FAIR(Findable, Accessible, Interoperable, Reusable) 원칙과 함께 이러한 요인들을 고려해, 개방된 데이터가 어떤 특성을 가지고 있는지에 대해 명확하게 파악하는 것도 데이터 공유와 활용을 촉진하기 위한 과정으로써 선행돼야 할 것이다.

그리고 해외뿐만 아니라 우리나라에서도 공적 자금으로 수행된 연구의 데이터는 가능한 적은 제한으로 공개 및 공유할 것을 추구하고 있지만, 아직까지 과학기술 데이터는 자신의 연구 활동을 통해 생산됐기 때문에 연구자들이 개인의 소유로 인식하는 경우가 많다(최명석 외, 2019). 이와 같은 상황에서 데이터(연구데이터)를 의무적으로 제출할 것을 강요하기보다 연구자들에 대한 보상과 인정이 필요하다. 다시 말해, 연구자들이 과학적 발견과 새로운 지식 창출을 위해 데이터를 검토하고 가공·변환·결합하는 등 연구 과정에 대해 인센티브를 제공하는 방안에 대해 고려할 수도 있을 것이다. 공적 자금으로 지원받은 연구를 통해 게재한 논문이나 출원한 특허는 연구자의 실적으로 인정받고 있으며, 기술이전의 경우에도 「기술의 이전 및 사업화 촉진에 관한 법률 시행령」에서 개발한 기술을 이전하거나 사업화해 얻은 기술료의 100분의 50 이상을 연구자에게 보상금으로 배분할 것을 명시하고 있다. 반면, 데이터 공유에 대한 보상 체계는 아직까지 구체적으로 존재하지 않는다. 이는 (디지털) 데이터가 거의 무료로 유통될 수 있다는 측면에서 기존에 연구자의 노력을 간과한 채 데이터 공유를 너무 쉽게 인식하는 경향이 있었다고 볼 수 있다. 향후 데이터를 가치 있는 '정보 자원'으로 바라보는 인식의 전환이 요구되며(정준화, 2020), 이를 통해 고품질의 데이터가 지속적으로 축적된다면 그 가치는 더욱 커질 것이다.

이 외에도 데이터 공유의 한계를 극복하고 데이터 공유와 활용을 장려하기 위해서는 지금보다 훨씬 많은 노력이 요구될 것이다.

향후 과제

2020년 8월 5일에 '데이터 3법(「개인정보 보호법」, 「정보통신망법」, 「신용정보법」)' 개정안이 시행됐고, 2022년 4월 20일에 데이터 경제 촉진을 위해 「데이터 산업진흥 및 이용촉진에 관한 기본법(데이터산업법)」이 시행됐다. 그리고 2022년 7월 5일에 산업데이터의 생성·활용 활성화를 위한 「산업 디지털전환 촉진법(산업디지털전환법)」이 시행됐으며, 윤석열 정부는

모든 데이터가 연결되는 디지털 플랫폼을 기반으로 사회문제를 해결하고 새로운 가치를 창출하는 '디지털플랫폼정부'를 강조하고 있다. 이와 같이 데이터 경제 시대에 대응해 다양한 법안과 정책 등이 마련되고 있다.

하지만 데이터와 관련된 법안들이 기존의 법률들과의 충돌, 데이터 거래·활용으로 인한 분쟁 등 일부 혼란을 야기할 수 있다. 예를 들어 개인정보의 경우 「개인정보 보호법」이 우선적으로 적용돼 데이터 활용을 저해하거나 글로벌 빅테크 기업들의 데이터 독점을 용인할 수도 있을 것이다. 또한 「데이터산업법」에서 규정한 범정부 데이터 정책 컨트롤타워 역할을 수행할 국가데이터정책위원회에 대한 기대와 함께, 과거 실행력에 한계를 드러낸 4차산업혁명위원회를 답습할 수 있다는 우려도 제기된다. 물론 2023년 1월에 제2차 국가데이터정책위원회를 통해 '제1차 데이터산업 진흥 기본계획'과 '인공지능 일상화 및 산업 고도화 계획'을 중점적으로 논의해 정책 방향을 제시했지만, 국가 데이터 정책을 추진할 때 디지털플랫폼정부위원회, 공공데이터전략위원회 등과 원활한 협력 체계를 구축해야 한다. 디지털플랫폼정부와 관련해서도 단지 데이터의 연결 실적에 치중하기보다 데이터의 실질적인 활용성을 강조해야 하며, 공급자 중심으로 데이터를 제공하는 포털(Portal)이 아니라 플랫폼(Platform)에 대한 이해를 바탕으로 데이터 공유·활용이 이뤄져야 할 것이다.

이에 따라, 「데이터산업법」 등의 제정·시행 목적을 궁극적으로 달성할 수 있도록 부처·기관 간의 유기적인 연계뿐만 아니라, 다양한 관점에서 관련 법률들을 충분히 검토하고 조율하는 과정이 필요하다. 이를 효과적으로 추진하기 위해서는 무엇보다 국가데이터정책위원회의 역할과 책임이 매우 중요할 것이다. 그리고 데이터 경제 활성화를 위해 연구자를 포함해 공공과 민간, 더 나아가 모든 국민들이 데이터 공유·활용에 대한 인식을 제고함으로써 '공유'와 '협력'의 문화를 만들어 나가야 할 것이다. 우리나라는 해외보다 상대적으로 집단지성, 클라우드 소싱 등이 원활하지 않은 편인데, 이는 결국 우리나라의 문화와도 관련이 높다. 문화는 노력 없이 결코 조성되지 않으며, 국가와 기관 차원에서 다양한 캠페인과 성공 사례, 직·간접적인 경험을 꾸준히 제공해 국민들이 '공유'와 '협력'에 대한 필요성을 스스로 인식할 수 있어야 한다. 앞에서 언급한 데이터 공유와 관련해 연구자에게 금전적 또는 비금전적 인센티브 제공도 연구자들이 직접 자신의 데이터를 공유해보고 이에 대한 보상과 인정을 받을 기회를 마련해주는 일환으로 볼 수 있다. 물론 데이터를 안전하게 공유하고 잘 활용하기 위해 보안에 대한 기술적 지원, 데이터 품질 및 연계를 위한 표준, 데이터의 경제적 가치 측정, 개인정보 보호와 활용 간의 균형 등 다양한 이슈에 대해서도 지속적으로 고려해야 할 것이다.

마지막으로 국내외 보고서를 살펴봤을 때 산업 전반에 걸쳐 데이터 관련 직무에 대한 수요가

증가함에도 불구하고, 우리나라는 필요 인력이 부족한 것을 확인할 수 있다. 즉 데이터를 직접 분석하고 활용할 수 있는 데이터 과학자(Data Scientist)를 양성해야 한다. 데이터 경제 시대에 데이터가 핵심 자원이 되겠지만, 데이터 분석 역량을 보유함으로써 효율성 제고, 의사결정 지원 등 부가가치를 극대화할 수 있을 것이다. 특히 과학기술 분야처럼 전문성이 큰 분야일수록 관련 분야에 대한 지식의 격차가 존재하기 때문에, 데이터 분석을 별도로 맡긴다면 실패할 확률이 증가한다. 조직의 영업비밀이나 민감한 데이터의 경우에도 담당자가 스스로 데이터를 분석할 필요가 있다. 데이터 분석과 활용의 핵심은 데이터 기술을 다루는 역량과 더불어, 주어진 상황을 가장 잘 이해하고 문제나 해결 방향 등을 판단할 수 있는 역량과도 큰 관련이 있을 것이다. 참고로 최근 챗GPT 활용 시에도 '프롬프트(명령어)'를 얼마나 잘 입력하느냐에 따라 결과물이 완전히 달라질 수 있기 때문에, 프롬프트 엔지니어(Prompt Engineer)의 중요성이 강조되고 있다(권기대, 2023).

앞으로 데이터 공유·활용이 원활하게 이루어진다면 데이터 경제가 더욱 활성화될 수 있을 것이며, 전 산업의 디지털전환과 함께 코로나19와 같은 감염병, 기술패권 경쟁, 글로벌 공급망 파편, 탄소중립, 저출산·고령화 등 다양한 이슈를 해결하는데 있어서도 데이터를 중심으로 대응해 나가야 할 것이다.

■ 참고문헌

- 권기대 (2023), 「챗GPT 혁명: 산업과 투자의 지형을 뒤흔드는 인공지능의 진화」, 베가북스
- 문희진·정도범·주원균·정용일 (2019), 「빅데이터 분석, 사회 문제 해결의 길을 열다」, 한국과학기술정보연구원, KISTI ISSUE BRIEF, 제5호
- 삼일회계법인 (2020), 「코로나19가 가져올 구조적 변화: 디지털 경제 가속화」, Samil Issue Report, 2020.04.
- 정준화 (2020), 「데이터 정책 거버넌스의 현황과 과제」, 국회입법조사처, 이슈와 논점, 제1787호
- 제러미 리프킨 (2014), 「한계비용 제로 사회」, 민음사.
- 최명석·김현우·이상환 (2019), 「국가 연구데이터 공유의 시작, 데이터 관리계획(Data Management Plan)」, 한국과학기술정보연구원, KISTI ISSUE BRIEF, 제13호
- Buch, R., Beheshti-Kashi, S., Nielsen, T. A. S. and Kinra, A. (2018), "Big Data Analytics: A Case Study of Public Opinion Towards the Adoption of Driverless Cars", International Conference on Dynamics in Logistics.
- Newman, D. (2011), "How to Plan, Participate and Prosper in the Data Economy", Gartner Research.
- OECD (2021), "Measuring the economic value of data", OECD Going Digital Toolkit, Measurement Note.
- The Economist (2017), "The world's most valuable resource", May 2017.
- WEF(World Economic Forum) (2021), "Towards a Data Economy: An enabling framework", World Economic Forum, White Paper, Aug 2021.
- Zhao, L. and Chen, F. (2018), "Big Data Analytics for Social Event Forecasting", IEEE Conference on Big Data 2018, Dec 2018.