

AI-데이터 시장에서의 데이터 최적 공유와 제약 조건*

신 위 회**

논문 초록

AI 생태계 내 데이터 시장에서 발생하는 경제주체 간 데이터 공유는, 사회적 혜택과 사회적 비용이 동시에 존재하므로, 사회 후생을 극대화하기 위해서는 이 둘 사이의 균형 관계를 명확히 이해하는 것이 중요하다. 본 논문에서는 AI 학습 데이터 최적 공유 문제와 제약 조건에 대한 경제학적 고찰을 하기 위해, 데이터 공유에 대한 경제학적 메커니즘을 구체화하고, 정책적으로 데이터 공유를 촉진할 시 고려해야 할 요소들에 대해 검토한다. 데이터 공유가 사회 후생을 증가시키며 AI-데이터 생태계의 선순환을 일으키기 위해서는 공유되는 데이터의 경제적 특성, 사업자 간 경쟁 관계, 거래 메커니즘, 시장 환경 등에 대한 조건이 존재한다. 이러한 조건이 충족되는 경우, 데이터 공유는 다양한 AI 도구를 통해 기업의 생산성과 혁신의 수를 증가시킬 수 있으며, 이는 경제 성장에도 기여할 수 있다. 그러나 해당 조건이 충족되지 않을 시, AI 생태계 내 데이터 생산이 줄어들거나, 부정적인 외부효과가 발생할 수 있다.

핵심 주제어: 인공지능, 데이터, 데이터 공유

경제학문헌목록 주제분류: L1, L4, L5

투고 일자: 2025. 3. 31. 심사 및 수정 일자: 2025. 6. 16. 게재 확정 일자: 2025. 6. 23.

* 본 논문은 한국경제학회 2025 경제학공동학술대회 'AI(인공지능) 관련 경제환경 변화에 대한 대응' 특별세션에서 발표되었습니다. 세심한 심사를 해주신 익명의 두 심사자분들과 편집위원께 깊은 감사를 드립니다. 또한 본 논문은 저자가 2023-2024년에 참여한 전남대학교 산학협력단의 네이버 주식회사 연구과제 '플랫폼 사업자의 데이터 개방 의무화의 타당성에 관한 연구' 최종보고서를 수정하고 발전시킨 것임을 밝힙니다.

** 전남대학교 경제학부 조교수, e-mail: wiroypsu@gmail.com

I. 서 론

대규모 언어 모델 및 데이터 기반 기계학습을 바탕으로 서비스가 이루어지는 생성형 AI 서비스가 세계적으로 주목을 받음에 따라, AI에 의한 의사 결정과 그를 통해 이루어지는 서비스의 필수적인 생산 요소로써 인공지능 학습 데이터(이하 ‘데이터’)의 중요성이 대두되고 있다. 이러한 흐름은 데이터를 수집하고 보유하는 플랫폼이 속한 시장의 네트워크 효과 및 쏠림현상과 같은 플랫폼 시장의 구조적인 특징과 이어져, 데이터 독과점 문제에 대한 경쟁정책적 우려가 제기되기도 한다. AI-데이터 생태계의 이러한 특성들로 인해, AI 산업의 발전이 상대적으로 빠르지 않은 국가에서는, 특정 사업자가 보유한 데이터를 타 사업자가 AI 학습을 위해 사용할 수 있도록 공개하여 자국의 AI 산업 혁신을 촉진시키려는 정책적 노력을 진행하고 있다. 즉, 기업 간 데이터 공유를 통해 AI를 개발하거나 활용하는 기업의 성장을 촉진하고, AI 시장의 구조를 개선하고자 하는 것이다.¹⁾

그러나 이러한 정책적 논의는 데이터 생산자 및 데이터 개방·공유 메커니즘에 참여하는 다양한 경제주체들의 유인을 충분히 고려하지 않을 경우, 해당 AI-데이터 생태계에 부정적 외부효과를 발생시킬 가능성이 존재한다. 그러므로 데이터 개방·공유와 관련한 데이터 시장에서의 경쟁과 혁신, 사회적 후생 관점에서의 정당성 논의 등이 경제학적 관점에서 이루어질 필요가 있다. 따라서 본 논문에서는 AI-데이터 시장에서의 학습 데이터 최적 공유 문제와 제약 조건에 대한 경제학적 고찰을 하고자 한다. 이를 위해, 데이터 공유에 대한 경제학적 메커니즘을 구체화하고, 데이터 공유를 정책적으로 추진할 시 고려해야 할 요소들에 대해 검토한다.

이러한 논의를 구성하기 위해 본 논문에서는 경제 성장과 같은 데이터 공유로 인해 이끌어낼 수 있는 긍정적인 효과를 먼저 설명한 후, 데이터 공유에서 발생할 수 있는 부정적인 외부 효과를 살펴 본다. 외부 효과에 대한 분석을 보다 구체적으로 하기 위해서 본 논문은 데이터 시장을 데이터 생산 시장과 활용 시장으로 구분하

1) EU의 경우, Digital Market Act(DMA)를 통해 게이트키퍼로 지정된 플랫폼의 데이터 공유를 의무화하고 있으며, Data Act(DA)를 통해 B2B 데이터 공유를 추진하고 있다. 한국의 경우 데이터 경제의 지속가능한 발전과 생태계의 선순환을 위한 전기통신사업법 일부개정법률안이 2021년 발의된 바 있으며, 최근 산업부는 산업데이터 은행(Data Bank) 설립 추진을 밝히기도 했다.

며, 해당 시장에 참여하는 경제주체의 효용의 합을 사회적 후생으로 정의한다. 따라서 활용을 위한 데이터 거래에서의 데이터 수요 기업과 공급 기업의 수익 뿐만 아니라, 플랫폼과 같은 데이터 생산 시장에서 사용자에게 지급되는 보상, 프라이버시 침해와 같은 외부효과 등도 데이터 시장을 판단하는 효율성 구성 요소로 포함될 수 있다.

본 논문의 구체적인 구성은 다음과 같다. 제Ⅱ장에서는 본 논문에서 이루어지는 논의의 기본적 틀을 정립하기 위해, 데이터 거래 시장에 대해 구체적으로 살펴 본다. 이를 위해 재화로서 데이터가 어떻게 공급되는지에 대한 과정을 설명하고, 데이터의 생산과 활용을 구분하는 단계적 시장을 정의하며, 해당 시장에서의 사회 후생 구성 요소를 분석한다. 이를 통해, AI-데이터 생태계 내의 데이터 최적 공유를 위해 어떠한 사항들이 고려되어야 하는지에 대해 추론한다. 제Ⅲ장에서는 기존의 경제학 연구를 바탕으로 데이터가 최적으로 공유될 때 경제성장과 어떻게 연결되는지에 대한 논리에 대해 알아본다. 제Ⅳ장에서는 데이터 공유에 어떠한 외부효과가 존재할 수 있는지에 대해 분석함으로써 데이터 최적 공유를 위한 조건들을 구체적으로 검토한다. 이를 위해 AI 학습 데이터가 가지는 경제적 특성, 시장 경쟁, 프라이버시, 거래 비용, 정책적 효과 등을 고려한다. 마지막으로 제Ⅴ장에서는 데이터 공유에 대한 논의를 종합하여 정리하고, 데이터 공유 정책에 대한 시사점과 본 논문이 지니는 한계점에 대해 논의한다.

Ⅱ. AI 학습 데이터 거래 시장

데이터는 생성과 보유의 주체에 따라 공공 부문과 민간 부문으로 구분될 수 있다. 본 논문은 민간 부문의 사업자 간 데이터 공유 문제를 중점적으로 논의하고자 한다. 이를 위해 본 연구에서는 데이터 거래 메커니즘에서 데이터 보유자의 자발적인 선택권(가격, 거래 대상 등)이 주어진 경우를 공유로 정의하며, 정책이나 시스템 등에 의해 강제성이 부여된 거래를 개방으로 정의한다. 또한 많은 수의 경제주체가 데이터를 생성하며 특정한 활동을 하는 디지털 매개체를 플랫폼으로 정의한다. 이는 일반적인 인터넷 플랫폼 뿐만 아니라, 오프라인에서 사용자 데이터를 수집하고 디지털 서비스를 제공하는 제조사와 연결된 플랫폼을 포함하는 개념이다.²⁾

1. 데이터의 종류 및 거래의 가치 사슬

1) AI와 데이터의 상호 의존성과 데이터의 종류

AI는 데이터를 활용하여 작업을 수행하고 학습하며 진화한다. AI를 목적에 맞게 학습시키고, 학습한 AI를 다시 미세조정 (fine-tuning) 하기 위해서는 데이터가 필수적이며, 이때 사용되는 데이터의 양과 품질은 AI 시스템의 신뢰성과 품질을 결정짓는다(이용 외, 2022). 데이터가 다양하고 포괄적일수록 AI는 더 나은 성능을 발휘할 수 있으며, 이로 인해 데이터가 AI의 ‘학습 연료’로 일컬어지기도 한다(Sehgal, 2023). 데이터의 경우 AI가 없어도 기존의 통계 및 계량경제 모형을 통해 특정 문제에 대한 분석(예측, 추천, 인과관계 도출)을 할 수 있지만, AI를 통해 빅데이터²⁾와 비정형 데이터와 같이 보다 다양한 데이터를 쉽게 관리하거나 분석할 수 있다는 점에서 AI와 데이터는 공생관계에 있다고 볼 수 있다.

AI 학습에 사용되는 데이터는 형태에 따라 정형 데이터와 비정형 데이터로 구분할 수 있다. 정형 데이터의 경우 미리 정의된 형식을 사용하여 이미 구조화되어 있으며, 정량적 성격을 지닌다. 이러한 높은 수준의 체계성은 기계학습 알고리즘 및 기존의 연산, 계량 모델의 적용을 용이하게 한다. 정형 데이터의 예로는 온라인 구매 데이터(시간 기록, 구매 금액, 연계된 계정 정보, 구매 품목, 지불 정보 및 확인 번호), 다양한 시스템에서 생성되는 수치화 가능한 기록측정 데이터, 거래 데이터, 생산 데이터 등을 들 수 있다.

비정형 데이터는 일반적으로 정성적 데이터로 분류되며, AI의 도움 없이 기존의 정해진 데이터 도구와 방법으로 처리와 분석을 하기 힘들다. 텍스트, 오디오, 비주얼 데이터(예: 텍스트 메시지, 소셜 미디어 포스트, 기계가 생성하는 비디오·이미지 센서 데이터 등)가 이에 속하며, 일반적으로 원형 그대로 저장되기 때문에 많은 저장 공간을 필요로 한다.

형태가 아닌 데이터가 지니고 있는 주요 정보나 출처로도 데이터의 분류가 가능하다. 대표적으로는 소비자(사용자) 데이터가 있다. 소비자 데이터는 ‘식별되거나 식별 가능한 소비자와 관련된 모든 정보’로 정의되며, 그 예로는 사용자가 인터넷에

2) 예: 현대자동차의 현대 디벨로퍼스.

3) 거대한 규모, 빠른 속도, 높은 다양성을 특징으로 하는 데이터(Chen, 2024).

서 생성하는 콘텐츠, 활동 데이터, 소셜 데이터, 위치 데이터, 인구통계적 데이터, 개인 식별 정보 등이 있다(Barker, 2021). 이러한 사용자 데이터는 AI 비즈니스 모델에 직접 적용이 가능하여 상업적 활용도가 높다.⁴⁾ 따라서 다양한 디지털 서비스를 통해 수집되는 개별 소비자 데이터의 경우 상당한 경제적 가치가 있으며, 이는 기업에게 이를 수집할 큰 인센티브를 부여한다.

2) 데이터 거래의 가치 사슬

AI 학습 목적에 맞게 데이터 아키텍처(Data Architecture)에 의해 체계적으로 관리된 데이터는 경제적 가치를 지니며, 데이터 판매자-소비자 간 거래가 이루어진다. 데이터가 수요자에 의해 생성형 AI의 학습 또는 기타 다양한 심층 분석 용도에 적합한 형태로 갖추어지기까지는 많은 과정(생성, 수집, 처리, 저장, 보안, 분석, 유통 등)을 거치며, 이러한 과정에 상당한 비용과 기술이 요구된다(〈Figure 1〉참고).

과학기술정보통신부에 의해 이루어지는 데이터 산업 현황 조사에 의한 국가 통계 역시 이러한 데이터 가치 사슬을 반영하여, 데이터 산업을 ‘데이터 처리 및 관리 솔루션 개발·공급업, 데이터 구축 및 컨설팅 서비스업, 데이터 판매 및 제공 서비스업’으로 구분하고 있다. 2021년 기준 국내 데이터 산업 시장 규모는 22조 8986억원이며(22년 추정치는 25조원 이상), 판매 및 제공 서비스업이 약 50%의 비중을 차지하고, 다음으로 구축 및 컨설팅 서비스업(약 37%), 처리 및 관리 솔루션 개발·공급업(13%)의 순이다(한국데이터산업진흥원, 2023). 가장 높은 비중을 차지하는 판매 및 제공 서비스업에서는(판매·중개가 아닌) 정보 제공 서비스업에서의 매출이 대부분(82.3%)을 차지한다.

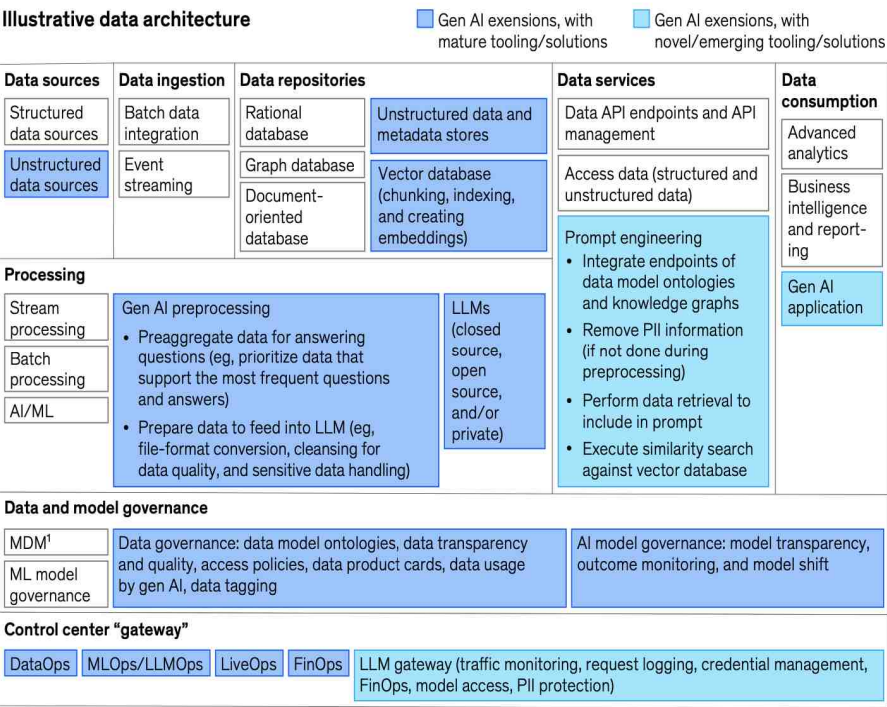
데이터 보유자로부터 데이터의 수요자에게 이전이 발생하는 데이터 거래는 수요-

4) 강화학습을 통한 온라인 광고 알고리즘, 매출 및 이익 증가 요인 식별 등에 활용이 가능하다. 구체적으로는, 방문한 적이 있거나 구매한적이 있는 웹사이트 정보(혹은 자주 구매된 상품 목록 또는 비슷한 유형의 소비자들이 구매한 물품 목록)에 의해 연관성이 있는 제품, 서비스 등을 광고의 형태로 더 자주 보여주는 알고리즘을 강화학습으로 구현할 수 있으며, 그 광고가 클릭되는 경우 알고리즘이 옳게 설정된 것으로 학습하고 더욱 유사한 방식으로 광고를 내보낼 수 있게 된다. 또한 소비자의 구매를 증가시키는 요소들을 소비자의 타입, 페이지에 머무는 시간 등을 통해 판별하여, 이를 이커머스 디자인이나 마케팅 전략, A/B 테스트에 활용하기도 한다.

공급 주체 간 원하는 형태로 직거래를 하는 경우와 거래 플랫폼을 통한 거래가 있다.⁵⁾ 거래 계약의 형태로는 권리 주체가 변경되는 양도 계약과 일정 범위 내에서 이용이 허락되는 라이선스 계약으로 구분되며, 가격 책정은 원가 기준, 경쟁사 가격 기준, 가치 기준 등을 통해 이루어 진다.

〈Figure 1〉 Data architecture

Upgrades are needed within the existing data architecture to enable generative AI.



¹Master data management.

Source: Caserta et al. (2023).

2. 데이터의 단계적 시장

거래를 통한 데이터 공유 문제에 대해 경제학적 접근을 하기 위해서는 해당 시장

5) 가격이 존재하는 데이터 상품 외에도 제공에 따른 금전적인 급부가 존재하지 않는 데이터 상품이 존재할 수도 있다.

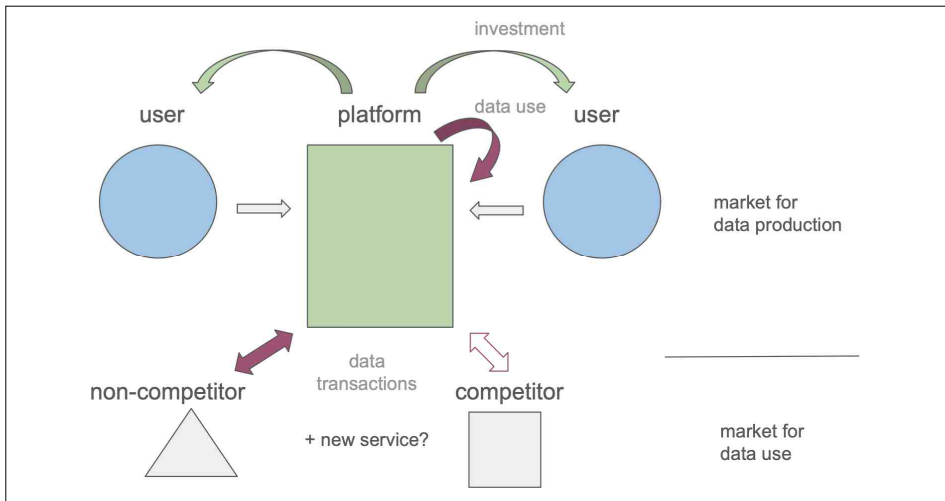
의 주요 경제주체와 그들의 선택, 경제주체의 효용을 구성하는 요소에 대한 정의가 필요하다.

앞 소절에서 논의한 데이터 가치 사슬에서 보았듯이, 데이터는 소비되기 전까지 다양한 단계를 거친다. 이를 참고하였을 때 데이터 시장은 크게 2단계로 구분될 수 있다. 1차 시장은 데이터 생산 시장으로서 데이터가 거래되기 이전, 사용자에 의해 데이터가 생성되고 유의미한 형태로 처리되는 과정의 시장을 뜻하며, 2차 시장은 실질적으로 데이터가 거래되어 AI 서비스를 위해 활용되는 시장을 의미한다. 1차 시장의 주요 경제주체로는 데이터를 수집하는 플랫폼 사업자와 플랫폼 사용자가 있다. 해당 시장에서는 사용자가 양적·질적으로 우수한 데이터를 생성할 수 있도록 서비스와 시스템 측면에서 플랫폼 사업자가 충분한 인센티브를 제공하는 것이 중요하다. 1차 시장에서 생산된 플랫폼 데이터는 2차 시장에서 거래되며, 이 때 플랫폼 사업자는 공급자가 되며, 플랫폼 데이터를 이용하고자 하는 AI 기업 및 기관은 수요자가 된다. 2차 시장의 수요자는 플랫폼 사업자와 경쟁 관계에 있는 기업일 수도 있으며, 협력하여 시너지를 낼 수 있는 보완적 관계의 사업자, 혹은 공공기관일 수도 있다. 더불어, 1차 시장에서의 플랫폼 사업자가 2차에서 자사 플랫폼 데이터를 스스로 소비하여 (데이터가 생성된 서비스 외 기타 영역에서) AI 비즈니스를 하는 것 또한 가능하다(〈Figure 2〉 참고).⁶⁾ 2차 시장의 효율성을 위해서는 공급자와 수요자가 서로의 이익에 맞게 원활히 매칭될 수 있는 것이 중요하며, 이러한 매칭을 위한 중개자가 존재할 수도 있다.

2차 시장은 거래 시장인 만큼, 시장 참여자의 전략적 선택이 시장 결과에 큰 영향을 미친다. 공급자의 경우 판매 혹은 공개하고자 하는 데이터의 범위 및 구성, 가격 책정 방법, 가격 수준, 수요자 선택, 거래 플랫폼 등을 선택하여야 하며, 수요자는 공급자가 설정한 가격에 대한 수락 여부(협상이 가능한 경우 가격 수준) 등을 결정하게 된다. 거래 플랫폼(중개자)의 경우 공급자 그룹의 구성, 중개 수수료 등을 선택한다.

6) Martens et al. (2020)은 이러한 2차 시장을 수요자와 데이터의 성격에 따라 세 가지로 구분한다(플랫폼 사업자의 플랫폼 데이터가 주요 투입 요소가 되는 기존 서비스 시장, 플랫폼 사업자의 플랫폼 데이터가 주요 투입 요소가 되는 새로운 서비스 시장, 플랫폼 사업자와 기타 사업자 모두의 데이터가 필요한 새로운 서비스 시장).

〈Figure 2〉 A two-tier data market



3. 데이터 시장에서의 효율성 구성 요소

2절에서 논의한 데이터의 단계적 시장의 정의에 따라, 시장의 효율성을 결정짓는 전체 사회 후생의 수준은 1, 2차 시장에서 활동하는 경제주체들의 효용의 합으로 볼 수 있다. 본 절에서는 각 경제주체 효용의 구성 요소를 검토함으로써 AI-데이터 생태계 내의 데이터 최적 공유를 위해 어떠한 사항들이 고려되어야 하는지에 대해 설명한다.

1차 시장의 사용자 그룹 효용의 경우 사용자가 어떠한 활동을 하며 데이터를 생성하고 있는 플랫폼이 사용자에게 제공하는 서비스의 가치 및 가격, 생성한 데이터 활용으로 인해 발생하는 프라이버시 침해 정도가 효용 수준을 결정한다. 1, 2차 시장 모두에 속해있는 플랫폼 사업자의 경우 각 시장에서의 수익이 주요 기준이며, 이는 1, 2차 시장에서 제공하는 서비스 및 거래로부터의 매출, 데이터 수집·처리 비용, 데이터를 수요자에게 유통·공급하는데 발생하는 비용에 의해 결정된다. 2차 시장의 수요자 그룹의 경우 플랫폼 데이터 활용을 통해 발생하는 AI 사업에서의 매출 혹은 공적 가치, 플랫폼 데이터에 대한 지불 가격이 수익을 구성하는 주요 요소이다.

1차 시장의 결과와 2차 시장의 결과는 서로 연결되어 있다. 1차 시장에서 생산된 플랫폼 데이터의 수준은 2차 시장의 가격 및 거래량에 영향을 미치며, 2차 시장의

결과는 1차 시장의 데이터 생성 및 처리 과정에서의 투입 수준에 영향을 미친다. 데이터 시장에 대한 기존 연구에서는 이러한 1, 2차 시장의 상호 연결성을 고려하여 경제주체들의 다양한 선택, 경제 성장, 프라이버시 문제 등에 대해 논의한다.

Ⅲ. 데이터 공유와 경제 성장

AI 생태계 내 데이터 시장에서 자율적 균형으로 나타나는 경제주체 간의 데이터 공유가 사회적 최적 수준에 미치지 못할 경우, 이는 시장 실패로 정의될 수 있다. 이러한 시장 실패 환경에서 적극적인 정책적 개입 혹은 일정한 조건 하에 완전한 데이터 개방의 제도화에 찬성하는 의견들은 1, 2차 시장에서 아직 실현되지 않은 잠재적 사회 가치에 주목하며, 정책적 개입이 없을 경우 추가적인 데이터 개방으로 인해 발생할 수 있는 사적 혹은 공공의 경제적 기회를 놓친다고 본다. 이러한 논리들은 데이터가 가지는 일반적인 특성인 범위의 경제와 규모의 경제, 그리고 비경합성에 주목한다. 여러 종류의 데이터가 개방되어 합쳐짐으로써 하나의 큰 데이터셋이 구성되어 데이터의 다양성과 양이 증가하면, 범위의 경제를 통해 데이터가 보다 다양한 곳에서 의미있어지며, 이를 통해 새로운 경제적 기회가 생겨날 수 있기 때문이다. 또한, 규모의 경제를 통해 각각의 데이터가 기존에 활용되던 AI에서 정확성이 증가하게 된다. 이러한 긍정적인 흐름이 AI-데이터 생태계에서 가능한 이유는 데이터가 비경합적이므로, 데이터의 손실 없이 다수의 사용자가 동일한 데이터셋을 사용할 수 있기 때문이다(OECD, 2016).

Jones and Tonetti (2020)는 거시 경제 모형을 통해 비경합적 데이터 공유가 사회 후생 증가와 경제 성장을 이끌 수 있는 경로를 설명하고 있다. 이를 AI 생태계에 적용해보면 다음과 같은 메커니즘이 가능하다. 1, 2차 시장에서 발생하는 C2B (Consumer to Business) 데이터 공유, B2B (Business to Business) 간의 데이터 공유로 인해, 기업은 AI 학습을 위해 보다 다양하고 많은 데이터를 활용할 수 있게 된다. 이렇게 증가한 데이터 풀을 통해 향상된 AI 성능으로 기업은 생산의 효율성을 높이고, 경제 내 상품과 서비스의 다양성을 높일 수 있다. 즉, 기업의 대규모 데이터 풀에 대한 접근은 AI를 통해 기업의 생산성과 혁신의 수를 증가시킬 수 있으며, 이는 경제 성장과 소비자 후생 증가에도 기여할 수 있다.

물론 이 과정에서 개인의 프라이버시 및 대체재 출현에 따른 기업의 경쟁 수준

증가에 대한 염려로 인해, 1차 시장에서 공유되는 데이터의 양은 줄어들 수 있다. 만약 그러한 압력이 강하다면, 해당 경제주체에 의해 공유되는 데이터의 양은 줄어들 것이다. 이러한 데이터 축적(data hoarding) 현상은 개인과 기업의 사적 후생을 증가시키지만, 동시에 혁신과 경제 성장을 둔화시킬 수도 있다.

Jones and Tonetti (2020)는 사회 후생을 극대화하는 사회 최적 수준의 데이터 공유를 달성할 수 있는 메커니즘에 대한 논의를 포함하고 있다. 이들의 모형에 따르면, 데이터 공유 환경에 대한 일정 가정이 만족하면, 개인에게 데이터 소유권(데이터 거래 권한)을 할당하여, 개인들이 데이터를 직접 거래하는 것이 최적 공유 수준을 달성할 수 있다고 본다. 이는 소비자가 프라이버시에 대한 우려와 모든 이해관계자에게 데이터를 판매하여 얻을 수 있는 경제적 이득을 균형있게 고려(trade-off에 대한 고려)할 수 있는 위치에 있기 때문이다. 즉, 기업이 데이터를 소유할 경우, 소비자의 프라이버시를 존중하지 않는 데이터 남용이 일어날 수 있으며, 이는 개인들로 하여금 다시 1차 시장에서 데이터 공유 수준을 낮추게 하는 유인이 될 수 있다. 동시에 Jones and Tonetti (2020)는 데이터 공유로 인해 각 기업의 경쟁 부담이 증가하거나, 개인의 프라이버시 우려가 너무 높은 상황에서는 기업이 데이터 거래 권한을 갖는 것이 사회 후생 측면에서 나은 메커니즘일 수 있음을 설명한다.

제도를 통한 적극적인 데이터 공유에 찬성하는 논리는 경제 성장과 관련한 시장 실패 이외에도 산업정책과 경쟁정책적 접근으로 부터 찾을 수 있다. 산업정책적인 접근은 자국 산업·기업의 보호와 데이터 공유로 발생할 수 있는 잠재적 경제적 기회를 구체화 하는 접근으로, EU의 Digital Market Act와 Data Act를 예로 들 수 있다. 산업정책적 접근은 제도를 통해 해외 빅테크 기업들이 보유한 데이터를 강제적으로 개방하게 하여, 자국 기업의 데이터 접근 수준을 높이고, 궁극적으로는 디지털 경제 내 해외 기업에 대한 의존성을 줄이고자 한다. 일정한 조건 내 데이터 개방 강제를 통해, 데이터를 활용하고 있지 못한 자국 경제주체들의 데이터 접근과 데이터를 통한 혁신 활동을 향상시키고자 하는 것이다. 이러한 정책들은 데이터 개방에 대한 혜택은 특정 산업 내 주요 대기업의 전·후방에 존재하는 중소 기업군에게 돌아갈 것이라고 발표하며, 데이터 접근을 통한 중소기업의 디지털 전환, 창업 활동의 증가, 새로운 서비스 및 산업의 탄생, 이를 통한 일자리 창출 등을 기대한다. 또한 데이터 개방을 통해 공공 의료, 국가 행정, 재난 방지 등에 활용되는 AI 개발을 장려할 수 있다.

적극적인 데이터 공유를 지지하는 경쟁정책적 접근으로는, 데이터를 필수설비로 보고 시장지배적 사업자가 데이터 접근을 거절하는 것을 경쟁제한적 행위로 보아야 한다는 의견이 존재한다(방정미, 2020).⁷⁾ 또한 데이터 개방을 통해 소수 빅테크 기업에 의한 독과점적 시장구조 개선과 같은 정책적 효과를 거둘 수 있을 것으로 예상된다. 즉, 소수 빅테크 기업에게만 의미있는 데이터가 집중되어 보유되어 있는 상태는 시장의 진입장벽을 형성하고, 데이터 관리의 고비용, 기계학습 능력, 플랫폼의 네트워크 효과에 의해 ‘이용자 피드백 순환구조⁸⁾’와 ‘현금화 피드백 순환구조⁹⁾’를 촉진시킴으로써, 거대 사업자에서 초거대사업자가 될 가능성이 높아질 수 있다는 것이다(최난선휘, 2017).

IV. 데이터 공유의 외부 효과와 최적 공유를 위한 조건

1. 시장 경쟁의 관점

1) 1차 시장에서 수집되는 데이터의 경제적 특성

1차 시장에서 플랫폼에 의해 수집되는 사용자 데이터는 1차 시장에서의 플랫폼 서비스와 2차 시장에서의 AI 서비스의 경쟁력을 결정짓는 주요한 생산 요소이다. 1차 시장에서는 데이터의 품질과 양에 따라 양면 시장으로서의 플랫폼의 본질적 기능인 매칭, 사용자별 맞춤형 서비스의 수준이 올라가게 된다. 또한 II장의 인공지능과 데이터의 상호 의존성 논의에서 살펴 보았듯이, 플랫폼이 보유하고 있는 데이터의 수준(e.g. 다양성, 포괄성, 규모 등)은 2차 시장에서 AI 서비스의 수준을 결정짓게 된다. 따라서 1차시장에서 수집되는 데이터는 1, 2차 시장에서의 플랫폼 혹은

7) 필수설비이론의 법리를 적용하기 위한 요건은 다음과 같다(선지원, 2022; 이효석, 2022). 데이터를 보유하고 있는 기업이 시장지배적 지위에 있는 사업자여야 하며, 경쟁기업이 사용하기를 원하는 데이터에 대한 접근 통제권을 가지고 있어야 한다. 이 때 경쟁기업이 해당 데이터 없이는 경쟁이 불가능해야 하며, 동시에 경쟁기업이 해당 데이터의 사용을 대체할 만한 수단을 보유하고 있지 않아야 한다. 물론 해당 산업 환경에서 데이터에 대한 접근 허용이 기술적으로 가능해야 한다는 전제 하에 이는 성립된다.

8) 대규모 이용자 보유 ⇒ 정보 보유 ⇒ 서비스 품질 향상 ⇒ 고객 유인.

9) 대규모 이용자 보유 ⇒ 정보 보유 ⇒ 맞춤형 광고를 통한 수익 창출 ⇒ 투자 촉진 ⇒ 서비스 품질 향상 ⇒ 고객 유인.

인공지능 서비스 사업자 간의 차별성을 결정짓는 핵심 요소라고 볼 수 있다.

더불어 1차 시장에서 수집되는 데이터는 투자와 관리가 필요한 자산의 성격을 지닌다. II장의 데이터 가치사슬 논의에서 보았듯이, 데이터를 생성, 수집, 처리, 저장, 보안, 분석 하는 단계에서 상당한 투자와 관리 비용이 소요되며, 특히 데이터 규모가 커질 수록 데이터 관리에 필요한 관련 AI 기술 혁신이 필요하다. 또한 플랫폼의 경우 시장에서 경쟁하고 사용자를 유지하여 데이터를 생성하기 위해서는, 사용자에게 제공하는 플랫폼 서비스 시스템의 편의성과 혜택에 상당한 투자 비용이 들어가기도 한다. 플랫폼 경쟁에서는 플랫폼 서비스의 네트워크 효과로 인해 사용자(개인 소비자 및 다양한 제품·서비스 공급 사업자)를 고착시키기 위한 노력이 매우 중요하기 때문이다. 그러므로 서비스 a를 위해 서비스 b를 무료로 제공해야 하기도 하며, 이는 다면 플랫폼으로의 진화를 불러일으킨다. 이는 플랫폼 서비스 b의 사용자-i 데이터 생성을 위해, 서비스 a에서 자금이 공급될 수 있음을 의미하며, 플랫폼 데이터 생성의 기여도가 단순화 되기 어려운 특성을 보여준다.

2) 대체성에 따른 데이터 공유 제한의 중요성

Martens et al. (2020)은 비교적 차별적인 데이터(이하 d1)를 보유하고 있는 플랫폼이 생산하는 서비스(이하 s1)가, 2차 시장에서 타 사업자의 서비스·제품(이하 s2) 혹은 데이터와 어떠한 관계를 지니고 있는지에 따라 데이터 공유와 관련한 유인 및 최적 선택이 달라짐을 설명한다. 이에 따르면, 플랫폼은 데이터를 수요하는 타 사업자가 플랫폼과 경쟁관계에 있는 대체재를 생산할 경우, s2 생산을 위한 데이터 공유 유인이 없다. 만약 s2가 s1의 보완재 성격을 지니고 있다면, 이는 s1의 판매 증가를 가져올 수 있으므로, s2 생산 촉진을 위한 데이터 공유 유인이 발생할 수 있다. 또한 s2가 대체성 혹은 보완성 측면에서 중립적이고, 공공의 성격(예: 도시 교통 관리)을 지닐 경우, 데이터 판매를 통해 플랫폼은 추가 수익을 거둘 수 있으므로, 데이터 공유 유인이 존재한다고 볼 수 있다.

이를 종합하여 정리하면 s1과 s2가 대체성이 높은 경우, 플랫폼 공유로 인한 범위의 경제가 실현되지 않을 수 있으며, 특히 s2로 인해 하부(downstream) 서비스 시장의 경쟁이 심화될 경우, 이는 사회 후생 손실을 수반할 수 있음을 의미한다. 이러한 논의는 수직적 통합 측면으로도 확장이 가능하다. 타 사업자가 s2 생산을

위해, d1을 대체할 수 있는 데이터셋이 있는지의 여부와 d1의 가격 수준, 플랫폼이 d1을 활용하여 s2 시장에 직접 진출(혹은 타 사업자와 합병)하였을 때의 이익과 데이터 판매 수익에 따라 데이터 공유의 유인 및 플랫폼의 s2 시장 진입 선택이 달라질 수 있다.¹⁰⁾

따라서 데이터는 공유 상대에 따라 경제적 가치가 달라질 수 있으며, 과도하게 공유될 경우 데이터의 가치에 부정적인 영향을 미칠 수 있다. 이는 AI-데이터 생태계의 데이터 공유 문제에서 데이터가 클럽재의 성격을 지님을 뜻한다. 그러므로 데이터가 비경합적 성격을 지니고 있더라도, 경제적 가치 극대화를 위해서는 제한된 접근에 의한 어느 정도의 배제성이 필요하다. 즉, 비경합 데이터의 무제한 공유는 전체 시장의 사회 후생 측면에서 최적이지 아니다. 이와 비슷한 논의로, 비경합적 데이터 공유와 경제 성장을 설명한 Jones and Tonetti (2020)에서도, 사업자의 데이터 공유 유인 감소, 광범위한 데이터 공유로 인한 데이터 생산 과소 투자 가능성이 산업 조직적 측면에서 실제 경제에 해당 모형을 적용할 때 중요한 고려 요소임을 밝히고 있다.

실제로 EU의 Data Act의 경우 시장에서의 경쟁 관계를 고려하여 사업자 간 데이터 공유 원칙에 예외를 두고 있다. 해당 제도는 데이터 보유자와 관련 시장에 있거나 새로운 제품을 개발하는 사업자들의 혁신을 촉진하는 것이 목적임을 명시하며, 경쟁관계가 데이터 공유 의무 여부의 주요 기준이 되는 제도를 도입하여, 경쟁관계에 있는 사업자들의 제품 개발을 위해 데이터가 활용되는 것을 제한하고 있다.¹¹⁾

10) 관련한 추가적인 이론 연구로는 Tsoy and Konstantinov (2023)가 있다. Tsoy and Konstantinov (2023)는 기존의 산업조직 과점 모형에 머신러닝 이론을 결합한 게임 모형을 구축하여, 시장 경쟁도가 충분히 낮지 않거나 제품 간 대체성이 높은 경우, 머신러닝을 위한 학습 데이터 공유는 기업들의 최적 선택이 될 수 없으며, 시장 참여자의 수익에 부정적인 영향을 미침을 보인다.

11) 이와 관련한 Data Act 조항의 구체적 예는 다음과 같다. “(32) this Regulations aims to avoid undermining the investment incentives for the type of connected product from which the data are obtained, for instance, by the use of data to develop a competing connected product which is considered to be interchangeable or substitutable by users, in particular on the basis of the connected product’s characteristics, its price and intended use. ... Prohibiting the use of data accessed under this Regulation for developing a competing connected product protects data holders’ innovation efforts.” (Regulation 2023/2854).

시장 경쟁의 관점에서 데이터 공유를 논의한 본 절의 내용을 종합하면, 플랫폼에서 수집되어 AI 학습을 위해 쓰이는 데이터는 클럽재의 성격을 지니므로, 대체성 혹은 경쟁도가 낮은 경제주체 간의 제한적인 데이터 공유만이 사회 후생 증가를 위한 정책으로 장려될 수 있다. 이러한 성질을 간과한 데이터 개방 의무화 정책은 플랫폼 사업자의 1차 데이터 시장에 대한 투자와 혁신, 그리고 잠재적 데이터 보유자와 생산자의 혁신 유인을 낮추어 단기적으로는 생산자 잉여의 감소, 장기적으로는 데이터 경제의 혁신 투자 및 데이터 생산 저하를 일으켜 소비자와 생산자 잉여의 감소를 가져와, 정책의 본래 기능을 상실할 가능성이 존재한다.

2. 프라이버시와 소비자 후생

사용자 데이터는 사용자의 플랫폼 이용에 따라 직·간접적으로 생성되므로 다양한 범위의 데이터를 사용자 계정에 따라 구조화하는 것이 가능하다. 따라서 플랫폼은 이를 통해 개인식별이 가능한 정형 데이터 및 패널 데이터 축적을 용이하게 할 수 있다. 이는 기계학습 및 인과 분석, 상업적 활용에 더욱더 용이한 데이터를 생산할 수 있다는 장점이 있지만, 프라이버시 보호 측면에서 취약성이 존재한다는 단점이 있다. Kearns and Roth (2019)는 가명처리 등의 데이터 프라이버시 보호 처리에도 불구하고, 데이터 결합으로 인해 개인식별이 쉽게 일어날 수 있음을 설명한다.

플랫폼 데이터 활용의 주요 외부 효과 중 하나인 프라이버시 침해는 사용자(소비자) 효용 감소분의 구성 요소이자 소비자 후생의 중요 기준으로서, 경제학의 주요 연구 대상이기도 하다(Stigler, 1980; Posner, 1981; Taylor, 2004; Varian, 2009; Ichihashi, 2020).

Acemoglu et al. (2022)은 개인정보 추정(inference)의 기능을 하는 플랫폼 데이터의 성질을 고려했을 때, 사회 후생 측면에서 데이터 공유는 과도하게 일어날 수 있음을 지적하고 있다. 해당 논문은 데이터 공유에는 프라이버시와 관련한 부정적 외부효과가 존재한다는 가정에서 출발하여, 데이터 공유와 프라이버시 사이의 경제학적 관계에 대해 다음과 같이 설명한다. 개인의 특성에 대한 정보를 담고 있는 플랫폼 데이터의 경우, 사용자 a에 대한 데이터와 기타 다양한 데이터의 결합을 통해, a와 상관관계가 높은 사용자 그룹 B에 대한 여러 추정(estimation)이 가능하다.

그러므로 프라이버시에 대한 가치 평가가 낮은 사용자의 데이터 공유 선택으로 인해, 프라이버시 가치 평가가 높은 사용자 b의 프라이버시가 침해 당할 수 있다. 이때 다른 사용자의 데이터 공유로 인해 이미 개인의 프라이버시가 손상된 경우, b와 같은 사용자들은 자신의 데이터를 보호할 유인이 줄어든다. 또한 데이터의 추정 능력은 어느 수준을 넘어가면 한계 효용이 감소하기 때문에, 데이터 공유가 증가할수록, 데이터의 가치와 가격은 하락하게 된다. 따라서 데이터 시장에서는 사회 후생 측면에서 사용자 데이터의 과도한 공유가 나타날 수 있게되며, 극단적인 환경에서는 데이터 거래를 중지하는 것이 효율적인 상황이 나타날 수도 있다.

플랫폼 사용자가 프라이버시 침해와 관련한 효용 감소를 정확히 인식하지 못하고, 공유 결정을 내리게 된다는 행동경제학적 연구들도 존재한다. 데이터 생성자로서 사용자가 시간적 측면에서 근시안적이며 일관성이 결여된 경우, 단기적인 이익을 위해 장기에 나타나는 비용과 불확실성을 고려하지 않는 비합리적 선택을 할 수 있다는 연구 결과가 있다(Barker, 2021; Choi, Jeon and Kim, 2019). 또한 현상 유지 편향(status quo bias)이나 take it or leave it 서비스 형태를 지닌 플랫폼 사용자 사용자의 낮은 협상력으로 인해, 사용자의 원하지 않는 데이터 공유 선택이 발생할 가능성도 지적받고 있다(Costa-Cabral and Lynskey, 2017). 이러한 경제학적 연구 결과들에 따르면, 데이터 공유 정책은 프라이버시와 관련한 소비자 후생에 대해서는 어느정도 보수적인 접근을 하는 것이 필요하다.

3. 거래 비용의 관점

AI-데이터 생태계에서 데이터 공유의 최적 수준을 논의하기 위해서는 데이터 거래 계약에서 발생하는 거래 비용에 대한 고려 또한 중요하다. 플랫폼과 기타 사업자 간의 데이터 공유가 경쟁관계나 프라이버시 등과 같은 외부효과를 고려하여 효율성을 달성하는 최적 수준으로 자율적인 균형이 도출되더라도, 공유가 이루어지는 데이터 거래 메커니즘의 실행(implementation)이 원활하지 않을 경우, 현실에서의 데이터 거래는 사회적 효율을 달성할 수 없기 때문이다. 즉, 특정 수준 이상의 데이터 공유가 지속적으로 경제적 균형으로서 존재하기 위해서는, 수익과 거래 비용의 리스크가 낮아야 한다. 이를 위해 본질적으로 불확실성이 존재하는 데이터 공유 메커니즘에서 사회적 비용을 증가시키지 않기 위해 거래에서 어떠한 부분의 이

행에 주의가 필요한지에 대해 알아볼 필요가 있다.

데이터 공유가 사회적 효용을 달성하기 위한 거래 메커니즘의 첫 번째 조건으로 는 거래별 수익화와 효율적인 가격결정을 들 수 있다. 1, 2차 데이터 시장에 참여 하는 경제주체들이 데이터 공유를 통해 의도한 만큼의 효용을 가지기 위해선, 〈Figure 2〉에 나타나는 모든 거래에서, 참여자들이 제대로 된 수익을 거두는 것이 중요하다. 1차 시장의 사용자들은 데이터가 생성되는만큼 플랫폼 활용에서 제대로 된 금전적·비금전적 보상을 받을 수 있어야 하며, 2차 시장의 수요자들 경우에도, 지불한 비용만큼 AI 개발에서 플랫폼 데이터의 활용도가 높아야 한다. 1차 시장의 사용자의 경우 해당 보상(예: 플랫폼 서비스의 만족도) 수준이 낮으면, 타 플랫폼으로 이전할 수 있는 옵션이 있으며, 2차 시장 수요자의 경우, 데이터 거래 전 구입하고자 하는 데이터의 양과 질에 대한 정보 획득과 본인의 데이터 활용에 대한 이해가 가능하다. 따라서 이들에게는 데이터 거래에서 발생할 수 있는 불확실성이 상대적으로 낮은 편이다. 그러나 2차 시장에서 데이터를 공급하는 주체인 플랫폼은 수요자의 플랫폼 데이터 가치 평가와 관련한 지불 용의에 대해 불완전 정보(거래 이론에서의 수요자 타입)를 가지고 있으므로, 데이터 거래의 가격 설정에 있어 리스크가 존재한다. 이는 플랫폼의 경우 2차 시장에서의 데이터 거래를 통해, 가격 설정의 리스크 및 1차 시장의 비용과 기타 2차 시장에서의 기회비용을 상쇄할만큼 경제적 이익이 보장되어야 한다는 것을 의미한다. 따라서 AI-데이터 생태계에서 데이터 공유가 정책적으로 장려되기 위해서는 플랫폼의 데이터 공유를 통한 수익화 및 가격 설정에 있어서 특정 수준 이상의 자유도가 필요하다.

두 번째 조건으로는 데이터의 비배제성으로 인한 사후 거래 비용의 고려를 들 수 있다. 첫 번째 조건에서 논의한 2차 데이터 시장에서의 수요자 타입에 대한 불완전 정보 문제는 시장에서의 거래가 반복되고, 데이터 시장이 성숙해지면서, 장기적으로 전통적인 상품, 서비스 시장 수준의 효율성을 달성할 가능성이 있다. 즉, 거래에서 발생하는 사전적인 거래 비용, 탐색 비용, 교섭 비용은 시장 참여자의 학습을 통해 장기적으로는 낮아질 가능성이 존재한다. 그러나 데이터 거래는 사후적 측면에서 데이터 활용을 완벽히 모니터링 할 수 없고, 계약에서 지정한 용도에 정확히 한정적으로 사용되었는지 확인할 수 없는(non-verifiability) 이행 비용 측면의 리스크가 존재한다.

해당 리스크는 사적인 부분과 공적인 부분으로 나누어서 설명될 수 있다. 사적인

부분의 경우 데이터 사용에 있어 통제가 어려운 데이터의 비배제성으로 인해, 2차 활용 시장에서 공유와 관련한 데이터 보유자의 최적 선택이 실제로 최적이지 아닐 가능성이 높아진다. 실제로 이러한 부분은 EU 데이터 개방 정책의 대상인 유럽의 제조·서비스 업체의 실질적인 우려 사항이기도 하다. 데이터 개방을 통해 제3의 업체(애프터 마켓 사업자, 기타 데이터 사업자 등)에게 데이터가 공유될 수 있다는 것은 유럽 외 기타 국가 기업으로의 데이터 유출이 가능하다는 의미이며, 이는 경쟁사에게도 공유될 수 있음을 뜻하기 때문이다.^{12) 13)} 또한 공적인 부분으로는, 공공의 이익에 반하는 데이터 오용의 위험이 존재하므로 데이터 공유로 인한 직접적인 사회적 비용이 존재한다는 점을 들 수 있다. 플랫폼 데이터가 협력 사업자를 통한 유출에 의해 선거 운동에 활용된 페이스북-케임브리지 애널리티카 사례(Rehman, 2019)가 대표적인 예이다.

그러므로 플랫폼 데이터 공유는 사후적인 측면에서 이행 메커니즘과 관련한 사적, 공적 리스크가 존재하며, 사회 후생 증가를 위한 데이터 공유를 위해서는 이 부분에 대한 해결이 선행되어야 한다.

4. 산업정책과 경쟁정책적 관점

정책적 환경에 따라 산업정책과 경쟁정책은 III장에서의 논의와는 달리 사업자 간 데이터 공유를 장려하지 않을 수도 있다. 산업정책적 측면에서 가장 영향을 미치는 요소는 국내·외 사업자 간 경쟁구도이다. 게이트키퍼 및 AI 서비스를 주도하는 미국 빅테크들과 실질적으로 경쟁하고 혹은 도전하는 플랫폼 기업들이 존재하는 국가들의 경우, EU의 DMA, DA와 동일한 정책적 전략을 세우는 것은 산업정책적인 측면에서 바람직하지 않다. 해당 경우에는 데이터 공유의 강제성이 있는 제도에 비해 1차 시장에서 나타날 수 있는 부정적인 외부효과가 내생적으로 고려되는 사업자 간 자율적인 데이터 공유 협력이 더 나은 결과를 낼 수 있다. 따라서 이런 경우 국외 빅테크들의 자국 시장에 대한 지역화(localization)에 대항하여¹⁴⁾ 국내 기업의

12) 독일 기반의 제조업체(예: Siemens)와 데이터 서비스 업체(예: SAP)는 이러한 위험과 관련하여 Data Act에 큰 우려를 나타내며, 관련 서한을 유럽 집행위원회에 제출하였다(Chee, 2023).

13) Digital Europe (2023)의 보도자료, “CEOs call for urgent rethink on Data Act”.

로컬 특수성을 반영한 데이터를 주요한 경쟁력으로서 인식하고, 데이터 공유 측면에서는 플랫폼과 AI 학습 데이터의 기업 간 협력의 거래비용을 줄일 수 있는 측면의 간접적인 산업정책이 필요하다.

경쟁정책적으로는 데이터 공유로 인한 반경쟁적 시장 결과에 대한 우려를 들 수 있다. 기업 간 정보 공유는 암묵적 조정을 통해 경쟁 완화의 기능을 할 수도 있기 때문이다. 기업 간 데이터 공유는 정보 교환의 기능을 하며, 이는 기업 간 상호 작용 속도를 높이고, 시장의 투명성을 강화한다. 또한 경쟁 사업자 간 제품의 차별성을 낮추고, 생산 측면에서 나타나는 여러 특성 사이의 대칭성을 높인다. Ivaldi et al. (2007)에 따르면, 정보 교환을 통해 생기는 이러한 시장의 특성들은 기업 간 암묵적 담합을 촉진시키며 시장에 반경쟁적 영향을 미친다. 이러한 반경쟁적인 효과는 시장 구조 및 공급의 특성에 따라 더욱 강화될 수 있다.

실제로 2023년 3월 미국 법무부 반독점 부서의 도하 메키 수석부차관보는 의료산업에서 허용되어 온 정보교환 안전지대를 철회하며, 정보·데이터 교환에 대한 우려를 표명하였다(U.S. Department of Justice, 2023).¹⁴⁾ Rho and Shin (2025) 또한 데이터 공유를 통해 알고리즘에 의한 암묵적 담합이 더욱 촉진될 수 있는 가능성을 제기하고 있다.

V. 결론 및 정책적 시사점

AI 생태계 내 데이터 시장에서 일어나는 경제주체 간 데이터 공유는, 사회적 혜택과 사회적 비용이 동시에 존재하며, 사회 후생을 극대화하기 위해서는 이 둘 사이의 균형 관계를 명확히 이해하는 것이 중요하다. 이를 위해 공공 분야에서의 데이터 개방과 민간 분야에서의 데이터 공유를 구분할 필요가 있다. 민간 분야에서의 데이터 공유가 사회 후생을 증가시키며 AI-데이터 생태계에서 선순환하기 위해서

14) 예: 한국 사용자들에 대한 정보 수집 및 온라인 결제 특성 분석을 위한 구글의 신한카드 온라인 결제 데이터 구입 케이스.

15) “정보교환에 대한 지나치게 형식주의적인 접근방식은 가격 상승, 임금 억제 또는 혁신 억제로 이어질 수 있는 프레임워크를 허용하거나 심지어 지지할 위험이 있다. 특히 의료산업의 경우 점점 더 데이터 집약적인 산업이 되어가고 있으며, 제품이나 서비스를 개발하거나 제공하기 위해 머신러닝, 인공지능 및 기타 고급 도구의 힘에 의존하고, 이러한 현상은 해당 산업의 거래 및 경쟁에 영향을 미칠 수 있다.”(U.S. Department of Justice, 2023).

는, 공유되는 데이터의 경제적 특성, 사업자 간 경쟁 관계, 거래 메커니즘, 시장 환경 등에 대한 여러 조건이 전제된다. 그러한 조건이 충족되는 경우에는 별도의 제도가 존재하지 않아도, 플랫폼-기타 사업자 간에 AI 학습 데이터 공유 협력이 나타난다. 그러나 데이터 최적 공유를 위한 일정한 조건이 만족되지 않을 경우에는, 데이터 공유로 인해 1차 시장에서의 데이터 생산이 줄어들고 AI 서비스가 개발되는 2차 활용 시장에서 시장 참여자들의 장·단기적 경제적 손실이 발생하며, 혁신 활동의 감소가 발생할 가능성이 높다. 또한 데이터 거래 계약이 지니고 있는 불완전성으로 인해, 데이터 오용으로 인한 사회적 사후 손실이 발생하거나, 산업·경쟁정책적 측면에서 의도하지 않은 부정적 외부 효과가 발생할 수 있다.

따라서 경제주체의 유인체계를 고려하지 않는 정책적 데이터 개방은, 질적·양적 측면에서 1차 시장의 데이터 생산을 감소시키고, 장기적으로 2차 시장의 데이터 거래를 줄임으로써 경제에서 활용가능한 데이터 풀의 증가를 낮추어 사회 후생의 감소를 가져올 수 있다.

이러한 분석을 고려하였을 때, 민간 영역에서 AI 학습 데이터 공유에 대한 정책적 개입은 시장의 거래 비용을 낮추고, 데이터 생산 인프라 구축을 위한 지원에 집중되는 것이 바람직하다. 구체적으로 정부는 데이터 거래를 원하지만 탐색, 교섭, 이행 측면의 거래 비용으로 인해 거래가 성사되지 못하는 케이스에 대한 정책적 지원 체계를 마련하거나, 중소기업 간 범위의 경제를 높이기 위한 데이터 공유 지원 정책, 데이터 가치 사슬의 제고를 위한 인프라 투자 지원 정책을 수립해 볼 수 있다. 또한 사회적 가치가 존재하는 공공 의료, 행정, 재난 방지, 학문적 성격의 R&D와 같은 공공 분야는, 데이터 거래 시장에 참여하는 경제주체에 대한 부정적 외부효과가 최소화되는 지점에서 정부와 플랫폼, 데이터 생성주체 간의 협력이 추진될 수 있다.

한국의 경우 데이터 거래를 위한 데이터 플랫폼들이 기존에 존재하지만(예: 한국 데이터 거래소), 그 활용도가 제한적인 편이다. 데이터 제공 주체는 금융사나 공적 성격을 띤 기관인 경우가 대부분으로 일반 민간 사업자의 참여도가 낮다. 또한 플랫폼에서 데이터 보유자의 정보를 파악한 후, 실제 거래는 수요자-보유자 간 직거래로 하는 경우도 빈번하다(정원준, 2022). 이는 데이터 플랫폼이 시장으로서 제 기능을 하지 못하고, 수요와 공급을 하는 이용자들은 데이터 플랫폼을 통한 거래에 상당한 위험과 비용을 체감하고 있음을 시사한다. 따라서 한국의 경우 본 연구에서

논의한 바와 같이 데이터 시장 참여자들의 유인을 세심히 고려한 데이터 시장 설계에 대한 세부 정책이 필요하다. AI의 발전이 기존의 지도·비지도 학습에서 강화학습 단계로 빠르게 넘어가고 있는 현 시점에 강화학습에 필요한 인간의 선택과 그 결과에 대한 정보를 담고 있는 데이터의 활용이 AI 산업정책 측면에서 매우 중요하다는 점에서, 데이터 시장 설계에 관한 방향과 세부 정책은 신속히 제고되어야 한다.

본 논문은 AI-데이터 시장에 존재하는 다양한 경제주체의 유인을 고찰하며, 데이터 공유 정책 수립시 고려해야하는 지점들을 논의하였다. 그러나 본 논의는 데이터 가격 책정 방법, 데이터 공유로 발생할 수 있는 개인정보 보호 사건에서의 최적 책임 분담 방식 또는 반경쟁적 행위에 대한 모니터링 방법 등과 같은 다양한 개별 문제에 대한 구체적인 실행 방법을 엄밀히 검토하거나 제시하기에는 방법론적 한계가 존재한다. 이는 앞으로 경제학 이론을 포함한 다양한 방법으로 경제학 분야에서 연구가 될 것으로 기대한다.

■ 참 고 문 헌

1. 방정미, “데이터 경제활성화를 위한 필수요소법리의 재검토,” 『경제규제와 법』, 제13권 제1호, 2020, pp. 96-115.
(Translated in English) Bang, J., “Reassessment of Essential Facility Theory for Data Economy Stimulation,” *Journal of Law & Economic Regulation*, Vol. 13, No. 1, 2020, pp. 96-115.
2. 선지원, “데이터의 안전한 공유를 위한 법적 고찰,” 『경제규제와 법』, 제15권 제1호, 2022, pp. 172-192.
(Translated in English) Seon, J., “Legal Consideration for Safe Sharing of Data,” *Journal of Law & Economic Regulation*, Vol. 15, No. 1, 2022, pp. 172-192
3. 이 용·최명석·김성찬·이건우·장래영·이승우·이상환, “인공지능 학습 데이터 공유 활용 현황과 서비스 구축 방향,” KISTI Issue Brief, 제51호, 한국과학기술정보연구원, 2022.
(Translated in English) Lee, Y., M. Choi, S. Kim, K. Lee, L. Chang, S. Lee, S. Lee, *Status and Service Strategies on AI Learning Data Sharing*, KISTI Issue Brief, No. 51, Korea Institute of Science and Technology Information, 2022.
4. 이효석, “데이터 독점과 경쟁법의 적용,” 『법학논고』, 제77호, 2022, pp. 185-213.

- (Translated in English) Lee, H., "Application of Competition Law to Data Processing Activities of Dominant Companies," *Kyungpook National University Law Journal*, No. 77, 2022, pp. 185-213.
5. 정원준, "데이터 거래 및 유통 활성화를 위한 법제연구," 연구보고 22-12, 한국법제연구원, 2022.
(Translated in English) Chung, W., "A Study on the Legislation to Revitalize Data Transactions and Distribution," Research Report, 22-12, Korea Legislation Research Institute, 2022.
 6. 최난설현, "기업결합심사에 있어서 빅데이터의 경쟁법적 의미 -최근 해외 주요 기업결합 사례를 중심으로-, "『외법논집』, 제41권 제4호, 2017, pp. 323-343.
(Translated in English) Choi, N., "Merger Control, Big Data, Market Concentration, Foreclosure Effects, Barriers to Entry, Google/DoubleClick, Google/ITA, Facebook/WhatsApp, Microsoft/LinkedIn," *HUFS Law Review*, Vol. 41, No. 4, 2017, pp. 323-343.
 7. 한국데이터산업진흥원, "2022 데이터산업 현황조사 보고서," 과학기술정보통신부, 2023.
(Translated in English) Korea Data Agency, "2022 Data Industry Report," Ministry of Science and ICT, 2023.
 8. Acemoglu, D., A. Makhdoumi, A. Malekian, and A. Ozdaglar, "Too Much Data: Prices and Inefficiencies in Data Markets," *American Economic Journal: Microeconomics*, Vol. 14, No. 4, 2022, pp. 218-256.
 9. Barker, A., "Consumer Data and Competition: A New Balancing Act for Online Markets?" OECD Going Digital Toolkit Note, No. 5., 2021.
 10. Caserta, J., H. Harreis, K. Rowshankish, N. Srinidhi, and A. Tavakoli, "The Data Dividend: Fueling Generative AI," McKinsey Digital, 2023.
 11. Chee, F. Y., "EU Draft Data Act Puts Trade Secrets at Risk, Siemens, SAP Say," Reuters, 2023. 05. 09, <https://www.reuters.com/technology/siemens-sap-say-eu-draft-data-act-puts-trade-secrets-risk-2023-05-07/>.
 12. Chen, M., "Big Data," Oracle, 2024. 09. 23, <https://www.oracle.com/kr/big-data/what-is-big-data>.
 13. Choi, J., D. Jeon, and B. Kim, "Privacy and Personal Data Collection with Information Externalities," *Journal of Public Economics*, Vol. 173, 2019, pp. 113-124.
 14. Costa-Cabral, F., and O. Lynskey, "Family Ties: The Intersection between Data Protection and Competition in EU Law," *Common Market Law Review*, Vol. 54, No. 1, 2017, pp. 11-50.
 15. Digital Europe, "CEOs Call for Urgent Rethink on Data Act," Press Release, 2023. 05. 08, <https://www.digitaleurope.org/news/ceos-call-for-urgent-rethink-on-data-act/>.
 16. Ichihashi, S., "Online Privacy and Information Disclosure by Consumers," *American Economic Review*, Vol. 110, No. 2, 2020, pp. 569-595.
 17. Ivaldi, M., B. Jullien, P. Rey, P. Seabright, and J. Tirole, "The Economics of Tacit Collusion: Implications for Merger Control," *Contributions to Economic Analysis*, Vol. 282, 2007, pp. 217-239.
 18. Jones, C. I., and C. Tonetti, "Nonrivalry and the Economics of Data," *American*

- Economic Review*, Vol. 110, No. 9, 2020, pp.2819-2858.
19. Kearns, M., and A. Roth, Ch.1 Algorithmic Privacy: From Anonymity to Noise, *The Ethical Algorithm: The Science of Socially Aware Algorithm Design*, Oxford University Press, 2019.
 20. Martens, B., A. De Streel, I. Graef, T. Tombal, and N. Duch-Brown, "Business-to-Business Data Sharing: An Economic and Legal Analysis," JRC Digital Economy Working Paper, No. 2020-05, European Commission, Joint Research Centre (JRC), Seville, 2020.
 21. OECD, "Maximizing the Economic and Social Value of Data: Understanding the Benefits and Challenges of Enhanced Data Access," CDEP Meeting DSTI/CDEP(2016)4, OECD, 2016.
 22. Posner, R. A., "The Economics of Privacy," *The American Economic Review*, Vol. 71, No. 2, 1981, pp.405-409.
 23. Regulation 2023/2854, Regulation (EU) 2023/2854 of the European Parliament and of the Council of 13 December 2023 on harmonised rules on fair access to and use of data and amending Regulation (EU) 2017/2394 and Directive (EU) 2020/1828 (Data Act). <http://data.europa.eu/eli/reg/2023/2854/oj>.
 24. Rehman, ur I., "Facebook-Cambridge Analytica Data Harvesting: What You Need to Know," *Library Philosophy and Practice*, 2019, pp.1-11.
 25. Rho, S., and W. Shin, "When Data Sharing Meets Price Sharing," *Working Paper*, 2025.
 26. Sehgal, R., "AI Needs Data More Than Data Needs AI," *Forbes*, 2023. 10. 05.
 27. Stigler, G. J., "An Introduction to Privacy in Economics and Politics," *The Journal of Legal Studies*, Vol. 9, No. 4, 1980, pp.623-644.
 28. Taylor, C. R., "Consumer Privacy and the Market for Customer Information," *RAND Journal of Economics*, Vol. 35, No. 4, 2004, pp.631-650.
 29. Tsoy, N., and N. Konstantinov, "Strategic Data Sharing between Competitors. Advances in Neural Information Processing Systems," Vol. 36, 2023, pp.16483-16514.
 30. U.S. Department of Justice, "Principal Deputy Assistant Attorney General Doha Mekki of the Antitrust Division Delivers Remarks at GCR Live: Law Leaders Global 2023," 2023. 02. 02, <https://www.justice.gov/opa/speech/principal-deputy-assistant-attorney-general-doha-mekki-antitrust-division-delivers-0>.
 31. Varian, H. R., "Economic Aspects of Personal Privacy," *Internet Policy and Economics: Challenges and Perspectives*, 2009, pp.101-109.

Optimal Data Sharing in AI-data Markets*

Wiroy Shin**

Abstract

Data sharing within the AI ecosystem brings both social benefits and social costs. To understand the trade-off and study efficiency of the data market, I specify the economic mechanism of data sharing and explore factors that should be considered when promoting data sharing: economic characteristics of shared data, competition among firms in the AI ecosystem, and the transaction mechanism. If these conditions are met, data sharing can increase the number of innovations and productivity of the firms in the AI ecosystem, leading to economic growth. If data sharing policies are implemented without reflecting these factors, the production of data would decrease and negative externalities can occur in the AI ecosystem.

Key Words: AI, data, data sharing

JEL Classification: L1, L4, L5

Received: March 31, 2025. Revised: June 16, 2025. Accepted: June 23, 2025.

* This paper was presented at the special session on “Response to changes in the economic environment related to AI (Artificial Intelligence)” at 2025 Korea’s Allied Economic Associations Annual Meeting. This paper is a modified version of the final research report “A Study on Mandating Data Sharing of Platforms” conducted by the Chonnam National University Industry Cooperation Foundation, in which the author participated in 2023-2024.

** Assistant Professor, School of Economics, Chonnam National University, 77, Yongbong-ro, Buk-gu, Gwangju 61186, Phone: +82-62-530-1455, e-mail: wiroypsu@gmail.com