

(ISBN 979-11-6884-089-8 13500)

발 간 등 록 번 호
11-1430000-001911-01



AI 학습데이터의 활성화 방안 연구

- AI 학습데이터 공개 · 보호 · 활용을 중심으로 -

(A Study on the Utilization of AI Learning Data

- Focusing on the disclosure, protection, and utilization of AI learning data -)



제 출 문

특 허 청 장 귀 하

본 보고서를 “AI 학습데이터 활성화 방안 연구” 최종보고서로 제출합니다.

2022년 12월 21일

- 주관연구기관명 : 기율특허법률사무소
- 연구기간 : 2022.10.14~12.21
- 주관연구책임자 : 김윤명
- 참여연구원
 - 연구원 : 신무연
 - 연구원 : 정희원
 - 연구원 : 이영훈
 - 연구원 : 김정현

목 차

제1장. 서 론	1
1. 연구 목적	1
2. 연구 배경	2
3. 본 연구의 주요 내용	17
제2장. 현행법 상 AI 학습데이터의 보호 체계	18
제1절 AI발명과 학습데이터	18
1. 데이터의 법적 정의	18
2. AI발명과 학습데이터의 특성	19
3. 데이터의 구조	25
4. 학습데이터의 구축과 학습모델	29
제2절. 학습데이터 보호 및 활용 법제	31
1. 개관	31
2. 국내 법제의 현황과 문제점	31
제3절 특허법적 보호에 대한 검토	44
1. 특허법적 보호의 필요성 : 현행 법제하에서 데이터 보호의 한계를 중심으로 정리	44
2. 검토 내용	45
제3장. 기술공개 요건으로서 데이터 공개에 대한 연구	48
제1절 문제 제기	48
1. AI의 블랙박스화와 데이터 공개	48
2. AI발명의 구조와 데이터의 관계	50
3. 데이터 보호 법익과 데이터 공개의 충돌	51
제2절. 기술공개와 관련 제도	52

1. 기술공개제도의 의의	52
2. 기술공개로서 발명의 상세한 설명	55
3. 기술분야 및 평균적 기술자의 수준	59
4. 미완성 발명과 발명의 완성	60
5. 기술공개 기준 등에 대한 한계	63
제3절. IP5의 AI발명에 있어서 데이터 공개 논의	63
1. 각국의 AI발명 관련 기술공개 수준 검토	63
2. 미국	66
3. 일본	71
4. EU	76
5. 중국	81
제4절. 데이터 공개관련 주요 논의/사례/판례의 분석·검토	89
1. EPO심판원 심결	89
2. 영국 IPO의 데이터 공개에 대한 논의	93
3. WIPO 차원의 논의	94
4. 학술연구 부문에서의 논의	103
제5절. AI 발명의 데이터 공개 범위 및 방안	106
1. AI발명에 있어서 기술공개	106
2. 데이터 공개의 수준 및 범위	109
3. 인센티브의 부여 방안	117
4. 공개시의 기대효과	121
제6절. 특허제도 개선의 필요성 검토	123
1. 데이터 공개의 필요성에 대한 의견 수렴	123
2. 심사지침 개정에 대한 검토	125
3. 특허법제의 개정 사항	126
제4장. AI 학습데이터의 보호·활용을 위한 제도개선 기초 연구	129

제1절. 논의의 필요성	129
1. 발명의 투명성 확보	129
2. 발명의 반복재현성 확보	130
제2절. 데이터 관련 제도 방안 검토	132
1. 실무적인 검토사항	132
2. 유사 제도 검토	133
3. 데이터 기탁제도에 대한 비교·분석	140
제3절. 데이터의 보호·활용을 위한 제도개선	145
1. 데이터의 신뢰성 확보 방안 검토	145
2. 데이터 재현가능성 확보 방안 검토	145
3. 기탁제도 도입시에 고려되어야할 법률 문제	146
제4절. 데이터에 대한 독자적 권리화 방안	147
1. 독자적 권리화의 필요성	147
2. 데이터 특허화 방안	147
3. 데이터 등록제도의 도입	148
제5절. 소결	149
제5장. AI 학습데이터를 활용한 데이터산업 활성화 방안 연구	150
제1절. 데이터산업의 활성화	150
1. 데이터산업 현황	150
2. 데이터산업 활성화 방안	152
제2절. 데이터 거래 체계 및 거래 현황	153
1. 데이터 거래 체계	153
2. 데이터의 거래	155

제3절. 기탁된 데이터의 활용 방안	156
1. 데이터의 제공	156
2. 기탁된 데이터의 활용	157
3. 법적 근거의 마련	158
제4절. 데이터 산업법제와의 정합성 검토	161
1. 데이터보호 체계는 정합성을 갖는가?	161
2. 개선 방안	161
제5절. 데이터 이용활성화를 위한 제도적 가이드라인의 제시	162
1. 기탁 데이터의 분양	162
2. 산업재산정보 활용촉진법(안)과의 연계	162
3. 기계학습에서의 공정이용 가이드라인	163
제6장. 결 론	169
참고문헌	171

요 약 문

기계학습(machine learning)에서 중요한 요소가 알고리즘에서 데이터 중심으로 이동하고 있다. 학습데이터를 통하여 알고리즘은 규칙과 상관관계를 학습하게 됨으로써 인공지능 모델(AI 모델)이 된다. 학습과정에서 데이터는 인공지능 모델이나 시스템을 개발하는 데 결정적인 역할을 한다. 특허법상 기술공개는 법정 요건이기는 하지만 실질적으로 해당 기술의 충분한 공개여부를 판단하는 입장은 평균적인 기술자로서 하겠지만, 발명의 특성이나 요소기술의 수준에 따라 달라질 수 있는 사안이다. 그럼에도 불구하고, 인공지능 발명을 구성하는 다양한 요소 중에서 데이터가 차지하는 비중이 커지고 있으며 중요부를 차지하기 때문에 AI발명에 있어서 사용되는 데이터 구조 내지 데이터셋을 공개하거나 관련 기술의 공개는 중요한 일이 아닐 수 없다. 그렇지만, AI발명의 기술공개와 더 나아가 데이터 공개에 대해서는 별다른 논의가 이루어지지 않고 있다. 데이터 공개는 인공지능이 갖는 내재적 한계인 블랙박스(black box)화에 따른 투명성 확보, 그에 따른 사회적 후생 등에서도 유의미한 결과를 가져올 것으로 기대된다.

본 연구의 목적은 특허법상 발명의 상세한 설명에 부합하기 위한 기술적 공개요건으로서 데이터 공개를 위한 기준이나 제도를 설계하고자 하였다. 이를 위하여 IP5나 WIPO 등을 중심으로 논의 동향, 사례나 판례 등을 조사·분석하였다.

먼저, 제2장에서는 <현행법 상 AI 학습데이터의 보호 체계>에 대해 연구하였다. 인공지능의 학습을 위한 데이터 관련 정책을 위해 다양한 법제가 만들어졌다. 데이터의 법적인 정의에서부터, 관련 입법 현황을 정리하고자 한다. 아울러, 데이터의 특성 및 구조화와 관련된 내용을 살펴볼 필요가 있다. 특히, AI발명과 SW발명 등 일반발명과의 차이 등에 대해 살펴보고, 데이터셋에 대한 특허법적 보호 필요성에 대해 살펴보았다.

제3장에서는 <AI발명에 대한 특허법상 기술공개 범위>에 대해 연구하였다. 데이터 처리가 주요 구성요소를 이루는 AI발명의 기술공개 방법, 범위에 대해 살펴보았다. 각국의 특허법에서는 발명의 공개를 법정화하고 있다. 이는 기술공개에 따른 인센티브로써 특허권이라는 독점권을 부여하는 것이기 때문이다. 이러한 점에서 기술공개, 특히 인공지능 발명에서 데이터 등의 기술공개의 수준을 어느정도 까지 할 것인지는 주요한 과제이다. 이를 위해 각국에서 AI발명과 데이터 발명에 대해 어떠한 정책을 가지고 있는지에 대해 검토하였다. 다만, 데이터 공개는 발명자나 기업의 입장에서 영업비밀로 유지되는 경우에 있어서 정책적인 결정사항이기도 하기 때문에 자발적인 공개 내지 발명의 상세한 설명 등의 방법으로 공개방안을 제안하고자 하였다.

제4장에서는 <학습데이터의 보호 및 활용을 위한 제도개선 기초연구>로서 데이터의 재현가능성과 신뢰가능성을 확보하기 위한 제도에 대해 조사하였다. 데이터가 사용되는 발명에 있어서 신뢰가능성과 재현가능성은 중요한 과제가 아닐 수 없다. 전자는 데이터에 대한 신뢰성을 비롯하여 발명 자체에 대한 신뢰성을 확보할 수 있는 방안을 찾는 것이다. 후자인 재현가능성은 공개된 발

명을 통하여 평균적 기술자가 용이하게 실시할 수 있을 정도의 기술공개를 위한 요건으로서 데이터를 이용할 수 있는 방안을 찾고자 하는 것이다. 이를 위하여, 데이터를 공개하거나 제3의 기관에 기탁하는 방안을 고려할 수 있다.

제5장에서는 <AI 학습데이터를 활용한 데이터산업 활성화 방안>에 대해 연구하였다. 실상 데이터를 생산하거나 이용하는 산업계에서는 데이터의 활용에 대한 방점을 찍는 경우가 많다. 데이터 확보에 따른 비용이 소요되기 때문이다. 이러한 이유로 정부에서는 데이터댐 또는 데이터레이크(data lake) 정책을 펼치면서 다양한 학습용 데이터셋을 구축하고 있다. 이러한 데이터셋은 AI-HUB를 통하여 일반 국민에게 공개되고 있는 데이터는 인공지능 학습에서 중요하게 활용되고 있다. 이러한 공개된 데이터셋을 활용하는 것은 경쟁력을 갖출 수 있는 방안이 된다는 점에서 중요한 시사점을 가질 것이다.

S U M M A R Y

An important element in machine learning is moving from algorithm to data center. Algorithms become artificial intelligence models (AI models) by learning rules and correlations through learning data. In the learning process, data plays a crucial role in developing artificial intelligence models or systems. Although disclosure of technology is a legal requirement under the Patent Act, the position of determining whether the technology is sufficiently disclosed is an average technician, but it may vary depending on the characteristics of the invention or the level of elemental technology. Nevertheless, since data accounts for an increasing proportion of the various elements that make up artificial intelligence inventions, disclosure of data structures or datasets used in AI inventions or disclosure of related technologies is important. However, there is no discussion on the technology disclosure of AI invention and further data disclosure. Data disclosure is expected to have significant results in securing transparency due to black boxization, which is an inherent limitation of artificial intelligence, and social welfare.

The purpose of this study was to design standards or systems for data disclosure as technical disclosure requirements to conform to the detailed description of the invention under the Patent Act. To this end, discussion trends, cases, and precedents were investigated and analyzed, focusing on IP5 and WIPO.

Chapter 2, the "protection system of AI learning data under the current law" was studied. Various laws have been created for data-related policies for learning artificial intelligence. From the legal definition of data, I would like to summarize the current status of related legislation. In addition, it is necessary to look at the contents related to the characteristics and structure of data. In particular, the differences between AI inventions and general inventions such as SW inventions were examined, and the necessity of patent legal protection for datasets was examined.

Chapter 3, the scope of technology disclosure under the Patent Act on AI inventions was studied. We looked at the technology disclosure method and scope of AI invention, where data processing is a major component. Patent laws in each country legalize the disclosure of inventions. This is because the exclusive right of patent rights is granted as an incentive for technology disclosure. In this respect, it is a major task to what level of technology disclosure, especially in artificial intelligence inventions, will be made. To this end, we reviewed what policies each country has for AI invention and data invention. However, since data disclosure is also a policy decision when it is maintained as a trade secret from the perspective of the inventor or company, it was intended to propose a disclosure plan through voluntary disclosure or detailed explanation of the invention.

Chapter 4 investigated the system to secure the reproducibility and reliability of data as <Basic Research on System Improvement for the Protection and Utilization of Learning Data>. Reliability and reproducibility are important challenges in the invention in which data is used. The former is to find a way to secure the reliability of the invention itself, including the reliability of data. The latter

reproducibility is to find a way to use data as a requirement for technology disclosure that can be easily implemented by the average technician through the disclosed invention. To this end, a method of disclosing data or depositing data to a third institution may be considered.

Chapter 5, the data industry revitalization plan using AI learning data was studied. In fact, industries that produce or use data often focus on the use of data. This is because it costs money to secure data. For this reason, the government is building various learning datasets while implementing data dam or data lake policies. These datasets are being disclosed to the general public through AI-HUB, which is importantly used in artificial intelligence learning. Utilizing these public datasets will have important implications in that it becomes a competitive method.

제1장. 서론

1. 연구 목적

인공지능이 하는 일이 많아지면서, 인공지능을 학습시키기 위한 데이터의 중요성이 커지고 있고, 이를 위한 데이터 확보는 중요한 일이 되고있다. 정부는 관련 법제를 입법화하여 데이터를 효과적으로 확보하기 위해 필요한 근거를 뒷받침하고 있다. 인공지능 알고리즘의 고도화는 데이터에 기반한 기계학습을 통하여 이루어지고 있다. 기계학습이 이루어진 후에 생성된 인공지능모델(또는 학습모델이라고도 함)은 다양한 데이터와 매개변수, 가중치 등에 따라 다양한 결과를 만들어내게 된다. 인공지능모델을 거치면서 제공된 데이터나 인간에 의하여 처리된 다양한 정보나 비기술적 요인에 의해서도 그 결과가 달라질 수 있다. 동일한 인공지능모델이라고 하더라도, 데이터에 따라 달라지는 것이기 때문에 재현가능성이나 투명성을 위해서 인공지능학습 과정에서 소요된 데이터 등의 리소스에 대한 공개의 필요성이 커지고 있다. 무엇보다, 기계학습 과정이나 인공지능모델이 내린 결과에 대해서는 그 원인이나 인과관계를 확인할 수 있는 것이 아닌 결과에 따른 상관관계만을 확인할 수 있을 뿐이다. 따라서, 내부적인 처리과정이 윤리적인지도 확인하기 어렵다. 이러한 점이 인공지능의 투명성이 강조되고 있는 이유이기도 하다.

인공지능의 투명성을 확보하기 위하여 인공지능 윤리가 강조되거나 책무(accountability)가 강조되기도 한다. 실제 각국은 AI관련 규제법을 제안하고 있는 것도 이러한 이유 때문이다. EU는 AI Act, AI Liability Directive, Product Liability Directive의 입법화를 추진 중에 있으며, 미국은 Algorithm Accountability Act를 2019년, 2022년 각각 제안한 바 있다. 일본은 플랫폼공정화법을 제정하였으며, 우리나라는 알고리즘 규제에 대한 다양한 법안이 국회에 발의되어있는 상황이다. 이러한 법안은 알고리즘의 투명성 확보를 위한 사업자의 책무를 강조하고 있다. 이러한 상황에서 인공지능에 대한 투명성 확보 및 공개된 기술에 대한 재현가능성의 확보는 시대적인 요구 사항이기도 하다. 개발자들은 자발적으로 알고리즘이나 데이터를 공개하거나 학술단체에서는 공개된 논문이나 기술에 대한 재현가능성을 체크할 수 있는 방안을 제시하기도 한다. 이러한 과정이 인공지능에 대한 투명성을 확보할 수 있는 방안이라고 보기 때문이다.

본 연구는 투명하고 신뢰가능한 인공지능 기술의 발전과 사회적인 후생을 향상시키고, 산업발전을 위한 특허법의 목적과 이를 달성하기 위한 기술공개제도를 AI발명, 특히 데이터에 적용하는 방안에 대해 살펴보고자 한다. 특허법은 기술보호만이 아닌 기술을 활용하여 다양한 혁신을 가져온다는 점에서 데이터와 관련한 공개방안에 대한 연구는 의미가 있다. 데이터의 생애주기에 따른 보호에 대한 고민은 AI발명의 보호를 넘어서, 좀더 구체적으로는 데이터의 중요성에 비추어 볼 때 데이터의 구조 또는 구조화된 데이터를 넘어서 데이터 또는 데이터셋의 보호와 맞물려 충분한 기술공개(sufficiency of disclosure)와도 밀접한 관련이 있기 때문이다. 무엇보다, 기술공개는 해당 발명에 대한 인센티브로서 독점권을 부여함으로써 이루어지는 것이기 때문에 실질적인 공

개를 통하여 제3자가 공개된 기술내용을 통하여 발명을 쉽게 실시할 수 있을 정도로 공개된다면 특허법의 목적은 달성된 것으로 볼 수 있다. 그렇지만, 충분한 공개에 대해서는 특허법이나 심사 지침에서 개별적인 사안에 대해 명확하게 기준을 제시하기는 어렵다. 기술공개는 법정 요건이기는 하지만 실질적으로 해당 기술의 충분한 공개여부를 판단하는 입장은 평균적인 기술자로서 하겠지만, 발명의 특성이나 요소기술의 수준에 따라 달라질 수 있는 사안이기도 하다. 그럼에도 불구하고, 인공지능 발명을 구성하는 다양한 요소 중에서 데이터가 차지하는 비중이 커지고 있으며 중요부를 차지하기 때문에 인공지능 발명에 있어서 사용되는 데이터 구조 내지 데이터셋을 공개하거나 관련 기술의 공개는 중요한 일이 아닐 수 없다. 그렇지만, 각국의 정책당국은 인공지능 발명의 기술공개뿐만 아니라 더 나아가 데이터 공개에 대해서는 별다른 관심을 두고 있지는 않다.

본 연구는 기술공개로서 데이터의 공개 및 이에 따른 범위와 기준 등에 대해 살펴보고, 이를 뒷받침할 수 있는 데이터 기탁(deposit)에 관한 제도에 관하여 연구한다. 이를 위하여 IP5, WIPO를 중심으로 AI발명의 성립성과 이를 판단하는 과정에서 이루어지는 충분한 기술공개 및 데이터 관련 공개에 관한 논의를 살펴보고자 하였다.

2. 연구 배경

가. 왜, 데이터가 중요한가?

인공지능의 고도화를 위한 기계학습에서 중요한 요소는 알고리즘에서 데이터 중심으로 이동하고 있다. 인공지능 기술은 특정 과제를 해결하기 위하여 주어진 문제점을 찾아내고, 이를 해결하기 위하여 데이터를 이용하여 반복학습하며, 그 학습을 통하여 스스로 최적의 해결방안을 도출하는 것을 말한다. 인공지능은 기본적인 알고리즘은 사람이 프로그래밍하지만, 이를 고도화하기 위해서는 데이터에 기반한 기계학습(machine learning) 과정을 거치게 된다. 이 과정에서 데이터가 사용되며 대량의 데이터를 활용함으로써 보다 지능화된 결과를 얻게 된다.

무엇보다, 빅데이터는 인공지능 시스템을 개발하는 데 결정적인 역할을 한다. AI 시스템은 빅데이터를 소화하고 처리한 후, 기본 AI 모델은 규칙과 관계를 학습하고 숨겨진 패턴을 표면화한다. 컴퓨팅 능력과 GPU 등 병렬처리는 인공지능 시스템이 거대하고 복잡한 양의 데이터를 분석하고 학습할 수 있게 해주었다. 따라서, 데이터 수집 방법과 데이터 엔지니어링 및 데이터 분석 방법은 인공지능 시스템의 중요한 부분이다.¹⁾

학습데이터를 통하여 알고리즘은 규칙과 상관관계를 학습하게 됨으로써 인공지능 모델(AI 모델)이 된다. 학습과정에서 데이터는 인공지능 모델이나 시스템을 개발하는 데 결정적인 역할을 한다. 즉, AI 시스템은 데이터를 소화하고 처리한 후, 규칙과 관계를 학습하고 숨겨진 패턴을 밝

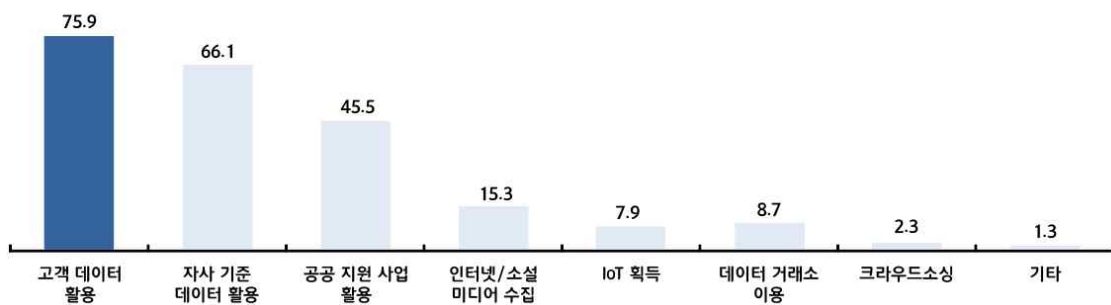
1) Mehdi Poursoltani, Disclosing AI Inventions, TEXAS INTELLECTUAL PROPERTY LAW JOURNAL Vol.29, 2021, p. 49.

혀낸다. 컴퓨터의 컴퓨팅 파워와 병렬처리는 인공지능 시스템이 거대하고 복잡한 양의 데이터를 분석하고 학습할 수 있게 한다. 이처럼 데이터 수집, 데이터 엔지니어링 및 데이터 분석 방법은 인공지능 시스템의 중요한 부분이다.²⁾ 데이터가 많을수록 학습모델(AI모델)³⁾의 품질은 우수해지며, 이 과정에서 편향이나 오류가 줄어들 수 있다. 데이터가 적을수록 편차가 커지게 되기 때문에 “AI기술을 활용해 정확한 결정을 내리려면 양질의 데이터가 많아야 한다. 기계는 인간에 의해 학습된다. 그러므로 편향된 결과가 도출되었다면 기계가 편견을 생성한 게 아니고 인간이 제공한 지식이나 데이터의 문제”⁴⁾이다. 결국, AI가 내리는 결론은 인간의 의도적인 것이기는 하지만, 악의적인 의도성을 배제함으로써 AI가 내린 결론에 대한 신뢰성을 확보하고, 또한 투명하고도 객관적인 결론을 이끌 수 있는 방법은 양질의 데이터를 확보하여 학습용 데이터로 활용하는 것이다.

데이터를 확보하기 위해서는 인터넷이나 IoT 등의 기술을 통하여 데이터를 수집하거나 크롤링(crawling)한다. 이렇게 수집된 원시데이터(raw data)를 가공하여 의미있는 데이터로 만들어내는 정제과정을 거치는데, 이를 레이블링(labeling)이라고 한다. 정제된 데이터는 학습과정에서 활용되면서 그 결과로 학습모델(AI모델)이 만들어진다. 인공지능을 활용한다는 것은 학습된 AI모델을 활용하여 데이터를 입력하고 결과로서 출력되는 과정을 말한다. AI모델의 성능을 높이기 위해서는 데이터의 확보가 무엇보다 중요한 이유이며, 다음 그림은 데이터 확보 방식에 대한 실태조사 결과이다. 실질적으로 플랫폼사업자들은 고객들의 데이터를 활용하는 경우가 많은 것으로 파악된다.

[그림 1-1] 학습용 데이터의 확보 방식

[Base= 모집단 전체, N=1,365, 단위: % 복수응답]



<출처 : 인공지능산업 실태조사, 2022>

2) Mehdi Poursoltani, Disclosing AI Inventions, TEXAS INTELLECTUAL PROPERTY LAW JOURNAL Vol.29, 2021, p.48.
 3) AI학습모델에는 선형 회귀, 로지스틱 회귀, 선형 판별 분석, 의사 결정 트리, 나이브 베이즈, K-Nearest Neighbors, 학습 벡터 양자화, 지원 벡터 머신, 배깅 및 랜덤 포레스트, 심층 신경망과 같은 가장 인기 있는 기계 학습 알고리즘을 포함하는 여러 유형의 모델 등이 있다. Vladimir Fedak, Top 10 Most Popular AI Models, DZone (Nov. 8, 2018), <https://dzone.com/articles/top-10-most-popular-ai-models>.
 4) 이수호, AI인사이트, 한빛미디어, 2022, 247면.

학습데이터의 확보가 쉽지 않기 때문에 많은 사업자들은 고객 데이터를 활용하거나 공공지원 사업을 활용하는 경우가 많다. 또한, 데이터 거래소를 이용하는 경우도 있다.⁵⁾ 이처럼, 기업은 양질의 데이터를 확보하기 위해서는 작지 않은 비용을 투자한다. 실제, 기계학습 과정에서 알고리즘의 구현에는 5% 정도의 예산이 소요된다면 데이터 관련 분야에서는 95% 정도의 예산이 소요된다고 한다.⁶⁾ 이처럼 데이터를 확보하고 처리하는데 기업은 예산을 소요하는 것으로 이해할 수 있다. 따라서, 그러한 투자는 그 자체를 보호받거나 그 결과를 보호받아야 한다.

데이터의 중요성으로 인하여 정부는 다양한 데이터 정책과 데이터 관련 입법을 진행하였다. 저작권법, 부정경쟁방지법 등 전통적인 지식재산 관련 법제에서도 데이터의 보호 및 활용이 가능하도록 하였으며, 저작권법은 데이터 확보를 위한 공정이용(fair use)이 가능하도록 하고 있다. 여기에 더하여 데이터 산업법제가 마련되었다. 과학기술정보통신부 소관의 데이터 산업진흥 및 이용촉진에 관한 기본법(이하, 데이터 산업법이라 함)이나 산업통상자원부 소관의 산업 디지털 전환 촉진법이 그것이다. 물론, 다른 부처에서도 데이터 관련 법안은 준비 중에 있다. 이처럼 정부는 데이터의 중요성을 인식하고 데이터의 보호 및 이용활성화를 꾀하고 있다. 다만, 각각의 법제가 추구하는 가치와 방법의 차이로 인하여 보호받을 수 있는 영역의 차이가 있는 것은 사실이다.

나. 데이터 공개 현황 및 효과

인공지능이 지금처럼 발전할 수 있었던 것은 많은 개발자들이나 기업들이 인공지능 알고리즘 및 이와 관련된 SW와 데이터를 공개했기 때문이다. 특히, 알고리즘이나 데이터는 깃허브(GitHub)에, 논문 등 학술자료는 Arxiv.org 등의 사이트에 공개됨으로써 누구나 해당 알고리즘이나 데이터를 선행기술(prior art)로써 활용하고 있다. 더욱이, 미흡한 공개일 경우에는 제3자가 데이터셋을 추가하여 해당 내용을 보완하기도 한다. 지금처럼, 인공지능이 발전하게 된 것은 특허법상 기술 공개와 더불어 자발적인 공개와 같은 오픈소스(open source) 정신이 반영되었기 때문이라고 생각된다. 인공지능 발명품이나 인공지능 소프트웨어를 오픈소스로 만드는 것도 가능하다. 오픈소스 커뮤니티의 부상은 인간이 한편으로는 본질적으로 창의적이고 다른 한편으로는 사회적이고 관대하다는 것을 보여준다. 자원봉사자들은 이타주의나 교육이나 사회적 인정에 대한 열망과 같이 이익보다 더 부드러운 인센티브 때문에 그러한 공동체에 가입한다. 대규모 공개 데이터베이스는 정보 공유를 촉진한다. 지식의 발전과 접근의 증가, 그리고 모든 사람의 복지를 증진시키는 것이 지식재산법의 근본적인 목표이다. 오픈소스 데이터의 데이터베이스화와 자발적인 지식 공유는 이러한 목표를 위한 노력으로 볼 수 있다. 또한 이러한 공유는 또한 선행기술의 원천을 제공함으로써 비실무적 기업이 지대추구(rent-seeking) 행위⁷⁾에 관여하는 것을 방지할 수 있다.⁸⁾ 다음은 비

5) 소프트웨어정책연구소, 2021 인공지능산업 실태조사, 과학기술정보통신부, 2022, 52면.

6) 샌딕 우담찬다니, 데이터 민주화와 셀프서비스 데이터, 에이콘, 2022, 27면.

7) 지대추구행위란 기존의 부에서 자신의 몫을 늘리는 방법을 찾으면서도 새로운 부를 창출하지는 않는 활동을 말한다.

8) Yanisky-Ravid, Shlomit and Liu, Xiaoqiong (Jackie), When Artificial Intelligence Systems Produce

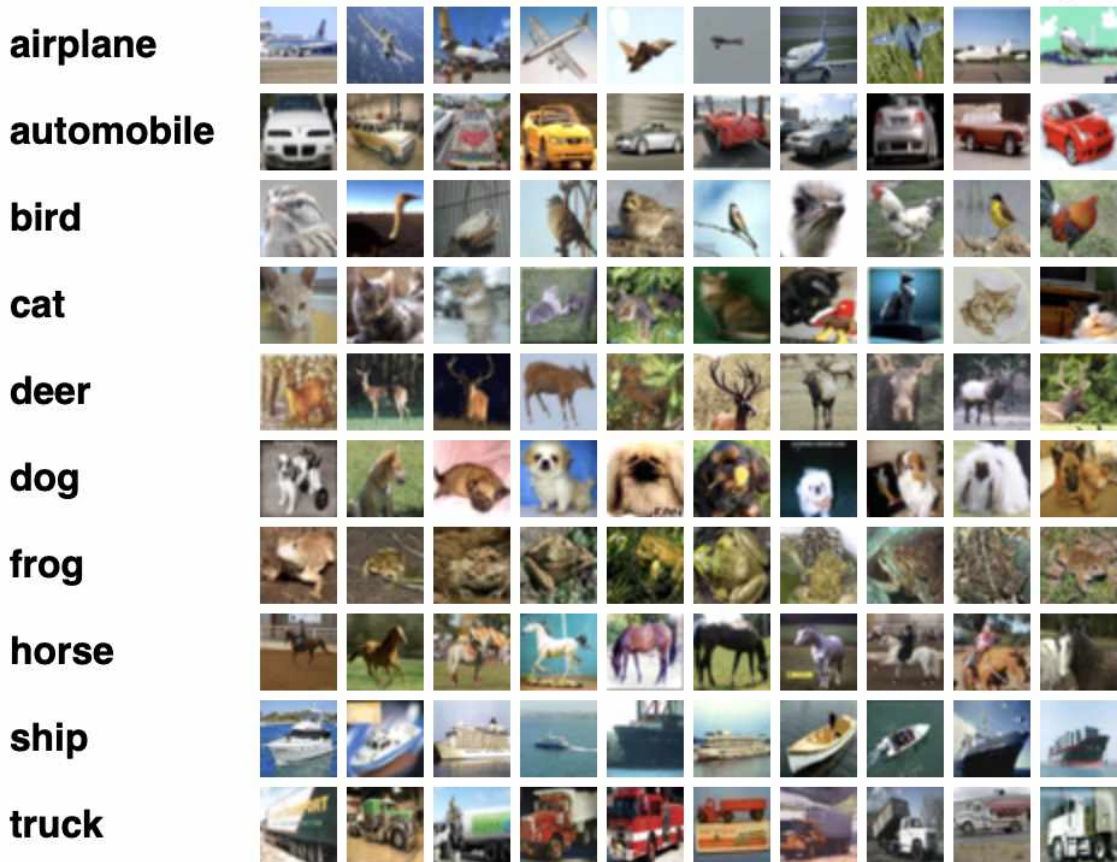
전연구 분야에서 대표적으로 사용되는 공개된 데이터셋의 현황이다.⁹⁾ 물론, 이외에도 다양한 사이트에서 학습용 데이터셋이 공개되어 있으며 이를 활용하여 다양한 AI모델이 만들어지고 있다.

(1) 데이터 공개 현황

1) CIFAR-10

CIFAR-10 데이터셋은 머신러닝 연구에 가장 활발히 사용되는 데이터셋으로, 2009년에 딥러닝의 아버지라고 불리는 제프리 힌튼 교수팀이 구축했다. 10개의 클래스(airplane, automobile, bird, cat, deer, dog, frog, horse, ship, truck)로 이루어진 6만 개의 이미지 데이터셋이다.

[그림 1-2] CIFAR-10 데이터셋



<출처 : CIFAR-10>

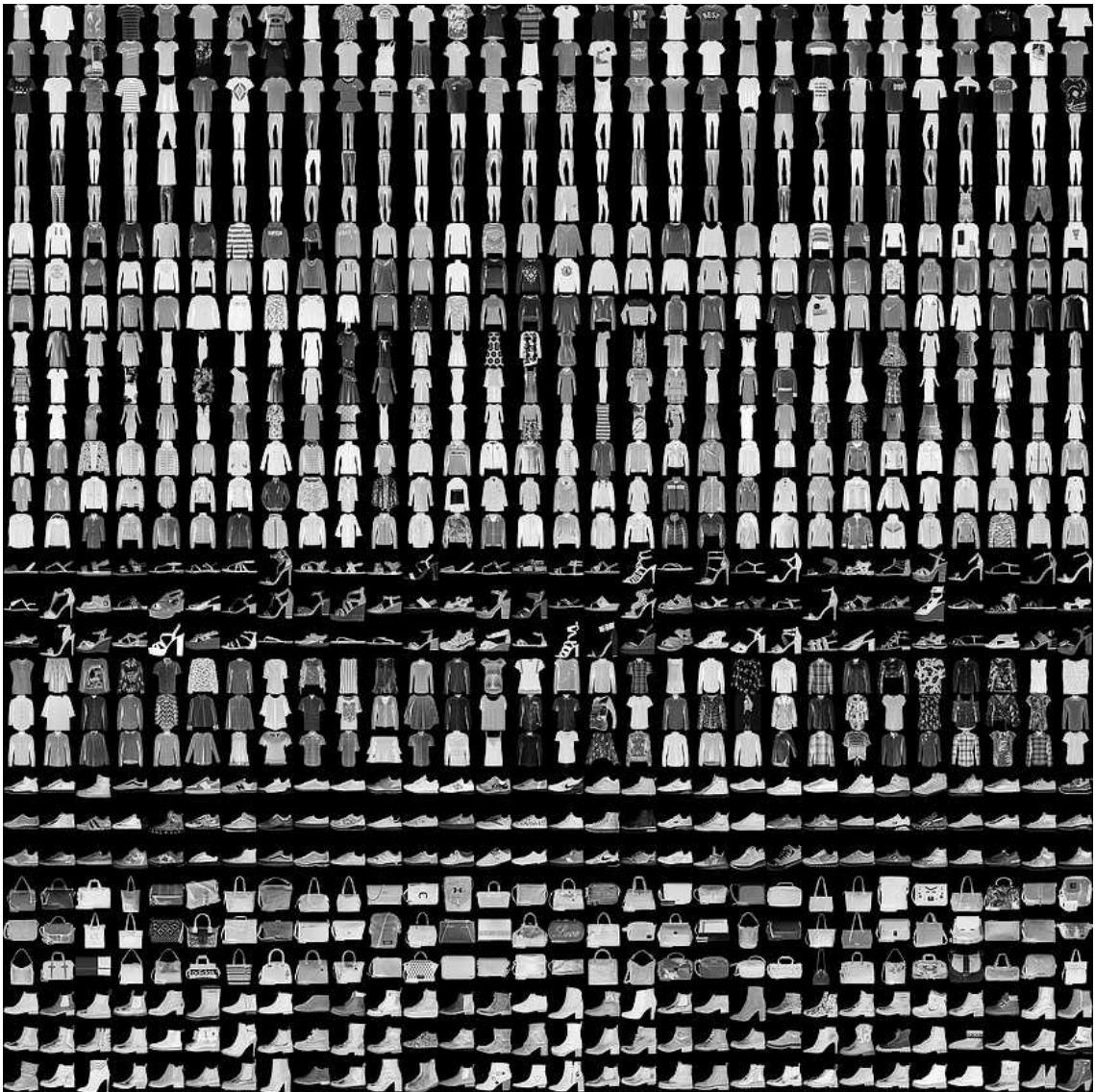
Inventions: The 3A Era and an Alternative Model for Patent Law , 39 Cardozo Law Review, 2215-2263 (2018), p. 52

9) Olivia Ha, 컴퓨터 비전 데이터에 대한 모든 것, SuperbAI, 2021. <https://www.superb-ai.com/ko-blog/everything-about-computer-vision-data> <2022.11.22 접속>

2) Fashion MNIST

손글씨 우편번호를 빠르게 읽기 위해 우편봉투로부터 숫자를 직접 추출하여 만든 손글씨 숫자 데이터셋인 NIST를 일반화, 표준화하여 발전시킨 MNIST에 기원을 두는 데이터셋이다. 손글씨 숫자를 패션과 관련된 이미지로 대체한 것으로, Fashion MNIST는 10개의 카테고리로 분류되는 70,000개의 흑백 이미지로 구성된 데이터셋이다.

[그림 1-3] Fashion MNIST 데이터셋



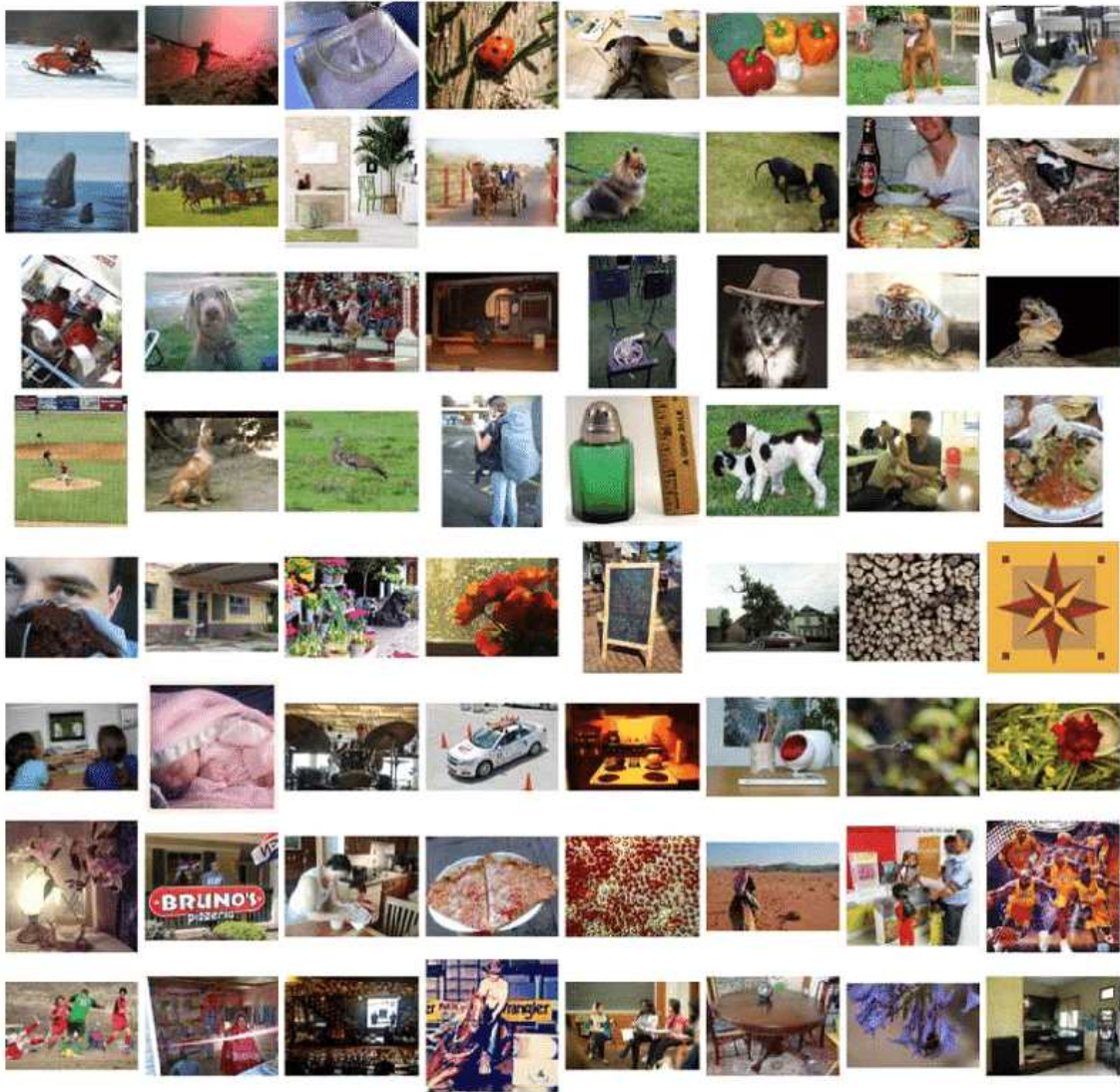
<출처 : Fashion MNIST>

3) ImageNet

ImageNet은 1,400만 개 이상의 공개 이미지 데이터셋으로, 스탠포드 대학의 Fei-Fei Li 교수 등

이 컴퓨터 비전 분야에 있어 학습 데이터의 부족 문제 해결을 위해 2009년에 시작한 데이터셋이다. 그 분류가 매우 다양해서 일상생활에서 볼 수 있는 거의 모든 종류의 이미지를 얻을 수 있다.

[그림 1-4] ImageNet 데이터셋



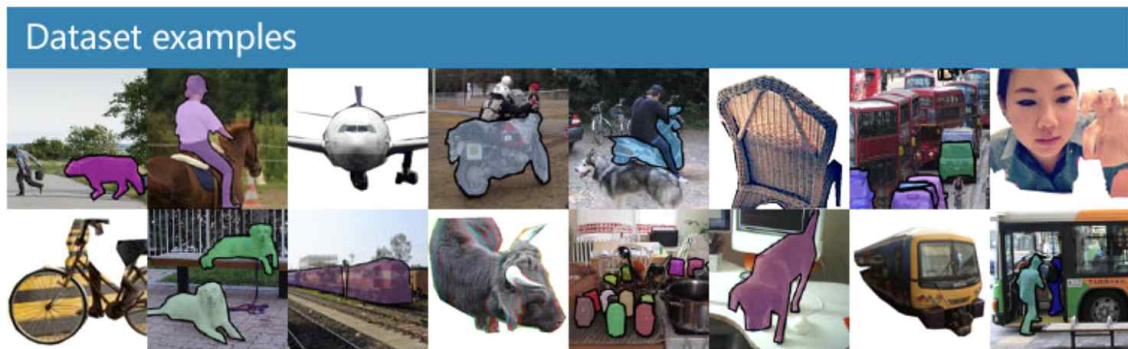
<출처 : ImageNet>

4) MS COCO

마이크로소프트사(MS)에서 구축한 COCO는 Common Objects in COntext의 약자로 딥러닝 프로그램을 교육하는 데 사용되는 가장 인기 있는 오픈소스 데이터베이스 중 하나로, Object Detection, Segmentation, Keypoint Detection 등에 쓰인다. COCO 데이터는 이미지를 설명하는 캡션도 함께 제공하여, 이미지를 설명하는 문장을 학습시켜 유사한 이미지가 주어졌을 때 설명을

자동으로 생성할 수 있도록 하는 데 큰 기여를 하고있다.

[그림 1-5] MS COCO 데이터셋

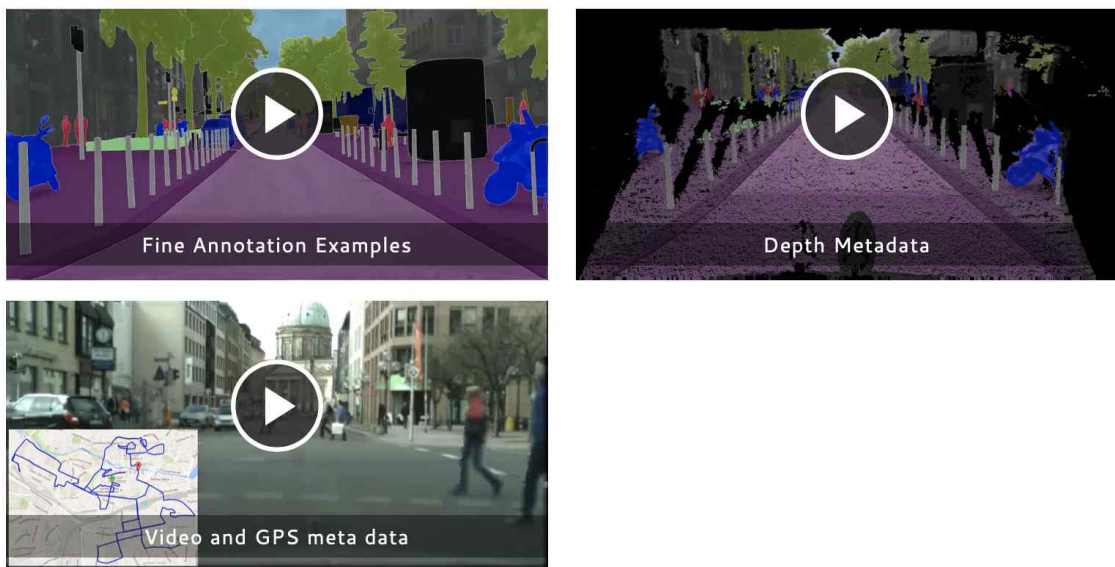


<출처 : MS COCO>

5) Cityscapes

자율주행에서 사용될 수 있는 도로환경과 관련된 데이터셋이다. 50개 도시의 거리 장면에서 기록된 다양한 스테레오 비디오 시퀀스 세트를 포함하는 새로운 대규모 데이터셋으로 도시 장면에 대한 의미론적 이해에 중점을 두고 있다.

[그림 1-6] Cityscapes 데이터셋



<출처 : Cityscapes>

6) Open image

구글이 머신러닝을 위해 2016년에 공개한 이미지에 주석이 달린 데이터셋으로 약 190만개에

전문 라벨러들이 라벨링을 검수한 이미지들을 포함하고 있다. 데이터셋은 V1부터 계속 업데이트 되어 2020년 2월 기준으로 가장 최신 버전인 V6버전까지 공개되었다. 이미지 상의 객체간 관계를 보여주는 라벨링, 이미지를 설명하는 캡션과 음성 나레이션이 추가되어 이미지와 캡션 사이를 연결하는 복잡한 모델을 만드는 데 유용하게 사용되고 있다.

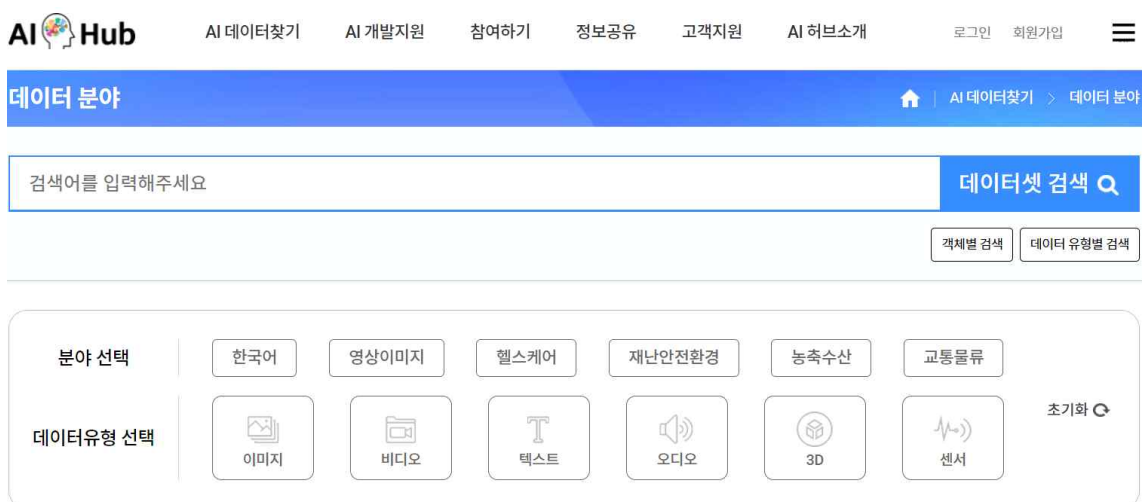
(2) 데이터셋 플랫폼 현황

국내외적으로 많이 활용되고 있는 데이터 포털을 통해 공개되는 데이터셋 플랫폼 현황은 다음과 같다. 다만, 아래는 예시적인 것이기 때문에 이외에도 다양한 데이터셋 플랫폼을 통하여 데이터셋이 공개되어있음을 알 수 있다.¹⁰⁾

1) AI허브¹¹⁾

AI허브는 정부의 데이터댐이라는 디지털뉴딜을 통하여 구축한 학습용 데이터셋을 제공한 사이트이다. 누구나 회원으로 가입하는 경우에는 이용할 수 있도록 개방하고 있다. 현재, 9억건 이상의 데이터셋이 구축되어있으며, 다양한 분야에서 제공 중이다. 다만, AI허브에서는 해외 이용자는 그 이용에 제한을 두고 있다.

[그림 1-7] AI허브 데이터 찾기



<출처 : AI 허브(2022)>

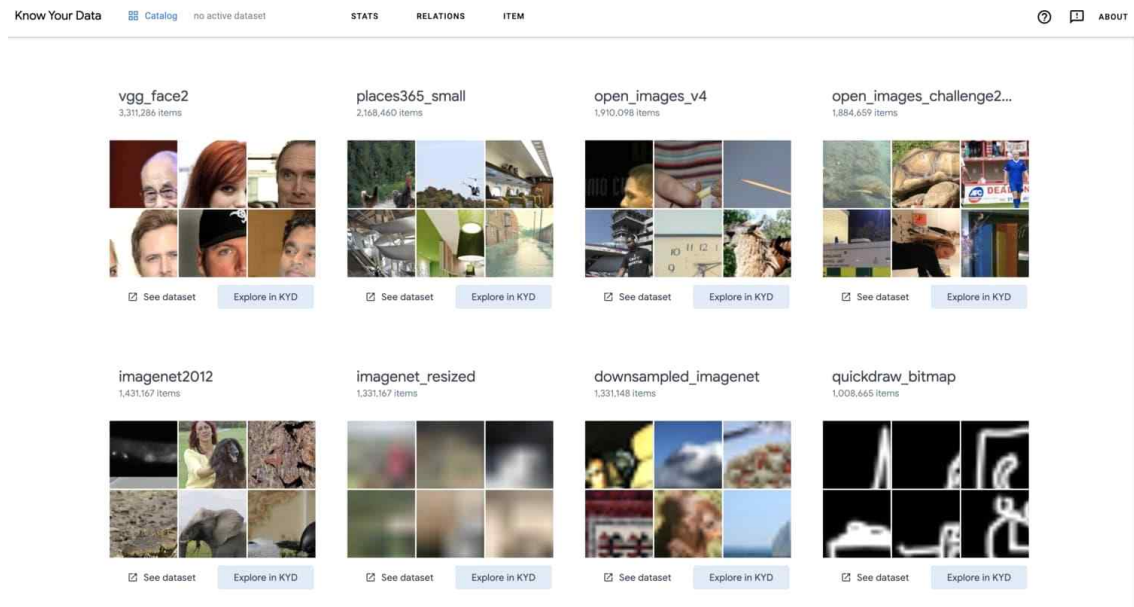
10) 전세계적으로 사용되고 있는 데이터셋을 정리 분석해놓은 책으로는 반병현, 142가지 데이터셋, 생능북스, 2022를 참조

11) ai-hub.or.kr에서 직접 데이터셋을 다운로드 받아 사용할 수 있다. AI허브에 대한 활용가이드나 현황에 대해서는 AI-Hub 활용가이드, 지능정보화진흥원, 2022, 15~16면 참조.

2) 텐서플로우 데이터셋¹²⁾

텐서플로우는 구글에서 제공하는 사이트로 인공지능 개발 SW는 물론 데이터셋을 같이 공개함으로써, 누구나 AI관련 서비스를 개발할 수 있도록 하고 있다.

[그림 1-8] 텐서플로우 데이터셋



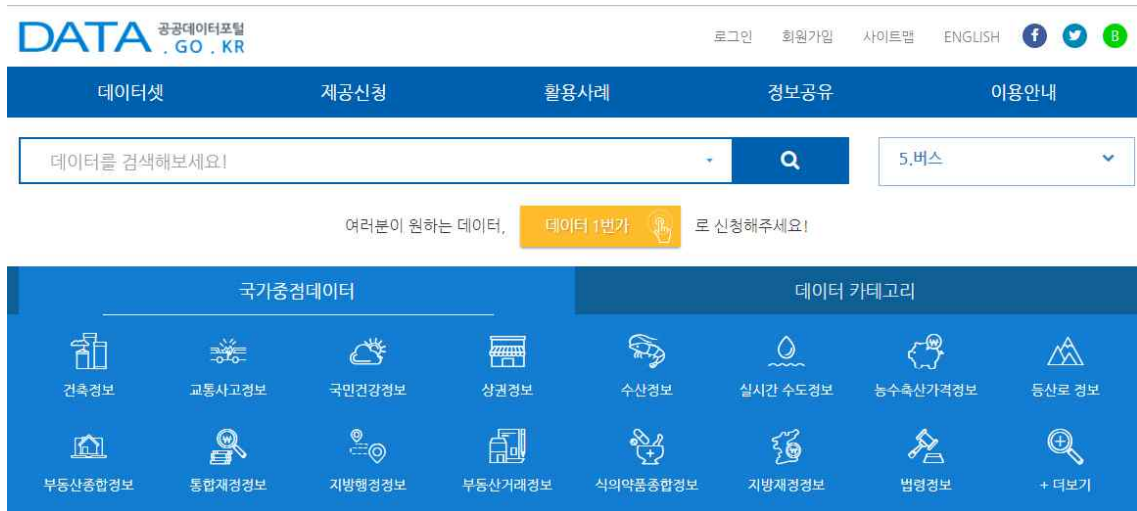
<출처 : 구글>

3) 공공데이터 포털

공공데이터의 제공 및 이용 활성화에 관한 법률에 의해서 구축된 공공데이터를 제공하는 사이트이다. 정부 및 공공기관에서 보유하고 있는 데이터를 민간에 개방함으로써 누구나 이용할 수 있도록 하고 있다. 참고로, 동 법에서는 공공데이터를 데이터베이스, 전자화된 파일 등 공공기관이 법령 등에서 정하는 목적을 위하여 생성 또는 취득하여 관리하고 있는 광(光) 또는 전자적 방식으로 처리된 자료 또는 정보로서 「전자정부법」 제2조 제6호에 따른 행정정보, 「지능정보화 기본법」 제2조 제1호에 따른 정보 중 공공기관이 생산한 정보, 「공공기록물 관리에 관한 법률」 제20조 제1항에 따른 전자기록물 중 대통령령으로 정하는 전자기록물, 그 밖에 대통령령으로 정하는 자료 또는 정보 중 어느 하나에 해당하는 것을 말한다.

12) 텐서플로우 tf.data.Dataset 사용 방법은 아래 링크에서 확인할 수 있다.
<https://hiseon.me/data-analytics/tensorflow/tensorflow-dataset/> <2022.11.26. 방문>

[그림 1-9] 공공데이터포털



<출처 : 공공데이터포털>

공공데이터 포털은 오픈 API방식으로 제공하며, 이용자가 공공데이터 포털에 API를 요청을 하고 제공받은 API를 통하여 데이터를 이전받게 된다.

4) 구글 데이터셋 검색엔진

구글이 제공하는 데이터셋에 특화된 검색 서비스로서, 공개된 데이터셋 플랫폼에서 관련된 내용을 검색해준다.

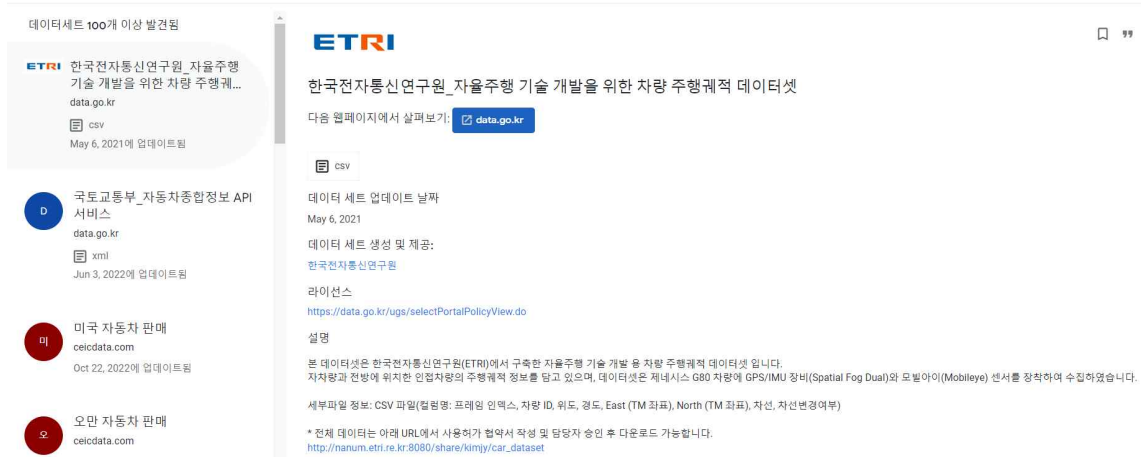
[그림 1-10] 구글 데이터셋 검색 화면



<출처 : 구글>

해당 사이트(datasetsearch.research.google.com)에서 특정 키워드를 입력하면 관련 데이터셋을 제공하는 사이트에 연동된다. 예를 들면, 구글데이터셋 검색엔진에서 ‘자동차 관련 데이터셋’을 입력할 경우 다음과 같이 제공 사이트를 보여준다.

[그림 1-11] 구글 데이터셋 검색 결과



<출처 : 공공데이터포털>

구글 데이터셋 검색 결과에서 하단부의 data.go.kr을 클릭하면 공공데이터 포털을 통하여 실제 데이터셋을 다운로드 받아 이용할 수 있다. 아래 검색 결과는 전자통신연구원(ETRI)에서 제공하는 자율주행용 데이터셋에 해당하는 정보 현황은 다음과 같다. 공공데이터포털을 통하여 제공하는 데이터셋을 확인할 수 있다.

[그림 1-12] 공공데이터 포털 세부 화면

파일데이터명	한국전자통신연구원_자율주행 기술 개발을 위한 차량 주행궤적 데이터셋_20191204		
분류체계	과학기술 - 과학기술진흥	제공기관	한국전자통신연구원
관리부서명	자율주행지능연구실	관리부서 전화번호	042-860-5157
보유근거		수집방법	
업데이트 주기	수시 (1회성 데이터)	차기 등록 예정일	
매체유형	텍스트	전체 행	5
확장자	CSV	키워드	정보/통신,정보이론,인공지능
데이터 한계		다운로드(바로가기)	1902
등록일	2019-12-04	수정일	2021-05-06
제공형태	공공데이터포털에서 다운로드(원문파일등록)		
설명	<p>본 데이터셋은 한국전자통신연구원(ETRI)에서 구축한 자율주행 기술 개발 용 차량 주행궤적 데이터셋입니다. 자차량과 전방에 위치한 인접차량의 주행궤적 정보를 담고 있으며, 데이터셋은 제네시스 G80 차량에 GPS/IMU 장비(Spatial Fog Dual)와 모빌아이(Mobility e) 센서를 장착하여 수집하였습니다.</p> <p>세부파일 정보: CSV 파일(컬럼명: 프레임 인덱스, 차량 ID, 위도, 경도, East (TM 좌표), North (TM 좌표), 차선, 차선변경여부)</p> <p>* 전체 데이터는 아래 URL에서 사용자가 협약서 작성 및 담당자 승인 후 다운로드 가능합니다. http://nanum.etri.re.kr:8080/share/kimjy/car_dataset</p>		

<출처 : 공공데이터포털>

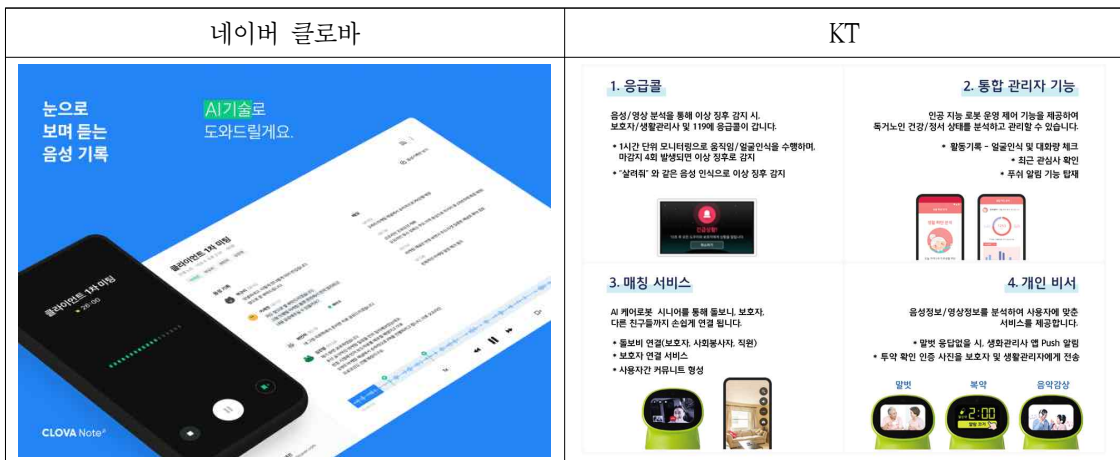
위 그림에서는 자율주행차 개발을 위한 차량 주행궤적 데이터셋에 관한 사항으로,¹³⁾ 데이터셋

을 제공하는 제공기관, 매체유형, 확장자인 파일형태, CSV에 포함된 세부적인 항목 등 제공되는 데이터셋과 관련된 공개된 정보에 대해 확인할 수 있다.

(3) 공개에 따른 효과

데이터를 공개하는 것은 인공지능 기술의 발전을 위한 것이라는 점은 부인하기 어렵다. 수많은 연구나 개발과정에서 데이터셋을 생성하거나 찾는 것은 어려운 일이다. 따라서, 이러한 과정에서 수집하는 리소스를 줄일 수 있다면 본질적인 기술투자를 이끌어 낼 수 있다는 점에서 데이터 공개는 정책적인 측면에서 이루어졌다. 국내 대표적인 공공데이터 포털이나 AI허브를 통해서 공개되는 데이터셋은 다양한 기업에서 활용하고 있기 때문에 데이터 개방을 통하여 얻는 이익은 크다고 할 것이다.

[그림 1-13] AI허브 데이터셋 활용 사례



<출처 : 각 사이트(2022)>

위 그림 좌측은 네이버의 음성기록 앱인 클로바노트는 한국어 대화 요약 데이터(' 20년 구축)를 활용하여 제작된 것으로, 네이버의 음성인식 기술을 기반으로, 회의, 강의 등 녹음한 대화를 텍스트로 볼 수 있는 AI 서비스이다. 위 그림 우측은 KT에서 제공하는 시니어 돌봄 서비스로 한국어 방언 발화(전라도) 데이터(' 20년 구축)를 활용하여 제작된 것으로, 고령층, 기저질환자 등 취약계층의 정서적 안정 도모 및 고독사 예방을 위한 AI 스피커로 개발된 것이다. 이와 같이 공개된 데이터셋은 누구나 자유롭게 이용할 수 있기 때문에 다양한 서비스의 개발을 이끌어낼 수

13) 참고로, 테슬라의 경우는 전 세계의 도로를 주행하고 있는 120만대 이상의 테슬라 차량으로부터 방대한 양의 데이터를 수집하고 있다. 2021년 1월까지 수집된 주행 데이터는 82억 킬로미터에 달한다고 한다. 무엇보다, 동영상 데이터가 100만개 이상이고, 깊이, 속도값 등을 라벨링한 데이터가 60억 개 이상에 달한다고 한다. SuperbAI, 한국기업들은 Data Centric AI에 얼마나 준비되어 있을까?, 2021, 9면.

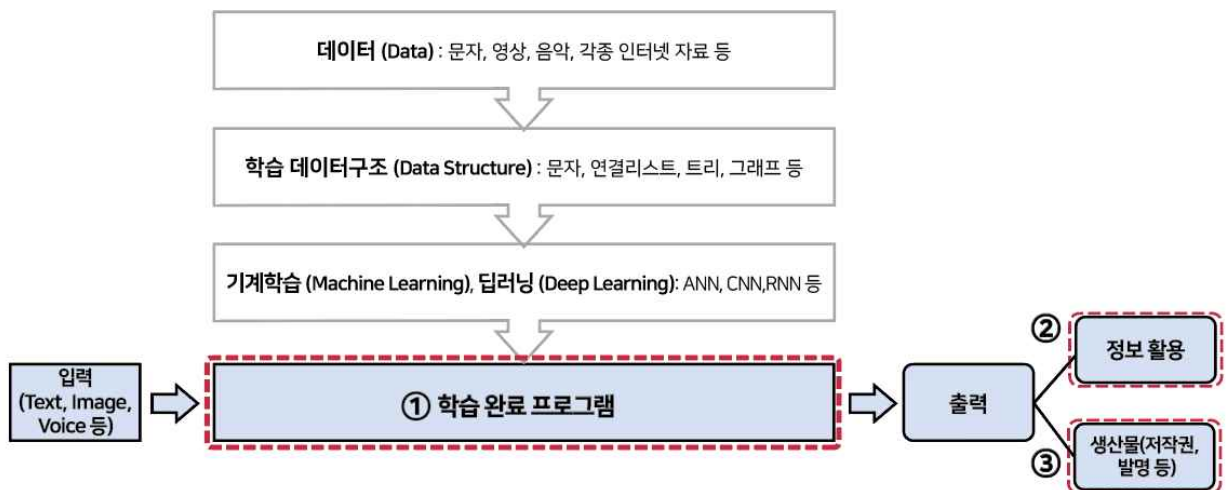
있다는 장점을 갖는다.

이처럼, 공개된 데이터셋을 활용하기 때문에 제3자가 용이하게 해당 서비스에 대해 검증할 수 있다. 누군가가 데이터를 이용하여 왜곡된 결과를 만들어낼 경우에 공개된 데이터셋을 활용할 경우에는 오류나 문제점을 상대적으로 용이하게 파악할 수 있다는 점에서 데이터의 신뢰성이나 인공지능의 신뢰성을 확보할 수 있다. 물론, 발명특허에서도 구현가능성을 높일 수 있다.

다. 데이터 공개와 학습데이터 활성화를 위한 논의

AI발명은 물론 일반적인 AI기술에서도 데이터의 중요성은 크다. 데이터는 AI기술과 학습된 AI 모델에 있어서도 중요한 요소이다. 특히, AI발명에 있어서 데이터는 발명의 구성요소로서 포함된다는 점에서 데이터의 중요성을 보여준다. 다음 그림은 데이터 학습이 이루어진 학습모델의 과정을 보여주는 것으로, 데이터가 어떠한 상태로서 학습되는지를 보여준다.

[그림 1-14] 데이터 학습이 이루어진 AI모델¹⁴⁾



<출처 : 특허와 상표, 2021>

AI발명은 특정 과제를 수행하기 위하여, i) 문자, 영상, 음악, 각종 인터넷 자료 등의 데이터를 수집하고, ii) 데이터 사이의 관계를 조직적·체계적으로 구분하기 위한 학습 데이터구조를 구축하며, iii) 학습 데이터구조에 적합한 알고리즘을 적용하여 반복적으로 기계학습을 함으로써 학습 모델이 완료된 컴퓨터 프로그램이다.¹⁵⁾ 이러한 절차에 따라 AI 학습이 이루어진 프로그램이나 모델은 학습모델 또는 AI모델이라고 한다. 데이터 자체의 보호는 저작권법이나 부정경쟁방지법으로 가능하다. 다만, 데이터가 구조화되는 경우에는 기술적 사상의 구현이라는 점에서 발명의

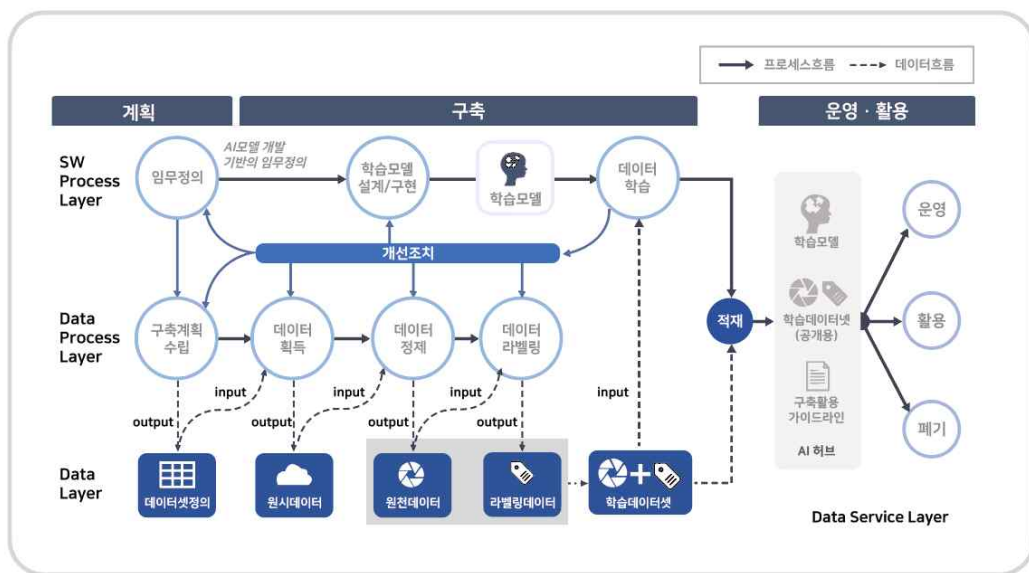
14) 이해진, 특허와 상표, 2021.2.24.일자.

15) 이해진, 특허와 상표, 2021.2.24.일자.

로써 성립할 수 있다. 따라서, 학습데이터의 집합이라고 할 수 있는 데이터셋(data set)은 발명의 요소가 되거나 또는 발명 자체의 대상이 될 수 있다. 이러한 방식은 학습데이터에 대한 특허법적 관점에서 보호방안을 찾고자하는 것으로, 정책 및 법률전문가들이 갖는 데이터에 대한 오해는 데이터셋에는 기술적 사상이 부가된 것으로 이해하지 않는다는 점이다. 데이터셋은 학습과 테스트 용으로 구분할 수 있지만, 실상은 학습, 검증, 테스트용 등 3가지로 구분하며, 특히 검증 데이터셋은 하이퍼 파라미터를 결정하는데 쓰인다.¹⁶⁾

실상 사실(fact)이 아닌 데이터는 데이터의 집합만이 아닌 데이터를 학습용으로 가공하는 과정에서 구조화되며 이는 알고리즘 내지 SW로서 데이터셋이라는 점을 이해할 필요가 있다. 이러한 특성에 따라, 데이터의 성질이 학습용 데이터셋으로 바뀌는 경우에는 특허법적 접근 가능성이 있는지 생각해볼 필요가 있다.¹⁷⁾ 데이터의 수집과 처리에 따라, 데이터 생애주기에 따른 절차와 기계학습 과정을 도시하면 다음과 같다.

[그림 1-15] 인공지능 학습용 데이터 생애주기



<출처 : NIA(2021)>

데이터의 생애주기에 따른 보호에 대한 고민은 AI발명의 보호를 넘어서, 좀더 구체적으로는 데이터의 중요성에 비추어 볼 때 데이터의 구조 또는 구조화된 데이터를 넘어서 데이터 또는 데이터셋의 보호와 맞물려 충분한 기술공개(sufficiency of disclosure)와도 밀접한 관련이 있다. 무엇보다 기술공개는 해당 발명에 대한 인센티브로서 독점권을 부여함으로써 이루어지는 것이기 때

16) 조성준 외, 빅데이터 커리어 가이드북, 길벗, 2022, 259~260.

17) 이규호, 인공지능 학습용 데이터셋 보호를 위한 특허법상 주요 쟁점 연구, 산업재산권 제64호, 2020 참고.

문에 실질적인 공개를 통하여 제3자가 공개된 기술내용을 통하여 발명을 쉽게 실시할 수 있을 정도이어야 한다. 그렇지만, 충분한 공개에 대해서는 기술적 특성에 따라 특허법이나 심사지침에서 개별적인 사안에 대해 명확하게 기준을 제시하기는 어렵다. 이처럼, 기술공개는 법정 요건이기는 하지만 실질적으로 해당 기술의 충분한 공개여부를 판단하는 입장은 평균적인 기술자로서 하겠지만, 발명의 특성이나 요소기술의 수준에 따라 달라질 수 있는 사안이다.

그럼에도 불구하고, 인공지능 발명을 구성하는 다양한 요소 중에서 데이터가 차지하는 비중이 커지고 있으며 중요부를 차지하기 때문에 AI발명에 있어서 사용되는 데이터 구조 내지 데이터셋을 공개하거나 관련 기술의 공개는 중요한 일이 아닐 수 없다. 그렇지만, AI발명의 기술공개와 더 나아가 데이터 공개에 대해서는 별다른 관심을 두고 있지는 않다.¹⁸⁾

본 연구에서는 이를 위하여 IP5¹⁹⁾를 중심으로 AI발명의 성립성과 이를 판단하는 과정에서 이루어지는 데이터관련 기술적 사항이 어떻게 구현되는지, 심사지침이나 가이드에서는 어떠한 기재요건을 제시하고있는 지를 살피는 것은 중요한 단초가 될 것이다. 다만, IP5는 특허정책의 일관성을 위하여 어느 정도 수준을 맞추고 있다. 따라서, 기술적 특성에 따른 제도에 대한 논의는 어느 일국에서 선도할 수 있는 것은 아니라고 본다. IP5가 공통된 주제와 수준을 가지고 논의를 통하여 일관된 기준을 수립하는 것이 합리적이라고 보기 때문이다. 이를 통해서 확립된 기준이 WIPO나 WTO를 통하여 글로벌 환경에 적용되는 것이 바람직하다. 다행히 2020년부터 WIPO에서는 AI발명에 대한 기술공개 및 이와 관련하여 학습데이터의 기탁 등에 대한 이슈에 대해 대응하고 있다.

3. 본 연구의 주요 내용

본 연구에서 다루는 주된 내용은 데이터의 보호 및 데이터 공개라는 점이다. 데이터 공개는 인공지능이 갖는 내재적 한계인 블랙박스(black box)²⁰⁾에 따른 투명성 확보, 그에 따른 사회적 후생 등에서도 유의미한 결과를 가져올 것으로 기대된다. 이러한 관점에서 AI발명에서 사용되는 학습데이터의 구조화된 기술적 사항은 물론 데이터셋이 갖는 논리적 구조로서 기술적 사항은 알고리즘으로 볼 수 있다는 점에 주안점을 두고자 한다. 이러한 논리적인 구성은 실질적으로 데이터셋이 갖는 정책적인 보호의 필요성을 넘어 발명에서 주요 구성요소를 이룬다는 점에서 고찰이 필요하다고 보기 때문이다. 각 장별로 살펴보고자하는 주요 내용은 다음과 같다.

18) Tabrez Ebrahim, Artificial Intelligence Inventions & Patent Disclosure, Iowa Legal Studies Research Paper No. 2021-48, p.148; Vijay/Devesh et al., AI and Indian Patent Law-Sufficiency of Disclosure for Artificial Intelligence-based Patents, International Journal of Mechanical Engineering, Vol. 7 (Special Issue, Jan.-Feb. 2022), p.341.

19) 대한민국, 미국, EU, 중국, 일본 등 5개국의 특허청의 협의체를 말한다.

20) 기계학습으로 구축된 심층 신경망에 의사결정과정을 묻는다면, 활물적 판단 과정을 설명해야할 것이다. 심층 신경망의 경우 활물적 판단이 여러번 중첩되었기 때문에 그 판단 수식을 설명해도 사람이 이해할 수 없다. 그런 의미에서 블랙박스 시스템이라고 한다. 김진형, AI최강의 수업, 매일경제신문사, 2020, 274면.

제2장에서는 <현행법 상 AI 학습데이터의 보호 체계>에 대해 연구하였다. 인공지능의 학습을 위한 데이터 관련 정책을 위해 다양한 법제가 만들어졌다. 데이터의 법적인 정의에서부터, 관련 입법 현황을 정리하고자 한다. 아울러, 데이터의 특성 및 구조화와 관련된 내용을 살펴볼 필요가 있다. 특히, AI발명과 SW발명 등 일반발명과의 차이 등에 대해 살펴보고, 데이터셋에 대한 특허 법적 보호 필요성에 대해 살펴보았다.

제3장에서는 <AI발명에 대한 특허법상 기술공개 범위>에 대해 연구하였다. 데이터 처리가 주요 구성요소를 이루는 AI발명의 기술공개 방법, 범위에 대해 살펴보았다. 각국의 특허법에서는 발명의 공개를 법정화하고 있다. 이는 기술공개에 따른 인센티브로써 특허권이라는 독점권을 부여하는 것이기 때문이다. 이러한 점에서 기술공개, 특히 인공지능 발명에서 데이터 등의 기술공개 수준을 어느정도 까지 할 것인지는 주요한 과제이다. 이를 위해 각국에서 AI발명과 데이터 발명에 대해 어떠한 정책을 가지고 있는지에 대해 검토하였다. 다만, 데이터 공개는 발명자나 기업의 입장에서 영업비밀로 유지되는 경우에 있어서 정책적인 결정사항이기도 하기 때문에 자발적인 공개 내지 발명의 상세한 설명 등의 방법으로 공개방안을 제안하고자 하였다.

제4장에서는 <학습데이터의 보화 및 활용을 위한 제도개선 기초연구>로서 데이터의 재현가능성과 신뢰가능성을 확보하기 위한 제도에 대해 조사하였다. 데이터가 사용되는 발명에 있어서 신뢰가능성과 재현가능성은 중요한 과제가 아닐 수 없다. 전자는 데이터에 대한 신뢰성을 비롯하여 발명 자체에 대한 신뢰성을 확보할 수 있는 방안을 찾는 것이다. 후자인 재현가능성은 공개된 발명을 통하여 평균적 기술자가 용이하게 실시할 수 있을 정도의 기술공개를 위한 요건으로서 데이터를 이용할 수 있는 방안을 찾고자 하는 것이다. 이를 위하여, 데이터를 공개하거나 제3의 기관에 기탁하는 방안을 고려할 수 있다. 다만, 권리발생 요건으로서 기탁이나 등록이 아닌 발명의 재현가능성을 위한 목적으로서 제한될 필요가 있다. 이를 위하여 NFT나 블록체인의 기술을 활용하는 방안도 고려해볼 수 있다.

제5장에서는 <AI 학습데이터를 활용한 데이터산업 활성화 방안>에 대해 연구하였다. 실상 데이터를 생산하거나 이용하는 산업계에서는 데이터의 활용에 대한 방점을 찍는 경우가 많다. 데이터 확보에 따른 비용이 소요되기 때문이다. 이러한 이유로 정부에서는 데이터댐 또는 데이터레이크(data lake) 정책을 펼치면서 다양한 학습용 데이터셋을 구축하고 있다. 이러한 데이터셋은 AI-HUB를 통하여 일반 국민에게 공개되고 있는 데이터는 인공지능 학습에서 중요하게 활용되고 있다. 이러한 공개된 데이터셋을 활용하는 것은 경쟁력을 갖출 수 있는 방안이 된다는 점에서 중요한 시사점을 갖는다.

제2장. 현행법 상 AI 학습데이터의 보호 체계

제1절 AI발명과 학습데이터

1. 데이터의 법적 정의

가. 데이터의 정의

데이터는 정형화되어 데이터베이스(DB)에 담겨진 정보라기 보다는 수집된 원정보(raw data)로 이해될 수 있다. 원정보는 데이터 필드로 나누어진 것이 아닌 전체가 하나의 단위로 비정형화된 경우가 많다. 특히, IoT를 통하여 획득되는 산업데이터는 그 자체로 가치를 갖기보다는 목적에 맞게 가공하는 절차를 거치면서 정형화되어 간다. 따라서, 원데이터에 대한 가공, 분석 등의 절차를 거치면서 용도가 정해지게 되므로, 개발된 데이터란 이와 같은 절차를 거친 일종의 목적에 따라 가공된 데이터로 이해할 수 있다.

개발된 데이터에 대한 법적 논의는 한 건의 데이터에 대한 것이라기 보다는 데이터셋이나 데이터베이스 형태에 대한 것이기 때문에 사실상 데이터 관련 법제라기 보다는 데이터베이스에 대한 보호 가능성을 어떻게 정책적으로 가져갈 것인지에 방점을 두게 된다. 무엇보다, 데이터 관련 법제의 기본은 저작권법상 데이터베이스인지 여부 및 그에 따른 데이터베이스제작자의 권리와 관련성이 크다. 다만, 우리 법제는 저작권법 이외에 콘텐츠산업 진흥법에서 콘텐츠를 개발한 사업자에 대한 법적 보호도 아울러 규정하고 있기 때문에 예외적으로, 콘텐츠산업 진흥법에 따른 보호를 받게된다. 최근 부정경쟁방지법이 개정되면서 데이터의 무단이용에 대한 부정경쟁 법리가 도입되는 등 다양한 정책목표를 달성하기 위한 데이터 법제의 개정 또는 제정이 이루어지고 있다.

본 연구도 데이터 및 데이터의 집합체인 데이터베이스를 전제하는 개념으로써 개발된 데이터를 상정한다. 물론, 개별 데이터에 대한 보호 체계의 논의도 필요하나 데이터의 보호 논의에서 데이터는 사실상의 정보에 해당할 가능성이 높기 때문에 그 자체의 보호보다는 가공 또는 대규모의 비정형 데이터에 대한 보호 가능성을 어떻게 찾아갈 것인지에 대한 문제의식에 가깝다.

나. 데이터의 보호 체계

데이터 보호의 기본 원칙은 비정형 데이터 또는 정형화된 데이터에 대하여 물권적인 권리로서 접근할 것인지, 아니면 채권적인 권리로서 접근할 것인지에 대한 논의이다. 물론, 데이터를 가공하면서 인적 또는 물적인 상당한 투자가 이루어질 가능성이 있다.

데이터 관련 법제에 따른 보호 체계에 대하여는 저작권법 체계와 부정경쟁방지법 체계로 나누어볼 수 있다. 전자는 데이터셋 또는 데이터베이스 제작자에게 저작권 내지 물권 유사의 권리를

부여하는 것이다. 후자는 데이터셋을 무단으로 이용하는 자에 대한 채권적 권리를 행사할 수 있도록 하는 것이다. 대표적으로 금지청구권이나 손해배상청구권을 행사하게 된다.

현행 법제하에서는 저작권법이 물권적 권리를 부여하고 있으며, 부정경쟁방지법이 채권적 권리를 부여하고 있다. 여기에 더하여, 콘텐츠산업 진흥법도 부정경쟁방지법과 동일한 법리에 따른 보호가 가능하다. 다만, 콘텐츠산업 진흥법에서는 저작권이 있는 콘텐츠에 대해서는 저작권법이 우선 적용된다는 우선순위에 대한 규정을 두고 있으나, 부정경쟁방지법상 우선순위에 관하여는 별도 입법정비를 하지 않았기 때문에 이에 대한 법체계 정비가 필요한 상황이다.

최근 제정된 산업 디지털전환 촉진법에서도 산업데이터의 보호에 대해 부정경쟁법리를 차용하고 있다. 다 같은 데이터인데도 불구하고 소관하는 법률에 따른 적용법리 또는 보호요건이 다르게 적용되는 것이 합리적인 산업정책인지는 의문이다. 왜냐하면, 지식재산 법제와 데이터보호 법제는 정책적인 목표가 상이하기 때문에 그동안 보호받지 못하던 데이터의 보호범위가 넓어지는 경향에 서로 다른 법리나 정책목표에 따라 지식재산법제가 추구했던 퍼블릭도메인(public domain)의 영역에 대한 자유로운 이용이 제한되는 결과를 가져오기 때문이다. 산업정책은 산업상 필요성을 목적으로 하므로, 지식재산에 적용되던 법리가 상대적으로 밀려날 수 있음을 보여준다.

2. AI발명과 학습데이터의 특성

가. AI발명의 구성

일반적으로 AI발명을 구성하는 AI시스템은 (1) 데이터 수집 기술과 특정 목적 또는 개선을 위해 어떤 종류의 데이터를 수집해야 하는지, (2) 데이터를 모델에 공급하는 전처리 단계와 특징 추출, (3) 학습 방법과 반복적인 입력을 모델에 공급하는 과정, (4) 모델 그 자체, (5) 기존 모델의 사용자 정의 및 개선, (6) 모델 출력의 후처리 및 해석 등 세부적인 내용을 담고 있다. 따라서 AI발명은 언급된 부분 중 하나 또는 많은 부분을 포함할 수 있다.²¹⁾ 이러한 구성요소에 따른 인공지능은 알고리즘, 데이터, HW 등으로 정리할 수 있다. 데이터나 알고리즘 측면에서 보면 AI발명은 수집 또는 가공된 원천데이터, 학습데이터 구조, 딥러닝의 인공신경망 구조, 학습전 파라미터 알고리즘, 학습용프로그램 및 알고리즘, 학습완료모델, 학습후 파라미터 알고리즘, 추론프로그램, 응용 입력데이터, 산출물 등으로 구현된 수단 또는 단계로 구성된다.²²⁾

AI발명의 내부적인 처리는 인공신경망 구조에서 이루어지기 때문에 이를 확인할 수 있는 방법은 없으며, 파라미터²³⁾ 또한 마찬가지이다. 입출력데이터의 경우에는 학습된 AI모델에 입력하여

21) Medhi Poursoltani, Disclosing AI Inventions, TEXAS INTELLECTUAL PROPERTY LAW JOURNAL Vol.29, 2021, pp. 49~50.; Oleksii Kharkovyna, Machine Learning vs Traditional Programming, towards data science (Apr. 22, 2019), <https://towardsdatascience.com/machine-learning-vs-traditionalprogramming-c066e39b5b17>.

22) 권지현, 비기술적요소가 포함된 AI발명의 진보성 판단기준, 법학논총 제48집, 2020, 35면.

산출물을 생성하는 것이라는 점에서 이용자가 제공하거나 그에 따른 결과라는 점에서 어느 정도 확인이 가능할 수 있다. 이러한 이유 때문에 AI의 투명성에 대한 논의가 이루어지고 있으며, 특허법에서 A발명에 있어서 I알고리즘이나 데이터를 공개하려는 이유 중 하나가 될 수 있다. 물론, 특허법의 가장 큰 이유는 특허권이라는 독점권을 부여하는 기술공개에 따른 조건으로 부여하는 것이지만 인공지능 발명이 현실에서 사용될 경우에 나타날 수 있는 투명성이나 설명가능성 이슈 및 그에 따른 법적인 문제를 차단할 수 있는 방법으로도 활용될 수 있다는 점에서 AI발명의 공개의 의의를 찾을 수 있다.²⁴⁾

나. AI발명의 특성과 차이

(1) 발명의 성립

발명이란 자연법칙을 이용한 기술적 사상의 창작으로서 고도(高度)한 것을 말한다. 자연법칙을 이용하지 않을 경우에는 발명으로서 성립하지 아니한다. 각국의 발명에 대한 정의나 규정 방식은 차이가 있으나 대체로 자연법칙을 이용할 것을 요건으로 하고 있다. 인공지능 발명을 포함하여 소프트웨어(SW)나 비즈니스모델 특허 등 자연법칙을 이용하지 않을 경우에는 발명의 성립성 요건이 충족되지 않는다. 이러한 요건에 따라 각국의 특허제도는 운용되고 있다. 우리나라에서 AI 관련 발명이 특허대상이 되기 위해서는 특허법 제2조 제1호에서 규정한 자연법칙을 이용한 기술적 사상의 창작으로 고도한 것이라는 발명의 성립 요건을 만족해야 하고, 만약 발명의 성립성 요건을 만족하지 않는 경우에는 산업상 이용할 수 있는 발명이 아니므로 특허법 제29조 제1항 본문의 규정에 의하여 특허를 받을 수 없다. 또 미국에서는 미국특허법 제101조의 사법적요건과 판례법에 의한 사법적 예외요건이 제시한 특허적격성 요건을 만족하여야 한다. 일본에서는 우리와 같은 내용의 발명의 정의 규정을 두고 있는 특허법 제2조 제1호에서 규정한 발명해당성 요건을 만족해야 한다. EU 특허조약(European Patent Convention : EPC)에는 발명의 정의규정이 없지만, EPC 제52조 제1항 및 제2항에서 규정하는 불특허 대상에 포함되지 않아야 한다는 특허적격성 요건을 만족하여야 한다.²⁵⁾ 중국도 EPC와 유사하게 불특허 대상의 예외 경우를 두고 있다.

23) 학습완료 파라미터란 단순한 수치행렬로서 표현되는 것이 통상적이고, 인간의 사상이나 감정이라는 주관적 요소를 포함하지 않는 수치의 나열이다. 그러므로, 인간의 사상이나 감정이라는 창작적 표현이 아니라서 구체적으로는 저작권법의 보호대상이 되지 않는다고 한다. 차상욱, 데이터법, 세창출판사, 2022, 215면.

24) 인공지능은 블랙박스(black box)로 불리운다. 그 누구도 그 처리과정을 이해할 수 없으며, 단순히 결과로서 추론만이 가능하기 때문이다. 인공지능의 투명성이 강조되는 주된 이유다. 알고리즘을 왜곡하거나 또는 편견(bias)이 들어갈 경우에는 그 결과의 공정성(fairness)을 담보하기 어렵다. 이 때문에 인공지능을 견제하기 위한 정책이나 법제도 이외에 다양한 기술적 기법들이 제시된다. 알고리즘 소스를 공개하거나 감사(auditing) 기법을 도입하는 것도 하나의 방안으로 논의되고 있다. 블랙박스로서 인공지능은 그 과정을 설명하지 못하기 때문에 설명가능한 인공지능(explainable AI)이 새로운 방안으로 제시되기도 한다. 김윤명, 블랙박스를 열기위한 인공지능법, 박영사, 2022, 29면.

25) 함영욱, AI 관련 발명의 성립성 판단기준에 관한 연구, 한양대학교 박사학위논문, 2022, 3면.

(2) AI발명의 특성

AI발명은 통상의 프로그램발명과 같이 소프트웨어 시스템 설계와 알고리즘은 동일하지만, AI가 반복학습을 할 수 있게 학습데이터 구조가 가공·구축되어야 하고, 그러한 학습데이터 구조를 가지고 학습하고 또 그 학습의 결과를 판단하여 문제점을 인지하고, 그 문제점을 해결하기 위하여 계속 반복학습을 수행하여 최적의 목표값을 컴퓨터 등 하드웨어에 의하여 도출된다는 반복학습 개념(아이디어로서의 알고리즘)이 구체적으로 제시되어야 한다.²⁶⁾

AI발명은 기본적으로 프로그램발명의 유형으로써 물건(장치)발명과 방법발명에 속하고, 그 객체는 하나의 프로그램 또는 시스템으로 상거래의 대상이 된다. AI가 적용된 제품은 프로그램의 유형으로써 자유롭게 유통되는 상거래의 성격을 가지고 있다 보니, 다른 특허발명에 비하여 권리 범위의 해석이 복잡하다. 특히 특허출원에 있어서는 발명의 설명에 관한 실시예에 블록도 및 흐름도를 근거로 AI 학습프로그램이 수행하는 학습방법을 구체적으로 기재하여야 한다. 또한 AI의 학습방법도 AI라는 기술적 사상이 주체이지만, 실제로 AI의 반복학습은 AI가 아니라 컴퓨터 등의 하드웨어에 의하여 수행된다는 것을 명확히 특허출원의 명세서에 기재하여야 한다.²⁷⁾ 이렇게 AI 발명은 SW와 HW를 기반으로 하면서도, 학습과정에서 데이터를 중심으로 이루어진다는 점에서 그 특징이 있으며, 내부적인 처리과정을 개발자도 알 수 없다는 점에서 블랙박스(black box)라고 칭하여진다.

(3) SW발명 등과의 차이

1) SW발명

일반적인 표현으로써 SW발명 또는 발명의 컴퓨터 구현(computer implemented invention)은 기술적 사상의 창작으로서 고도한 것을 말하는 발명을 컴퓨터로 구현하는 것을 말한다.²⁸⁾ SW발명에 대한 많은 논란에도 불구하고 많은 나라에서 SW발명은 인정되고 있는 것이 사실이다. 컴퓨터 프로그램 언어 자체는 물론 컴퓨터 프로그램은 컴퓨터를 실행하는 명령에 불과한 것으로 컴퓨터 프로그램 자체는 발명이 될 수 없다.²⁹⁾ 대법원도 “특허법 제2조 제1호는 자연법칙을 이용한 기술적 사상의 창작으로서 고도한 것을 발명으로 정의하고 있으므로, 출원발명이 자연법칙을 이용한 것이 아닌 때에는 같은 법 제29조 제1항 본문의 산업상 이용할 수 있는 발명의 요건을 충족

26) 권태복, 제4차 산업혁명과 특허전략, 한국지식재산연구원, 2019, 27면.

27) 권태복, 제4차 산업혁명과 특허전략, 한국지식재산연구원, 2019, 28~29면.

28) 김윤명, SW특허는 기술발전을 이끄는가?, 홍익법학 제15권 제4호, 2014, 836~837면.

29) “특허권은 그 보호대상을 기술적 사상(idea) 좀 더 구체적으로는 기술적 사상에 대한 응용(application)을 그 대상으로 하고 있고, 이에 반하여 저작권에서 사상은 보호대상이 되지 아니하며 사상의 표현(expression)을 그 보호대상으로 하고 있으므로, 보호대상에서 근본적인 차이점이 존재한다”고 한다. 김관식, 컴퓨터프로그램의 특허법상 보호에 관한 특허법 제2조 개정시안의 법적 의의 및 과제, 법학연구 제23권 제1호, 2012, 554면.

하지 못함을 이유로 그 특허출원을 거절하여야 한다. 특히, 정보기술을 이용하여 영업방법을 구현하는 이른바 영업방법(business method) 발명에 해당하기 위해서는 컴퓨터상에서 소프트웨어에 의한 정보처리가 하드웨어를 이용하여 구체적으로 실현되고 있어야 한다. 한편, 출원발명이 자연법칙을 이용한 것인지 여부는 청구항 전체로서 판단하여야 하므로, 청구항에 기재된 발명의 일부에 자연법칙을 이용하고 있는 부분이 있더라도 청구항 전체로서 자연법칙을 이용하고 있지 않다고 판단될 때에는 특허법상의 발명에 해당하지 않는다.”³⁰⁾고 판시한 바 있다. 이를 따른 특허청 심사지침에서도 “컴퓨터 프로그램에 의한 정보처리가 하드웨어를 이용해 구체적으로 실현되는 경우에는 해당 프로그램과 연동해 동작하는 정보처리장치(기계), 그 동작 방법 및 해당 프로그램을 기록한 컴퓨터로 읽을 수 있는 매체는 자연법칙을 이용한 기술적 사상³¹⁾의 창작으로서 발명에 해당한다”³²⁾고 규정하고 있다.

[표 2-1] 컴퓨터 관련 발명의 범주

방법 발명	컴퓨터를 사용한 발명이 시계열적으로 연결된 일련의 처리 또는 조작. 즉 단계로 표현할 수 있을 때, 그 단계로 특정된 방법	
물건 발명	물건	- 컴퓨터를 사용한 발명이 복수의 기능 요소로 표현할 수 있을 때 그 기능 요소로 특정된 물건
	매체	- 프로그램을 설치하고 실행하거나 유통하기 위해 사용되는 프로그램을 기록한 컴퓨터로 읽을 수 있는 매체 - 기록된 데이터 구조로부터 컴퓨터가 수행하는 처리 내용이 특정되는 구조를 가진 데이터를 기록한 기록 매체 - 하드웨어와 결합되어 특정과제를 해결하기 위하여 매체에 저장된 컴퓨터 프로그램

<출처 : 특허청, 2022>

2) 일반적인 발명

SW 자체의 발명의 성립성에 대해서는 자연법칙의 이용가능성에 대한 논란에서 시작한다. 발명이란 자연법칙을 이용한 기술적 사상의 창작으로서 고도한 것을 말하기 때문에 기본적으로 자연법칙을 이용한 것인지의 판단이 우선시 된다.³³⁾ 이는 쾰러(Kohler)에 의해서 정의된 것으로, 일본의 특허법 및 이를 계수한 우리나라의 특허법에 정의규정으로 남아있다.³⁴⁾ 고도한 것을 요건으

30) 대법원 2008. 12. 24. 선고 2007후265 판결.

31) 기술적 사상이 존재하지 않는 가운데 데이터, 수학적 공식, 추상적 사업방식 그 자체만으로는 아무리 컴퓨터 기록매체에 저장된 형태로 출원되었다 하더라도 발명에 해당한다고 볼 수 없다고 한다. 정상조·박준석, 지식재산권법, 홍문사, 2013, 83면.

32) 특허청, 특허·실용신안 심사기준, 2014, 3103-3104면.

33) 특허법 제2조 제1호는 자연법칙을 이용한 기술적 사상의 창작으로서 고도한 것을 “발명”으로 정의하고 있고, 위 특허법 제2조 제1호가 훈시적인 규정에 해당한다고 볼 아무런 근거가 없으므로, 자연법칙을 이용하지 않은 것을 특허출원하였을 때에는 특허법 제29조 제1항 본문의 ‘산업상 이용할 수 있는 발명’의 요건을 충족하지 못함을 이유로 특허법 제62조에 의하여 그 특허출원이 거절된다. 대법원 1998. 9. 4. 선고 98후744 판결.

34) “Kohler는 특허권을 국가의 행정행위에 의한 행정법상의 독점권으로 보던 과거 시각에서 벗어나

로 하는 것은 실용신안의 고안과의 차별성을 두기 위한 것이다.³⁵⁾

SW특허가 독립적으로 운용될 경우에 프로그램 자체가 지시·명령의 집합이라는 점에서 자연법칙을 이용한 것인지에 대한 발명의 성립성 요건에 대한 논란이 있다.³⁶⁾ 특히 1998년 심사지침에서는 매체에 고정된 SW특허를 인정함으로써 실질적으로 컴퓨터프로그램 분야에서의 자연법칙의 이용이라는 요건은 거의 유명무실해졌다고 평가된다. 그러나, Bilski 판결이나 Alice 판결에서 볼 때 미연방 대법원은 발명의 성립성에 대해 엄격하게 판단하고 있으며 단순한 추상적 아이디어 자체의 보호는 부정하고 있는 것으로 판단된다. 이러한 점에서 볼 때 미국 판례가 특허법의 초기 형태로 회귀한다는 극단적인 주장은 합리성이 떨어지지만, 미국 특허청과 사법부의 특허요건을 완화하는 지침과 판단과는 사뭇 다른 결이다. 우리 대법원은 “구 특허법 제2조 제1호는 자연법칙을 이용한 기술적 사상의 창작으로서 고도한 것을 ‘발명’으로 정의하고 있으므로, 출원발명이 자연법칙을 이용한 것이 아닌 때에는 구 특허법 제29조 제1항 본문의 ‘산업상 이용할 수 있는 발명’의 요건을 충족하지 못함을 이유로 그 특허출원이 거절되어야 하는 바 특히 정보 기술을 이용하여 영업방법을 구현하는 이른바 영업방법(business method) 발명에 해당하기 위해서는 컴퓨터상에서 SW에 의한 정보처리가 하드웨어를 이용하여 구체적으로 실현되고 있어야 한다”³⁷⁾고 판시하는 한편 “출원발명이 자연법칙을 이용한 것인지 여부는 청구항 전체로서 판단하여야 하므로 청구항에 기재된 발명의 일부에 자연법칙을 이용하고 있는 부분이 있더라도 청구항 전체로서 자연법칙을 이용하고 있지 않다고 판단될 때에는 특허법상의 발명에 해당하지 않는다”³⁸⁾고 보고 있다.

3) 정리

기본적인 SW발명과 구별되는 AI발명의 특징은 “AI 관련 발명은 기계학습이 필수적이다. 일반

처음으로 사권으로 분명하게 파악함으로써 독일에서 근대 특허법학의 초석을 놓은 학자였는 바, 그는 발명을 자연력을 이용하여 자연을 제어함으로써 기능적 효과를 야기하고 그로써 인간의 요구를 충족시키는 것으로, 기술적으로 표현된 사상의 창작이라고 정의하였다”고 한다. 박준석, 우리 특허법상 ‘발명’의 개념에 관한 고찰, 서울대학교 법학 제54권 제3호, 2013.9, 801면. 그러나 독일 특허법에서는 쾰러의 정의규정은 채용된 바 없다.

35) 실용신안법이 정하는 실용적 고안이라 함은 물품의 형상 구조 또는 조합에 관한 자연법칙을 이용한 기술적 사상의 창작으로서 특허법이 정하는 자연을 정복하고 자연력을 이용하여 일정한 효과를 창출하고 이에 따라 인간의 수요를 충족하는 기술적 사상의 고도의 창작인 발명과 그 성질에서 같으나 다만 고도의 것이 아닌 점에서 다를 뿐이다(실용신안법 제3조 및 특허법 제5조 참조). 따라서 실용신안법에 의하여 장려, 보호, 육성되는 실용신안은 물품의 특수한 형상에 그치는 것이 아니라 그 실용성 즉 실용적 가치 나아가 그 기술적 고안이 대상이 되는 것이며 기술적 사상의 창작으로서 그 작용효과가 등록의 적부를 가리는 주요기준이 되는 것이라고 풀이할 것이다. 대법원 1983.11.22. 선고 83후42 판결[거절사정].

36) “컴퓨터프로그램이란 컴퓨터로 하여금 어떠한 작업을 어떠한 방법으로 수행하라고 지시하는 명령어들의 집합에 지나지 않을 뿐이고, 그 자체로 어떠한 자연법칙을 이용하는 것이 아니기 때문이다. 컴퓨터 프로그램이 특허법상 발명으로 보호받을 수 있는지를 둘러싼 미국 판례의 변화는 이에 관한 인식의 변천을 잘 반영해준다”. 조영선, 특허법 2.0, 박영사, 2018, 15면.

37) 대법원 2003. 5. 16. 선고 2001후3149 판결 등 참조.

38) 대법원 2008.12.24. 선고 2007후265 판결[거절결정(특)].

적인 SW발명은 코딩의 완료로써 SW발명으로 성립할 수 있지만, AI 관련 발명은 학습용 데이터 구조를 근거로 인공지능기술 기반 딥러닝 학습을 통하여 학습완료모델을 도출하는 것이 필수적으로 요구된다”³⁹⁾는 점이다. 결국, 일반적인 SW는 알고리즘을 사람이 개발하여 고도화하지만 AI발명은 데이터를 학습하여 알고리즘을 고도화한다는 점에서 가장 큰 차이를 갖는다. 따라서, AI 발명의 성립성 과정에서 데이터구조에 관한 구성을 갖추고 있는 발명이라면 데이터발명으로도 인식할 수 있다.

다. 학습데이터의 특성

학습데이터는 기계학습에 있어서 기본적인 요소이며, 경제적 가치가 높은 자산으로서 성질을 갖는다. 본 연구의 특성에 따른 학습데이터는 발명의 구현 요소를 구성한다는 면에서 보면 관리 내지 구조화에 있어서 기술적, 관리적 특징이 있다. 이를 정리하면 아래와 같다.⁴⁰⁾

첫째, 인공지능 학습용 데이터는 크게 텍스트, 음성, 이미지, 동영상 등 주로 비정형 데이터가 다수를 차지하고 있으며, 데이터의 저장구조가 전통적인 DB에 저장하는 방식이 아닌 인공지능 학습에 필요한 데이터 파일 단위로 관리되는 특성이 있다. 둘째, 인공지능 학습용 데이터는 임무 정의, 획득, 정제, 라벨링 등의 구축 단계를 거치면서 목적에 부합하는 데이터를 획득하고, 정제 과정을 통해 기계학습에 적합한 데이터로 가공되며, 정제된 데이터에 학습을 위한 참값(Ground Truth)을 부여하는 등 각 구축 단계별로 고려해야 할 품질 요구사항이 상이한 특성을 갖고 있다. 셋째, 인공지능 학습용 데이터의 품질은 80~90% 이상이 구축과정에서 결정되는 특성이 있다. 그 이유는 인공지능 학습용 데이터는 목적성을 갖는 데이터이기 때문이다. 즉, 인공지능을 통해 해결하려는 문제가 무엇인지를 정의하고, 문제 해결에 필요한 데이터를 획득하여, 목적에 맞게 정제하고, 학습에 필요한 정확한 참값을 부여해야 목적에 부합하는 학습모델의 성능을 확보할 수 있기 때문이다. 넷째, 학습된 인공지능 모델의 성능 측정 결과로 데이터의 품질 수준을 평가하려는 특성이 있다. 인공지능 학습모델을 구현하는 작업은 흔히 데이터와 구분하여 소프트웨어 구현의 영역으로 다루고 있지만, 앞서 언급한 바와 같이 인공지능 학습을 목적으로 구축되는 데이터의 특성상 인공지능 학습에 따른 성능이 낮은 경우, 인공지능 학습용 데이터의 품질의 낮다고 판단하는 것이 사용자가 바라보는 품질에 대한 시각이다. 따라서, 인공지능 학습용 데이터의 품질 확보를 위해서는 인공지능 학습용 데이터의 특성을 고려해야 하며, 데이터베이스에 저장하는 구조화된 데이터(Structured Data)를 대상으로 하는 일반적인 데이터 품질관리와는 차별화된 품질관리 체계를 확보하는 것이 중요하다.

39) 함영욱, AI 관련 발명의 성립성 판단기준에 관한 연구, 한양대학교 박사학위논문, 2022, 28면.

40) 한국지능정보화진흥원, 인공지능 학습용 데이터 품질관리 가이드라인 v1.0, 2021, 11~12면.

[표 2-2] 인공지능 학습용 데이터와 DB 데이터의 비교

구분	인공지능 학습용 데이터	DB 데이터
구축 목적	인공지능 모델의 학습	정보 검색 및 활용
데이터 유형	소리, 영상, 이미지, 자연어 등 비정형 데이터	텍스트 데이터 중심의 정형데이터
데이터 구조	원천 데이터와 라벨링 데이터가 쌍으로 구성	테이블과 컬럼으로 구성
데이터 구축절차	입무정의-> 획득-> 정제-> 라벨링-> 학습-> 저장	수집-> 정제-> 변환-> 저장
데이터 관리방식	파일형태로 저장 (폴더+JSON파일)	범용적인 DBMS에 저장 (테이블+컬럼 구조)

<출처 : NIA, 2021>

3. 데이터의 구조

현행 심사기준에서도 데이터의 구조에 대한 특허발명을 인정하고 있다. 데이터 구조가 일반적인 데이터 자체와는 다른 특성이 있기 때문이다. 그렇다면, 데이터와 데이터 구조화 내지 구조화된 데이터와는 어떤 차이와 특성이 있는지에 대한 고찰이 필요하다. 참고로, 「기술분야별 심사 실무 가이드」 <제10장 컴퓨터 관련 발명>에서 데이터 구조를 데이터 요소 사이의 상호관계를 표시한 데이터의 논리적 구조로, 데이터 기록매체를 기록된 데이터 구조로 말미암아 컴퓨터가 하는 처리 내용이 특정되는 ‘구조를 가진 데이터를 기록한 컴퓨터로 읽을 수 있는 매체’로 정의하고 있다.

가. 데이터의 구조화

데이터 구조를 다루는 것이 ‘데이터 구조론’으로, 데이터 처리 시스템에서 취급하는 데이터 객체들을 기억 공간 내에 표현하고 저장하는 방법과 데이터 상호간의 관계를 파악하여 수행할 수 있는 연산과 관련된 알고리즘을 연구하는 학문이다. 데이터 구조는 다음과 같이 몇 가지로 유형화된다.

- 선형 구조(linear structure, sequential structure): 데이터 상호간에 1:1의 관계를 가진 것-> 연결 리스트, 연결 리스트, 스택, 큐 등
- 비선형 구조(non-linear structure): 데이터 상호간에 1:n 또는 n:m의 관계를 가진 것-> 트리, 그래프
- 파일 구조(file structure): 레코드의 집합체로 이루어지는 특수한 형태의 데이터 구조

위와 같은 유형의 데이터 구조를 선택하는 기준은 데이터의 양, 데이터의 활용 빈도, 데이터의 갱신 정도, 데이터 처리를 위하여 사용가능한 기억 용량, 데이터 처리 시간의 제한, 데이터 처리를 위한 프로그래밍의 용이성 등을 들 수 있다.

나. 구조화된 데이터

(1) 구조화된 데이터의 개념

구조화된 데이터라는 개념은 파일 또는 기록 내의 고정 필드에 있는 데이터를 나타낸다. 구조화된 데이터는 일반적으로 관계형 데이터베이스(RDBMS)에 저장된다. 숫자와 텍스트로 구성될 수 있으며 RDBMS 구조 내에 있는 한 자동으로 또는 수동으로 소싱할 수 있다. 데이터 구조는 포함할 데이터 유형과 저장 및 처리 방법을 정의하는 데이터 모델 생성에 따라 달라질 수 있다. 구조화된 데이터에 사용되는 프로그래밍 언어는 SQL(Structured Query Language)이며, 1970년대에 IBM에서 개발한 SQL은 관계형 데이터베이스를 처리한다. 구조화된 데이터의 일반적인 예로는 이름, 주소, 신용카드 번호, 지리적 위치 등이 있다. 반면, 구조화되지 않은 데이터란 구조화되지 않은 거의 모든 데이터를 말한다. 구조화되지 않은 데이터는 기본 내부 구조를 가질 수 있지만 사전 정의된 방식으로 구조화되지 않는다. 데이터 모델이 없어서 데이터가 기본 형식으로 저장된다. 구조화되지 않은 데이터의 일반적인 예로는 리치 미디어, 텍스트, 소셜 미디어 활동, 감시 이미지 등이 있다.

(2) 구조화된 데이터와 구조화되지 않은 데이터의 비교

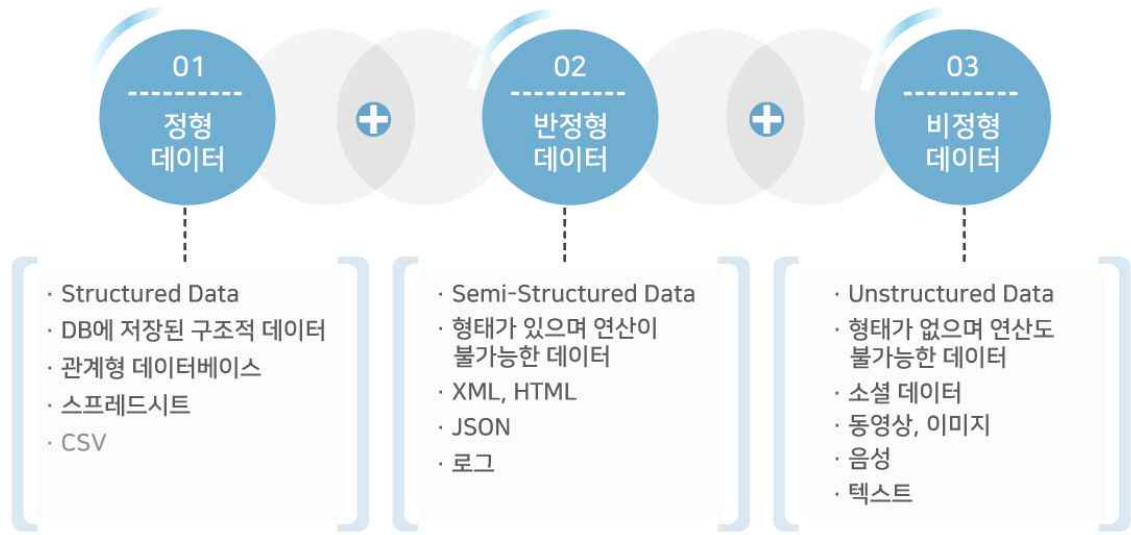
구조화된 데이터와 구조화되지 않은 데이터를 비교하면 다음과 같이 5가지 주요 차이점을 발견할 수 있다.⁴¹⁾

1) 정의된 데이터와 정의되지 않은 데이터 비교

구조화된 데이터는 명확하게 정의된 데이터 유형의 구조인 반면, 구조화되지 않은 데이터는 일반적으로 기본 형식으로 저장된다. 구조화된 데이터는 행과 열에 있으며 사전 정의된 필드에 매핑할 수 있다. 관계형 데이터베이스에서 구성되고 액세스하기 쉬운 구조화된 데이터와 달리 구조화되지 않은 데이터에는 사전 정의된 데이터 모델이 없다.

41) <https://www.integrate.io/ko/blog/structured-data-vs-unstructured-data-ko/> <2022.11.10. 방문>

[그림 2-1] 데이터 유형



<출처 : NIA, 2022>

데이터의 유형을 정리하면, 위의 그림과 같다. 실상 데이터는 정제과정을 거치면서 가치를 높이기 때문에 유형에 따른 구분은 데이터가 갖는 속성이나 구조에 따른 유형화라는 점에서 의의가 있다. 따라서, 기계학습을 위해서 필요한 유형은 정형이나 반정형 또는 비정형 데이터가 아니라 실질적으로 데이터셋으로서 구조화가 가능한 형태인지가 의의가 있으며, 비정형 데이터라고 하더라도 그 속성마다 필요한 학습모델이 있기 때문이다. 다만, 데이터의 표준화에 있어서 고려될 사항은 데이터 속성에 따른 데이터셋의 포맷을 구체화하는 것이다.

2) 정성적 데이터와 정량적 데이터 비교

구조화된 데이터는 정량적 데이터이다. 즉, 일반적으로 하드 숫자 또는 셀 수 있는 항목으로 구성된다. 분석 방법에는 회귀(변수 간의 관계 예측), 분류(확률 추정), 데이터 클러스터링(서로 다른 속성 기반)이 있다. 반면, 구조화되지 않은 데이터는 정성적 데이터로 분류되며 기존의 틀과 방법을 사용하여 처리 및 분석할 수 없다. 예를 들어 비즈니스 컨텍스트에서 정성적 데이터는 고객 설문 조사, 인터뷰, 소셜 미디어 상호 작용에서 얻을 수 있다. 정성적 데이터에서 유용한 정보를 추출하려면 데이터 마이닝 및 데이터 스테킹(data staking)⁴²⁾과 같은 분석 기술이 필요하다.

3) 데이터 하우스의 스토리지와 데이터 레이크의 스토리지 비교

구조화된 데이터는 종종 데이터 웨어하우스에 저장되고 구조화되지 않은 데이터는 데이터 레이크에 저장된다. 데이터 웨어하우스는 ETL 파이프라인⁴³⁾을 통한 데이터 여정의 엔드포인트이

42) 데이터 스테킹이란 여러 가지 모델들의 예측값을 최종 모델의 학습 데이터로 사용하는 예측하는 방법을 말한다.

다. 반면에 데이터 레이크(data lake)⁴⁴⁾는 데이터가 원래 형식으로 저장되거나 기본 정리 프로세스를 거친 후 저장되는 일종의 거의 무한한 저장공간이다. 둘 다 클라우드 사용 가능성이 있다. 구조화된 데이터는 더 적은 스토리지 공간을 필요로 하는 반면, 구조화되지 않은 데이터는 더 많은 스토리지 공간을 필요로 한다. 예를 들어 작은 이미지라도 여러 페이지의 텍스트보다 더 많은 공간을 차지한다. 데이터베이스의 경우 구조화된 데이터는 일반적으로 관계형 데이터베이스(RDBMS)에 저장되지만, 구조화되지 않은 데이터에 가장 적합한 것은 소위 비관계형 또는 NoSQL⁴⁵⁾ 데이터베이스이다.

4) 분석의 용이성

구조화된 데이터와 구조화되지 않은 데이터의 가장 중요한 차이점 중 하나는 분석에 얼마나 적합한지 여부이다. 구조화된 데이터는 사람과 알고리즘 모두에서 검색하기 쉽다. 반면 구조화되지 않은 데이터는 본질적으로 검색하기가 더 어려우므로 이해하기 쉽게 처리해야 한다. 사전 정의된 데이터 모델이 없어서 관계형 데이터베이스에 적합하지 않기 때문에 분해하기가 까다롭다. 구조화된 데이터를 위한 정교한 분석 틀은 다양하지만, 구조화되지 않은 데이터를 마이닝하고 정렬하기 위한 대부분의 분석 틀은 아직 개발 단계에 있다. 사전 정의된 구조가 없으면 데이터 마이닝이 까다로워지고 리치 미디어, 블로그, 소셜 미디어 데이터, 고객 커뮤니케이션과 같은 데이터 소스를 처리하는 방법에 대한 모범 사례를 개발하는 것이 쉽지 않다.

5) 사전 정의된 형식과 다양한 형식 비교

구조화된 데이터의 가장 일반적인 형식은 텍스트와 숫자이다. 구조화된 데이터는 데이터 모델에 미리 정의되어 있다. 반면 구조화되지 않은 데이터는 다양한 모양과 크기로 제공된다. 이는 오디오, 비디오, 이미지에서 이메일 및 센서 데이터에 이르기까지 모든 것으로 구성될 수 있다. 구조화되지 않은 데이터에 대한 데이터 모델은 없다. 따라서 기본 형식으로 저장되거나 변환이 필요하지 않은 데이터 레이크에 저장된다.

다. 데이터셋과 데이터구조

43) 데이터 파이프라인과 ETL 파이프라인이라는 용어는 종종 같은 의미로 사용되기도 하지만 ETL 파이프라인을 데이터 파이프라인의 하위 범주로 생각해야 한다. 데이터 파이프라인은 다양한 데이터 소스에서 원시 데이터를 수집한 다음 분석을 위해 데이터 레이크 또는 데이터 웨어하우스와 같은 데이터 저장소로 이전하는 방법이다. 일반적으로 데이터는 데이터 저장소로 이동하기 전에 데이터 처리 과정을 거친다. 여기에는 적절한 데이터 통합과 표준화를 보장하는 필터링, 마스킹, 집계와 같은 데이터 변환이 포함된다. 이 과정은 데이터셋의 대상이 관계형 데이터베이스인 경우 특히 중요하다. 이 유형의 데이터 저장소에는 기존 데이터를 새 데이터로 업데이트하기 위해 정렬(즉, 데이터 열 및 유형 매칭)이 필요한 정의된 스키마가 있다. <https://www.ibm.com/kr-ko/topics/data-pipeline>. <2022.12.10. 방문>

44) 데이터레이크에 대해서는 알렉스 고렐릭, 엔터프라이즈 빅데이터 레이크, 에이콘, 2020 참조

45) Not Only SQL의 약어로, 관계형 DBMS의 특성과 다른 특성들을 함께 지원한다는 의미이다.

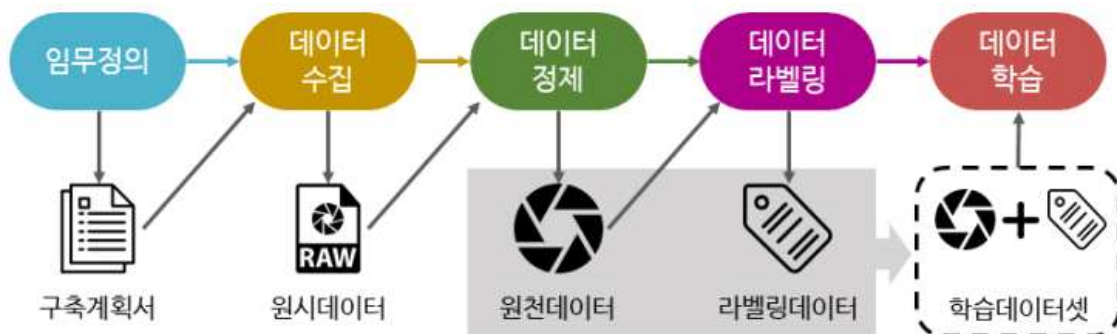
데이터는 수집된 원데이터(raw data)를 포함하여 이를 레이블링 등의 가공을 거친 학습데이터와 이들의 집합이라고 할 수 있는 학습데이터셋으로 나누어볼 수 있다. 데이터는 가치평가가 들어가지 않는 개념이기 때문에 법적으로 볼 때도 다양하게 정의될 수 있다. 데이터 산업법에서는 데이터를 “다양한 부가가치 창출을 위하여 관찰, 실험, 조사, 수집 등으로 취득하거나 정보시스템 및 소프트웨어 진흥법 제2조 제1호에 따른 소프트웨어 등을 통하여 생성된 것으로서 광 또는 전자적 방식으로 처리될 수 있는 자료 또는 정보”로 정의하고 있다. 데이터의 집합을 데이터셋이라고 할 수 있으며 데이터 구조는 데이터를 저장하는 방식이며 여기에는 정형, 반정형, 비정형 데이터의 특성에 따른 구조로 나뉜다.

4. 학습데이터의 구축과 학습모델

가. 학습데이터의 구축

학습데이터의 구축 과정은 다음 그림과 같이 나타낼 수 있다.⁴⁶⁾ 데이터의 구축은 최종적으로 데이터 학습을 목적으로 한다. 데이터셋을 구성하는 과정은 정제와 라벨링을 거치고 데이터셋으로 만들어지면서 알고리즘성이 추가된다. 이러한 특성은 데이터 자체의 속성을 벗어난 데이터간의 논리적으로 구조화되는 과정이라고 할 수 있다.

[그림 2-2] 인공지능 학습용 데이터의 구축과정



<출처 : NIA, 2021>

위와 같이, 데이터셋의 구축과정을 정리하면 다음과 같다. 첫째, 임무정의에서는 인공지능이 기계학습을 통해 해결하고자 하는 문제를 명확하게 정의하고, 문제 해결에 필요한 학습용 데이터를 구체적으로 정의하고 설계한다. 둘째, 데이터 수집에서는 인공지능의 기계학습에 필요한 데이터를 현실 세계에서 직접 생성하거나, 이미 보유하고 있는 조직이나 시스템 등으로부터 인공지능 학습에 필요한 데이터의 수집 시 포함되는 개인정보나 저작권 등 법률적 제약이 없도록 원시데

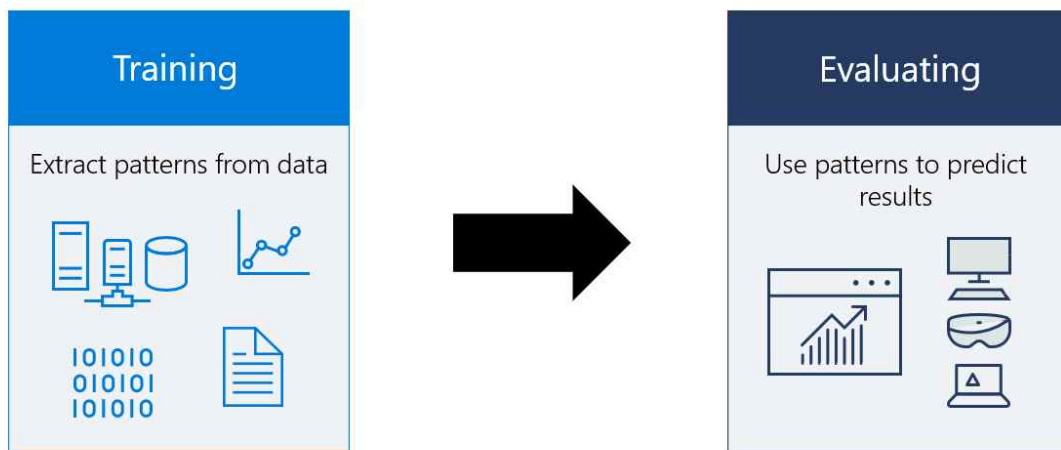
46) 한국지능정보화진흥원, 인공지능 학습용 데이터 품질관리 가이드라인 v1.0, 한국지능정보화진흥원, 2021, 10면.

이터를 확보한다. 여기서 원시데이터는 기계학습을 목적으로 획득 단계에서 수집 또는 생성한 음성, 이미지, 영상, 텍스트 등의 데이터를 의미한다. 셋째, 데이터 정제에서는 획득한 원시데이터를 기계학습에 필요한 형식이나 크기로 맞추고, 데이터의 중복을 제거하며, 원시데이터 획득 시 포함된 개인정보를 비식별화하여 처리하는 등의 과정을 통해 원천데이터를 확보한다. 이 단계에서 확보된 원천데이터는 원시데이터가 라벨링 작업에 투입하기 위해 필요한 정제 작업을 수행한 데이터로 라벨링 데이터가 부여되지 않은 상태의 데이터를 의미한다. 넷째, 데이터 라벨링에서는 인공지능이 기계학습에 활용할 수 있도록 기능이나 목적에 부합하는 라벨링 데이터를 원천데이터에 부착한다. 여기서 라벨링 데이터란, 원천데이터에 부여한 참값(Ground Truth), 파일형식, 해상도 등의 데이터 속성과 설명, 주석 등이 포함된 집합을 의미한다. 마지막으로, 데이터 학습에서는 원천데이터와 라벨링 데이터의 묶음인 학습데이터셋을 이용하여 사전에 정의된 인공지능 알고리즘을 학습시키고, 학습된 인공지능 모델의 성능을 향상시키거나 보정한다.

나. 데이터 학습에 따른 학습모델

기계학습 모델은 특정 유형의 패턴을 인식하도록 학습된 프로그램으로서 AI모델을 말한다. 학습모델은 데이터셋에 대해 모델을 학습하여 해당 데이터로 추론하고 학습하는 데 사용할 수 있는 알고리즘을 제공한다. 그러나 우리나라는 일본이나 유럽과 달리 학습모델 발명을 학습모델 청구항으로 기재해 특허로 출원하는 것을 허용하지 않는다. 모델을 학습한 후에는 이를 사용하여 이전에 표시되지 않은 데이터를 추론하고 해당 데이터에 대한 예측을 만들 수 있다.

[그림 2-3] 학습된 AI모델



<출처 : Microsoft, 2022>

예를 들어, 얼굴 식에 따라 사용자의 감정을 인식할 수 있는 애플리케이션을 제작하는 경우를

가정해 본다. 특정 감정으로 각각 태그가 지정된 얼굴 이미지를 제공하여 모델을 학습한 다음, 사용자의 감정을 인식할 수 있는 애플리케이션에서 해당 모델을 사용할 수 있다.⁴⁷⁾

제2절. 학습데이터 보호 및 활용 법제

1. 개관

데이터셋을 포함한 개발된 데이터에 대한 법적 논의는 한 건의 데이터에 대한 것이라기 보다는 데이터베이스 형태에 대한 것이기 때문에 사실상 데이터 관련 법제라기 보다는 데이터베이스에 대한 보호 가능성을 어떻게 정책적으로 가져갈 것인지에 방점을 두게 된다. 결국, 데이터 관련 법제의 기본은 저작권법상 데이터베이스인지 여부 및 그에 따른 데이터베이스제작자의 권리와 관련성이 크다. 다만, 우리 법제는 저작권법 이외에 콘텐츠산업진흥법에서 콘텐츠를 개발한 사업자에 대한 법적 보호도 아울러 규정하고 있기 때문에 예외적으로, 콘텐츠산업진흥법에 따른 보호를 받게된다. 최근 부정경쟁방지법이 개정되면서 데이터의 무단이용에 대한 부정경쟁법리가 도입되는 등 다양한 정책목표를 달성하기 위한 데이터법제의 개정 또는 제정이 이루어지고 있다. 데이터 보호의 기본 원칙은 비정형 데이터 또는 정형화된 데이터는 물권적인 권리로써 접근할 것인지, 아니면 채권적인 권리로써 접근할 것인지에 대한 논의이다. 물론, 데이터를 가공하면서 인적 또는 물적의 상당한 투자가 이루어질 가능성이 있다. 보호체계에 대해서는 저작권법 체계와 부정경쟁방지법 체계로 나누어볼 수 있다. 전자는 데이터 또는 데이터베이스 제작자에게 저작권 내지 물권 유사의 권리를 부여하는 것이다. 후자는 데이터를 무단으로 이용하는 자에 대한 채권적 권리를 행사할 수 있도록 하는 것이다.

이하에서는 국내외 데이터셋관련 법제 현황과 그 문제점에 대해 살펴보고자 한다.

2. 국내 법제의 현황과 문제점

가. 저작권법

(1) 저작권법에 따른 데이터 보호

저작권법은 데이터베이스에 대해 소재를 체계적으로 배열 또는 구성한 편집물로서 개별적으로 그 소재에 접근하거나 그 소재를 검색할 수 있도록 한 것으로 정의하고 있다. 또한, 데이터베이스를 제작한 사람을 말하는 데이터베이스제작자에 대하여 데이터베이스의 제작 또는 그 소재의 갱신·검증 또는 보충에 인적 또는 물적으로 상당한 투자를 한 자로 정의하고, 관련 규정을 두고

47)

<https://learn.microsoft.com/ko-kr/windows/ai/windows-ml/what-is-a-machine-learning-model>. <2022.11.11. 접속>

있다.

데이터베이스제작자의 권리에 대해 저작권법은 그의 데이터베이스의 전부 또는 상당한 부분을 복제·배포·방송 또는 전송할 권리를 가진다고 규정하고 있다. 데이터 필드나 레코드 등에 저장된 형태와 같이, 데이터베이스의 개별 소재는 해당 데이터베이스의 상당한 부분으로 간주되지 아니한다. 다만, 데이터베이스의 개별 소재 또는 그 상당한 부분에 이르지 못하는 부분의 복제 등이라 하더라도 반복적이거나 특정한 목적을 위하여 체계적으로 함으로써 해당 데이터베이스의 통상적인 이용과 충돌하거나 데이터베이스제작자의 이익을 부당하게 해치는 경우에는 해당 데이터베이스의 상당한 부분의 복제등으로 본다. 이 장에 따른 보호는 데이터베이스의 구성부분이 되는 소재의 저작권 그 밖에 이 법에 따라 보호되는 권리에 영향을 미치지 아니한다. 또한, 이 장에 따른 보호는 데이터베이스의 구성부분이 되는 소재 그 자체에는 미치지 아니한다(제93조).

데이터베이스제작자의 권리제한을 제한하고 있는데, 데이터베이스제작자의 권리의 목적이 되는 데이터베이스의 이용에 관하여는 제23조, 제28조부터 제34조까지, 제35조의2, 제35조의4, 제35조의5, 제36조 및 제37조를 준용한다. 즉, 저작권재산권 제한규정의 일부를 준용함으로써 데이터베이스의 제작자의 권리가 제한되도록 하고 있다. 또한, 다음 각 호의 어느 하나에 해당하는 경우에는 누구든지 데이터베이스의 전부 또는 그 상당한 부분을 복제·배포·방송 또는 전송할 수 있다. 다만, 해당 데이터베이스의 통상적인 이용과 저촉되는 경우에는 그러하지 아니하다(제94조).

i) 교육·학술 또는 연구를 위하여 이용하는 경우. 다만, 영리를 목적으로 하는 경우에는 그러하지 아니하다.

ii) 시사보도를 위하여 이용하는 경우

보호기간에 대해 보면, 저작권은 저작자 사후 70년간 보호를 받지만 데이터베이스제작자의 권리는 데이터베이스의 제작을 완료한 때부터 발생하며, 그 다음 해부터 기산하여 5년간 존속한다. 데이터베이스의 갱신등을 위하여 인적 또는 물적으로 상당한 투자가 이루어진 경우에 해당 부분에 대한 데이터베이스제작자의 권리는 그 갱신등을 한 때부터 발생하며, 그 다음 해부터 기산하여 5년간 존속한다(제95조). 저작권법상 데이터베이스제작자의 권리에 대한 5년간의 보호기간은 콘텐츠산업 진흥법에서도 동일하게 규정되어있으며,⁴⁸⁾ EU DB지침에서 15년간으로 규정하고 있다는 점이 반영된 것이다. 문제는 데이터베이스에 대한 갱신이 지속적으로 이루어질 경우, 그 보호기간의 제한이 없다는 점이다. 경우에 따라서는 상표권과 마찬가지로 제한없이 보호가 가능하다.

48) 콘텐츠산업진흥법 제37조(금지행위 등) ① 누구든지 정당한 권한 없이 콘텐츠제작자가 상당한 노력으로 제작하여 대통령령으로 정하는 방법에 따라 콘텐츠 또는 그 포장에 제작연월일, 제작자명 및 이 법에 따라 보호받는다라는 사실을 표시한 콘텐츠의 전부 또는 상당한 부분을 복제·배포·방송 또는 전송함으로써 콘텐츠제작자의 영업에 관한 이익을 침해하여서는 아니 된다. 다만, 콘텐츠를 최초로 제작한 날부터 5년이 지났을 때에는 그러하지 아니하다.

기본적으로 데이터베이스는 대한민국 국민이 제작한 것이라면 법적 보호를 받는다. 또한, 데이터베이스의 보호와 관련하여 대한민국이 가입 또는 체결한 조약에 따라 보호되는 외국인이 아닌 경우에도 또한 보호가 가능하다. 그렇지만, 데이터베이스에 관한 입법이 글로벌하게 이루어지고 있지는 않다. 따라서, 앞에서 본 바와 같이 보호되는 외국인의 데이터베이스라도 그 외국에서 대한민국 국민의 데이터베이스를 보호하지 아니하는 경우에는 그에 상응하게 조약 및 이 법에 따른 보호를 제한할 수 있다. 미국의 경우에는 명시적으로 데이터베이스에 대한 보호법이 존재하지 않기 때문에 상호주의 원칙에 따라 국내에서 저작권법에 따른 보호가 이루어지기 어렵다.

(2) 저작권법에 따른 한계

1) 부정경쟁법리의 발전과 퍼블리도메인

저작권법은 창작성 여부와 상관없이 투자가 이루어진 경우에는 보호가 가능하다. 사실정보의 집합물인지 여부는 상관이 없다. 다만, 예외적으로 저작권재산권의 제한규정을 준용함으로써 일정한 경우에는 이용이 가능하기 때문에 데이터베이스의 무조건적 보호입법은 아니다. 저작권법에 따른 데이터의 보호 법리는 창작성 내지는 노력과 투자에 따른 경우에 한정된다. 따라서, 데이터에 대한 저작권법의 경우에는 투자와 노력이 요구되지만, 부정형 데이터의 경우에는 검색의 용이성이 인정되지 못할 수 있기 때문에 데이터베이스 제작자의 법리에 따른 보호가 어렵다. 그러한 점 때문에 부정경쟁법리를 적용했다. 즉, 대법원은 “경쟁자가 상당한 노력과 투자에 의하여 구축한 성과물을 상도덕이나 공정한 경쟁질서에 반하여 자신의 영업을 위하여 무단으로 이용함으로써 경쟁자의 노력과 투자에 편승하여 부당하게 이익을 얻고 경쟁자의 법률상 보호할 가치가 있는 이익을 침해하는 행위는 부정한 경쟁행위로서 민법상 불법행위에 해당한다.”⁴⁹⁾라고 판단하였다. 위 대법원의 법리를 반영하여 개정된 부정경쟁방지법 제2조 제1호(차)목은 위 대법원 결정의 취지를 반영하여 “그 밖에 타인의 상당한 투자나 노력으로 만들어진 성과 등을 공정한 상거래 관행이나 경쟁질서에 반하는 방법으로 자신의 영업을 위하여 무단으로 사용함으로써 타인의 경제적 이익을 침해하는 행위”를 부정경쟁행위의 하나로 추가하였다.

부정경쟁방지법의 적용 범위에 포함되지 않았던 새로운 유형의 부정경쟁행위에 관한 규정을 신설한 것이다. 이는 새로이 등장하는 경제적 가치를 지닌 무형의 성과를 보호하고 입법자가 부정경쟁행위의 모든 행위를 규정하지 못한 점을 보완하여 법원이 새로운 유형의 부정경쟁행위를 좀 더 명확하게 판단할 수 있도록 함으로써, 변화하는 거래관념을 적시에 반영하여 부정경쟁행위를 규율하기 위한 보충적 일반조항이다. (카)목은 그 보호대상인 ‘성과 등’의 유형에 제한을 두고 있지 않으므로, 유형물뿐만 아니라 무형물도 이에 포함되고, 종래 지식재산권법에 따라 보호받기 어려웠던 새로운 형태의 결과물도 포함될 수 있다. ‘성과 등’을 판단할 때에는 위와 같

49) 대법원 2010. 8. 25.자 2008마1541 결정.

은 결과물이 갖게 된 명성이나 경제적 가치, 결과물에 화체된 고객흡인력, 해당 사업 분야에서 결과물이 차지하는 비중과 경쟁력 등을 종합적으로 고려해야 한다. 이러한 성과 등이 ‘상당한 투자나 노력으로 만들어진’ 것인지는 권리자가 투입한 투자나 노력의 내용과 정도를 그 성과 등이 속한 산업분야의 관행이나 실태에 비추어 구체적·개별적으로 판단하되, 성과 등을 무단으로 사용함으로써 침해된 경제적 이익이 누구나 자유롭게 이용할 수 있는 이른바 공공영역(public domain)에 속하지 않는다고 평가할 수 있어야 한다. 또한 (카)목이 정하는 ‘공정한 상거래 관행이나 경쟁질서에 반하는 방법으로 자신의 영업을 위하여 무단으로 사용’ 한 경우에 해당하기 위해서는 권리자와 침해자가 경쟁관계에 있거나 가까운 장래에 경쟁관계에 놓일 가능성이 있는지, 권리자가 주장하는 성과 등이 포함된 산업분야의 상거래 관행이나 경쟁질서의 내용과 그 내용이 공정한지, 위와 같은 성과 등이 침해자의 상품이나 서비스에 의해 시장에서 대체될 수 있는지, 수요자나 거래자들에게 성과 등이 어느 정도 알려졌는지, 수요자나 거래자들의 혼동가능성이 있는지 등을 종합적으로 고려해야 한다.⁵⁰⁾ 이처럼 부정경쟁방지법은 기본적으로 타인의 성과를 무단으로 사용함으로써 침해된 경제적 이익이 누구나 자유롭게 이용할 수 있는 이른바 공공영역에 속하지 않는다고 평가할 수 있다면 자유로운 이용이 가능함을 확인하고 있다.

2) 부정형의 데이터의 보호

무엇보다, 비정형 내지 부정형 데이터는 데이터베이스의 정의에 부합하지 않을 수 있기 때문에 저작권법의 적용이 배제될 수 있다. 이러한 경우에는 저작권법의 기본적인 취지상 공공영역에 두어 누구나 자유롭게 이용할 수 있도록 두는 것이 정책적으로 타당하다. 물론, 데이터는 저작권과 상관없거나 사실정보이지만, 뉴스기사와 같은 사실정보와는 성질이 다르다는 점에서 보호범위의 문제가 아닌 산업정책적 판단에 따른 보호가능성을 제시하기도 한다. 이러한 주장에 따라 부정경쟁방지법에 따른 비정형데이터의 보호가 이루어진 것이다.

창작자 보호의 예외인 인접권으로써 보호는 비정형 데이터의 보호가능성을 적용할 수 있다. 다만, 저작권법은 명시적으로 보호에 대한 데이터베이스 요건을 요구하지만 검색이 용이하지 않는 경우에는 보호가능성이 높지 않기 때문에 저작권법의 보호가능성은 낮다. 저작권법이 명시적으로 창작자 원칙에 따른 보호주의를 취하고, 예외적으로 투자에 대한 보호주의를 취하고 있기 때문에 부정형의 데이터 보호를 취하지 않는 것은 합리적이다.

나. 데이터산업법

(1) 산업정책에 따른 데이터

데이터산업법으로 알려진 데이터 산업진흥 및 이용촉진에 관한 기본법(이하, 데이터산업법이라

50) 대법원 2020. 3. 26. 선고 2016다276467 판결

함)이 2022년 시행되었다.⁵¹⁾ 데이터산업법은 데이터의 생산, 거래 및 활용 촉진에 관하여 필요한 사항을 정함으로써 데이터로부터 경제적 가치를 창출하고 데이터산업 발전의 기반을 조성하여 국민생활의 향상과 국민경제의 발전에 이바지함을 목적으로 한다.

동 법에서는 데이터를 “다양한 부가가치 창출을 위하여 관찰, 실험, 조사, 수집 등으로 취득하거나 정보시스템 및 소프트웨어 진흥법에 따른 소프트웨어 등을 통하여 생성된 것으로서 광(光) 또는 전자적 방식으로 처리될 수 있는 자료 또는 정보”로 정의하고 있다. 이에 따라, 디지털 여부와 상관없이 존재하는 모든 정보 또는 자료가 데이터가 될 수 있다.

무엇보다, 데이터 산업법은 데이터컨트롤타워로서 국가데이터정책위원회를 국무총리 산하에 두도록 하고 있다. 국가데이터정책위원회가 데이터 관련 업무를 승계하는 것이 정책적으로 타당하다. 실무단은 과학기술정보통신부에 두도록 하고 있다.

(2) 데이터 자산의 보호

데이터의 활용과 관련하여 데이터 안심구역을 지정하도록 하고 있다. 과학기술정보통신부장관과 관계 중앙행정기관의 장은 누구든지 데이터를 안전하게 분석·활용할 수 있는 구역을 지정하여 운영할 수 있으며, 데이터안심구역 이용을 지원하기 위하여 미개방 데이터, 분석 시스템 및 도구 등을 지원할 수 있다. 또한, 미개방 데이터 지원을 위하여 필요한 경우에는 정부 및 지방자치단체, 공공기관, 민간법인 등에게 데이터 제공을 요청할 수 있도록 하고 있다. 다만, 데이터안심구역에 대한 제3자의 불법적인 접근, 데이터의 변경·훼손·유출 및 파괴, 그 밖의 위협에 대하여 대통령령으로 정하는 바에 따라 기술적·물리적·관리적 보안대책을 수립·시행하여야 한다(데이터 산업법 제11조). 특이한 사항은 데이터의 일반적 정의와 달리, 데이터생산자가 인적 또는 물적으로 상당한 투자와 노력으로 생성한 경제적 가치를 가지는 데이터를 ‘데이터자산’으로 정의하고 있다는 점이다.

[표 2-3] 데이터 산업법상 데이터자산의 보호

<p>제12조(데이터자산의 보호) ① 데이터생산자가 인적 또는 물적으로 상당한 투자와 노력으로 생성한 경제적 가치를 가지는 데이터(이하 “데이터자산”이라 한다)는 보호되어야 한다.</p> <p>② 누구든지 제1항에 따른 데이터자산을 공정한 상거래 관행이나 경쟁질서에 반하는 방법으로 무단 취득·사용·공개하거나 이를 타인에게 제공하는 행위, 정당한 권한 없이 데이터자산에 적용한 기술적 보호조치를 회피·제거 또는 변경하는 행위 등 데이터자산을 부정하게 사용하여 데이터생산자의 경제적 이익을 침해하여서는 아니 된다.</p> <p>③ 제2항에 따른 데이터자산의 부정사용 등 행위에 관한 사항은 「부정경쟁방지 및 영업비밀보호에 관한 법률」에서 정한 바에 따른다.</p>
--

51) [시행 2022. 4. 20.] [법률 제18475호, 2021. 10. 19., 제정].

이는 저작권법상 데이터베이스 및 데이터베이스제작자의 권리와 유사한 구조를 갖는다.⁵²⁾ 또한 누구든지 데이터자산을 공정한 상거래 관행이나 경쟁질서에 반하는 방법으로 무단 취득·사용·공개하거나 이를 타인에게 제공하는 행위, 정당한 권한 없이 데이터자산에 적용한 기술적 보호조치를 회피·제거 또는 변경하는 행위 등 데이터자산을 부정하게 사용하여 데이터생산자의 경제적 이익을 침해하여서는 아니 된다는 점 등이다. 그렇지만, 데이터자산 부정사용 등 행위에 관한 사항은 부정경쟁방지법에서 정한 바에 따르도록 함으로써 실효적인 규제는 데이터 산업법에서 제외토록 하고 있다. 이는 정책적 합의에 따른 입법으로 볼 수 있으나, 데이터자산에 대한 규정이 포함됨으로써, 데이터 소유권에 대한 논의가 확대될 것으로 보인다. 그렇지만, 데이터 소유권을 민법상 개념으로 포섭하기에는 너무 불확정적이라는 점에서 한계를 갖는다. 이러한 논란에 대해 “데이터에 소유권을 인정할 수 있을까 아니면 무형적 정보로서의 비경합적, 비배타적 속성과 공유재적 성격에 기반해 자유사용체제가 바람직한가? 권리를 인정한다면 누구에게 얼마동안 인정할 것인가? 부여하는 권리의 성격도 직접적 배타적 지배하에 이익을 얻을 수 있는 민법상 ‘물권(物權)’에 준하는 권리(침해 금지, 방해배제 청구권 인정)로 인정하여야 할 것인지 아니면 행위규제형 채권적 권리로만 인정할 것인지 등 다양한 논의가 진행되고 있다. 또한 개인정보보호법상 ‘개인정보 자기결정권’이나 저작권법상 ‘데이터베이스제작자의 권리’와는 어떻게 차별화 할 것인지? 데이터를 결합한 경우에 권리귀속은 어떻게 할 것인지 등 논란이 많다. 데이터 자산 부정 취득 행위 등을 규율하는 방식은 일본의 한정데이터 부정사용금지 규정에서 차용한 것으로 보이나 우리 전체적 법체계에 부합하지 않는다.”⁵³⁾고 평가된다. 부정경쟁행위규제에 대해서는 부정경쟁방지법에서 직접 규율하는 것이 법체계에 맞다.⁵⁴⁾

또한, 정부는 기계학습 등 데이터 기반의 정보분석을 활성화하기 위하여 데이터의 수집, 가공 등 정보분석에 필요한 사업을 지원할 수 있도록 하고있다. 정보분석을 위하여 데이터를 이용하는 경우에 그 데이터에 포함된 저작권법에 따른 저작물 등의 보호와 이용에 관하여는 같은 법에서 정하는 바에 따르도록 하고있다(제13조). 따라서, 정보분석을 위한 저작권법 전면개정안이 통과되면 해당 규정을 따르게 된다.⁵⁵⁾ 그렇지 못하더라도, 정보분석을 위해 필요한 경우라면 공정이용 규정의 해석적용이 가능하다. 데이터 산업법에 명시적으로 이러한 사항을 규정하고 있다는 점에

52) 데이터베이스제작자는 데이터베이스의 제작 또는 그 소재의 갱신·검증 또는 보충에 인적 또는 물적으로 상당한 투자를 한 자를 말하는데, 데이터베이스제작자는 그의 데이터베이스의 전부 또는 상당한 부분을 복제·배포·방송 또는 전송할 권리를 가진다. 만약, 데이터베이스제작자의 권리를 복제·배포·방송 또는 전송의 방법으로 침해한 자는 3년 이하의 징역 또는 3천만원 이하의 벌금에 처하거나 이를 병과할 수 있다

53) 김원오, 데이터 산업법 제정안에 관한 소고, 산업재산권 제68호, 2020, 197면.

54) 김원오, 데이터 산업법 제정안에 관한 소고, 산업재산권 제68호, 2020, 206면.

55) 도종환의원 대표발의(의안번호 7440, 2021.1.15.) 저작권법 전면개정안 제43조(정보분석을 위한 복제·전송) ① 컴퓨터를 이용한 자동화 분석기술을 통해 다수의 저작물을 포함한 대량의 정보를 분석(규칙, 구조, 경향, 상관관계 등의 정보를 추출하는 것)하여 추가적인 정보 또는 가치를 생성하기 위한 것으로 저작물에 표현된 사상이나 감정을 향유하지 아니하는 경우에는 필요한 한도 안에서 저작물을 복제·전송할 수 있다. 다만, 해당 저작물에 적법하게 접근할 수 있는 경우에 한정한다.

② 제1항에 따라 만들어진 복제물은 정보분석을 위하여 필요한 한도에서 보관할 수 있다.

서 이중규정에 따른 혼란이 예상된다(제56조) 하나,56) 법률의 우선순위를 조정하고 있다는 점에서 큰 문제는 아니다. 우선순위에 관한 것으로 데이터 산업법상 몇 가지 조항은 저작권법, 개인정보보호법이나 부정경쟁방지법 등 타 법령을 따르도록 하고 있다. 즉, 데이터 생산, 거래 및 활용 촉진에 관하여 다른 법률에 특별한 규정이 있는 경우를 제외하고는 이 법으로 정하는 바에 따른다. 개인정보, 저작권 및 공공데이터에 관하여는 각각 개인정보 보호법, 저작권법, 공공데이터법 등 다른 법률에서 정하는 바에 따른다(제7조).

동법에서는 데이터 공정거래를 위하여 표준계약을 마련토록 하고 있다. 과학기술정보통신부장관은 데이터의 합리적 유통 및 공정한 거래를 위하여 공정거래위원회와 협의를 거쳐 표준계약을 마련하고, 데이터사업자에게 그 사용을 권고할 수 있다(제21조). 데이터 산업법을 위반하는 행위로 인하여 자신의 영업에 관한 이익이 침해되어 손해를 입은 자는 그 위반행위를 한 자에 대하여 위반행위로 인한 손해의 배상을 청구할 수 있다. 이 경우 그 위반행위를 한 자는 고의 또는 과실이 없음을 입증하지 아니하면 책임을 면할 수 없다. 법원은 이 법을 위반한 행위에 관한 소송에서 손해의 발생은 인정되나 손해액을 산정하기 곤란한 경우에는 변론의 취지 및 증거조사 결과를 고려하여 상당한 손해액을 인정할 수 있다(제42조). 다만, 데이터사업자는 손해배상책임의 이행을 위하여 보험 또는 공제에 가입하거나 준비금을 적립하는 등 필요한 조치를 할 수 있다(제43조). 이는 개인정보 보호법 등에서도 유사하게 규정하고 있는 것으로, 사업자의 부담을 덜어 주기 위한 제도이다.57)

아울러, 데이터 생산, 거래 및 활용에 관한 분쟁을 조정하기 위하여 데이터분쟁조정위원회를 두도록 하고 있다.58) 다만, 공공데이터의 제공거부 및 제공중단과 관련한 분쟁은 공공데이터법 및 데이터기반행정법에 따르고, 개인정보와 관련한 분쟁은 개인정보 보호법에 따르며, 저작권에 관련한 분쟁은 저작권법에 따른다(제34조). 사실상 분쟁조정 규정은 저작권, 공공데이터, 개인정보, 콘텐츠, 국가지식정보 등에 관한 다양한 법률에서 두고 있기 때문에 실효성에 대해서는 의문이다.

(3) 데이터 산업법에 따른 한계

현행 법률에서는 데이터 관련 법제가 다수 존재하기 때문에 데이터 산업법과 같이, 데이터 생

56) 김원오, 데이터 산업법 제정안에 관한 소고, 산업재산권 제68호, 2020, 220면.

57) 개인정보 보호법 제39조의9(손해배상의 보장) ① 정보통신서비스 제공자등은 제39조 및 제39조의2에 따른 손해배상책임의 이행을 위하여 보험 또는 공제에 가입하거나 준비금을 적립하는 등 필요한 조치를 하여야 한다.

② 제1항에 따른 가입 대상 개인정보처리자의 범위, 기준 등에 필요한 사항은 대통령령으로 정한다.

58) 데이터 생산, 거래 및 활용과 관련한 피해의 구제와 분쟁의 조정을 받으려는 자는 위원회에 분쟁의 조정을 신청할 수 있다. 다만, 다른 법률에 따라 분쟁조정이 완료된 경우는 제외한다(제35조). 조정은 당사자가 조정안에 동의한 경우나 당사자가 위원회에 조정합의서를 제출한 경우에 성립한다. 조정의 효력은 민사소송법에 따른 재판상 화해와 동일한 효력을 갖는다(제36조). 물론, i)분쟁조정 신청이 취하되거나 당사자 어느 한 쪽이 분쟁의 조정에 응하지 아니하는 경우, ii) 당사자가 위원회의 조정안을 거부한 경우에는 조정이 성립하지 아니한다.

해주기 관련 법률은 필요하지만 다른 법률과의 관계를 위해 우선순위를 규정하다 보니, 정작 데이터 산업법은 데이터사업 지원 등 산업정책적 측면에서 정책지원을 위한 역할로 한정된다는 점이 한계이다. 이는 기본법이 갖는 한계이기도 하지만, 향후 데이터 산업법 개정과정에서 기본법이 추구해야 할 가치와 이념을 보다 명확히 함으로써 데이터에 관한 기본법으로서 역할을 충실히 할 필요가 있다.

데이터 보호에 관한 원칙이 부정경쟁법리를 채택함으로써 행위규제 형식으로 규정된다는 점에서는 어느정도 일관성을 갖는다고 볼 수 있다. 그렇지만, 세부적인 요건에서는 차이가 있기 때문에 데이터법제 관련하여 일관되게 보호방식을 채택하기 위한 입법은 필요하다. 데이터 산업법이 기본법적 역할을 하기는 하지만, 데이터 산업 전체를 관통하는 기본법에서는 부족하기 때문에 데이터 산업법을 제정하고, 그에 따라 데이터 산업법 등 산업법제를 정비할 필요성이 크다.

다. 부정경쟁방지법

(1) 부정경쟁방지법에 따른 보호

데이터의 중요성이 커짐에 따라 지식재산 보호의 일반법으로서 역할을 하는 부정경쟁방지법의 역할도 또한 확대되고 있다. 개정법에 따라 (파)목으로 옮겨졌지만, 일반조항으로서 역할을 하던 (카)목을 대신하여 데이터의 부정사용을 금지하는 행위유형을 추가함으로써 부정형 데이터의 보호 근거를 마련하게 되었다. 저작권법상 데이터베이스제작자의 권리와 같은 물권적인 권리가 아닌 행위유형에 따른 불법사용 행위를 규제하는 채권적 권리를 부여한 것이다. 부정경쟁방지법의 입법취지는 다음과 같다.

“4차 산업혁명, 인공지능 등 디지털시대의 근간인 데이터의 중요성이 날로 커지고 있고, 빅데이터를 활용해 경제적 부가가치를 창출하고 있음. 하지만, 데이터를 부정한 수단으로 취득해서 부당하게 이익을 얻거나 데이터 보유자에게 손해를 끼치는 행위에 대한 제재는 미흡한 실정임. 대법원은 타인이 영업 목적으로 공개한 데이터를 무단으로 수집하여 제3자와 거래하거나, 상업적 목적으로 활용한 행위에 대해서 이 법의 보충적 일반조항(제2조 제1호 카목)을 근거로 ‘부정경쟁행위’로 판결한 바 있음. 그러나 이는 향후 발생할 수 있는 다양한 형태의 데이터 무단 수집·이용·유통 행위를 적절히 제재하기에 한계가 있음. 따라서 ‘데이터 부정사용행위’를 법률에 명확하게 규정하여 부정경쟁행위와 마찬가지로 제재함으로써 건전한 데이터 시장질서를 확립하려는 것임.”

주요 내용은 부정경쟁방지법에서 보호하는 데이터를 데이터 기본법 제2조 제1호에 따른 데이터⁵⁹⁾ 중 “업으로서 특정인 또는 특정 다수에게 제공되는 것으로, 전자적 방법으로 상당량 축

적·관리되고 있으며, 비밀로서 관리되고 있지 않은 기술상 또는 영업상의 정보”로 정의하고 있다. 데이터를 부정하게 사용하는 행위를 부정경쟁행위의 유형으로 신설하면서, 구체적인 금지행위로 다음의 4가지 행위유형을 규정하고 있다(제2조 제1호 카목).⁶⁰⁾

- i) 접근권한이 없는 자가 데이터를 부정하게 취득하거나 그 취득한 데이터를 사용·공개하는 행위
- ii) 접근권한이 있는 자가 부정한 목적으로 데이터를 제3자에게 제공하거나 사용·공개하는 행위
- iii) i) 또는 ii)가 개입된 사실을 알고 데이터를 취득하거나 그 취득한 데이터를 사용·공개하는 행위
- iv) 데이터의 기술적 보호조치를 무력화하는 행위

특히, 데이터에 적용된 기술적 보호조치를 무력화하는 경우에는 형사처벌을 받도록 하고 있다.⁶¹⁾ 부정경쟁 행위유형의 하나인 데이터에 적용된 기술적 보호조치를 무력화하는 등의 금지대상 행위는 정당한 권한 없이 데이터의 보호를 위해 적용한 기술적 보호조치를 회피·제거 또는 변경하는 것을 주된 목적으로 하는 기술·서비스·장치 또는 그 장치의 부품을 제공·수입·수출·제조·양도·대여 또는 전송하거나 이를 양도·대여하기 위하여 전시하는 것을 말한다. 다만, 기술적 보호조치의 연구·개발을 위하여 기술적 보호조치를 무력화하는 장치 또는 그 부품을 제조하는 경우에는 처벌받지 않는다.

59) 데이터 기본법 제2조(정의) 이 법에서 사용하는 용어의 뜻은 다음과 같다.

1. “데이터”란 다양한 부가가치 창출을 위하여 관찰, 실험, 조사, 수집 등으로 취득하거나 정보시스템 및 소프트웨어 진흥법 제2조제1호에 따른 소프트웨어 등을 통하여 생성된 것으로서 광(光) 또는 전자적 방식으로 처리될 수 있는 자료 또는 정보를 말한다.

60) 부정경쟁방지법 제2조(정의) 1. “부정경쟁행위”란 다음 각 목의 어느 하나에 해당하는 행위를 말한다.

카. 데이터(데이터 기본법 제2조제1호에 따른 데이터 중 업으로서 특정인 또는 특정 다수에게 제공되는 것으로, 전자적 방법으로 상당량 축적·관리되고 있으며, 비밀로서 관리되고 있지 않은 기술상 또는 영업상의 정보를 말한다. 이하 같다)를 부정하게 사용하는 행위로서 다음의 어느 하나에 해당하는 행위

- 1) 접근권한이 없는 자가 절취·기망·부정접속, 그 밖의 부정한 수단으로 데이터를 취득하거나 그 취득한 데이터를 사용·공개하는 행위
- 2) 데이터 보유자와의 계약관계 등에 따라 데이터에 접근권한이 있는 자가 부정한 이익을 얻거나 데이터 보유자에게 손해를 입힐 목적으로 그 데이터를 사용·공개하거나 제3자에게 제공하는 행위
- 3) 1) 또는 2)가 개입된 사실을 알고 데이터를 취득하거나 그 취득한 데이터를 사용·공개하는 행위
- 4) 정당한 권한 없이 데이터의 보호를 위해 적용한 기술적 보호조치를 회피·제거 또는 변경(이하 “무력화”라 한다)하는 것을 주된 목적으로 하는 기술·서비스·장치 또는 그 장치의 부품을 제공·수입·수출·제조·양도·대여 또는 전송하거나 이를 양도·대여하기 위하여 전시하는 행위. 다만, 기술적 보호조치의 연구·개발을 위하여 기술적 보호조치를 무력화하는 장치 또는 그 부품을 제조하는 경우에는 그러하지 아니하다.

61) 부정경쟁방지법 제18조(벌칙) ③ 다음 각 호의 어느 하나에 해당하는 자는 3년 이하의 징역 또는 3천만원 이하의 벌금에 처한다.

1. 제2조 제1호(아목, 차목, 카목 1)부터 3)까지, 타목 및 파목은 제외한다)에 따른 부정경쟁행위를 한 자

(2) 부정경쟁방지법에 따른 한계

부정경쟁방지법상 금지되는 행위유형으로써 기술적 보호조치는 저작권법 보다는 구체적이거나,⁶²⁾ 적용예외에 있어서는 보다 협소하다. 저작권법 제140조의2⁶³⁾에 따른 예외규정에 비하여 부정경쟁방지법은 기술적 보호조치의 연구·개발을 위하여 기술적 보호조치를 무력화하는 장치 또는 그 부품을 제조하는 경우만 한정적으로 허용되기 때문이다. 또한, 데이터에는 개인정보, 저작물 등 다양한 유형의 정보가 포함될 수 있다. 이번 개정으로 데이터 내에 개인정보가 포함되어 있는 경우에는 개인정보 보호법이 추가되어 부정경쟁방지법 보다 우선 적용되도록 함으로써 다

-
- 62) 저작권법 제2조(정의) 28. “기술적 보호조치”란 다음 각 목의 어느 하나에 해당하는 조치를 말한다.
가. 저작권, 그 밖에 이 법에 따라 보호되는 권리의 행사와 관련하여 이 법에 따라 보호되는 저작물등에 대한 접근을 효과적으로 방지하거나 억제하기 위하여 그 권리자나 권리자의 동의를 받은 자가 적용하는 기술적 조치
나. 저작권, 그 밖에 이 법에 따라 보호되는 권리에 대한 침해 행위를 효과적으로 방지하거나 억제하기 위하여 그 권리자나 권리자의 동의를 받은 자가 적용하는 기술적 조치
- 63) 저작권법 제104조의2(기술적 보호조치의 무력화 금지) ① 누구든지 정당한 권한 없이 고의 또는 과실로 제2조제28호가목의 기술적 보호조치를 제거·변경하거나 우회하는 등의 방법으로 무력화하여서는 아니 된다. 다만, 다음 각 호의 어느 하나에 해당하는 경우에는 그러하지 아니하다.
1. 암호 분야의 연구에 종사하는 자가 저작물등의 복제물을 정당하게 취득하여 저작물등에 적용된 암호 기술의 결함이나 취약점을 연구하기 위하여 필요한 범위에서 행하는 경우. 다만, 권리자로부터 연구에 필요한 이용을 허락받기 위하여 상당한 노력을 하였으나 허락을 받지 못한 경우로 한정한다.
 2. 미성년자에게 유해한 온라인상의 저작물등에 미성년자가 접근하는 것을 방지하기 위하여 기술·제품·서비스 또는 장치에 기술적 보호조치를 무력화하는 구성요소나 부품을 포함하는 경우. 다만, 제2항에 따라 금지되지 아니하는 경우로 한정한다.
 3. 개인의 온라인상의 행위를 파악할 수 있는 개인 식별 정보를 비공개적으로 수집·유포하는 기능을 확인하고, 이를 무력화하기 위하여 필요한 경우. 다만, 다른 사람들이 저작물등에 접근하는 것에 영향을 미치는 경우는 제외한다.
 4. 국가의 법집행, 합법적인 정보수집 또는 안전보장 등을 위하여 필요한 경우
 5. 제25조제3항 및 제4항에 따른 학교·교육기관 및 수업지원기관, 제31조제1항에 따른 도서관(비영리인 경우로 한정한다) 또는 공공기록물 관리에 관한 법률에 따른 기록물관리기관이 저작물등의 구입 여부를 결정하기 위하여 필요한 경우. 다만, 기술적 보호조치를 무력화하지 아니하고는 접근할 수 없는 경우로 한정한다.
 6. 정당한 권한을 가지고 프로그램을 사용하는 자가 다른 프로그램과의 호환을 위하여 필요한 범위에서 프로그램코드역분석을 하는 경우
 7. 정당한 권한을 가진 자가 오로지 컴퓨터 또는 정보통신망의 보안성을 검사·조사 또는 보정하기 위하여 필요한 경우
 8. 기술적 보호조치의 무력화 금지에 의하여 특정 종류의 저작물등을 정당하게 이용하는 것이 불합리하게 영향을 받거나 받을 가능성이 있다고 인정되어 대통령령으로 정하는 절차에 따라 문화체육관광부장관이 정하여 고시하는 경우. 이 경우 그 예외의 효력은 3년으로 한다.
- ② 누구든지 정당한 권한 없이 다음과 같은 장치, 제품 또는 부품을 제조, 수입, 배포, 전송, 판매, 대여, 공중에 대한 청약, 판매나 대여를 위한 광고, 또는 유통을 목적으로 보관 또는 소지하거나, 서비스를 제공하여서는 아니 된다.
1. 기술적 보호조치의 무력화를 목적으로 홍보, 광고 또는 판촉되는 것
 2. 기술적 보호조치를 무력화하는 것 외에는 제한적으로 상업적인 목적 또는 용도만 있는 것
 3. 기술적 보호조치를 무력화하는 것을 가능하게 하거나 용이하게 하는 것을 주된 목적으로 고안, 제작, 개조되거나 기능하는 것
- ③ 제2항에도 불구하고 다음 각 호의 어느 하나에 해당하는 경우에는 그러하지 아니하다.
1. 제2조제28호가목의 기술적 보호조치와 관련하여 제1항제1호·제2호·제4호·제6호 및 제7호에 해당하는 경우
 2. 제2조제28호나목의 기술적 보호조치와 관련하여 제1항제4호 및 제6호에 해당하는 경우

른 법률과의 관계를 명확히 규정하였다(제15조).⁶⁴⁾ 데이터에 관한 다양한 법률의 적용가능성 때문에 타법과의 관계 설정은 무엇보다 중요하기 때문이다. 법률의 우선순위에 따라, 기술적 보호 조치의 시행과정에서 저작권법과의 관계 설정 또한 중요하다.

라. 산업 디지털전환 촉진법에 따른 보호

(1) 입법취지

산업데이터 생성·활용의 활성화와 지능정보기술의 산업 적용을 통하여 산업의 디지털 전환을 촉진하기 위한 산업 디지털전환 촉진법이 국회를 통과하였다. 산업 디지털전환 촉진법은 3개 법안이 병합된 것으로, 제안이유는 다음과 같다.

“최근 4차 산업혁명 가속화와 글로벌 공급망 재편의 흐름 속에서 인공지능·빅데이터와 같은 디지털 기술을 제조업 등 산업 전반에 적용하는 산업의 디지털 전환이 전 세계적으로 확산되고 있음. 특히, 코로나19로 인해 생산·소비 과정이 비대면 방식으로 급격히 전환되면서 산업데이터 활용의 중요성이 더욱 높아짐. 그간 우리나라는 의료·금융·통신 등 서비스 분야 중심으로 개인 정보의 활용은 높았던 반면, 산업 밸류체인 전 과정에서 발생하는 산업데이터는 과도한 기업 영업비밀 보호 경향, 다양한 형태와 방대한 범위 등의 특성으로 인해 활용에 어려움이 있었음. 산업 전반에 걸친 디지털 기술 적용 활성화를 통해 밸류체인 전반을 혁신하고 고부가가치화하기 위해, 산업데이터 활용·보호 원칙을 제시하여 데이터 등 디지털 기술 활용에 따른 기업의 법적 불확실성을 해소하는 것이 시급하며, 산업 디지털 전환 정책을 종합적으로 수립·시행하기 위한 정부 내 추진 체계와 민간의 디지털 전환 활동을 강력하게 뒷받침할 수 있는 지원 제도 관련 규정 마련도 긴요함. 이에 개인정보보호법 등 기존 권리보호 법령에서 규정하고 있지 않은 산업데이터의 개념을 정의하고 이에 관한 활용·보호 원칙을 제시하여 기업의 불확실성을 해소하고 산업데이터의 활용을 활성화하고자 함.”

(2) 산업데이터의 정의

산업 디지털전환 촉진법은 산업데이터를 산업활동에서 생성 또는 활용되는 것으로서 광(光) 또는 전자적 방식으로 처리될 수 있는 모든 종류의 자료 또는 정보로 정의하고 있다. 또한, 산업데이터 생성을 산업활동 과정에서 인적 또는 물적으로 상당한 투자와 노력을 통하여 기존에 존재하지 아니하였던 산업데이터가 새롭게 발생하는 것으로 정의하고 있다. 참고로, “산업데이터 생성의 요건으로 인적 또는 물적으로 상당한 투자와 노력을 투입할 것을 명시하고 있는데, 이는 인

64) 부정경쟁방지법 제15조(다른 법률과의 관계) ① 특허법, 실용신안법, 디자인보호법, 상표법, 농수산물 품질관리법, 저작권법 또는 개인정보보호법에 제2조부터 제6조까지 및 제18조제3항과 다른 규정이 있으면 그 법에 따른다.

적 또는 물적으로 상당한 투자를 한 자를 데이터베이스제작자로 인정하고 있는 저작권법 규정에 비하여 보다 엄격한 요건을 설정한 것” 65)이라고 한다.

일종의 2차적 데이터로 볼 수 있는 산업데이터의 활용을 통하여 독자성을 인정할 수 있는 새로운 산업데이터가 발생하는 경우를 산업데이터의 생성에 포함하고 있다. 산업데이터를 활용하여, 새로운 가치를 만들어내는 데이터도 산업데이터와 동일하게 보는 것이 타당한 것인지는 의문이다.

(3) 산업데이터 보호 원칙

산업데이터를 생성한 자는 해당 산업데이터를 활용하여 사용·수익할 권리를 가진다. 산업데이터를 활용하여 경제적 가치를 얻을 수 있는 권리, 즉 데이터권을 법정권리화 하고 있다. 더욱이, 산업데이터를 2인 이상이 공동으로 생성한 경우 각자 해당 산업데이터를 활용하여 사용·수익할 권리를 가진다. 다만, 당사자 간의 약정이 있는 경우에는 그에 따른다. 이는 저작권법의 공동저작물 원리를 차용한 것이다. 또한, 산업데이터가 제3자에게 제공된 경우 산업데이터를 생성한 자와 제3자 모두 해당 산업데이터를 활용하여 사용·수익할 권리를 가진다. 다만, 당사자 간의 약정이 있는 경우에는 그에 따른다. 공유 내지 이용허락을 받은 경우를 전제하는 것으로, 이러한 경우에는 독자적으로 사용하거나 수익할 수 있는 권리를 갖도록 한 것이다.

문제는 “인적 또는 물적으로 상당한 투자와 노력을 투입한 자가 명확한 경우도 있겠으나, 여러 단계에 걸쳐 용역 계약을 체결하거나 다수의 기업이 데이터 생성에 관여하는 경우 등에는 산업데이터를 생성한 자가 명확하지 않아 분쟁이 발생할 여지가 있어 보” 66)인다는 지적이다. 산업데이터가 생성되는 과정이나 이용과정에서 명확하게 권리관계가 설정될 수 있을지 의문이다. 그렇기 때문에 산업데이터의 권리화에 대해서는 물권적인 권리가 아닌 채권적 형태로 권리화한 것임을 알 수 있다. 권리화된 산업데이터는 누구든지 공정한 상거래 관행이나 경쟁질서에 반하는 방법으로 침해하여 사용해서는 아니 된다. 이 경우, 공정한 상거래 관행이나 경쟁질서에 반하는 방법인지 여부를 판단할 때에는 산업데이터 활용의 목적 및 성격, 산업데이터의 활용이 그 산업데이터의 현재 또는 잠재적 가치에 미치는 영향 등을 종합적으로 고려하여야 한다. 기본적으로 부정경쟁법리가 차용된 것으로, 침해여부를 판단하는 기준은 저작권법상 공정이용(fair use) 요건을 차용하였다.⁶⁷⁾ 공정이용 첫 번째 요건인 이용의 목적 및 성격, 네 번째 요건인 저작물의 이용

65) 채수근, 산업디지털 전환 촉진법안, 산업의 디지털 전환 및 지능화 촉진에 관한 법안 검토보고서, 국회산업통상자원중소벤처기업위원회, 2020.

66) 채수근, 산업디지털 전환 촉진법안, 산업의 디지털 전환 및 지능화 촉진에 관한 법안 검토보고서, 국회산업통상자원중소벤처기업위원회, 2020.

67) 저작권법 제35조의5(저작물의 공정한 이용) ① 제23조부터 제35조의4까지, 제101조의3부터 제101조의5까지의 경우 외에 저작물의 통상적인 이용 방법과 충돌하지 아니하고 저작자의 정당한 이익을 부당하게 해치지 아니하는 경우에는 저작물을 이용할 수 있다.

② 저작물 이용 행위가 제1항에 해당하는지를 판단할 때에는 다음 각 호의 사항등을 고려하여야 한다.

1. 이용의 목적 및 성격
2. 저작물의 종류 및 용도

이 그 저작물의 현재 시장 또는 가치나 잠재적인 시장 또는 가치에 미치는 영향을 고려토록 하고 있다.

산업데이터의 활용과 그에 따른 이익배분이 문제될 수 있다는 우려 때문에 계약체결에 관한 사항을 두고 있다. 다만, 계약체결을 위한 노력의무(best effort)에 관한 규정이므로, 강제성은 없다는 점이 한계이다. 산업데이터 생성 또는 활용에 관여한 이해관계자들은 산업데이터의 원활한 활용과 그에 따른 이익의 합리적인 배분 등을 위한 계약을 체결하도록 노력하여야 하며, 합리적인 이유 없이 지위 등을 이용하여 불공정한 계약을 강요하거나 부당한 이익을 취득하여서는 아니 된다. 제3자가 산업데이터를 활용할 경우에는 산업데이터를 사용·수익할 권리를 가지는 자는 산업데이터의 무결성·신뢰성을 확보하고 산업데이터가 분실·도난·유출·위조·변조 또는 훼손되지 아니하도록 하며, 산업데이터를 활용한 제품·서비스가 위해를 발생시키지 않도록 노력하여야 한다. 기술적 보호조치를 확보토록 하고 있다. 다만, 부정경쟁방지법이나 저작권법과 같이 구체적인 요건이나 정의를 둔 것은 아니다. 사업자는 이에 준하는 수준의 조치를 해야 할 것으로 이해된다. 아울러, 산업데이터를 활용하는 등의 과정에서 고의 또는 과실에 의하여 타인에게 손해를 입힌 자는 그에 대한 손해를 배상할 책임을 진다.

(4) 산업데이터 활용 촉진

산업통상자원부장관은 산업데이터의 합리적 유통 및 공정한 거래 등 원활하고 안전한 산업데이터 생성·활용 환경을 보장하고 기업 등의 산업데이터 생성·활용 활성화를 위하여 필요한 지원을 할 수 있다. 합리적 유통 및 공정한 거래를 위해서는 유통시스템이나 거래 과정에서 필요한 표준계약서 등을 준비하는 것이 일반적이다.

산업통상자원부장관은 산업데이터 활용 및 보호 원칙을 준수하도록 하고, 계약의 체결을 촉진하기 위하여 관계 중앙행정기관의 장과 협의를 거쳐 산업데이터 활용 계약에 관한 지침을 마련할 수 있다. 관계 중앙행정기관과 협의하여 계약 지침을 마련할 수 있도록 하고 있으나 보통 표준계약서에 관한 규정을 두는 법률에서는 공정거래위원회와 협의토록 하고 있다. 일례로, 콘텐츠 관련 표준계약서의 경우, 문화체육관광부 장관과 공정거래위원장이 협의하도록 하고 있다.⁶⁸⁾

(5) 우선순위

데이터의 활용에 관하여 다른 법률의 보호규정이 있는 경우 동 제정안이 아닌 다른 법률의 규

3. 이용된 부분이 저작물 전체에서 차지하는 비중과 그 중요성

4. 저작물의 이용이 그 저작물의 현재 시장 또는 가치나 잠재적인 시장 또는 가치에 미치는 영향

68) “산업데이터에 대한 권리를 다수가 공동으로 소유하는 경우 그 활용·수익에 관하여 다툼이 생길 수 있으므로, 산업데이터 활용 및 이익의 합리적인 배분에 관한 계약을 체결하도록 하고 정부 차원에서 계약에 관한 가이드라인을 마련하는 내용은 타당한 측면이 있음. 다만, 공정거래위원회는 공정한 상거래 관행 및 경쟁질서 유지를 위한 가이드라인을 마련함에 있어 공정위와의 협의 절차를 거치도록 하여 그 실효성을 높일 필요가 있다는 입장” 이었다. 채수근, 산업디지털 전환 촉진법안, 산업의 디지털 전환 및 지능화 촉진에 관한 법안 검토보고서, 국회산업통상자원중소벤처기업위원회, 2020.

정을 우선적으로 적용받게 된다. 예컨대 산업데이터에 포함되어 있는 개인정보는 개인정보 보호법의 보호 규정을 적용받고, 타인의 산업데이터를 무단으로 활용하는 행위가 부정경쟁방지법에 따른 부정경쟁행위에 해당하는 경우 해당 법률에 따른 제재를 받게 된다. 법률간 우선순위를 설정하는 방식은 다른 법률에 의해 보호받지 못하는 경우에 보충적으로 적용되는 규정을 둠으로써 산업데이터에 관한 기본적인 보호 규정을 마련하고자 하는 동 제정안의 취지에 부합한다. 다만, 제정안에 따른 산업데이터에는 제조업뿐만 아니라 정보서비스업, 보건업 등 다양한 업종이 포함되어 있어⁶⁹⁾ 개인정보가 포함된 데이터가 당사자의 동의 없이 활용될 우려가 있으므로, 개인정보 보호와 관련해서는 개인정보 보호법을 우선 적용한다는 것을 명확히 규정할 필요가 있다.

제3절 특허법적 보호에 대한 검토

1. 특허법적 보호의 필요성 : 현행 법제하에서 데이터 보호의 한계를 중심으로 정리

앞서 살펴본 바와 같이, 데이터 관련 법제는 다양하다. 먼저, 데이터의 내용을 보호하는 저작권법은 데이터의 집합을 의미하는 데이터베이스 제작자의 권리를 보호기도 한다. 부정경쟁방지법은 부정경쟁행위 유형을 통하여 보호를 한다. 데이터 산업법은 데이터 자산에 대한 보호를 한다. 다만, 데이터 자산이 민법상의 권리라기 보다는 데이터를 자산적 가치의 척도화할 수 있는 채권적인 권리라는 점에서 보호가 제한적이다.

[표 2-4] 데이터의 종류 및 보호 현황

구분	비공개 데이터	공개 데이터
정형 데이터	부정경쟁방지법에 따라 영업비밀로서 보호 ①민사상 손해배상청구권, 금지청구권, 신용회복청구권 ②형사처벌(해외유출	저작권법에 따라 편집저작물 또는 데이터베이스로서 보호 -(창작성 有) 사망 다음해부터 70년 -(창작성 無) 제작 완료/갱신한 때로부터 5년간 ①민사상 손해배상청구권, 금지청구권, 명예회복청구권 ②형사처벌(5년/5천만원, 병과 가능)
비정형 데이터	15년/15억원, 국내유출 10년/5억원)	부정경쟁방지법에 따른 데이터부정경쟁행위 유형으로 보호

<출처 : 특허청(2021)자료 & 연구진 수정>

69) 산업발전법 시행령 [별표1]에 따른 산업의 범위(제2조 관련)에는 소프트웨어 개발 및 공급업, 방송업, 연구개발업, 인력공급 및 고용알선업, 신용조사 및 추심대행업, 정보서비스업, 온라인 교육 학원, 병원, 의원, 기타 보건업 등이 포함되어 있다.

그렇지만, 데이터가 기술적인 수단으로 활용되는 과정에서 발생할 수 있는 흠결을 고려하지 않을 수 없다. 저작권법은 기술적 사상이 체화된 데이터의 구조에 대해서는 보호범위에 포함되지 아니한다. 또한, 데이터 구조는 발명을 통하여 구현되기 때문에 영업비밀성을 갖추기 어렵다. 이러한 점에서 데이터의 보호가능성에 더하여 활용가능성에 대한 정책적인 판단이 필요하다.

2. 검토 내용

데이터를 활용하는 발명에서 데이터 구조는 물건발명으로서 인정되고 있으나 데이터셋은 견해가 갈린다. 즉, 현행 특허법은 데이터 자체를 보호할 수 없다는 것이 일반적인 견해이다.⁷⁰⁾ 데이터는 자연법칙을 이용한 기술적 사상이라고 보기 어렵기 때문이다. 다만, 명세서의 발명의 기재를 통하여 보완할 수 있는 가능성 또는 심사기준을 개정할 수 있는지 여부를 검토할 수 있을 것이다. 따라서, 데이터 자체는 특허법적 보호의 대상에서 제외된다. 다만, 데이터셋은 구조화된 데이터의 집합이라는 점에서 데이터 구조 자체화는 다른 기술적 사상으로 볼 수 있다는 점에서 차별성을 가질 수 있다.

가. 현행 심사기준

인공지능 관련 발명에 대한 특허청의 심사기준은 인공지능 관련 발명의 상세한 설명에 학습데이터를 나열만 하고 있을 뿐 입력 데이터와 학습된 모델의 출력 데이터 간의 상관관계를 구체적으로 기재하고 있지 않고, 통상의 기술자가 출원시의 기술상식으로 발명의 설명에 기재된 실시예를 통해서 그 상관관계를 추정하기 어려운 경우에는 등록요건 중 하나인 실시가능 요건을 만족하지 않은 것으로 판단될 수 있다.

이러한 점에서 학습데이터는 인공지능발명에 없어서는 안되는 필수 요소이고 인공지능이 탑재된 기계를 생산하는 데에만 사용되는 물리적 객체라고 봐도 무방할 것이다. 우리나라 인공지능 관련 발명 심사기준은 학습모델 자체에 특징이 있는 경우, 청구항에 학습모델에 관해 구체적으로 특정해 그 특징에 신규성 및 진보성이 인정된 경우에 등록을 허용하도록 가이드하고 있다.

나. 심사기준의 해석과 적용

(1) 심사기준의 규정

1) 물건발명

심사기준에서 AI발명을 포함한 SW발명을 물건발명이나 방법발명으로 보고 있다. 인공지능 관

70) 정원준 외, AI 학습데이터의 특허법적 보호방안 및 산업계 영향 분석, 특허청, 2021 참조.

런 발명은 그 발명을 구현하는 복수의 기능으로 표현할 수 있을 때 그 기능으로 특정된 물건(장치)의 발명으로 청구항에 기재할 수 있다. 또한 인공지능 관련 발명은 아래와 같이 『컴퓨터프로그램 기록매체 청구항』, 『기록매체에 기록된 컴퓨터프로그램 청구항』, 『데이터 구조 기록매체 청구항』 형식에 추가하여, 학습모델이나 학습모델을 이용하는 물건을 청구하는 경우에는 『기록매체에 저장된, 학습모델을 구현한 컴퓨터프로그램 청구항』, 『학습모델을 이용하는 물건(장치) 청구항』 형식으로 기재할 수 있다.

데이터 구조 기록매체 청구항 데이터 구조 기록매체, 즉 기록된 데이터 구조로 말미암아 컴퓨터가 수행하는 처리 내용이 특정되는 ‘구조를 가진 데이터를 기록한 컴퓨터로 읽을 수 있는 기록매체’, 또는 ‘데이터 구조(data structure)를 기록한 컴퓨터로 읽을 수 있는 기록매체’는 물건의 발명으로 청구항에 기재할 수 있다.

(예) 컴퓨터에서 실행되는 A 구조, B 구조, C 구조, …를 가진 데이터 구조를 기록한 컴퓨터로 읽을 수 있는 기록매체

2) 방법발명

인공지능 관련 발명은 시계열적으로 연결된 일련의 처리 또는 조작, 즉 단계로 표현할 수 있을 때 그 단계를 특정하는 것에 의해 방법의 발명으로 청구항에 기재할 수 있다

(2) 심사기준의 해석과 적용

데이터의 구조화에 대해서도 청구항의 기재방식에 따라 방법발명 내지 물건발명으로 등록되고 있다는 점에서 데이터의 핵심적인 내용은 보호범위에 포함되어있음을 알 수 있다. 그렇지만, 데이터 자체의 보호가능성에 대해서는 명확하지 않다는 견해이다. 과연 그러한 것인지 살펴보고자 한다.

우선 데이터의 속성을 보면 데이터 자체는 정형화되거나 비정형화된 정보의 구조라는 점에서 정보 그 자체로 볼 수 있으나, 데이터의 집합으로써 데이터셋을 보면 데이터간 연계 구조가 발생한다는 점에서 데이터와는 논리적 성질이 다르다. 즉, 데이터셋은 일종의 SW로서 또는 알고리즘으로서의 성격을 갖게 된다는 점이다.

여기서 일반적인 데이터와 데이터셋의 차이점에 대해 2가지의 고려할 사항이 제기된다.

첫째, 데이터의 집합체인 데이터셋이 논리적인 구조를 갖는다는 점에서 지금의 데이터 관련 발명이 물건발명 내지 방법발명으로 등록되는 것과 크게 다르지 않게된다는 점이다. 즉, 구조로 이루어진 데이터셋을 청구항으로 기재할 경우에는 이를 거절할 수 있는 논리가 부재하다는 점이다.

둘째, 데이터셋을 물건발명으로 등록받을 경우에는 물건성을 암묵적으로 인정하고있다는 점에

서 물건발명에 따른 간접침해의 적용가능성을 부인하기 어렵다는 것이다. 아울러, 특허법 제94조에 따른 실시의 경우에도 고의성을 배제할 경우에는 이미 온라인을 통하여 데이터 관련 발명이 실시되는 경우라면 적용가능성을 부인하기 어렵다. 더 나아가, 특허법 제127조에 따른 간접침해 규정을 적용할 수 있다고 해설될 여지는 있다.

현행 법체계 내에서도 단수로서 데이터는 모르겠지만, 데이터의 집합인 데이터셋의 성격상 논리적인 구조체로서 이루어진다는 점에서 데이터 발명을 거부할 실익은 없다고 판단된다. 이러한 해석론을 부정하는 것은 이미 다양한 형태로 AI, SW, 데이터 발명이 물건발명의 카테고리에 발명특허로서 등록이 이루어지고 있다는 현실을 외면하고 있는 것은 아닌지 살펴볼 필요가 있다. 또한, 발명자의 입장에서 명확한 가이드라인을 제시하는 것이 법적안정성과 예측가능성을 높일 수 있다는 점에서 심사기준을 개정하는 것은 필요하다. 현재의 데이터 내지 구조화된 데이터를 넘어서 데이터셋에 대한 방식도 명식적으로 제시하는 것이다. 아울러, 기술공개 요건에 대해서도 살펴보겠지만, 보다 구체화된 공개요건을 제시할 수 있을 것이다. 데이터셋을 발명으로 포함할 경우에는 데이터셋의 일부나 데이터 구조 등에 대해 설명함으로써 용이하게 실시할 수 있도록 하는 것이다.

다. 특허법상 데이터 발명

데이터 내지 데이터 발명에 대해 명시적으로 특허법을 개정하자는 주장이 발견된다.⁷¹⁾ 이러한 주장은 특허정책에 부정적인 영향을 줄 가능성이 크다. 더욱이, 수많은 분야의 발명에서도 각각의 주장을 펼 수 있을 것이나 이는 특허정책상 효용성이 크지 않다. 물론, 주장에 대한 합리성을 떠나 학자적 견해의 제시라는 점에서 긍정적이라고 볼 수 있다.

기술공개에 따른 요건으로 데이터에 대한 특허정책은 우리나라만의 이슈라기 보다는 글로벌 환경에서 공통된 관심사라고 보기 때문에 IP5 내지 WIPO 차원에서 대응하는 것이 바람직하다고 판단된다. 다행히도 WIPO에서는 어느정도 대응하고 있음을 확인하였다. 이에 대해서는 다음 장에서 구체적으로 살펴보고자 한다.

71) 이규호, 인공지능 학습용 데이터셋 보호를 위한 특허법상 주요 쟁점 연구, 산업재산권 제64호, 2020.07; 권지현, AI발명에 있어서 데이터의 물건특허 인정방안, 서울法學 제28권 제4호, 2021.

제3장. 기술공개 요건으로서 데이터 공개에 대한 연구

제1절 문제 제기

1. AI의 블랙박스화와 데이터 공개

가. 기계학습의 특성⁷²⁾

기계학습의 대표적인 알고리즘인 딥러닝은 인공신경망구조(ANN)⁷³⁾에 따라 이루어지기 때문에 여러 가지 효과를 가져오면서도 문제점을 동시에 갖게 되었다. 즉, 딥러닝은 뇌의 신경망(neural network)처럼 알고리즘으로 구현하여 학습하는 것으로, “데이터를 군집화하거나 분류하는데 사용되는 방법론으로써 하나의 데이터를 입력해주면 컴퓨터가 스스로 학습하면서 비슷한 데이터들의 패턴을 찾아내어 분류⁷⁴⁾하는 방식”⁷⁵⁾으로 이해할 수 있다. 인공신경망 기반의 딥러닝은 인공지능을 위한 논리를 만드는 게 아니라 논리를 만들도록 설계해두고 기계가 스스로 학습하도록 하는 것을 의미하며, 데이터에 담긴 다양한 경험(특징)을 통하여 이를 기반으로 논리를 만들어가는 것이다. 딥러닝은 데이터에서 지식을 추출하는 방식이 아닌 데이터에서 특징(feature)을 분석하여 이를 신경망에 저장하는 방식으로, 신경망에 저장된 특징값은 데이터를 통하여 학습한 결과물이다. 이렇게 활용되는 딥러닝은 “데이터 분석 및 예측, 영상인식, 음성인식, 자연어 처리, 보안 등 다양한 관련 분야의 기술수준을 향상시켰다”⁷⁶⁾고 평가된다. 그렇지만, 기술적 활용과 다르게 법적 쟁점이 될 수 있으며, 이는 입출력 과정 이외에 중간에 있는 은닉층(隱匿層)에서 진행되는 내용에 대해 누구도 알 수 없기 때문이다.⁷⁷⁾

이러한 딥러닝에서 가장 중요한 요소는 데이터로서, 데이터의 질은 학습결과를 달라지게 만들

72) 김윤명, 블랙박스를 열기위한 인공지능법, 박영사, 2022, 64~66면.

73) 인공신경망에 대해서는 다음 사이트 참조.

https://velog.io/@e_sin528/%EB%94%A5%EB%9F%AC%EB%8B%9D-%EC%9D%B8%EA%B3%B5%EC%8B%A0%EA%B2%BD%EB%A7%9DANN%EC%9D%98-%EC%A2%85%EB%A5%98
<2022.11.22. 접속>

74) 기계학습의 분류는 실제 학습하는 것을 말하며, “컴퓨터가 스스로 높은 차원인 특징을 획득하고, 그것을 바탕으로 이미지를 분류할 수 있게 된다”는 의미이다. Yutaka Matsuo, 인공지능과 딥러닝, 동아엠앤비, 2015, 150면.

75) 또한, 데이터를 분류하는 기계학습 알고리즘들은 현재에도 많이 활용되고 있으며, 대표적으로는 의사결정나무, 베이지안망, 서포트벡터머신(SVM), 인공신경망 알고리즘을 들 수 있다. 이 중에서도 딥러닝은 인공신경망(ANN) 알고리즘을 기반으로 입력 계층(Input Layer)과 출력 층(Output Layer) 사이에 복수의 은닉 계층(Hidden Layer)이 존재하는 심층 신경망(Deep Neural Network) 이론이 등장하면서 현재의 딥러닝 알고리즘이 탄생하였다. 정보화진흥원, near&future 제19호, 2016, 24면.

76) 김인중, 기계학습의 발전 동향, 산업화 사례 및 활성화 정책 방향, 이슈리포트 2015-017, 소프트웨어정책연구소, 2015, 4면.

77) “중간 은닉층에서 이루어지는 오류보정과 가중치 변경 등을 전문가가 아닌 법관 등 법 집행기관이 이해하기에는 이를 뒷받침할 만한 과학적 추론이나 설명이 충분할 수 없다”고 한다. 양종모, “인공지능 이용 범죄예측 기법과 불심검문 등에의 적용에 관한 고찰”, 형사법의 신동향 통권 제51호, 2016.6, 233면.

고,⁷⁸⁾ 특히 데이터가 사용됨으로써 효과를 높이게 된다.⁷⁹⁾ 빅데이터의 의미와 가치에 대해 “데이터에 수많은 상관관계(correlations)가 존재한다는 것이다. 내 소비 패턴, 학습 패턴, 인터넷 검색 패턴, 병원 방문 패턴, 전화 이용 패턴 등이 종합되면, 나에 대해서 (심지어 자신도 모르는) 많은 상관관계가 생길 수 있다. 이런 데이터가 나에 대해서만이 아니라 다른 사람에 대해서도 수집이 되고, 횡적으로 비교가 되면 훨씬 더 많은 상관관계가 생기게 된다. 여기에 패턴으로부터 스스로 배우는 기계학습(machine learning) 알고리즘이 결합하면, 내 행동에 대해 예측이 가능해진다.”⁸⁰⁾는 평가가 가능하다.

데이터의 양과 의도성 및 그에 담겨있는 상관관계는 그만큼 데이터가 중요하게 된 이유이다. 물론, 플랫폼사업자는 데이터의 수집이나 활용에 큰 무리는 없을 것이다. 이용자가 제공한 다양한 정보를 이용할 수 있는 권한을 갖기 때문이다. 예를 들면, 구글이나 페이스북, 네이버 등 대부분의 플랫폼사업자는 이용약관을 통하여 이용자 정보를 무상으로 이용할 권한을 확보하고 있다.⁸¹⁾ 전통적인 제조기업과 달리 플랫폼 기업은 데이터 확보에 시간과 비용을 들이지 않아도 된다. 물론, 전통적인 제조기업도 IoT기술을 도입하여 보다 용이하게 산업데이터를 확보할 수 있게 되었다.⁸²⁾

인공지능 알고리즘에서 실질적으로 데이터가 사용되는 것은 기계학습 과정으로 이 과정은 블랙박스화에 따라, 다음과 같은 특성이 있다. 실질적으로 내부적인 처리과정을 외부에서 확인하기 어렵다는 점, 심사관이 파악하기 어렵다는 점, 블랙박스화에 따른 사회적인 편익이 저해될 수 있다는 점이 문제로 제기된다.

나. 블랙박스화에 대응

데이터를 포함한 AI에 대한 공개를 주장하는 경우에는 전통적인 SW에 비하여 AI발명에 대해서는 AI의 블랙박스화에 따른 투명성 확보, 신뢰성을 확보 등을 이유로 알고리즘이나 소스코드를 포함하여 더욱 확대된 공개를 주장한다.⁸³⁾ 기본적으로 AI는 설명가능성이 부재하다. 데이터의 입출력에 따른 결과는 AI모델을 통하여 이루어지기 때문에 이 과정을 설명하는 것은 불가능한 영역이기 때문이다. 이러한 이유로 알고리즘이나 데이터의 공개는 발명의 재현성 확보라는 점에서

78) “인터넷 시대 도래와 빅데이터 기술 발전으로 인해 인공지능이 ‘정확성’과 ‘데이터 무제한 보유’ 등 두 가지 강점을 겸비하게 된 셈”이라고 한다. 삼성 뉴스룸, 2016.7.20.일자.

79) 김인중, 기계학습의 발전 동향, 산업화 사례 및 활성화 정책 방향, 이슈리포트 2015-017, 소프트웨어정책연구소, 2015, 5면.

80) 오요한 외, 인공지능 알고리즘은 사람을 차별하는가?, 과학기술학연구 제18권 제3호, 2018, 162면.

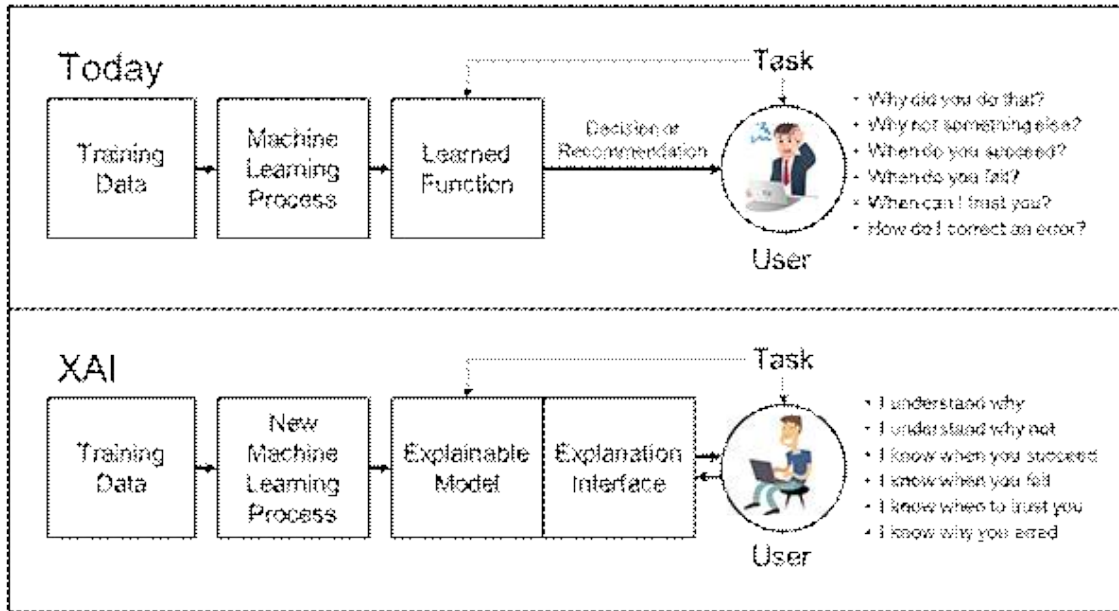
81) 특히, 플랫폼 기업은 이용자의 거래 관계를 통하여 얻는 데이터 또는 이용자가 플랫폼 서비스에 게시하는 다양한 사진 저작물과 동영상의 이미지, 또는 키워드 검색 등을 통하여 데이터를 획득하게 된다. 물론, 텍스트 형태의 게시글도 데이터 마이닝을 통하여 훌륭한 데이터가 되기 때문에 딥러닝의 효율성을 높일 수 있다.

82) 중견 및 중소 가전제품 제조사들은 한국정보통신기연합회를 중심으로 IoT가전을 활용하여 데이터를 수집 및 분석하는 플랫폼을 구축 및 운영하고 있다.

83) Tabrez Ebrahim, Artificial Intelligence Inventions & Patent Disclosure, Iowa Legal Studies Research Paper No. 2021-48, p.148.

용이실시를 위한 요건으로서 필요성을 논하는 것이다. 물론, 인공지능 자체에 설명가능한 알고리즘을 부가함으로써 인간이 이해할 수 있도록 하는 기술도 개발 중이다. 설명가능성이나 설명가능한 인공지능(eXplainable AI)⁸⁴⁾이란, 인공지능 모델이 특정 결론을 내리기까지 어떤 근거로 의사결정을 내렸는지를 알 수 있게 하는 것을 말한다.⁸⁵⁾

[그림 3-1] 설명가능한 인공지능의 Concept



<출처 : DARPA, 2018>

설명가능한 인공지능은 블랙박스화되고 있는 알고리즘에 대해 신뢰성과 투명성을 높이기 위한 기술적 접근이다. AI발명을 공개함에 있어서 데이터에 대한 공개도 고려할 수 있을 것이다. 예를 들면, AI발명의 특성이나 학습에 사용된 학습데이터의 특성에 따라, 알고리즘의 공개, 소스코드의 공개, 전체 데이터셋의 공개, 데이터 일부에 대한 샘플링, 일부 데이터셋의 공개, 공개된 사이트의 주소, 기탁처 등에 대한 사항을 고려할 수 있을 것이다. 이에 대해서는 뒷부분에서 구체적으로 다루고자 한다.

2. AI발명의 구조와 데이터의 관계

AI발명에 있어서 중요한 것은 AI발명에서 사용되는 기술의 공개에 관한 사항이며, 본 연구에서 탐구하고자 하는 구체적인 사항은 데이터가 포함되는 발명에 있어서 기술공개 수준과 범위이다. AI발명은 SW의 발명과 비교했을 때, 코딩이라는 면보다는 데이터 기반의 기계학습을 통하여

84) <https://www.darpa.mil/program/explainable-artificial-intelligence>. <2022.11.22. 접속>

85) 안재현, XAI 설명가능한 인공지능을 해부하다, 위키북스, 2020, 4면.

알고리즘이 구체화된다는 점에서 실상 기술자가 인지할 수 있는 요소기술을 얻는 것은 SW에 발명에 비해 낮을 수 있기 때문이다. 물론, AI발명의 내부적인 알고리즘은 더욱 고도화되고 있다는 점에서 산업상 활용가능성은 훨씬 높다.

그럼에도 불구하고, AI발명이 가지고 있는 한계점은 기술발전을 위한 특허법의 목적에 비추어 보았을 때 기술공개 수준과 범위가 높지 않을 수 있다는 점이다. 더욱이 데이터 처리를 필요로 하는 발명에 있어서 데이터의 구조화 내지 구조화된 데이터의 표현은 기술적(descriptive)이기 때문에 데이터 구조를 구체적으로 공개하는 것이 필요할 수도 있다. 표준화된 알고리즘을 사용하는 경우에는 입출력 데이터에 있어서 기술적 기여만 있다면 발명의 성립성에 있어서 만족할 수 있을 것이다.⁸⁶⁾ 따라서, 데이터나 데이터셋의 공개 또한 충분한 기술공개(sufficiency of disclosure)에 해당할 수 있을 것이다.

다만, 논의과정에서 고려할 사항은 실제 AI기업이나 AI관련 사업영역에서 기술공개에 대해 민감하게 반응할 수 있다는 점이다. 기술공개에 따른 독점권을 부여하는 것은 특허법이 추구하는 발명의 활성화를 위하여 발명자에게 부여하는 일종의 특권이라는 점이다. 그럼에도 불구하고 기술공개는 기업에 있어서 경쟁력을 확보하기 위한 기본적인 수단이며 이러한 수단을 공중에게 공개하는 것은 리스크로 작용될 수 있기 때문이다. 이러한 발명자의 우려에 대해서는 뒷부분에서 살펴보겠지만, 다양한 인센티브를 부여함으로써 해소할 수 있을 것으로 보인다. 또한 학습데이터셋과 관련하여 영업비밀 보호제도가 특허제도와 완전히 배치되는 것은 아니라는 점도 고려될 수 있다.

3. 데이터 보호 법익과 데이터 공개의 충돌

기술공개의 요소로서 데이터의 구조 내지 데이터셋(data set)의 공개가 요구될 경우에는 자칫 기술공개 자체를 거부할 가능성도 있다. 즉, 데이터 관련 발명을 보호하겠다는 선의의 취지가 오히려 데이터 공개라는 요건으로 인하여 부정적일 수 있다.

추측컨대, 데이터셋의 성격에 따라서 기업의 입장에서 일관되지 못할 수 있다. 먼저, 데이터가 개인정보를 침해할 수 있는 경우에는 위법성을 밝히는 것이 될 수 있다는 점이다. 물론 기업은 개인정보 보호법 등의 강력한 규제사항에 부합하게 데이터를 활용하겠지만 의도하지 않게 수집 및 가공되는 경우도 상정할 수 있기 때문이다.

86) 일반적으로 컴퓨터는 제어장치, 논리·연산장치, 기억장치 및 입출력장치로 구성되어 있고, 그 기계적 설비인 하드웨어는 독자적인 작업수행 능력이 없고 소프트웨어인 프로그램의 작업수행 지시에 따라 특정목적을 위한 제어·논리 및 연산·기억 등의 기능을 발휘하는 것이어서 하드웨어 자체는 범용성이 있다고 할 것이므로, 컴퓨터를 기능실현수단으로 이용한 장치발명의 출원에 있어서 그 장치에 고유한 독자적인 작업수행능력을 갖도록 특별히 고안된 하드웨어를 사용한다면 모르되 범용성이 있는 하드웨어를 사용하는 경우에는 하드웨어 자체의 구성에 관한 상세한 설명을 특허출원서에 일일이 기재하지 아니하더라도 컴퓨터와 관련된 기술분야에서 평균적 기술능력을 가진 자이면 하드웨어의 기능내용을 능히 이해할 수 있다고 보는 것이 타당하다. 대법원 1987. 9. 29. 선고 84후54 판결 [거절사정].

또한, 저작권법에 위배되는 경우도 있다는 점이다. 인터넷 상에 공개된 정보를 가공하여 학습 데이터로 가공한 경우도 있을 수 있다. 이러한 경우에는 데이터셋을 공개하는 것은 개인정보가 포함된 데이터의 공개와 마찬가지로 위법성을 밝히는 것이 될 수 있다는 점에서 회사에서는 부정적인 입장이 될 수 있다.

계약을 통하여 이용허락을 받은 데이터셋은 계약 내용에 따라 이용 범위와 공개 여부 등에 대하여 합의되어있을 것이다. 만약, 계약조건을 위반하여 공개해야 한다면 이는 의도하지 않게 계약위반이 될 수 있다. 따라서 당사자간 합의를 깨트리는 요건으로 작동할 수 있다는 점도 우려할 수 있는 사항이다.

끝으로, 기업의 기밀정보나 영업비밀에 해당하는 데이터가 포함되는 경우도 있다. 학습데이터가 기업의 경쟁력을 좌우하는 주요한 요소자산이 되고있다는 점에서 영업비밀로 보호받는 데이터를 공개하는 것은 비밀유지성을 상실할 수 있기 때문에 기업의 입장에서는 특허를 받을 수 있다고 하더라도 데이터의 공개를 선택하지는 않을 가능성이 크다.⁸⁷⁾ 이러한 이유 때문에 영업비밀로 유지하거나 또는 기술공개를 하지 않는 방안으로써 특허출원 자체를 보류할 가능성도 예상된다. 물론 데이터의 적시성이나 라이프사이클에 따른 효용성을 판단하겠지만, 기업으로서는 경쟁 업체에게 자신의 학습데이터를 공개하는 것에 대해서 우호적이라고 보기 어렵기 때문이다. 이러한 우려를 고려하여 각국에서는 어떠한 공개정책을 펴고 있는지 살펴보고 우리도 AI발명의 기술공개에 있어서 데이터를 포함하는 것에 대한 정책적 결정에 참고할 수 있을 것이다.

제2절. 기술공개와 관련 제도

1. 기술공개제도의 의의

현행 특허제도에 따르면 발명을 공중에게 공개함으로써 특허권이라는 독점권이 부여된다. 명세서에 부기된 기술적 사항의 구체적인 설명이 공개됨으로써 새로운 기술투자를 유인할 수 있다는 점이 발명제도가 갖는 선순환적인 구조이다.

가. 특허법의 목적⁸⁸⁾

기술 법제인 특허법은 발명을 보호·장려하고 그 이용을 도모함으로써 기술발전을 촉진하여 산업발전에 이바지함을 목적으로 한다. 특허법은 자연법칙을 이용한 기술적 사상의 창작으로서

87) ‘영업비밀’이란 공연히 알려져 있지 아니하고 독립된 경제적 가치를 가지는 것으로서 상당한 노력에 의하여 비밀로 유지된 생산방법, 판매방법 그 밖에 영업활동에 유용한 기술상 또는 경영상의 정보를 말하고, 여기서 ‘상당한 노력에 의하여 비밀로 유지된다’는 것은 그 정보가 비밀이라고 인식될 수 있는 표시를 하거나 고지를 하고, 그 정보에 접근할 수 있는 대상자나 접근 방법을 제한하거나 그 정보에 접근한 자에게 비밀준수의무를 부과하는 등 객관적으로 그 정보가 비밀로 유지·관리되고 있다는 사실이 인식 가능한 상태인 것을 말한다. 대법원 2010. 12. 23. 선고 2008다44542 판결 [저작권침해 금지등]

88) 김윤명, 발명의 컴퓨터 구현 보호체계 합리화를 위한 특허제도 개선방안 연구, 특허청, 2014 참조.

고도(高度)한 것인 발명을 보호하고 장려하고, 아울러 발명의 이용을 도모한다. 특허법의 목적은 여기에만 머무르지 아니한다. 발명의 보호·장려·이용을 통해서 기술발전을 촉진하는 것이 또 하나의 목적이다. 발명은 이용하는 과정에서 기술혁신을 이끌어낸다.⁸⁹⁾ 기술의 보호 및 장려를 통하여 얻은 발명을 이용함으로써 기술혁신을 가져오고 결과적으로 산업을 발전시키고 국민경제에 기여하도록 하는 것이 특허법의 궁극적인 목적이다.⁹⁰⁾ 이런 목적에 따라 공공의 이익을 위하여 발명에 독점권을 부여하고, 대신 해당 기술을 공개하고 실시하도록 의무를 부여하는 것이 특허제도의 취지이다. 이처럼, 특허제도는 발명가들이 특정기간 동안 독점적 이익을 얻을 권리의 대가로 자신들의 발명에 대한 광범위한 공개를 제공해야 한다는 요구이다. 독점성을 제공하는 것은 잠재적으로 치명적인 독점 비용을 발생시키지만, 목표는 특허의 기초가 되는 아이디어를 구축하는 데 필요한 정보를 다른 혁신자들에게 제공하여 일반 지식을 증가시키고 경제성장을 촉진함으로써 혁신을 촉진하는 것이다.⁹¹⁾ 또한, 특허제도는 공개하는 대가로 일정기간 동안 독점권을 부여하는 ‘공적 계약’ 이기도 하다.⁹²⁾ 따라서 공적 계약의 요건 중 하나인 기술공개가 제대로 이루어지지 않으면 계약의 해지사유에 해당한다고 볼 수 있다. 그렇기 때문에 특허법은 기술공개를 위해서 특허명세서를 작성토록 하고 있으며, 명세서의 기재사항을 자세히 법정요건으로서 규정하고 있다. 만약, 이러한 요건을 갖추지 않을 경우에는 거절 내지 무효사유에 해당하는 것이다. 물론 기술공개 수준이나 범위에 대해서는 논란이 있으며, 이는 첨단기술의 경우에는 더욱 논란이 커질 수 있을 것이다. 기술적 사상이 눈에 보이는 것이 아닌 눈으로 이해할 수 있는 수준의 문서로 구성되기 때문이다.

그럼에도 불구하고, 특허법이 발명특허만을 보호한다는 것으로 오해하고 있으나, 이는 특허제도를 운영하는 과정에서 정책이 보호와 이용이라는 형평을 맞추지 못한 점에서 기인한 것이다. 물론 특허제도는 모든 발명을 인정하는 것은 아니다. 산업상이용가능성이 있더라도, 공서양속에 위배되는 발명에 대해서는 특허의 대상에서 제외하기 때문이다. 기술발전과 공서양속이 비례하는 것은 아니지만, 때로는 기술이 보편적인 윤리에 벗어날 경우에는 정책적으로 배제시킬 수 있는 것은 사회질서의 유지를 위한 것임을 부인하기 어려울 것이다.

나. 기술공개 효과 : 기술혁신

특허법의 목적은 발명의 장려를 통하여 기술축진을 이끌어내는 것임을 살펴보았다. 기술 혁신

89) “특허발명의 강제실시에 관한 규정도 연구시험의 예외와 마찬가지로 특허법의 목적에 해당되는 발명의 이용 및 기술발전의 촉진을 효율적으로 달성하기 위해서 필연적으로 요구되는 제도적 장치”라고 한다. 정상조·박성수 공편, 특허법 주해 I, 박영사, 2010, 20면.

90) “특허법의 목적은 발명과 경쟁이 모순충돌의 관계에만 있는 것이 아니라 상호보완할 수 있다는 전제로 하고있다는 것”이라고 한다. 손경한 편저, 신특허법론, 법영사, 2005, 41면.

91) Romer, P. “Endogenous technological change.” *Journal of Political Economy* 98, no. 5, Part 2 (1990): S71-S102; Travis A. Dyer, et. al. The Effect of Patent Disclosure Quality on Innovation, September 14, 2020. Available at SSRN: <https://ssrn.com/abstract=3711128>; Fromer, Jeanne C., Patent Disclosure. *Iowa Law Review*, Vol. 94, 2009, p. 541.

92) 조용선, 특허법 2.0, 박영사, 2018, 31면.

이나 축진이 갖는 의미는 다양하다.⁹³⁾ 단순하게 개선이 이루어진 경우라도 그 가치가 높은 경우도 있지만, 수많은 자본이 투입된 경우라도 그 가치를 인정받기 어려운 경우도 있기 때문이다. 결국, 최종적인 가치의 판단은 시장에서 이루어질 수밖에 없다. 특허제도는 기술의 혁신을 가져오는가? 기술혁신은 다양한 기술의 공개를 통하여 아이디어를 얻고, 그 아이디어를 개량함으로써 보다 진보한 기술을 개발할 때 일어난다. 모든 기술이 기존에 없던 것이 아닌 기존에 존재한 기술을 개량함으로써 진보가 일어난다. 특허 출원된 기술적 사상은 공개되기 때문에 누구라도 이용할 수 있다. 그렇기 때문에 발명을 장려하고, 특허출원토록 함으로써 기술적 사상을 외부에 공개하고, 해당 기술의 존재 여부를 확인함으로써 중복개발의 문제를 해소할 수 있다. 기술공개는 기술과 산업발전이라는 사회적 기여를 한다는 점에서 보상을 받는 구조이다.⁹⁴⁾ 따라서, 특허법에서 요구하고 있는 공개조건을 만족해야하는 조건부 공적 계약인 것이다. 물론, 기술혁신의 판단은 발명특허의 신규성과 진보성의 유무로 확인할 수 있다. 특허법의 신규성과 진보성 등의 요건은 특허발명의 보호를 통하여 기술의 발전을 촉진하고 장려하는 기능을 수행한다.⁹⁵⁾ 특허제도는 새로운 기술을 창작한 자에게 일정기간 독점적 이익을 부여함으로써 창작자의 노고에 보답하고, 아울러 그 독점기간이 끝나면 누구나 당해 기술을 이용하도록 함으로써 산업발전을 도모한다. 인공지능 기술의 혁신도 어느 한 기업이나 하나의 기술이 아닌, 다양한 기술과 인력의 직간접적인 협력을 통하여 이루어지고 있다.

또한, 특허제도는 새롭고 유용한 발명의 공개를 장려하기 위해 고안되었으며, 발명의 완전하고 완전한 공개를 제공하는 것은 장기적으로 발명자에게 유익할 수 있다. 많은 경우, 데이터 공개는 종종 발명의 가능성을 입증하는 것이 요구되기 때문에 특허출원 과정에서 필수적인 사항이다. 공개요건을 충족하고 특허를 획득하기 위해서는 데이터 공개가 필요한 경우가 많다. 특허출원은 관련 분야의 숙련된 사람이 과도한 실험 없이 발명을 만들고 사용할 수 있도록 충분한 정보를 제공해야 한다. 경우에 따라, 이것은 발명자가 특허출원 또는 데이터 기탁으로서 발명에 대한 데이터 또는 다른 정보를 공개할 것을 요구할 수 있다.⁹⁶⁾

이처럼 기술공개는 기술 축적을 가져오고 축적된 기술을 활용함으로써 새로운 발명을 위한 밑바탕이 될 수 있다. 기술공개는 기술적인 측면에서의 효용성도 중요하며, 더 나아가서 사회적 후생에 있어서도 효과적이다. 기술공개는 독점권을 부여하기는 하지만 공개를 통하여 중복적인 투자를 억제할 수 있으며, 이는 중복투자에 따른 비용은 사회적 비용으로 전환될 수 있기 때문이다. 결국, 특허제도의 궁극적인 목표는 기술공개를 통해 새로운 디자인과 기술을 공공영역(public

93) 독점은 상대 기업의 시장진출을 차단하여 기술혁신을 저해하고 스스로도 경쟁이 요인을 사라지게 함으로써 기술혁신의 악순환을 가져올 수 있다. 물론 상대 기업도 기술혁신을 통해서 독점기업을 넘어설 가능성도 있겠지만, 현실적인 재원이 소요되는 기술개발에서 쉽지 않다고 본다.

94) 인류에 대한 기여라는 주장도 있다. 정우성, 특허문서론, 에이콘, 2017, 38면.

95) 정상조 박성수 공편, 특허법 주해 I, 박영사, 2010, 15면.

96) Hagen, Gregory R., AI and Patents and Trade Secrets (February 1, 2021). in Florian Martin-Bariteau & Teresa Scassa, eds., Artificial Intelligence and the Law in Canada (Toronto: LexisNexis Canada, 2021), Available at SSRN:<https://ssrn.com/abstract=3734654>

domain)에 도입하는 것이라고 하겠다.⁹⁷⁾

2. 기술공개로서 발명의 상세한 설명

가. 의의

특허는 발명의 내용을 공중인 일반 국민에게 공개하는 대가로 부여받는 독점권이다. 기술공개라는 제도적 취지를 충분히 달성하기 위해서는 발명의 상세한 설명에는 출원인이 알고 있는 발명의 모든 내용이 최대한 자세하고 공개되어야 한다. 즉, 기본적으로 쉽게 실시할 수 있을 정도로 기재되어야 한다.⁹⁸⁾ 발명의 상세한 설명은 제3자로 하여금 당해 발명을 재현할 수 있도록 정보를 제공하는 기능은 물론, 발명의 기술적 의의를 명확히 알리는 기능을 하여야 하며, 특히 화학·제약·유전공학 등의 분야에서 출원 명세서에 기재되는 발명의 효과는 후자의 역할과 밀접하게 관련되어 있다. 그러나 위 분야의 기술적 특성상 출원인이 당해 발명의 기술적 유용성을 모두 파악하거나 관련 데이터를 모두 손에 넣기 어려울 수도 있고, 선행 특허권자의 전략적 목적에 의해 의식적으로 발명의 효과가 불충분하게 기재된 명세서로 특허출원이 이루어지는 수도 있다. 이러한 출원에 기한 특허는 후행 발명자와의 사이에 분쟁을 양산하고 선행 특허권자의 부당한 권리 연장에 봉사할 우려가 있기 때문에 통제가 필요하지만, 한편으로 지나치게 엄격한 명세서 기재요건의 기준을 유지하면 자칫 이 분야에서 발명을 위한 연구·개발 투자를 위축시켜 산업 발전을 저해할 우려도 생긴다. 각 국은 이 문제의 조화로운 해결을 위해 저마다의 준칙을 운영하고 있으며, 구체적으로는 발명의 상세한 설명에 출원발명의 유용성을 어느 정도까지 구체적으로 기재하도록 요구할 것인가, 그러한 유용성을 뒷받침하는 데이터의 제출을 강제할 것인가 말 것인가, 데이터의 사후 보완제출은 허용될 수 있는가, 허용된다면 그 시기는 언제까지로 보아야 할 것인가를 주로 문제 삼는다.⁹⁹⁾

그러나 발명의 실체가 자세하게 개시될수록 경쟁자가 발명을 완벽하게 이해하여 시장에 진입

97) Bonito Boats, Inc. v. Thunder Craft Boats, Inc., 489 U.S. 141, 156 (1989); Valinasab, Omid, Big Data Analytics to Automate Patent Disclosure of Artificial Intelligence's Inventions (April 19, 2022). p. 8

98) 특허법 제42조 제2항 제1호 내지 제3호에서는 특허출원서에는 '발명의 명칭', '도면의 간단한 설명', '발명의 상세한 설명', '특허청구범위'를 기재한 명세서와 필요한 도면 및 요약서를 첨부하여야 한다고 규정하고, 같은 조 제3항에서는 위 '발명의 상세한 설명'에는 그 발명이 속하는 기술분야에서 통상의 지식을 가진 자가 용이하게 실시할 수 있을 정도로 그 발명의 목적·구성 및 효과를 기재하여야 한다고 규정하였으며, 같은 조 제4항에서는 위 '특허청구범위'에는 보호를 받고자 하는 사항을 기재한 항(청구항)이 1 또는 2 이상 있어야 하고, 그 청구항은 '발명의 상세한 설명'에 의하여 뒷받침될 것, 발명이 명확하고 간결하게 기재될 것, 발명의 구성에 없어서는 아니되는 사항만으로 기재될 것 등의 요건을 갖추어야 한다고 규정하고 있는바, 이는 특허출원된 발명의 내용을 제3자가 명세서만에 의하여 쉽게 알 수 있도록 공개하여 특허권으로 보호받고자 하는 기술적 내용과 범위를 명확하게 하기 위한 것이라 할 것이므로, 위 '발명의 상세한 설명'은 그 출원발명이 속하는 기술분야에서 보통 정도의 기술적 이해력을 가진 자, 평균적 기술자가 당해 발명을 명세서 기재에 의하여 출원시의 기술수준으로 보아 특수한 지식을 부가하지 않고서도 정확하게 이해할 수 있고 동시에 재현할 수 있는 정도로 기재되어야 할 것이고, 특허출원의 명세서가 위와 같은 요건을 구비하지 못한 경우에는 특허법 제62조 제4호에 의하여 특허거절사정의 사유가 된다. 대법원 1999. 12. 10. 선고 97후2675 판결 [거절사정(특)]

99) 조영선, 명세서 기재요건으로서의 발명의 효과, 인권과정의 Vol.427, 2012.8, 95면.

하거나 우회 또는 개량발명을 수행하기 쉬워지는 것도 사실이다. 대부분의 출원인은 그러한 결과를 달가워하지 않으며, 특허는 부여받으면서도 발명의 내용은 가급적 모호하게 표현하려할 것이다. 이러한 이율배반은 특허제도에 내재된 불가피한 속성이라고 할 수밖에 없고, 그 때문에 법이 관여하여 특허 부여를 위해 필수적으로 요구되는 발명의 공개정도와 방법에 관하여 기준을 제시하고 집행하지 않으면 안 된다.¹⁰⁰⁾ 이에 대한 비교법적 검토를 진행하고자 한다.¹⁰¹⁾

나. 기재요건 - 요구되는 공개의 정도와 방법

발명의 설명은 그 발명이 속하는 기술분야에서 통상의 지식을 가진 사람이 그 발명을 쉽게 실시할 수 있도록 명확하고 상세하게 적어야 한다(특허법 제42조 제3항 제1호). 발명의 설명에는 ① 발명의 명칭, ② 기술분야, ③ 발명의 배경이 되는 기술, ④ 해결하려는 과제, ⑤ 과제의 해결 수단, ⑥ 발명의 효과, ⑦ 도면의 간단한 설명, ⑧ 발명을 실시하기 위한 구체적 내용, ⑨ 그 밖에 그 발명이 속하는 기술분야에서 통상의 지식을 가진 자가 그 발명의 내용을 쉽게 이해하기 위하여 필요한 사항을 기재해야 한다(특허법 시행규칙 제21조 제3항).

2006년 특허법 이전에는, 발명의 상세한 설명에 발명의 목적·구성·효과를 기재하게 되어 있었지만, 발명의 유형에 따라서는 이런 형식성이 부적절할 수 있고, 이는 국제적 기준에도 맞지 않기 때문에 이런 제한이 없어지고 출원인으로서의 발명의 실질적 내용에 비중을 두어 이를 보다 적절하게 기술(記述)할 수 있게 되었다. 이처럼 발명의 설명에 대한 형식적 기준은 폐지가 되었지만, 발명의 내용을 설명함에 있어 이를 목적·구성·효과에 입각하여 기술하는 것이 여전히 유용하고 합리적인 경우도 많을 것이다. 실무상 특허법 제42조 제3항 제1호의 발명의 설명 기재방법을 위반한 경우를 법 제42조 제4항의 특허청구범위 기재방법을 위반한 경우와 함께 명세서 기재불비라고 부른다. 그리고 발명의 설명 기재방법을 위반한 경우를 특정하여 발명의 설명 기재 불비라고 부르기도 한다.¹⁰²⁾

다. 쉽게 실시(용이실시) 요건

(1) 수준

특허법 제42조 제3항과 제4항의 규정은 특허 출원된 발명의 내용을 제3자가 명세서만에 의하여 쉽게 알 수 있도록 공개하여 특허권으로 보호받고자 하는 기술적 내용과 범위를 명확하게 하기 위한 것이므로, 위 제3항의 '발명의 상세한 설명'은 특허 출원된 발명이 속하는 기술분야에서 보통 정도의 기술적 이해력을 가진 자(이하 '평균적 기술자'라 한다)가 당해 발명을 명세서 기재에 의하여 출원시의 기술수준으로 보아 특수한 지식을 부가하지 않고서도 정확하게 이해할 수

100) 조영선, 특허법 2.0, 박영사, 2018, 70면.

101) 정상조 박성수 공편, 특허법 주해 I, 박영사, 2010, 520면.

102) 조영선, 특허법 2.0, 박영사, 2018, 71면.

있고 동시에 재현할 수 있는 정도로 기재되어야 할 것이며, 위 제4항 제1호에서 규정하는 바와 같이 특허청구범위가 발명의 상세한 설명에 의하여 뒷받침된다고 하기 위해서는 평균적 기술자의 입장에서 볼 때 그 특허청구범위와 발명의 상세한 설명의 내용이 일치하여 그 명세서만으로 특허청구범위에 속한 기술구성이나 그 결합 및 작용효과를 일목요연하게 이해할 수 있어야 한다.¹⁰³⁾

실제, 상세한 설명의 기재요건은 발명의 기술분야에 따라 다를 수밖에 없으므로 이를 명확하게 정량적으로 제시하기는 힘들다. 이러한 이유로 제시된 제도가 미생물 기탁제도이다. 물론, 후술하는 실시예를 기재하는 경우에도 구체적인 설명을 가름할 수 있다는 점에서 상호보완적인 제도가 될 수 있으나, 실시예는 강제적인 것이 아닌 발명의 성격에 따라서 달리할 수 있다는 점에서 차이가 있다.

(2) 실시예 및 도면의 기재 여부

1) 선택설

발명은 기술적 사상으로서의 추상성을 갖고 있기 때문에 당해 기술분야에서 통상의 지식을 가진 자는 그 기재만으로 발명을 쉽게 실시할 수 없는 것이 일반적일 수 있으므로, 발명사상이 실제상 어떻게 구체화되어 구현되는가를 나타내는 실시예를 기재하는 것은 기술을 공개해야 하는 명세서의 특성상 대단히 중요한 것이다. 그러나 특허법에서는 실시예의 기재를 명세서의 필수 기재요건으로 규정하고 있지 아니하고, 또한 심사기준에서도 발명의 구성에는 그 발명이 속하는 기술분야에서 통상의 지식을 가진 자가 그 발명을 재현할 수 있도록 하기 위해 필요에 따라서는 발명의 기술적 수단을 구체화한 실시예를 기재하여야 하며, 실시예는 가능한 한 여러 가지를 기재하여야 한다고 규정되어 있는 바, 모든 발명의 명세서에 실시예 기재가 필수적 기재사항은 아니라고 할 수 있다.¹⁰⁴⁾ 실시예로서, 미생물 기탁이나 SW발명에서 소스코드의 공개는 의미있는 제도라고 생각된다.

2) 필요설

일반적으로 기계장치 등에 관한 발명에 있어서는 특허출원의 명세서에 실시예가 기재되지 않더라도 평균적 기술자가 발명의 구성으로부터 그 작용과 효과를 명확하게 이해하고 용이하게 재현할 수 있는 경우가 많으나, 이와는 달리 이른바 실험의 과학이라고 하는 화학발명의 경우에는 당해 발명의 내용과 기술수준에 따라 차이가 있을 수는 있지만 예측가능성 내지 실현가능성이 현저히 부족하여 실험데이터가 제시된 실험예가 기재되지 않으면 평균적 기술자가 그 발명의 효

103) 대법원 2004. 10. 14. 선고 2002후2839 판결.

104) 특허법원 1998.12.17. 선고 98허928 판결.

과를 명확하게 이해하고 용이하게 재현할 수 있다고 보기 어려워 완성된 발명으로 보기 어려운 경우가 많고, 특히 약리효과의 기재가 요구되는 의약의 용도발명에 있어서는 그 출원 전에 명세서 기재의 약리효과를 나타내는 약리기전이 명확히 밝혀진 경우와 같은 특별한 사정이 있지 않은 이상 특정 물질에 그와 같은 약리효과가 있다는 것을 약리데이터 등이 나타난 시험예로 기재하거나 또는 이에 대신할 수 있을 정도로 구체적으로 기재하여야만 비로소 발명이 완성되었다고 볼 수 있는 동시에 명세서의 기재요건을 충족하였다고 볼 수 있을 것이며, 이와 같이 시험예의 기재가 필요함에도 불구하고 최초 명세서에 그 기재가 없던 것을 추후 보정에 의하여 보완하는 것은 명세서에 기재된 사항의 범위를 벗어난 것으로서 명세서의 요지를 변경한 것이다.¹⁰⁵⁾

도면과 관련해서도 선택적인 사항이라고 할 수 있다. 즉, 특허법 제42조 제2항은 특허출원서에는 발명의 명칭, 도면의 간단한 설명, 발명의 상세한 설명, 특허청구범위를 기재한 명세서와 더불어 필요한 도면 및 요약서를 첨부하여야 한다고 규정하고 있는바, 도면은 특허출원서에 반드시 첨부되어야 하는 것은 아니고 도면만으로 발명의 상세한 설명을 대체할 수는 없는 것이지만, 도면은 실시예 등을 구체적으로 보여줌으로써 발명의 구성을 더욱 쉽게 이해할 수 있도록 해주는 것으로서 도면이 첨부되어 있는 경우에는 도면 및 도면의 간단한 설명을 종합적으로 참작하여 발명의 상세한 설명이 청구항을 뒷받침하고 있는지 여부를 판단할 수 있다.¹⁰⁶⁾ 따라서, 구체적인 발명의 효과를 확인하기 어려울 경우에는 실시예나 도면을 통해 보완 내지 보충함으로써 기재요건을 충족시킬 수 있을 것이다. 이러한 실시예는 약리데이터나 AI발명에서 알고리즘이나 SW발명에서 소스코드 등을 공개하는 것처럼 데이터셋을 공개하는 경우를 들 수 있다.

3) 정리

앞서 살펴본 필수설에 따르면 실시예가 필요한 발명과 그렇지 못한 경우가 구분된다. 물론, 강제사항은 아니기 때문에 선택적으로 발명자가 출원시에 기재할 수 있겠지만 실시예를 부기하는 것은 다양한 면에서 효율성을 갖는다고 생각된다. 특히, 실시예는 발명의 효과를 확인해 주는 역할도 하지만, 그러한 역할 때문에 실시예의 기재가 요구되는 것이 아니라 통상의 기술자가 발명을 용이하게 실시(반복재현)하는 데 필요한 지식이기 때문에 필요한 것으로 보아야 한다.¹⁰⁷⁾ 실시예는 필수 기재사항은 아니므로, 실시예가 없이 발명의 구성이나 그 효과를 용이하게 파악할 수 있는 경우에는 실시예가 없어도 발명의 상세한 설명의 기재불비라고 볼 수 없으나 반면, 실시예가 없으면 그 구성이나 효과를 파악하기가 힘든 경우에는 실시예가 필수적으로 기재되어야 할 것이다.¹⁰⁸⁾

AI발명에 있어서도 재현가능성을 높이거나 용이하게 실시할 수 있도록 하는 면에서 데이터와

105) 대법원 2001. 11. 30. 선고 2001후65 판결 .

106) 대법원 2006. 10. 13. 선고 2004후776 판결.

107) 김영, 특허의 적극적 요건 및 명세서 기재요건으로서의 발명의 효과, 특허소송연구 제2집, 특허법원, 2001, 121면.

108) 정상조 박성수 공편, 특허법 주해 I, 박영사, 2010, 529면.

관련된 사항의 기재는 바람직하다고 본다. 다만, 데이터의 전체적인 기재나 제출은 쉽지 않다는 점에서 필요성과 실효성은 차이가 있다.

(3) 배경기술의 기재 - 범용성이 있는 선행기술의 기재불비

배경기술이라 함은 발명의 기술상 의의를 이해하는 데에 도움이 되고 선행기술 조사 및 심사 에 유용하다고 생각되는 종래의 기술을 말한다.¹⁰⁹⁾ 미국은 판례법상 출원인의 선행기술에 대한 정보개시의무가 인정되어 오고 있으며 특허법 규칙에 성문규정이 마련되어 있다.¹¹⁰⁾ 일본 특허법 또한 출원인의 선행기술에 대한 정보개시의무를 인정하고 있고(일본 특허법 제36조 제4항 제2호) 유럽특허협약(EPC)에 따르면 심사관은 출원인에게 출원발명에 대하여 일정한 선행기술 정보의 제공을 요청할 수 있는바,¹¹¹⁾ 특허법 제42조 제3항 제2호는 이러한 예들을 참고한 것으로 보인다.¹¹²⁾ 배경기술의 기재가 부적법한 것으로 인정되는 경우에 심사관은 거절이유를 통지하고, 출원인이 그에 상응하는 보정을 수행하지 아니하면 당해 출원은 등록거절에 이른다(특허법 제62조 제4호, 제42조 제3항).

대법원은 “특허법(1990.1.13. 법률 제4207호로 개정되기 전의 것) 제8조 제3항에 의하면 특허 출원서에 첨부하여 제출하여야 하는 명세서에 기재될 발명의 상세한 설명에는 그 발명이 속하는 기술분야에서 통상의 지식을 가진 자가 용이하게 실시할 수 있을 정도로 그 발명의 목적·구성·작용 및 효과를 기재하도록 규정되어 있고, 같은 조 제4항에 의하면 그 명세서에 기재될 특허청구의 범위는 명세서에 기재된 사항 중 보호를 받고자 하는 사항을 1 또는 2 이상의 항으로 명확하고 간결하게 기재하도록 규정되어 있는 바, 이와 같은 규정의 취지는 특허출원된 발명의 내용을 제3자에게 공표하여 그 기술적 범위를 명확하게 하기 위한 것으로서, 그 발명이 속하는 기술분야에서 통상의 지식을 가진 자가 용이하게 실시할 수 있을 정도라고 함은, 특허출원 당시의 기술수준을 기준으로 하여 그 발명과 관련된 기술분야에서 평균적인 기술능력을 가진 자이면 누구든지 특허출원된 발명의 내용을 명확하게 이해하고 이를 재현할 수 있는 정도를 일컫는 것이므로, 그 발명이 이용하고 있는 어떤 기술수단이 특허출원 당시의 기술수준에 속하는 범용성이 있는 것으로서 그 구성을 명시하지 아니하더라도 이해할 수 있는 것일 때는 구태여 그 기술수단의 내용을 기재할 필요가 없다.”¹¹³⁾고 판시하였다.

3. 기술분야 및 평균적 기술자의 수준

가. 기술분야의 기재

109) 특허청, 특허·실용신안 심사기준, 2021.

110) 미국특허법 시행령 37 CFR 1.56.

111) EPC 제124조.

112) 조영선, 특허법 2.0, 박영사, 2018, 72면.

113) 대법원 1992. 7. 28. 선고 92후49 판결 [거절사정].

발명이 속하는 분야를 결정하는 기준은 크게 기능설(product-function)과 문제해결설(problem-solving)의 2가지로 나뉜다.¹¹⁴⁾ 기능설에 의하면 발명이 속하는 기술분야는 특허출원된 발명의 사용목적이 되는 산업이라고 보는 반면, 문제해결설은 특허출원된 발명이 어떠한 문제를 해결하려고 시도하였느냐가 기준이 된다.¹¹⁵⁾ 다만, AI발명은 융합적 요소가 크기 때문에 기술분야의 기재가 큰 의미를 갖는다고 보기는 어렵다.

나. 통상의 지식을 가진 평균적 기술자의 수준

평균적 기술자란 그 출원에 관한 발명이 속하는 기술분야에서 보통 정도의 기술적 이해력을 가진 자를 말한다.¹¹⁶⁾ 통상의 지식의 가진자의 결정시 특단의 전문가, 특별한 지식과 능력을 가진 대학교수, 당해 특수 분야의 발전에 현저한 영향을 미친 과학적 명성이 있는 전문가는 평균 수준 이상의 자이므로 이들의 지식을 기준으로 하여서는 아니되고 선행기술의 조사를 통하여 가상적인 제3의 객관적인 인물을 상정하는 것이 바람직하다.¹¹⁷⁾

평균적 기술자의 수준에 대해서는 구 특허법(1990. 1. 13. 법률 제4207호로 전문 개정되기 전의 것) 제8조 제2항, 제3항에 의하면 특허출원의 명세서에는 그 출원에 관한 발명이 속하는 기술분야에서 보통 정도의 기술적 이해력을 가진 자, 즉 평균적 기술자가 당해 발명을 명세서 기재에 기하여 출원시의 기술수준으로 보아 특수한 지식을 부가하지 않고서도 그 발명을 정확하게 이해할 수 있고 동시에 재현할 수 있는 정도의 설명이 필요하다고 할 것이므로, 특허출원의 명세서가 위와 같은 요건을 구비하지 못한 것이라면 이는 구 특허법 제69조 제1항 제1호에 의하여 특허무효의 사유가 된다.¹¹⁸⁾

4. 미완성 발명과 발명의 완성

가. 기재불비와 미완성 발명

미완성발명이라 함은 발명의 성립이라고 볼 수 있는 외관을 갖추었으나 형식상의 하자가 있는 발명을 말한다. 이러한 미완성발명에는 ① 단순한 문제나 착상의 제출 또는 희망의 표명에 그치고, 어떻게 이것을 실현하는가를 알 수 없는 구체성이 완전히 결여된 것, ② 해결수단은 제시되었지만 극히 막연한 제안에 불과하기 때문에 어떻게 해서 이것을 구체화할 것인가에 대한 설명이 분명하지 않은 것, ③ 해결수단은 제시되어 있지만, 그 수단만을 가지고는 목적을 달성할 수 없다고 인정 되는 것, ④ 어떤 기술적 과제를 완전히 해결하기 위해서는 복수의 구성요소의 결합이 필요하다는 새로운 착상을 얻어 연구를 추진한 것일지라도 그 중의 하나의 구성 요소가 해결

114) 조영선, 특허법 2.0, 박영사, 2018, 516면.

115) 임호, 특허법, 법문사, 2003, 264면.

116) 대법원 1997. 7. 25. 선고 96후2531 판결.

117) 반응성, 화학발명에서의 진보성 판단기준에 관한 연구, 충남대학교 석사학위논문, 2002, 10면.

118) 대법원 1996. 1. 26. 선고 94후1459 판결.

되지 않아 현재의 기술 수준으로는 실현불가능이며, 장래의 실현가능성도 분명하지 않은 것, ⑤ 구성이 구체적으로 제시되어 있어도 그 구성을 해결수단으로 인정하기 위해서는 실험 결과 등의 구체적인 뒷받침을 필요로 하는데도 불구하고, 그 뒷받침이 없는 것, ⑥ 새로운 물건을 창작했다고 해도 어떤 도움이 되는지가 분명하지 않은 것 등이 있다. 또한 ⑦ 미생물관련 발명시 출원 전에 미생물을 기탁하도록 한 미생물(微生物) 기탁요건을 갖추지 않은 발명 등도 미완성발명이다.¹¹⁹⁾

구 특허법 제42조 제3항은 발명의 상세한 설명에는 그 발명이 속하는 기술분야에서 통상의 지식을 가진 자가 용이하게 실시할 수 있을 정도로 그 발명의 목적·구성 및 효과를 기재하여야 한다고 정하고 있다. 이는 특허출원된 발명의 내용을 제3자가 명세서만으로 쉽게 알 수 있도록 공개하여 특허권으로 보호받고자 하는 기술적 내용과 범위를 명확하게 하기 위한 것이므로, 위 조항에서 요구하는 명세서 기재의 정도는 통상의 기술자가 출원시의 기술수준으로 보아 과도한 실험이나 특수한 지식을 부가하지 아니하고서도 명세서의 기재에 의하여 당해 발명을 정확하게 이해할 수 있고 동시에 재현할 수 있는 정도를 말한다. 이를 위반할 경우에는 미완성 발명 또는 기재불비로서 거절되거나 보완요청을 받게 된다. 여기에서 실시의 대상이 되는 발명은 청구항에 기재된 발명을 가리키는 것이라고 할 것이므로, 발명의 상세한 설명의 기재에 오류가 있다고 하더라도 그러한 오류가 청구항에 기재되어 있지 아니한 발명에 관한 것이거나 청구항에 기재된 발명의 실시를 위하여 필요한 사항 이외의 부분에 관한 것이어서 그 오류에도 불구하고 통상의 기술자가 청구항에 기재된 발명을 정확하게 이해하고 재현하는 것이 용이한 경우라면 이를 들어 구 특허법 제42조 제3항에 위배된다고 할 수 없다.¹²⁰⁾

나. 미생물 기탁과 발명의 완성¹²¹⁾

구 특허법 시행령 제1조 제2항, 제3항, 구 특허법 시행규칙 제31조의2 제1항의 규정을 종합하면, 미생물을 이용한 발명에 있어서 그 발명이 속하는 기술분야에서 통상의 지식을 가진 자가 그

119) 윤선희, 특허법, 법문사, 2019, 133면.

120) (1) 발명의 완성 여부는 명세서 기재요건 충족 여부와는 구별되어야 하는 것으로 미완성 발명은 A. 구성요건을 결여한 경우, B. 해결수단을 제시하지 못한 경우, C. 그 수단으로 실행하였을 때 효과가 없는 경우, D. 용도를 밝히지 못한 경우, E. 그 실시의 결과가 사회적으로 용납되지 않는 위험한 상태로 방치되는 경우, 등에 해당하므로 사건발명이 특허법 제42조 제3항의 기재요건을 충족하지 못하였다고 하여 이를 미완성 발명이라고 단정할 수는 없다. (2) 사건발명은 발명의 목적, 구성, 효과가 기재되어 있고, 활성화합물 및 제조방법, 구체적인 실시예, 약학 조성물의 제제화, 투여방법 및 투여량, 제제화 및 독성시험 결과가 기재되어 있으며, 투여량은 0.5mg 내지 500mg 투여량이 바람직하다고 기재되어 있으므로, 따라서, 평균적 기술자는 화합물의 구체적인 질병의 치료효과는 충분히 알 수 있고, 상기 기재로부터 평균적 기술자가 반복 실시하여 목적하는 기술적 효과를 얻을 수 있을 정도까지 구체적, 객관적으로 기재된 것으로 판단되는바, 사건발명은 완성된 발명으로 봄이 상당하다. (3) 비록, 약리효과가 구체적 화합물마다 구체적인 수치로서 기재되어 있지 아니하나, 화합물의 약리효과에 관한 정량적 기재가 발명을 보다 정확하게 이해하는데 도움이 되는 것이라 하여도, 약리효과에 대한 데이터와 같은 정량적 기재는 발명의 구성으로 볼 것은 아니어서 이러한 기재가 없다고 하여 출원 당시 미완성된 것이라 할 수 없다. 특허법원 2001.7.20. 선고 2000허7052 판결

121) 대법원 2005. 9. 28. 선고 2003후2003 판결.

미생물을 용이하게 입수할 수 없는 경우에는 특허청장이 지정하는 기관에 그 미생물을 기탁하고, 명세서에 당해 미생물의 기탁번호·기탁기관의 명칭 및 기탁연월일을 기재하는 외에 그 기탁사실을 증명하는 서면을 출원서에 첨부하지 아니하면 그 발명이 완성되었다고 할 수 없는 한편, 특허협력조약에 의한 국제출원으로서 특허절차를 위한 미생물 기탁의 국제적 승인에 관한 부다페스트조약에 따라 세계지적소유권기구 총장이 승인한 국제기탁기관에 기탁한 경우에는 1985. 2. 26.자 특허청 고시 제85-1호에 의하여 국제특허출원의 출원번역문 제출기간이 경과하였다고 하더라도 그 출원이 공개되기 전까지 그 미생물을 한국과학기술원 또는 사단법인 한국중균협회에 다시 기탁하고 그 기탁증명서를 제출하면 된다.¹²²⁾

구 특허법 시행령 제1조 제2항이 정한 '기탁사실을 증명하는 서면'은 미생물의 수탁기관이 발행하는 미생물수탁번호통지서나 수탁증 등과 같이 당해 미생물의 기탁사실을 객관적으로 증명하는 서면을 말하는 것이므로, 특허발명의 출원시에 제출된 명세서에 당해 미생물의 기탁번호·기탁기관의 명칭 및 기탁연월일을 기재하였다고 하더라도, 이는 구 특허법 시행규칙 제31조의2 제1항의 명세서 기재요건을 충족한 것으로 볼 수 있을 뿐, 이러한 출원서의 제출을 들어 위 시행령 제1조 제2항의 기탁사실을 증명하는 서면이 제출되었다고 할 수는 없다. 따라서, 이 사건 미생물이 미국의 ATCC에 기탁되어 있다는 사실만으로는 이 사건 특허발명의 우선일인 1984. 12. 4.경 이 발명이 속하는 기술분야에서 통상의 지식을 가진 자가 이 사건 미생물을 용이하게 얻을 수 있는 것이라고 할 수도 없으며, 특허발명의 심사단계에서 미생물 기탁증명서의 미제출을 간과한 채 특허가 되었다고 하여 그 출원절차상의 하자가 치유된다거나 출원에 있어서의 하자를 들어 특허의 효력을 부인하는 것이 금지된다고 보아야 할 아무런 근거가 없다.

다. 미생물기탁과 데이터 기탁의 유사점

AI의 기술적 요소인 인공지능망에서의 정보처리는 내부적인 과정을 알 수 없으며, 시간과 데이터의 처리에 따른 결과라는 점에서 미생물발명과 유사한 면이 있다. 이러한 점에서 서면 기재(written description), 실시가능성(enablement), 신규성 및 유용성(utility) 요건과 관련된 AI 기술의 블랙박스(black box) 난제를 해결하기 위해 AI모델에 대한 기탁요건을 제안한다. 기탁 요건은 미생물이 관련된 특허 출원에 대해 오랫동안 존재해 왔다. 발명이 미생물을 포함할 때, 박테리아, 효모, 곰팡이 또는 바이러스와 같은 물질의 구조나 구성 요소를 명확히 하는 것은 보통 불가능하다. 살아있는 유기체를 묘사하는 것의 어려움에 대응하여, 1977년 WIPO는 부다페스트 조약을 제

122) 이 사건 특허발명의 원 출원발명은 국제기탁기관인 미국의 ATCC에 이 사건 미생물을 기탁하였고, 국내에서는 그 출원 공개일인 1987. 2. 28. 이전임이 역수상 명백한 1986. 9. 12. 한국중균협회에 이 사건 미생물을 기탁하였으므로 위 기탁이 기탁 기준일 이후에 이루어진 것이라고 할 수 없음에도 불구하고 이와 달리 판단한 원심에는 미생물 기탁에 관한 법리를 오해한 위법이 있으나, 위 기탁일 이후에 원 출원발명에서 분할되어 출원된 이 사건 특허발명의 출원서에 위 기탁사실을 증명하는 서면이 별도로 첨부된 바는 없고, 이 사건 특허발명의 출원시에 제출된 명세서가 기탁을 증명하는 서면에 해당한다고 할 수 없음은 아래에서 보는 바와 같으므로 원심의 이러한 잘못은 이 사건 특허발명이 미완성 발명이어서 그 권리범위를 인정할 수 없다는 원심의 결론에 아무런 영향이 없다.

정하였다. 이에 따라 미생물에 대한 기탁원칙을 확정하였다. 미국을 포함한 부다페스트 조약에 동의하는 당사자는 해당 국가의 특허 절차의 일부로 미생물에 대한 기탁원칙을 채택할 의무가 있다. 출원인은 특허 출원 중 특허청이 지정한 기관에 미생물을 기탁하게 된다.¹²³⁾

미생물기탁에 대해서는 법리적으로 특허출원은 서면으로 하여야 한다는 서면주의의 보완 또는 예외로 인정하고 있기는 하지만, 특허법 시행령 제2조 및 제3조의 근거를 모범인 특허법에서는 찾을 수 없다.¹²⁴⁾ 그렇지만 미생물이 발명의 효과를 가져오는 것처럼 데이터가 발명의 효과를 가져오는 것은 내부적인 처리 과정을 확인할 수 없다는 점에서 유사한 면이 있다. 따라서, AI발명에 있어서 학습데이터의 공개에 대해서도 미생물 기탁제도와 유사성이 있다고 볼 수 있다. 이에 대해서는 후술하고자 한다.

5. 기술공개 기준 등에 대한 한계

기술공개는 여러 면에서 미흡한 점이 발견된다. 명세서 기재에 있어서 방법의 상세한 설명을 요구하지만 정량적인 기준처럼 구체화된 기준의 제시가 어렵다. 그렇지만, 기술공개는 법정 요건이기 때문에 기재불비 등을 이유로 거절사정 내지 무효사유에 해당할 수 있다. 그만큼 독점권을 부여하는 특허제도가 갖는 공개요건에 부합하지 못할 경우에는 독점권을 부여할 공익적 이익이 소멸하기 때문이다. 이처럼, 명확하게 기술공개 수준이나 범위를 제시하지 못하기 때문에 심사 실무에서도 추상적인 기재요건을 제시하고 있다. 평균적 기술자의 수준에 따라 달라질 수밖에 없는 한계이기도 하다. 평균적 기술자의 수준이 높을 경우에는 방법의 설명은 구체화되지 아니할 수 있다. 즉, 기술적 요건을 상세히 기재하지 않아도 그 발명이 의도하는 기술적 사항의 구현에 힘들이지 않아도 되기 때문이다. 반면, 평균적 기술자의 수준이 낮다면 방법의 설명은 구체화되어야 한다. 그렇다면 평균적 기술자의 수준을 어떻게 설정하는 것이 타당한 것인지 의문이다.

AI와 관련되어서는 기술 수준에 따른 기준 제시나 평균적 기술자에 따른 기준 설정의 한계, 도면 및 데이터의 선택적 제공에 관하여 고려할 사항들이 적지 않다. 데이터를 구성요소로 하는 발명에 있어서도 더욱 그러하다.

제3절. IP5의 AI발명에 있어서 데이터 공개 논의

1. 각국의 AI발명 관련 기술공개 수준 검토

가. AI발명의 기술공개

123) Hagen, Gregory R., AI and Patents and Trade Secrets (February 1, 2021). in Florian Martin-Bariteau & Teresa Scassa, eds., Artificial Intelligence and the Law in Canada (Toronto: LexisNexis Canada, 2021), Available at SSRN:<https://ssrn.com/abstract=3734654>

124) 한국특허법학회 편, 특허판례연구, 박영사, 2017, 47면.

각 국은 AI발명에 대해서는 컴퓨터프로그램 관련 발명 또는 컴퓨터 구현 발명으로 다루고 있다. 특허법에서는 명시적으로 SW특허나 컴퓨터구현 발명에 대해 규정한 것은 아니지만 해석상 또는 심사지침에 따라 인정하고 있다. AI발명은 세부적으로 발명의 구성에서는 차이가 있다고 하더라도, 기본적으로는 SW로서 프로그램 발명에 해당한다. 이러한 측면에서 AI발명에 대해서는 기존의 SW발명에 따라 보호체계를 갖는 것이며, 구체적인 심사지침에서는 심사방법에 대한 것이라는 점에서 차이가 있으나 SW라는 본질적인 차이는 없다. 무엇보다, 실질적으로 AI발명의 구체적인 내용은 SW라는 점과 더불어 데이터가 구성요소로서 포함된다는 점이다. 그렇기 때문에 AI 발명에서 실질적으로 요부에는 데이터의 수집, 처리, 가공 등의 프로세스에 대한 사항이 반드시 포함된다. 다만, 기술공개는 블랙박스화된 경우에는 해당 발명의 투명성, 재현가능성, 신뢰성 등에 대한 이슈가 제기될 수 있기 때문에 공개의 충분성에 대한 논란이 가능하다.

나. 데이터 공개에 대한 논의

AI발명의 기술공개에 대해서는 논의가 시작되고 있으나, 데이터 공개에 대해서는 IP5의 공식문서에서는 구체적으로 다루고 있는 사항은 발견되지 않는다. WIPO나 AI관련 학술논문이나 정책제안 형태로 데이터 기탁(deposit)에 대해 언급되고 있는 수준이다. 다만, 일부에서는 데이터를 공개하는 것에 대해 실효성이 없다는 이유로 부정적인 견해가 나오고 있으나,¹²⁵⁾ 발명의 재현가능성을 위한 용이한 실시를 위해서라도 데이터의 공개는 바람직하다고 생각된다.

이하에는 각국의 심사실무에서 AI발명에서 데이터와 관련된 사항을 검토한다. 결론적으로 각국의 실무에서는 데이터 공개에 대한 규정은 확인하기 어려웠으며, 데이터 구조발명의 구성요건이나 발명의 해당성에 대한 논의가 중심을 이루고 있음을 확인하였다.

2. 미국

가. 기술 공개(Disclosure Requirement)

(1) 일반원칙

미국 특허법 제112조는 명세서 기재요건을 정하고, 제112조 (a)항은 출원인이 명세서에 공개하여야 할 사항을 3가지로 나누어 기술하도록 하고 있다. 이 요건들은 (i) 실시가능성 요건(Enablement Requirement), (ii) 최상의 실시례 요건(Best mode Requirement), (iii) 서면기재요건(Written Description Requirement)으로 나뉜다. 미국 특허법이 명세서 기재요건을 요구하는 이유는 특허명세서를 보고 통상의 기술자가 기술된 내용에 따라 출원된 내용을 재연하여 이를 똑같

125) Liva Rudzite, Algorithmic Explainability and the Sufficient-Disclosure Requirement under the European Patent Convention, *Juridica International*, 31, 2022, p. 125-135.

이 만들고 사용할 수 있도록 하기 위함이다. 특허는 공개의 대가인데, 공개가 적절하게 이루어지지 못하면, 이에 대한 대가인 특허를 허여할 수 없게 되는 것이다. 미국 특허법 제112조의 요건을 충족하지 못하면, 출원된 특허에 대해서 특허를 허여할 수 없고, 등록이 된다고 하더라도 무효가 된다.¹²⁶⁾

(2) 실시가능성 요건(Enablement Requirement)

특허출원인은 특허명세서를 보고 출원된 기술영역의 통상의 기술자가 기술된 내용에 따라 출원된 내용을 재연하여 이를 똑같이 만들고 실시할 수 있도록 기재하여야 한다. 이를 실시가능성 요건이라고 한다. 발명자는 출원을 하면서 당해 발명을 구현하여 제조하는 방법과 사용하는 방법을 모두 개시하여야 한다. 예를 들어 화학물질특허를 출원하는 경우라면 화학물질의 제조방법과 사용방법이 모두 개시되어야 한다. 이를 통해서 특허를 사용하려는 자는 과도한 실험(undue experiment)이 없이도 당해 특허를 이용할 수 있게 되는 것이다.

위와 같이, 실시가능성 요건의 의미를 감안하면 중요한 판단기준이 되는 것이 과도한 실험의 요부이다. 이와 관련하여 중요한 기준을 제공하는 판례가 *In re Wands* 판결이다.¹²⁷⁾ 이 판결에서 CAFC는 과도한 실험요건의 충족여부의 판단기준을 제시하였다. 그 8가지 고려요소는 (i) 필요로 하는 실험의 양(the quantity of experimentation necessary), (ii) 지시 또는 안내가 주어진 양(the amount of direction or guidance presented), (iii) 적용되는 사례의 존부(the presence or absence of working examples), (iv) 발명의 성질(the nature of the invention), (v) 과거 기술의 정도(the state of the prior art), (vi) 같은 기술에서 비교되는 발명의 기술정도(the relative skill of those in the art), (vii) 해당 기술분야에서 예견가능한지 여부(the predictability or unpredictability of the art), (viii) 청구의 폭(the breadth of the claims)이다.

(3) 최적 실시례(The Best Mode Requirement)

특허권을 부여함으로써 특허권자의 기술공개에 대한 대가를 부여하지만, 한편으로 특허가 만료되었을 때에 공공영역에 있는 사람들이 발명을 개시한 것과 같이 실시할 수 있을 뿐만 아니라, 기간이 만료된 특허의 권리자와 상업적으로 대등한 지위에서 시장경쟁을 할 수 있도록 할 필요가 있다. 특허권자는 특허를 받기 위해서 공개를 하지만 정작 경쟁의 핵심이 되는 부분을 의도적으로 누락하려고 할 유인이 있다. 이러한 유인을 억제하기 위해서, 미국 특허법 제112조는 최적의 실시례를 기재할 것을 요구한다. 실시가능성 요건은 청구항에 기재되는 사항들이 일반인들도 그대로 실시할 수 있는 것을 목적으로 하지만, 최적의 실시례 요건은 그 실시가 최선의 실시여야 한다는 점까지 요구한다는 점에서 구별된다. 다만, 실시례 요건의 문제제기가 지속적으로 이어졌

126) *In re Glass*, 492 F.2d 1228, 181 U.S.P.Q.(BNA) 31(CCPA 1974) ; *Reiffin v. Microsoft Corp.*, 48 U.S.P.Q.2d(BNA) 1274(N.D. Cal. 1998).

127) 858 F.2d 731(Fed. Cir. 1988).

기 때문에 특허법 제282조(b)(3)(A)는 특허소송에서 특허권자가 최적의 실시례 요건을 위반하였다는 것을 무효사유로부터 제외하였다.¹²⁸⁾

(4) 서면 기재요건(The Written Description Requirement)

미국특허법 제112조 (a)항은 특허 명세서 (specification) 는 반드시 발명의 상세한 설명을 포함할 것을 요구한다. 같은 조 (b)항은 “명세서는 하나 또는 그 이상의 청구항들로 마쳐야 한다.” 고 규정하고 있는 것과 함께 명세서에 필수적 구성요건을 규정한 것이다. 명세서의 ‘상세한 설명’ 부분은 발명의 배경(Background of the Invention), 발명의 요약(Summary of the Invention), 발명의 상세한 설명(Detailed Description of the Invention) 등으로 구성된다.

실시가능성(enablement) 요건은 특허가 기간이 만료되면 발명이 지나친 실험 없이 제조되고 사용될 수 있도록 하는 것을 목적으로 하는 반면, 서면기재요건은 발명가가 출원일 현재 청구한 발명의 외연을 정하는 역할을 한다.

나. AI발명에 있어서 특허적격성에 대한 논의

최근의 SW관련 발명에 대한 엄격한 판단을 요구하는 경향에 따른 추정이기는 하지만, 미국은 AI발명에 대해 보수적인 입장이라고 하겠다. 무엇보다, 미국특허법에는 발명의 정의규정이 없을 뿐만 아니라 AI 관련 발명의 특허적격성에 관한 규정이 없지만, 특허를 받을 수 있는 발명에 대하여 규정하고 있는 제101조에 의하여 특허적격성이 판단되고 있다. 제101조에는 “신규하고 유용한(any new and useful) 방법(process), 기계(machine), 제조물(manufacture), 조성물(composition of matter), 또는 이에 대한 새롭고 유용한 개량을 발명하거나 발견한 자는 본 조의 조건 및 요건에 따라 특허를 획득할 수 있다.” 고 규정하고 있다. 따라서 미국특허법상 특허를 받을 수 있는 발명에 해당하는 것인지를 판단하는 특허적격성의 판단기준은 (i) 청구항에 기재된 발명(이하, “청구발명” 이라고 한다)이 “신규하고 유용한 방법(방법발명)”, (ii) 신규하고 유용한 기계(장치·시스템발명), (iii) 신규하고 유용한 제조물(물건·물질), (iv) 신규하고 유용한 조성물(물질발명), (v) 방법·기계·제조물·조성물에 대한 새롭고 유용한 개량발명이어야 하고, 그리고 (vi) 특허를 받을 수 있는 자는 발명하거나 발견한 자연인이라고 특정하고 있다.¹²⁹⁾

미국에서는 AI 관련 발명을 SW발명의 범주에 속하는 것으로 보고, SW발명관련 MPEP 및 특허적격성 판단 가이드라인이 적용되고 있다. SW발명은 컴퓨터 등의 하드웨어와 결합되어 특정단계의 단계나 수단이 알고리즘 형태로 청구범위에 기재된 경우를 말한다. AI기술을 포함하는 SW발명이 특허를 받을 수 있는 특허대상에 포함되는 것인가 아닌가를 판단하는 기준은 미국특허법 제101조의 사법적 요건과 판례법에 의한 사법적 예외요건을 판단하는 것이며, 이것이 AI 관련 발명의

128) 최승재 외, 신미국특허법, 법문사, 2020, 276면.

129) 함영옥, AI 관련 발명의 성립성 판단기준에 관한 연구, 한양대학교 박사학위논문, 2022, 68~69면.

특허적격성을 판단하는 기준이 된다.¹³⁰⁾

다. MPEP개정과 사법적 예외

(1) 사법적 예외

미국 특허법 제101조의 특허적격성 판단과 관련한 사법적예외요건(비사법적요건)이란 관례법에 의하여 확립된 것으로, 청구항에 기재된 SW발명이 (i) 자연법칙(laws of nature)에 해당하는 경우, (ii) 자연현상(natural phenomena)에 해당하는 경우, (iii) 추상적 아이디어(abstract idea)에 해당하는 경우에는 사법적예외(Judicial exception)에 해당하여 특허적격성을 만족하지 못하여 특허대상이 되지 못하는 것을 말한다. 이러한 사법적예외요건은 사법적예외로서 USPTO의특허심사절차매뉴얼(Manual of Patent Examining Procedure : MPEP)에서 규정하고 있다.¹³¹⁾

(2) 엘리스 판결에 따른 MPEP 개정

그 후 Alice Corp. v. CLS Bank 사건¹³²⁾에서 결제 위험을 완화하는 에스크로(escrow) 방법에 관한 발명의 특허적격성이 문제되었는데 CAFC 다수의견은 에스크로 시스템은 추상적 아이디어 이상의 ‘상당한 추가적 요소(significantly more)’가 없다는 이유로 특허적격성을 부정하였다. 그리고 연방대법원도 위 발명이 추상적 아이디어에 관한 것이고 ‘발명적 개념(inventive concept)’ 또는 ‘상당한 추가적 요소(significantly more)’도 없으므로 특허적격성이 부정된다고 판시하였다. 이러한 Alice 사건 이후 USPTO의 특허심사편람(Manual of Patent Examining Procedure, MPEP)에는 특허적격성 판단에 있어서 ‘Step 1 : 특허법 제101조에 해당하는가? → Step 2A : 사법적 예외사유(자연법칙, 자연현상, 추상적 아이디어)가 있는가? → Step 2B : 상당한 추가적 요소를 포함하고 있는가?’ 라는 기준을 제시하였고, 청구범위가 사법적 예외사유에 해당하더라도 상당한 추가적 요소를 포함하고 있는 경우 특허적격성이 인정된다고 보고 있다. 그 후 2019. 1. 7. 개정된 특허적격성 지침에는 위 Step 2A 부분을 개정하여 보다 구체적으로 ‘Step 2A prong one : 사법적 예외사유가 있는가? → Step 2A prong two : 실질적 응용(A Practical Application)이 있는가?’로 더 세분화된 판단기준을 제시하고 있다. 즉, 사법적 예외사유에 해당하더라도 실질적 응용이 있다면 특허적격성이 인정되고, 실질적 응용이 없다면 Step 2B의 ‘상당한 추가적 요소’를 포함하고 있는지 판단하도록 하고 있다.¹³³⁾

130) 함영욱, AI 관련 발명의 성립성 판단기준에 관한 연구, 한양대학교 박사학위논문, 2022, 69면.

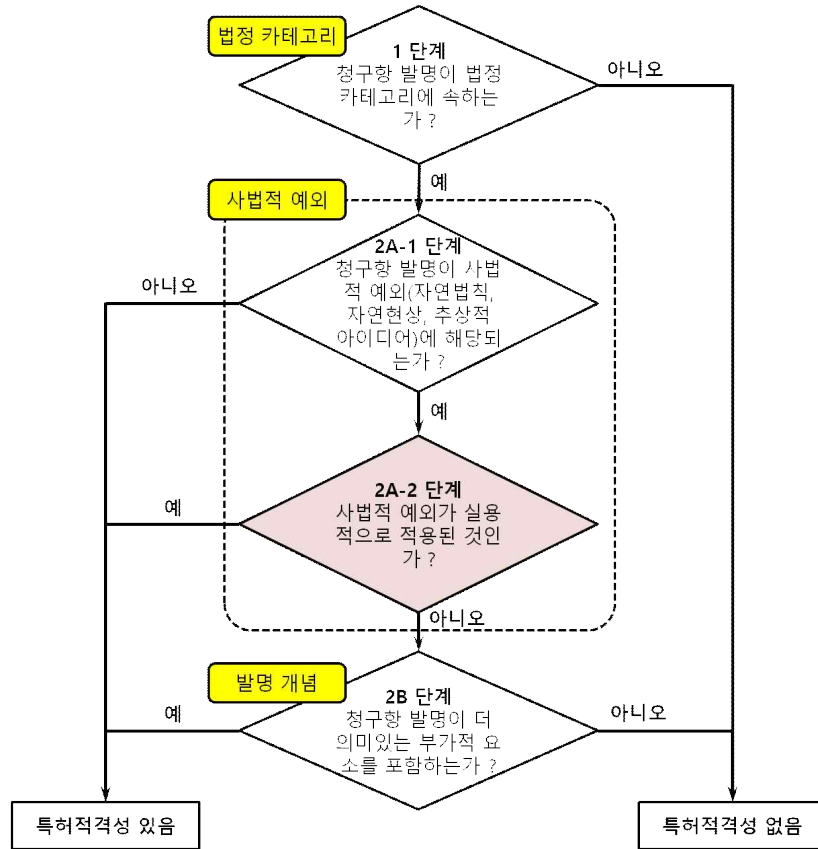
131) 함영욱, AI 관련 발명의 성립성 판단기준에 관한 연구, 한양대학교 박사학위논문, 2022, 73면.

132) 573 U.S. 208(2014).

133) 이혜진, 특허와 상표, 2021.03.22.일자.

<http://kpaanews.or.kr/news/view.html?section=91&category=120&no=4842> <2022.11.18.일자 접속>

[그림 3-2] 특허적격성 판단



<출처 : USPTO>

2019년 10월 17일의 현행 가이드스에서는 2014년에 선고된 Alice 판결에서 제시된 2단계 테스트에 대한 새로운 기준을 제시한다는 것을 명확히 하고있다. 현행 특허적격성 판단 가이드스에는 SW발명의 사법적 예외로 (i) 수학적 개념(수학적 관계, 수학 공식 또는 방정식, 수학 계산), (ii) 인간 활동을 조직하는 특정 방법(근본적인 경제적 원칙 또는 실행, 상업적 또는 법적상호작용, 개인행동 또는 사람들 사이의 관계 또는 상호작용관리), (iii) 정신적 프로세스(인간의 마음에서 수행되는 개념들) 등을 제시하고 있다. 다만, 현행 특허적격성 판단 가이드스에는 AI란 용어가 명확히 기재되어 있지않지만, AI 관련 발명을 SW발명의 범주에 속하는 것으로 보고, AI 관련 발명에도 동 가이드스를 적용하여 특허적격성을 만족하는지를 판단하고 있다. 즉 제4차 산업혁명 관련기술(AI 및 AI와 융합된 빅데이터, IoT, 블록체인 등)은 넓은의미에서 SW발명으로 알고리즘 형태의 프로그램발명이 대부분이므로, 미국 특허법 제101조의 사법적요건과 USPTO의 가이드스에 의한 사법적 예외요건을 근거로 특허적격성을 만족하는지가 판단되고 있다.¹³⁴⁾

134) 함영욱, AI 관련 발명의 성립성 판단기준에 관한 연구, 한양대학교 박사학위논문, 2022, 74~75면.

라. AI발명에 대한 가이드스

(1) AI발명에 대한 가이드스

미국은 AI 관련 발명, AI와 융합된 빅데이터, IoT, 블록체인 등의 제4차 산업혁명 관련 발명을 SW발명과 구분하여 별도의 특허적격성 판단기준을 제정하여 운영하는 것이 아니라, 통상의 MPEP 및 SW발명의 가이드스를 근거로 특허적격성을 만족하는지를 판단하고 있다. 물론 AI 관련 발명 등이 무엇인지에 대해서는 수많은 견해들이 난립하고 있고, 현행 미국특허법상 AI 관련 발명이 미국특허법 제101조의 규정을 충족하여 특허대상이 되기 위해서는 특허명세서의 ‘발명의 설명’ 및 청구범위를 어떻게 특정하여 기재하여야 하는 것인지가 쟁점이 되고 있는 것도 현실이다. 그러나 미국은 AI 관련 발명 등의 제4차 산업혁명 관련 발명은 인간의 인위적 행위가 아니라 인간의 개입에 의하여 컴퓨터 등 하드웨어에서 실현되는 프로그램을 토대로 하는 SW와 동일한 관점에서 단계 또는 수단이라는 특정 알고리즘 형태로 청구범위를 기재하는 경우, SW와 동일한 관점에서 특허적격성을 만족하여 특허대상이 될 수 있는 것으로 판단하고 있다.¹³⁵⁾

(2) 데이터에 대한 판단

MPEP 및 가이드스에 근거하여 SW발명으로서 AI발명의 특허적격성을 판단하고 있지만, MPEP 및 가이드스에는 구체적으로 AI발명 및 그 발명의 필수 구성요소인 데이터의 적격성 판단에 대해 별도의 구체적인 판단기준을 제시하고 있지는 않다. 그러나 실무적으로는 AI발명 및 데이터발명은 SW발명의 특허적격성 요건이 그대로 적용되고 있다. 따라서 데이터의 생성 및 수집 등은 방법발명 및 시스템발명(기계)으로 특허적격성 요건을 만족하여 특허대상이 되는 것은 물론이고, 또 데이터구조나 데이터세트로 기재한 경우에는 제101조의 사범요건에서 제시한 “새롭고 유용한 기계(물건)”에 해당하여 특허적격성 요건을 만족하기 때문에 특허대상이 되고 있다. 다만, “A 구조, B 구조, C 구조…의 특정 구조를 가진 데이터”와 같이 데이터를 권리주체(카테고리)로 기재한 경우, 그 카테고리의 데이터 물건발명은 “새롭고 유용한 기계(물건)”에 해당하여 특허적격성 요건을 충족하는 것인지는 명확하지 않다.¹³⁶⁾

마. AI발명 및 데이터 공개에 대한 논의

USPTO 차원에서 인공지능이 지식재산권 분야에 미치는 영향에 대해 2019년 조사를 시작하여 2020년 리포트를 발행하였으나, 데이터 공개나 기탁 등에 대해 구체적으로 논의된 자료는 공개된 바 없는 것으로 보인다.¹³⁷⁾

135) 함영욱, AI 관련 발명의 성립성 판단기준에 관한 연구, 한양대학교 박사학위논문, 2022, 85~86면.

136) 권지현, AI발명과 데이터의 실효적 특허보호방안, 法學研究 第24輯 第1號, 2021, 10~11면.

137) USPTO, Public Views on Artificial Intelligence and Intellectual Property Policy, October, 2020.

(1) 서면에 의한 설명

서면 기술의 요구 수준은 기술 분야에서 통상적인 기술을 가진 사람의 기술 수준, 청구된 발명의 성격 및 기술의 복잡성과 예측 가능성에 따라 결정된다. 서면 기술의 적절성을 평가하기 위해서는 공개 여부를 고려해야 한다. 출원의 신뢰성은 발명자가 출원일 현재 청구된 주제를 소유하고 있었다는 것을 평균적 기술자에게 합리적으로 전달한다. 그러나 발명자는 단순히 원하는 결과를 제공하는 것만으로는 요구되는 공시를 충족할 수 없다. 일부 기술 분야에서는 평균적 기술자의 지식 수준과 요구되는 서면 기술 수준 사이에 역상관이 존재한다. 소프트웨어 특허와 관련하여, 소프트웨어의 기본 코드의 구현은 일반적으로 평균적 기술자가 과도한 실험 없이 수행할 수 있다. 그러므로 소프트웨어의 기능의 공개는 서면 설명의 요구되는 기준을 적절하게 충족시킬 수 있다.¹³⁸⁾

(2) 발명의 실시요건¹³⁹⁾

청구항은 평균적 기술자가 과도한 실험 없이 동일한 발명을 제조하여 사용할 수 있도록 하여야 한다. 유효요건은 서면에 기재된 설명과 다르며, 별도로 충족되어야 한다. 서면 설명은 단순히 관련 기술 분야의 숙련된 사람이 청구항과 관련된 발명을 만들고 사용할 수 있도록 하는 것보다 더 광범위하다.

과도한 실험 요건은 기술 분야에 따라 달라진다. 일반적으로, 기술 분야가 복잡한 실험을 포함할 때, 기술 분야에서의 실험의 복잡성은 그것을 과도하게 만들지 않는다. Vasudevan 판결에서 연방 순회 항소법원은 “상당한 양의 실험이 있더라도 청구는 충분히 가능하다. 실험이 단지 일상적인 경우 또는 문제의 규격이 실험이 진행되어야 하는 방향과 관련하여 합리적인 양의 지침을 제공하는 경우에는 필요하다.” 고 판시하였다.

또한, In re Wands 판결¹⁴⁰⁾에서 CAFC는 과도한 실험요건의 충족여부를 위해 (i) 필요로 하는 실험의 양(the quantity of experimentation necessary), (ii) 지시 또는 안내가 주어진 양(the amount of direction or guidance presented), (iii) 적용되는 사례의 존부(the presence or absence of working examples), (iv) 발명의 성질(the nature of the invention), (v) 과거 기술의 정도(the state of the prior art), (vi) 같은 기술에서 비교되는 발명의 기술정도(the relative skill of those in the art), (vii) 해당 기술분야에서 예견가능한지 여부(the predictability or unpredictability of the art), (viii) 청구의 폭(the breadth of the claims) 등 8가지 판단기준을 제시하였다.

Everlight 사건에서는 청구된 과정의 가능한 모든 반복의 공개는 평균적 기술자가 청구의 전체

138) Mehdi Poursoltani, Disclosing AI Inventions, TEXAS INTELLECTUAL PROPERTY LAW JOURNAL Vol.29, 2021, p.51

139) Mehdi Poursoltani, Disclosing AI Inventions, TEXAS INTELLECTUAL PROPERTY LAW JOURNAL Vol.29, 2021, p.51~52

140) 858 F.2d 731(Fed. Cir. 1988).

범위를 실행할 수 있도록 하기 위해 필요하지 않다. 그러나 발명자는 “평균적 기술자에게 잘 알려진 것”을 공개할 필요가 없다. 추가적으로, 발명자는 평균적 기술자의 능력에 의존할 수 있다. 본 발명의 새로운 측면에 대해서는 동일한 것을 할 수 없음에도 불구하고, 그는 본 발명의 새로운 측면에 대해서는 동일하게 할 수 없다. 소프트웨어 발명의 요소들에 관한 앞의 항들에서 설명한 바와 같이, 알고리즘은 평균적 기술자가 청구항에 기재된 발명을 제조하여 사용할 수 있도록 하기 위해 개시되어야 하는 종래의 소프트웨어 발명의 가장 중요한 부분이다.

1984년 셔우드사건에서 CAFC이전의 관세특허심판원이 처음으로 이 문제를 다루었다. 법원은 “컴퓨터 프로그램을 작성하는 것은 창의적인 능력의 최고 수준을 요구하는 작업일 수도 있고, 사무적인 기술의 단조로운 사용만을 요구할 수도 있다. 두 극단의 차이는 (입력으로 시작하는) 정보와 (출력으로 시작하는) 정보 사이의 격차를 메우기 위한 수학적 방법론의 창조에 있다. 이러한 브리지 잭 도구들이 공개된다면, 이 기술을 수행하는 모든 사람들에게 알려진 하찮은 도구들의 공개를 요구할 설득력 있는 이유가 없어 보일 것이다.” 고 판시하였다. 셔우드사건에서는 법원이 소프트웨어 발명에 관한 공시기준을 하향 조정하여 필요한 공시를 원하는 출력물에 매핑하는 방법론과 단계로 제한하였다. 법원은 알고리즘의 기본 코드나 구현은 평균적 기술자에게 알려진 사무적인 업무와 같아서 과도한 실험 없이도 가능하다고 지적했다. 게다가, 법원은 컴퓨터 코드를 번역기로 유추했다. 법원은 다른 프로그래밍 언어로 코딩하는 것은 다른 언어로 쓰는 것과 비슷하며, 번역기에 의해 다른 언어로 번역될 수 있다고 설명했다.

3. 일본

가. 기술공개 요건¹⁴¹⁾

기본적으로 일본의 특허법상 기술공개 요건은 우리나라와 큰 차이는 없다. 출원서에는 명세서(specification)와 필요한 도면 및 요약서를 첨부해야 하고(특허법 제36조 제2항), 그 명세서에는 ① 발명의 명칭 ② 도면의 간단한 설명 ③ 발명의 상세한 설명 ④ 특허청구범위를 기재해야 한다(특허법 36조 제3항). 출원서에 첨부해야 하는 것 중에 가장 중요한 것은 명세서이다. 명세서는 발명의 내용을 개시하는 기술문헌으로서의 기능을 하고, 또한 권리 내용의 확정 즉 권리서로서의 기능을 한다. 명세서라고 하는 제도의 채택은 1852년 영국특허법이 성문법으로서 세계에서 최초로 이고, 이후 각국이 이 제도를 채택하고 있는데, 오늘날에는 특허제도에 없어서는 아니 될 사항으로 되어있다.

명세서는 (a) 도면의 간단한 설명에 대하여, 실용신안 등록출원에 있어서 도면은 필수적이지만, 특허 출원에 있어서는 필요한 경우만 첨부하면 된다. 첨부된 도면의 상세한 설명은 발명의 상세한 설명에서 행하여지고 여기에서는 평면도·입면도·단면도 등의 기재와 도면의 주요한 부분을

141) 中山信弘, 特許法, 法文社, 2001, 182~183면.

나타내는 부호 설명의 기재 정도로 충분하다. (b) 발명의 상세한 설명에 대하여, 이 기재는 권리 범위와 직접적인 관계는 없으나 기술의 내용을 개시한다는 중요한 의미를 가지고 있다. 그 기재는 당해 발명이 속하는 기술분야에 있어서 통상의 지식을 가진 자(평균적 기술자)가 용이하게 실시할 수 있을 정도로 명확하고 충분하게 기재해야 한다. 특허는 기술을 공개한 대가로써 부여되는 것이고, 기술공개는 평균적 기술자가 용이하게 실시할 수 있는 정도로 기재되어야 한다는 것은 당연하다. 평균적 기술자가 용이하게 실시할 수 없는 기재는 기술문헌으로서의 가치가 적고 출원의 거절이유가 되고(특허법 제49조 제4항), 비록 특허가 부여된다고 하더라도 무효사유가 된다(특허법 제123조 제1항 제4호).

AI의 발명에 있어서도 상세한 설명을 통하여 구체적인 기술사항이 공개되어야 하며, 용이하게 실시할 수 없는 정도라고 하면 이는 공개요건을 위반한 것으로 거절사정이나 무효화될 수 있다.

나. 발명해당성

일본특허법 제2조 제1호에는 “발명이라 함은 자연법칙을 이용한 기술적 사상의 창작으로서 고도한 것을 말한다.” 라고 정의하고 있으므로, 청구항에 기재된 발명이 특허법상의 발명이 되기 위해서는 이 규정을 만족해야 한다. 결국 AI 관련 발명도 특허를 받기 위해서는 반드시 제2조 제1호에서 규정하는 발명해당성을 만족하여야 한다.

일반적으로 청구항에 기재된 발명이 자연계에 존재하는 자연법칙 그 자체, 자연계에 이미 존재하는 물건이나 법칙을 단순히 찾아내는 발견, 자연법칙에 위배되는 것, 자연법칙을 이용하지 않은 경제법칙·수학공식·논리학적 법칙·작도법 등 자연법칙을 이용하지 않은 것, 게임 규칙과 같은 인위적인 약속, 영업계획·교수방법·과세방법 그 자체에 해당하는 인간의 정신활동, 개인의 숙련에 의해서 달성되는 기능, 단순한 정보의 제시, 수학적 과제의 해결방법 자체 또는 수학적 계산순서를 나타내는 것에 지나지 않는 SW 또는 컴퓨터에 정보처리를 수행하기 위한 지시의 순서에 지나지 않는 인간의 사고과정을 표현한 SW 등은 발명해당성을 만족하지 못하기 때문에 특허대상이 될 수 없는 것으로 보고 있다. 따라서 AI 관련 발명도 위에서 설명한 사항을 청구범위에 기재하는 경우에는 일본특허법 제2조 제1호에서 규정하는 발명해당성을 만족하지 못하기 때문에 특허대상이 될 수 없다.¹⁴²⁾

다. AI발명의 발명해당성 판단기준¹⁴³⁾

특히 「특허·실용신안심사핸드북」의 부속서 B 제1장에는 SW 관련 발명에대한 심사기준으로 「컴퓨터 SW 관련 발명의 심사기준¹⁴⁴⁾」을 규정하고 있고, 동 심사기준에 의하여 AI 관련 발명

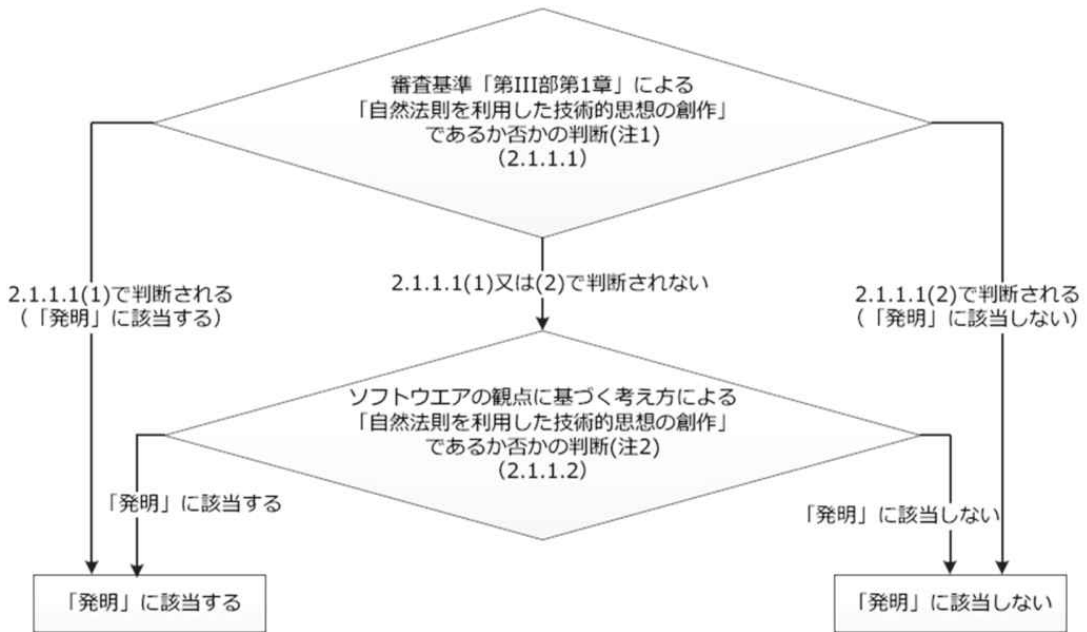
142) 함영욱, AI 관련 발명의 성립성 판단기준에 관한 연구, 한양대학교 박사학위논문, 2022, 105~106면.

143) 함영욱, AI 관련 발명의 성립성 판단기준에 관한 연구, 한양대학교 박사학위논문, 2022, 115~116면.

144)) 日本特許庁, 「特許・実用新案審査ハンドブック 附属書B 第1章 コンピュータソフトウェア関連発

이 발명해당성을 만족하는지가 판단되고 있다.

[그림 3-3] 발명해당성 판단



<출처 : 일본 특허청>

AI를 포함한 SW발명이 “자연법칙을 이용한 기술적 사상의 창작”에 해당하는지 여부가 「특허실용신안 심사기준」 “제III부 제1장 발명해당성 및 산업상이용가능성”에 의해 판단되지 않을 경우, 심사관은 아래에 제시된 기본적인 사고에 근거하여 판단해야한다고 규정하고 있다. 즉 SW발명이 “자연법칙을 이용한 기술적 사상의 창작”이 되는 기본적인 사고는 (i) SW발명 중 SW에 대해서는 “SW에 의한 정보처리가 하드웨어 자원을 이용해 구체적으로 실현되고 있다.”고 판단되는 경우에는 해당 SW는 “자연법칙을 이용한 기술적 사상의 창작”으로 본다. “SW에 의한 정보처리가 하드웨어를 이용해 구체적으로 실현되고 있다.”고 하는 것은 SW와 하드웨어 자원이 협동함으로써 사용목적에 따른 특유의 정보처리장치 또는 그 동작방법이 구축되는 것을 말하고, (ii) SW발명 중 SW와 협동하여 동작하는 정보처리장치 및 그 동작방법 및 SW를 기록한 컴퓨터 판독 가능한 기록매체에 관해서는 당해 SW가 위 (i)를 충족하는 경우에는 “자연법칙을 이용한 기술적 사상의 창작”에 해당하여 발명해당성을 만족하여 특허대상이 될 수 있는 것으로 판단한다. 이러한 발명해당성 판단에 관한 기본적 사고의 개념에 근거한 판단의 단계를 검토하면, 심사관은 위 기본적 사고에 근거하여 청구항에 기재된 SW발명에서 “SW에 의한 정보처리가 하드

明」, SW審査基準, 2021. 3.
https://www.jpo.go.jp/system/laws/rule/guideline/patent/document/cs_shinsa/cs_shinsa.pdf.
 <최종검색, 2022.11.18.>

웨어 자원을 이용하여 구체적으로 실현되고 있는지” 여부, 즉 SW와 하드웨어 자원이 협동함으로써 사용목적에 따른 특유의 정보처리장치 또는 그 동작방법이 구축되는지 여부에 따라 “자연법칙을 이용한 기술적 사상의 창작”의 요건을 판단한다. 여기서 심사관은 청구항에 기재된 SW 발명이 SW와 하드웨어 자원이 협동한 구체적 수단 또는 구체적 단계에 의해 사용목적에 따른 특유의 정보의 연산 또는 가공이 실현되고 있는지 여부를 판단해야 한다.

라. 데이터 구조와 관련한 발명해당성 판단¹⁴⁵⁾

(1) 심사기준의 적용

일본은 AI를 포함한 SW발명도 방법발명 또는 물건발명으로 청구항에 기재할 수 있다. 즉 SW의 방법발명은 시계열로 연결된 일련의 처리 또는 조작, 즉 ‘단계’로 표현할 수 있을 때에 그 ‘단계’를 특징하는 것에 의해 방법발명(물건을 생산하는 방법발명 포함)으로서 청구항에 기재한 것을 말한다. 또 SW의 물건발명은 당해 발명이 완수하는 복수의 기능으로 표현할 수 있을 때에 그 기능으로 특정된 물건의 발명으로서 청구항에 기재할 수 있으며, 프로그램 또는 구조를 갖는 데이터, 데이터구조도 물건발명으로 취급하고 있다. 따라서 데이터가 갖는 구조에 의하여 컴퓨터가 실시하는 정보처리가 규정되는 (i) ‘구조를 갖는 데이터’ 또는 ‘데이터구조’를 청구항의 말미에 기재하는 경우, (ii) ‘프로그램 또는 구조를 갖는 데이터, 데이터구조를 기록한 컴퓨터판독 가능한 기록매체’를 물건발명으로서 청구항에 기재하는 경우에는 발명해당성을 만족하여 특허대상이 될 수 있는 것으로 판단하고 있다.

(2) 데이터구조에 대한 판단

AI기술의 필수 구성요소인 데이터구조에 대한 발명해당성 판단에 있어서, 「컴퓨터 SW 관련 발명의 심사기준」은 ‘구조를 갖는 데이터’ 및 ‘데이터구조’가 프로그램과 같이 SW에 준하는 것, 즉 데이터가 갖는 구조가 컴퓨터의 처리를 규정하는 것이라고 하는 점에서 프로그램과 유사한 성질을 가지는 것인지를 판단하여야 한다고 규정하고 있다. ‘구조를 갖는 데이터’ 및 ‘데이터구조’가 프로그램에 준하는 것일 경우에는 이들은 컴퓨터 SW로 판단하고, 만약 ‘구조를 갖는 데이터’ 및 ‘데이터 구조’라 하더라도 프로그램에 준하는 것이 아닐 경우에는 발명해당성을 만족하지 못하기 때문에 특허대상인 SW발명이 아닌 것으로 판단한다. SW인 ‘구조를 갖는 데이터(구조를 갖는 데이터를 기록한 컴퓨터 판독 가능한 기록매체 포함)’ 또는 ‘데이터구조’가 “자연법칙을 이용한 기술적 사상의 창작”에 해당하는지 여부에 대해서는 앞에서 설명한 SW발명의 발명해당성 심사와같은 근거하여 판단한다. 또 SW인 ‘구조를 갖는 데이터’

145) 함영욱, AI 관련 발명의 성립성 판단기준에 관한 연구, 한양대학교 박사학위논문, 2022, 119~121면.

또는 ‘데이터구조’에 관해서는 「컴퓨터 SW 관련 발명의 심사기준」 2.1.1.2의 AI(SW)발명의 발명해당성 판단의 사고방식과 같은 관점에서 데이터가 가지는 구조가 규정하는 정보처리가 하드웨어 자원을 이용해 구체적으로 실현되고 있는지 여부에 대하여 판단하여야 한다. 즉 심사관은 청구항에 기재된 SW발명이 SW(프로그램에 준하는 데이터구조)와 하드웨어 자원이 협동한 구체적인 수단 또는 구체적 단계에 의해 사용목적에 따른 특유의 정보의 연산 또는 가공이 실현되고 있는지 여부를 판단하여, 발명해당성을 만족하는지 여부를 결정하고 있다.

마. 실무 가이드

출원인은 소프트웨어 관련 발명을 방법의 발명 또는 물건의 발명으로서 다음과 같이 청구항에 기재할 수 있다.

(1) 방법의 발명

출원인은 소프트웨어 관련 발명을 시계열로 연결된 일련의 처리 또는 조작, 즉 순서로서 표현할 수 있을 때에 그 순서를 특정함으로써, 방법의 발명(물을 생산하는 방법의 발명을 포함)으로서 청구항에 기재할 수 있다.

(2) 물건의 발명

출원인은 소프트웨어 관련 발명을, 그 발명이 완수하는 복수의 기능에 의해 표현할 수 있을 때에 그 기능에 의해 특정된 「물건의 발명」으로서 청구항에 기재할 수 있다. 출원인은 프로그램, 구조를 갖는 데이터 및 데이터 구조에 대해서는 이하와 같이 기재할 수 있다.

(i) 컴퓨터가 수행하는 복수의 기능을 특정하는 프로그램은 물건의 발명으로서 청구항에 기재될 수 있다.

예 1: 컴퓨터가 절차 A, 절차 B, 절차 C, ...를 실행하도록 하는 프로그램

예 2: 컴퓨터를 수단 A, 수단 B, 수단 C, ...로서 기능시키기 위한 프로그램

예 3: 컴퓨터가 기능 A, 기능 B, 기능 C, ...를 실현시키기 위한 프로그램

(ii) 데이터가 갖는 구조에 의해 컴퓨터가 행하는 정보 처리가 규정되는 「구조를 갖는 데이터」 또는 「데이터 구조」를, 「물건의 발명」으로서 청구항에 기재할 수 있다.

예 4: 데이터 요소 A, 데이터 요소 B, 데이터 요소 C, ...를 포함하는 구조를 갖는 데이터

예 5: 데이터 요소 A, 데이터 요소 B, 데이터 요소 C, ...를 포함하는 데이터 구조

(iii) 상기 (i)의 프로그램 또는 상기 (ii)의 구조를 갖는 데이터를 기록한 컴퓨터 판독 가능한 기

록 매체는 물건의 발명으로서 청구항에 기재 될 수 있다.

예 6: 컴퓨터가 절차 A, 절차 B, 절차 C, ...를 수행하도록 프로그램을 기록한 컴퓨터 판독가능한 기록 매체

예 7: 컴퓨터가 수단 A, 수단 B, 수단 C,... 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록 매체

예 8: 컴퓨터에 기능 A, 기능 B, 기능 C 등을 실현시키기 위한 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록 매체

예 9: 데이터 요소 A, 데이터 요소 B, 데이터 요소 C, ...를 포함하는 구조를 갖는 데이터를 기록한 컴퓨터 판독 가능한 기록 매체

4. EU

가. 기술공개 요건

EPC 제78조는 특허청구, 상세한 설명, 청구항, 상세한 설명 또는 청구항에서 인용되는 도면 및 요약서를 첨부토록 하고 있다. 특히, EPC 제83조는 EU 특허출원은 통상의 기술자에 의해 실시될 수 있도록 충분히 명확하고 완전한 방법(manner)으로 공개하여야 한다¹⁴⁶⁾고 규정하고 있다.

(1) 기술의 공개

상세한 설명은 발명의 실질적인 내용을 개시하는 부분으로, 그 발명의 기술분야에 대한 설명, 그 발명의 기술적 과제와 해결책, 발명의 실시예 등이 기재되어야 한다. 상세한 설명에는 그 발명이 평균적 기술자에 의해 충분히 명확하고 완전하게 수행될 수 있도록 개시하여야만 한다. 이 목적을 위한 통상의 기술자(a person skilled in the art)는 출원 그 자체의 가르침 및 거기에 있는 인용문헌뿐만 아니라 그 출원의 출원일에 그 기술분야에서 일반적인 상식을 알고 있는 일상적인 업무종사자(ordinary practitioner)이다. 통상의 기술자는 해당 기술분야에 대해 일반적인, 실험(experimentation) 및 일상적인 일에 대한 능력과 수단을 임의로 처리할 수 있는 것으로 여겨진다. 일반적인 상식은 일반적으로 해당 대상에 대한 핸드북, 전공논문(monographs) 및 교과서(textbooks)에 포함된 정보로 여겨질 수 있다. 예외로서, 만약 그 발명이 너무 새로워서 관련된 기술적 지식이 아직 교과서로부터 이용할 수 없는 조사 분야에 있으면, 특허명세서 또는 과학적 공개물에 포함된 정보도 될 수 있다. 공개의 충분성은 상세한 설명, 청구항 및 만약 있다면, 도면을 포함하여 전체적으로 출원을 기초로 판단되어야만 한다. 상세한 설명의 내용에 관련된 규정의 목적은 통상의 기술자가 청구된 발명을 실행할 수 있도록 충분한 기술적 정보를 출원이 포함하

146) Article 83 Disclosure of the invention. The European patent application shall disclose the invention in a manner sufficiently clear and complete for it to be carried out by a person skilled in the art.

도록 보장하고, 그리고 청구된 발명이 한 그 기술분야에의 공헌을 읽는 사람이 이해할 수 있도록 하는 것이다.¹⁴⁷⁾

(2) 공개의 충분성(sufficiency of disclosure)¹⁴⁸⁾

상세한 설명에는 출원된 발명을 수행하는 최소한 하나의 실시예에 대한 자세한 설명이 주어져야만 한다. 출원은 그 기술분야의 통상의 기술자에게 초점이 맞춰지므로, 잘 알려진 부수적인 특징의 상세 사항을 설명하는 것은 필요하지도 바람직하지도 않지만, 상세한 설명은 통상의 기술자가 그 발명을 실행하는 방법이 명백하게, 충분히 상세하게 그 발명의 수행을 위해 필수적인 특징을 설명하여 공개하여야만 한다. 단일의 예로 충분할 수 있으나, 청구항이 넓은 범위를 커버하는 경우, 그 출원은 청구항에 의해 보호되는 영역에 대해 확장되는 변형예 또는 다른 실시예들을 설명하거나 다수의 예를 상세한 설명이 제시하여야 한다. 하지만, 매우 넓은 분야일지라도 제한된 숫자의 실시예 또는 하나의 실시예로도 충분히 설명될 수 있는 경우의 예들이 있다. 이들 후자의 경우에 그 출원은, 실시예들 외에, 통상의 기술자가 발명적 기술을 필요로 함이 없이 또 과도한 부담 없이 일반적 상식을 사용하여 청구된 모든 분야에 대해 그 발명을 실시할 수 있도록 충분한 정보를 포함하여야 한다. 심사부가 그 출원이 충분한 공개를 결여하고 있다고 주장하는 경우, 그 발명이 실질적으로 그 청구된 범위 전체에 걸쳐 재현되고 수행될 수 있다는 것을 입증하는 책임은 출원인에게 있다. 발명이 유럽특허조약의 규정을 충분히 만족시키기 위해서는, 다양한 부분의 기능이 직접적으로 명백하지 않다면, 그 구조에 의해서뿐만 아니라 기능에 의해서도 설명되는 것이 필요하다. 실제로, 몇몇 기술적 분야에서(예를 들면 컴퓨터), 기능의 명확한 설명이 매우 상세한 구조의 설명보다 더 적절할 수 있다. 출원에 있어서, 상세한 설명에 충분한 공개를 제공하는 것은 출원인의 책임인데, 즉, 모든 청구항에서 청구된 것과 같은 발명에 대하여 유럽특허조약의 요건을 만족시키는 것이다. 만약 청구항이 파라미터에 의해 그 발명 또는 그의 특징을 정의한다면, 제출된 출원은 평균적 기술자가 어떤 방법이 사용되었는지를 알 수 없거나 또는 모든 방법이 동일한 결과를 낼지 알 수 없는 한, 그 파라미터값을 결정하기 위해 사용된 방법의 명백한 설명을 포함하여야 한다.

나. AI 발명의 특허적격성¹⁴⁹⁾

AI 관련 발명의 특허적격성은 EPC 제52조 제2항 및 제3항에 규정하고 있는 발명의 특허적격성을 기초로 판단하는 것이 원칙이다. 다만, EPC 제52조에는 발명의 특허적격성의 판단기준이 되는 발명의 정의를 명문으로 규정하고 있는 것이 아니라, 발명의 특허적격성이 없는 대상을 소극

147) 박동식, 유럽특허법, 세창출판사, 2009, 124면.

148) 박동식, 유럽특허법, 세창출판사, 2009, 127면.

149) 함영욱, AI 관련 발명의 성립성 판단기준에 관한 연구, 한양대학교 박사학위논문, 2022, 136~137면.

적 열거방식으로 제시하고 있다. EPC 제52조 제1항에는 특허를 받을 수 있는 발명의 특허요건을 규정하고 있고, 동조 제2항에는 발명의 특허적격성이 없어서 제외해야 한다는 대상을 열거하고 있으며, 동조 제3항에는 제2항의 적용을 다시 제한하고 있다. 즉, 제52조 제1항은 “산업상 이용 가능성이 있는 발명으로서 신규하고 진보성이 있는 발명에 대해서는 유럽특허를 부여해야 한다.” 고 규정하고 있고, 동조 제2항은 “(a) 발견, 과학적 이론, 수학적 방법, (b) 심미적 창작, (c) 계획, 규칙, 정신적 작용을 수행하는 방법, 게임을 하거나 영업을 하는 것 및 컴퓨터 프로그램, (d) 정보의 제시” 는 “제1항의 발명으로 간주되지 않는다.” 는 특허대상 제외를 규정하고 있으며, 동조 제3항은 “유럽특허 출원이나 유럽특허가 제2항에 열거된 대상이나 행위 그 자체에 관한 경우에만 특허적격성을 부인한다.” 고 규정하고 있다.¹⁵⁰⁾

따라서 EPC 제52조 제2항 및 제3항의 규정에 의하여 AI 관련 발명의 특허적격성은 다음과 같이 판단되고 있다. 첫째, 청구항에 기재된 AI 관련 발명이 “(a) 발견, 과학적 이론, 수학적 방법” 에 관한 것으로 추상적 아이디어에 해당하는 경우에는 특허적격성을 만족하지 못하기 때문에 특허를 받을 수 없다. 예를 들면, AI 관련 발명이 데이터구조를 이용한 딥러닝 학습 프로세스의 구체적인 알고리즘이 특정되어 있지 않고, 단순히 수학적 알고리즘 형태로 기재된 경우에는 특허적격성을 만족하지 못하는 것으로 판단한다. 둘째, 청구항에 기재된 AI 관련 발명이 “(b) 심미적 창작” 에 관한 것으로 추상적 아이디어에 해당하는 경우에는 특허적격성을 만족하지 못하기 때문에 특허를 받을 수 없다. 예를 들면, 청구항에 AI 딥러닝학습에 의하여 도출된 최적의 학습완료모델에 이용자가 생각한 특정한 도형, 그림, 숫자, 문자 등을 입력데이터로 입력하여 생성된 회화 또는 동물 모습은 AI의 생성물이지만, 이는 미적 창작물에 해당하기 때문에 특허적격성을 만족하지 못하는 것으로 판단한다. 셋째, 청구항에 기재된 AI 관련 발명이 (c) 정보의 제시에 관한 것으로 추상적 아이디어에 해당하는 경우에는 특허적격성을 만족하지 못하기 때문에 특허를 받을 수 없다. 예를 들면, 청구항에 기재된 AI 관련 발명으로 데이터구조가 기재되어 있지만, 실제로는 데이터 요소 사이의 상호관계를 표시한 데이터의 논리적 구조로 볼 수 없고 단순히 데이터 정보를 제시한 것에 불과한 경우에는 추상적 아이디어에 해당하여 특허적격성을 만족하지 못하는 것으로 판단한다. 넷째, 청구항에 기재된 AI 관련 발명이 “(d) 계획, 규칙, 정신적 작용을 수행하는 방법, 게임을 하거나 영업을 하는 것 및 컴퓨터프로그램” 에 관한 것으로 추상적 아이디어에 해당하는 경우에는 특허적격성을 만족하지 못하기 때문에 특허를 받을 수 없다. 예를 들

150) EPC Article 52 (Patentable Inventions)(1) European patents shall be granted for any inventions, in all fields of technology, provided that they are new, involve an inventive step and are susceptible of industrial application.(2) The following in particular shall not be regarded as inventions within the meaning of paragraph 1:(a) discoveries, scientific theories and mathematical methods;(b) aesthetic creations;(c) schemes, rules and methods for performing mental acts, playing games or doing business, and programs for computers;(d) presentations of information.(3) Paragraph 2 shall exclude the patentability of the subject-matter or activities referred to therein only to the extent to which a European patent application or European patent relates to such subject-matter or activities as.

면, 청구항에 인간의 인위적 행위가 구성요소로 기재되어 있는 경우, 또는 청구항에 프로그램 그 자체가 특정되어 있는 경우에는 추상적 아이디어에 해당하여 특허적격성을 만족하지 못하는 것으로 판단한다.

다. 인공지능발명을 위한 심사지침의 개정

(1) 2018년 심사지침

2018년 심사지침을 개정하여 인공지능 발명에 대한 심사 근거를 마련하고 있다. 심사지침 3.3.1에서는 인공지능과 기계학습에 대한 기준을 제시하고 있다.

인공지능 및 기계학습은 신경망, 유전자 알고리즘, 서포트 벡터 머신, k-평균, 커널 회귀 및 판별 분석과 같은 분류, 클러스터링, 회귀 및 차원 축소를 위한 계산 모델 및 알고리즘을 기반으로 한다. 이러한 계산 모델 및 알고리즘은 학습데이터를 기반으로 학습될 수 있는지 여부에 관계없이 본질적으로 추상적인 수학적 특성이다. 따라서 심사지침 3.3에서 제공하는 지침은 일반적으로 이러한 계산 모델 및 알고리즘에도 적용된다.

지원 벡터 머신, 추론 엔진 또는 신경망과 같은 용어는 문맥에 따라 추상 모델 또는 알고리즘을 참조할 수 있으므로 그 자체로 반드시 기술적 수단의 사용을 의미하지는 않는다. 이는 청구된 주제가 전체적으로 기술적 특성을 갖는지 여부를 검토할 때 고려되어야 한다. 인공지능과 기계학습은 다양한 기술 분야에서 응용 프로그램을 찾는다. 예를 들어, 불규칙한 심장 박동을 식별하기 위한 목적으로 심장 모니터링 장치에서 신경망을 사용하는 것은 기술적 기여를 한다. 낮은 수준의 기능(예: 이미지의 가장자리 또는 픽셀 특성)을 기반으로 하는 디지털 이미지, 비디오, 오디오 또는 음성 신호의 분류는 분류 알고리즘의 일반적인 기술적 응용 프로그램이다. 인공지능과 기계학습을 사용할 수 있는 기술적 목적의 추가 예는 심사지침 3.3의 목록에서 찾을 수 있다. 그러나 텍스트 내용에 대해서만 텍스트 문서를 분류하는 것은 그 자체로 기술적 목적이 아니라 언어적 목적으로 간주된다. 분류 알고리즘이 견고성과 같은 가치 있는 수학적 속성을 가지고 있는 것으로 간주될 수 있더라도 최종 분류가 기술적 용도로 사용된다는 표시 없이 추상 데이터 레코드 또는 심지어 통신 네트워크 데이터 레코드를 분류하는 것도 그 자체로 기술적 목적이 아니다. 분류 방법이 기술적인 목적을 제공하는 경우 학습셋을 생성하고 분류자를 학습시키는 단계는 해당 기술적 목적을 달성하는 데 도움이 되는 경우 발명의 기술적 특성에 기여할 수도 있다.

EPO는 심사지침에 의하여 AI와 기계학습은 계산적 모델과 알고리즘을 기반으로 하고 있어 그 자체로 추상적인 수학적 성격을 지닌다고 언급하며 특허적격성을 만족하지 못하기 때문에 특허 대상이 될 수 없다는 입장이지만, AI 관련 발명이 구체적인 기술적 특성 및 그 기술적 특성의 기여도에 의하여 달성되는 효과가 있는 경우에는 한정적으로 특허적격성을 만족하는 것으로 하여 특허대상에 포함시키고 있다.¹⁵¹⁾

(2) 2021년 심사지침

2021년 개정에서는 데이터 관련 발명에 대해서도 심사근거를 마련하고 있다. 즉, EPO는 AI 관련 발명의 특허적격성 판단기준을 추가하는 EPO 심사지침을 개정하여 공표하였다. 이번에 개정된 EPO심사기준의 Part G 중에서 AI 관련 발명과 관련한 EPC 제52조의 특허적격성 판단에 대한 것으로는 데이터관리시스템 및 정보검색의 심사에 관한 판단기준으로 Part G 제2장 3.6.4(데이터베이스 관리 시스템 및 정보검색)을 추가하였다.

특히, 2021년 3월 시행된 EPO 심사지침의 Part G 제2장 3.6.3에는 “매체 또는 전자파로 상징되는 컴퓨터로 구현된 데이터구조(A computer-implemented datastructure data structure) 또는 데이터포맷(data format)은 전체적으로 기술적특성을 가지며, 따라서 기술의 의미 내에서 EPC 제52조 제1항의 발명에 해당한다.” 는 판단기준을 제시하고 있다.

라. 데이터구조 및 데이터포맷의 특허적격성

EPO심사지침 3.6.3에 의한 데이터구조 또는 데이터포맷 발명의 특허적격성 판단기준을 검토하여 보면, (i) 데이터구조 또는 데이터포맷은 기술적 효과를 생성하는 경우 발명의 기술적 특성에 기여한 것으로 판단하고, (ii) 이러한 판단은 데이터구조 또는 데이터포맷이 기능데이터(functional data)인 경우, 즉 데이터를 처리하는 장치의 작동을 제어하는 것과 같은 기술 시스템에 기술적 기능이 있는 경우에 발생할 수 있는 것으로 판단하고, (iii) 기능데이터는 본질적으로 장치의 해당 기술의 기능을 구성하는 것으로 판단하고, 반면 인지데이터(Cognitive data)는 내용과 의미가 인간 사용자에게만 관련되고 기술적 효과를 생성하는데 기여하지 않는 데이터로 특허적격성을 만족하지 못하기 때문에 특허대상이 될 수 없는 것으로 판단하고 있다. 이러한 판단기준에 의하면, 데이터구조 또는 데이터포맷은 인지데이터(즉, 사용자에게 정보를 전달하기 위한 것이 아님)로 특성화되지 않을 수 있지만, 그럼에도 불구하고 기술적 기여를 하지 않는 특징을 가질 수 있다. 예를 들면, 프로그램의 구조는 단순히 프로그래머의 작업을 용이하게 하는 것을 목표로 할 수 있으며, 이는 기술적 기능을 제공하는 기술적 효과가 있는 것은 아니다. 따라서 추상 논리적 수준의 데이터모델과 기타 정보 모델은 그 자체로 기술적 특성이 없는 것으로 판단하고 있다. 따라서 EPO는 청구항에 기재된 데이터구조 발명이 헤더와 콘텐츠 섹션이 있는 전자 메시지가, 헤더의 정보가 수신 메시지 시스템에 의해 자동으로 인식되고 처리되는 명령으로 구성되어 있고, 그 처리는 콘텐츠 요소를 조합하여 최종 수신자에게 표시하는 방법을 결정하는 수단으로 구성된 경우, 헤더에 이러한 명령을 제공하는 것은 전자 메시지의 기술적 특성에 기여하는 반면, 인지데이터를 나타내는 콘텐츠 섹션의 정보는 기술적 특성에 기여하지 않는 것으로 보고, EPC 제52조의 특허적격성을 만족하는지를 판단하게 된다.

151) 함영욱, AI 관련 발명의 성립성 판단기준에 관한 연구, 한양대학교 박사학위논문, 2022, 144면.

마. 알고리즘 등에 대한 공개

EPO는 <Update of legal aspects of AI and patents>¹⁵²⁾라는 제목의 보고서에서 발명의 공개라는 챕터에서 다음과 같이 밝히고 있다.

인공지능 발명의 출원은 특정한 인공지능 시스템 및/또는 인공지능 알고리즘이 특정한 기술적 구현에 어떻게 적용되고 있는지 또는 그것이 인정된 기술적 구현에 어떻게 적용되는지에 대해 상세히 설명해야 한다. 기술 분야 기술 분야는 명세서에 제시되어야 하며 객관적인 기술 문제는 명세서에서 직접 도출되어야 한다. 이 정보는 숙련된 사람의 지식 수준을 결정할 수 있게 한다. AI 발명을 공개하는 데 있어 인식된 어려움은 때때로 미생물 기탁과 유사하게 AI 애플리케이션이나 학습데이터에 대한 공탁 시스템을 만들어 해결해야 한다고 제안된다. 그러나 이러한 시스템은 알고리즘의 경우 적합하지 않을 것으로 보인다. 첫째, 주제와 보호범위에 따라 전체 알고리즘을 항상 공개할 필요는 없다. 둘째, 미생물과 달리 알고리즘은 서면으로 설명할 수 있다. 따라서 (특별 취급이 필요한 미생물의 공개를 보호하기 위한) 부다페스트 조약에 따른 미생물 공탁에 대한 근거는 알고리즘에 적용되지 않는다. 그것은 어떤 부가가치도 제공하지 않을 것이다. 셋째, 이러한 기탁제도의 도입은 국제적인 승인을 필요로 할 것이며 시스템의 관리는 복잡할 수 있다.

5. 중국

가. 기술공개 요건

(1) 특허법상 요건

중국 특허법은 명세서는 발명 또는 실용신안에 대하여 명확(淸楚)하고 완전(完整)한 설명을 하여야 하며, 해당 기술영역의 기술자가 실현할 수 있는지를 기준으로 한다. 필요한 때에는 첨부도면이 있어야 한다. 요약서는 발명 또는 실용신안의 기술요점을 간단하고 명료하게 설명하여야 한다고 규정하고 있다(제26조 제3항)¹⁵³⁾. 이 규정은 국가가 수여하는 독점권을 획득하는 전제조

152) Update of legal aspects of artificial intelligence and patents, CA/PL 5/20, Munich, 23.10.2020

https://www.epo.org/modules/epoweb/acdocument/epoweb2/468/en/CA-PL_5-20_en.pdf
<2022.11.22. 접속>

153) 중국특허법 제26조 ①발명 또는 실용신안 특허를 출원하는 자는 명세서, 설명서, 요약서, 청구항 등의 서류를 제출하여야 한다.

②명세서에는 발명 또는 실용신안의 명칭, 발명자의 성명, 출원인의 성명 또는 명칭, 주소 및 기타 사항을 명시해야 한다.

③명세서는 해당 기술 분야의 기술자가 실현할 수 있는 것을 기준으로 발명 또는 실용신안에 대한 명확하고 완전한 설명을 제공해야 하며 필요한 경우 도면이 있어야 한다. 요약서는 발명 또는 실용신안의 기술적 요점을 간략하게 설명해야 합니다.

④청구항은 명세서를 근거로 특허 보호의 범위를 명확하고 간결하게 정의해야 합니다.

⑤유전자원에 의존하여 완성된 발명 및 창작의 경우 출원인은 특허출원서류에 해당 유전자원의 직접 출

건으로서 출원인이 명세서에서 발명의 내용을 충분히 공개해야 한다는 것을 의미이며, 동항에서 “명확“하다는 것은 ① 내용이 명확해야 한다는 것과 ② 기술용어가 정확해야 한다는 것을 의미하며, “완전“하다는 것은 해당 기술영역의 기술자가 발명을 이해하고 실시할 수 있을 정도로 기재한 것을 의미한다. 요약서는 발명 또는 실용신안의 기술요점을 간단하고 명료하게 설명하여야 한다“고 규정하고 있다. 심사지침에서 요약서는 300자 이내에서 작성하도록 하고 있으며, 요약서가 300자를 초과하는 경우 출원인에게 통지하여 감축하도록 하거나 심사관이 감축할 수 있다. 심사관이 감축한 경우에는 출원인에게 통지하여야 한다. 요약서가 심사지침이 규정하고 있는 사항을 만족하지 않더라도 거절이유에 해당하지 않는다.¹⁵⁴⁾

(2) 특허법 실시세칙에 따른 규정

특허법 실시세칙 제17조 ① 발명 또는 실용신안 특허출원의 명세서는 발명 또는 실용신안의 명칭을 명확하게 기재해야 하고, 그 명칭은 출원서 중의 명칭과 일치해야 한다. 명세서는 아래의 내용을 포함해야 한다. 1. 기술영역 : 보호를 요구하는 기술방안이 속하는 기술영역을 명확히 기재 한다. 2. 배경기술 : 발명 또는 실용신안의 이해, 검색, 심사에 유용한 배경기술을 명확히 기재하고, 가능한 경우 이러한 배경기술을 반영하는 서류를 인용하여 증명한다. 3. 발명내용 : 발명 또는 실용신안이 해결하고자하는 기술문제 및 그 기술문제를 해결하는데 채용한 기술방안을 명확히 기재하고, 선행기술과 발명 또는 실용신안을 대비하여 유익한 효과를 명확히 기재한다. 4. 첨부도면의 설명 : 명세서에 첨부도면이 있는 경우는, 각 도면에 대하여 간략한 설명을 작성한다. 5. 구체적인 실시방식 발명 또는 실용신안을 실현하는 우수한 방식을 출원인이 알고 있는 것에 대하여 상세하고 명확하게 기재하고, 필요한 때에는 예를 들어 설명하며, 첨부 도면이 있는 경우에는 첨부 도면과 대비한다. ② 발명 또는 실용신안 특허출원인은 전항 규정의 방식과 순서에 따라 명세서를 작성해야 하고, 명세서의 각 부분 앞에는 제목을 명확히 기재해야 한다. 그 발명 또는 실용신안의 성질이 기타 방식이나 순서를 이용하여 작성하면 명세서의 지면을 절약할 수 있고 타인이 그 발명 또는 실용신안을 정확하게 이해할 수 있게 하는 경우는 제외한다. ③ 발명 또는 실용신안의 명세서는 규범적으로 용어를 사용해야 하고 문장이 명확해야 하며, “청구항에 기재한 . . . 바와 같이”의 인용어구를 사용해서는 안 되며, 상업성 선전용어를 사용해서도 안 된다. ④ 특허출원이 하나 또는 여러 개의 아미노산 또는 뉴클레오티드 서열을 포함하는 경우에는, 명세서는 국무원전리행정부문 규정에 부합하는 서열표를 포함해야 한다. 출원인은 당해 서열표를 명세서의 독립적인 부분으로 하여 제출해야 하고, 국무원전리행정부문의 규정에 따라 당해 서열표의 컴퓨터가 읽을 수 있는 형식의 부분을 제출해야 한다. ⑤ 실용신안특허출원 명세서는 보호를 요구하는 물품의 형상, 구조 또는 그 결합을 표시하는 첨부도면이 있어야 한다.¹⁵⁵⁾

처와 원출처를 설명하여야 하며, 출원인이 원출처를 설명할 수 없는 경우에는 그 사유를 기재하여야 한다.

154) 김태수 외, 중국특허법, 한빛지적소유권센터, 2015, 133면.

(3) 정리

이 규정은 특허법 제26조에서 규정된 발명의 상세한 설명에 기재된 발명의 명칭과 청구항에 기재된 발명의 명칭이 일치해야 한다는 것과 명세서에는 기술영역, 배경기술, 발명내용, 첨부도면의 설명, 구체적인 실시방식을 기재해야 한다는 등의 기재방식을 규정하고 있다. 실무적으로, 청구항의 발명의 명칭과 상세한 설명의 발명의 명칭이 일치하지 않기 때문에, 특허법 실시세칙 제17조를 이유로 심사의견통지서를 발부하는 경우가 종종 발생한다. 이는 단일성의 거절이유를 해소하기 위하여 청구항을 삭제하였지만, 삭제된 청구항에 기재된 명칭을 상세한 설명에서 그대로 사용하는 경우가 대부분이다.¹⁵⁶⁾

나. 인공지능발명을 위한 심사지침의 개정

중국은 2020년 2월 1일부터 개정된 특허심사지침을 시행 중에 있다. 국가지식산업권국은 <특허심사지침>의 제2부분 제9장에 제6절을 추가하였다. 제6절은 계산법(알고리즘) 특징, 또는 영업규칙 및 방법(BM) 특징이 포함된 발명특허 출원심사의 관련 규정이다. 인공지능, 인터넷+¹⁵⁷⁾, 빅데이터 및 블록체인 등과 관련된 발명특허 출원은 일반적으로 알고리즘 또는 영업규칙 및 방법 등 지적 활동의 규칙과 방법 특징이 포함되므로, 본 절에서는 특허법 및 그 실시세칙에 근거하여 이러한 유형의 출원심사의 특수성에 대해 규정한다.

다. 인공지능 알고리즘 관련 발명의 심사기준

알고리즘적 특징 또는 업무 규칙 및 방법 특징을 포함하는 특허출원의 설명은 그 기술적 문제를 해결하기 위해 본 발명에 의해 채택된 해결책을 명확하고 완전하게 설명해야 한다. 기술적 특징을 통합하는 것 외에도, 상기 솔루션은 기능적으로 지원하고 기술적 특징과 상호작용하는 알고리즘적 특징 또는 비즈니스 규칙 및 방법적 특징을 추가로 포함할 수 있다. 또한, 기술적 특징이 기능적으로 지원하고 기술적 특징과 상호작용하는 알고리즘적 특징 또는 비즈니스 규칙 및 방법적 특징과 함께 작동하는 방법을 설명해야 하며, 유리한 효과를 생성해야 한다. 또한, 설명은 선행기술에 비해 본 발명의 유리한 효과를 명확하고 객관적으로 나타내야 한다. 사용자 경험이 객관적으로 개선된 경우(사용자 관점에서), 이는 또한 상세한 설명에서 언급될 수 있으며, 이 개선된 사용자 경험이 어떻게 본 발명의 기술적 특징뿐만 아니라, 기능적으로 지원하는 알고리즘적 특징 또는 사업 규칙 및 절차적 특징에 의해 야기되거나 생성되는지를 설명한다.¹⁵⁸⁾

155) 김태수 외, 중국특허법, 한빛지적소유권센터, 2015, 133면.

156) 김태수 외, 중국특허법, 한빛지적소유권센터, 2015, 133면.

157) 중국의 인터넷 플러스(互聯網+)란 인터넷 플랫폼, 정보통신 기술을 활용해 인터넷을 전(全) 산업과 융합시켜 새로운 경제발전 생태계를 창조하는 전략을 의미한다.

158) WIPO, FURTHER STUDY ON THE SUFFICIENCY OF DISCLOSURE (PART I), SCP/34/5, SEPTEMBER 1, 2022, p. 24.

특허심사지침은 ‘알고리즘 발명특허의 심사규정’이라는 표제하에 인공지능 알고리즘 관련 발명의 심사기준, 특허적격성 및 특허등록요건 등에 대해 규정하고 있다. 특허심사지침은 다음과 같은 심사기준을 정하고 있다. 청구한 해결방안 즉, 청구범위에서 한정한 해결방안에 대해 심사를 진행해야 하는바, 기술적 특징과 알고리즘 특징을 단순히 분할해서는 안 되며, 청구범위에 기재된 모든 내용을 하나의 전체(整体)로 보고, 관련된 기술적 수단, 해결하려는 기술적 문제 및 발생한 기술적 효과에 대해 분석해야 한다는 기준을 설정하고 있다.¹⁵⁹⁾

(1) 특허법 제25조 제1항 제2호에 따른 심사

청구범위가 추상적인 알고리즘 또는 단순한 영업규칙 및 방법과 관련되고, 어떠한 기술특징도 포함하지 않는 경우, 이 청구범위는 특허법 제25조¹⁶⁰⁾ 제1항 제2호에 규정된 지적 활동의 규칙과 방법(智力活动的规则和方法)에 속하므로, 특허권이 부여될 수 없다. 예를 들어, 어떤 추상적인 알고리즘을 토대로 하고 어떠한 기술특징도 포함하지 않는 수학적 모형 구축 방법은, 특허법 제25조 제1항 제2호 규정에 따라 특허권이 부여될 수 없는 경우에 속한다. 또 다른 예로, 고객의 소비 한도에 따른 캐쉬백 방법의 경우, 이 방법에 포함된 특징이 모두 캐쉬백 규칙과 관련된 영업 규칙과 방법 특징이고, 어떠한 기술특징도 포함하지 않는 경우, 특허법 제25조 제1항 제2호 규정에 따라 특허권이 부여될 수 없는 경우에 속한다.

청구범위에 알고리즘 특징 또는 영업규칙 및 방법 특징 외에, 기술특징도 가지고 있고, 청구범위가 전체적으로 일종의 지적활동의 규칙과 방법이라고 볼 수 없는 경우, 특허법 제25조 제1항 제2호에 따라 특허권 부여의 가능성을 배제해서는 안 된다.

(2) 특허법 제2조 제2항에 따른 심사

보호를 요구하는 청구범위가 전체적으로 특허법 제25조 제1항 제2호의 특허권 부여를 배제하는 경우에 속하지 않는 경우, 특허법 제2조 제2항¹⁶¹⁾에서 서술된 기술방안에 속하는지 여부를 심사해야 한다.

159) 임춘광, 중국 특허법에서 인공지능 알고리즘의 특허적격성, 아주법학15(2), 2021, 495~496면.

160) 중국특허법 제25조 ①다음 항목에 대해서는 특허권을 부여하지 않는다.

(1) 과학적 발견

(2) 지적 활동의 규칙 및 방법

(3) 질병의 진단 및 치료방법

(4) 동물 및 식물 품종

(5) 원자핵 변환 방법 및 원자핵 변환 방법으로 얻은 물질

(6) 평면 인쇄물의 패턴, 색상 또는 이 둘의 조합에 대한 주요 표시 역할을 하는 디자인

②전항 제4호에 열거된 제품의 생산방법에 대해서는 이 법의 규정에 따라 특허권을 부여할 수 있다.

161) 중국특허법 제2조 ①이 법에서 말하는 발명 및 창작이란 발명, 실용신안 및 디자인을 말한다.

②발명은 제품, 방법 또는 그 개선에 대해 제안된 새로운 기술방안을 말한다.

③실용신안은 제품의 모양, 구조 또는 조합에 대해 제안된 실용에 적합한 새로운 기술방안을 말한다.

④디자인은 제품의 전체 또는 부분적인 모양, 패턴 또는 그 조합, 색상과 모양, 패턴의 조합으로 만들어진 미적 감각이 풍부하고 산업 응용에 적합한 새로운 디자인을 말한다.

알고리즘 특징 또는 영업규칙 및 방법 특징을 포함하는 청구범위가 기술방안에 속하는지 여부에 대해 심사를 진행할 때, 청구범위에 기재된 모든 특징을 전체적으로 고려해야 한다. 해당 청구범위가 해결하고자 한 기술적 과제에 대해 자연법칙을 이용한 기술수단을 사용했고, 이로 인해 자연법칙에 부합하는 기술효과를 취득하였음을 기재한 경우, 해당 청구범위가 한정하고 있는 해결방안은 특허법 제2조 제2항이 규정한 기술방안에 속한다. 예를 들면, 청구범위에 알고리즘의 각 절차가 해결하고자 하는 기술문제와 밀접한 관계를 구현하고, 알고리즘이 처리하는 데이터가 기술분야에서 확실한 기술적 함의를 가진 데이터이며, 알고리즘의 수행이 자연법칙을 이용하여 어떤 기술적 과제의 해결과정을 구현하여 기술효과를 취득했다는 내용이 포함될 경우, 일반적으로 해당 청구범위가 한정하고 있는 해결방안은 특허법 제2조 제2항이 규정한 기술방안에 속한다.

(3) 신규성 및 진보성 심사

알고리즘 특징 또는 영업규칙 및 방법 특징을 포함한 발명특허출원에 대해 신규성을 심사할 때, 청구범위에 기재된 모든 특징을 고려해야 한다. 이때 모든 특징은 기술특징, 알고리즘 특징 또는 영업규칙 및 방법 특징을 포함한다.

기술특징을 포함하는 동시에 알고리즘 특징 또는 영업규칙 및 방법 특징을 포함하는 발명특허출원에 대해 진보성 심사를 진행할 때, 기술특징과 기능적으로 서로 상호 지원하고 상호작용관계가 존재하는 알고리즘 특징 또는 영업규칙 및 방법 특징과 기술된 기술특징을 하나의 전체로서 고려해야 한다. ‘기능적으로 서로 상호 지원하고 상호 작용관계가 존재한다’는 의미는 알고리즘 특징 또는 영업규칙 및 방법 특징이 기술특징과 긴밀히 결합되어 어떤 기술문제를 해결하기 위한 기술수단을 공동으로 구성하고 이에 대응하는 기술효과를 취득할 수 있다는 뜻이다.

예를 들면, 청구범위 중의 알고리즘이 구체적인 기술분야에 응용되어 세부적인 기술문제를 해결할 수 있을 경우 해당 알고리즘 특징은 기술특징과 기능이 서로 상호 지원하고 상호작용관계가 존재한다고 인정되며, 이때 알고리즘 특징은 사용한 기술수단의 구성부분으로 볼 수 있으므로, 진보성 심사를 진행할 때 해결방안에 대한 기술된 알고리즘 특징의 기여도를 고려해야 한다.

또 예를 들면, 청구범위 중의 영업규칙 및 방법 특징의 실행에 있어 기술수단의 조정 혹은 개선이 요구되는 경우, 해당 영업규칙 및 방법 특징은 기술특징과 기능적으로 서로 상호 지원하고 상호작용관계가 존재한다고 인정할 수 있으므로, 진보성 심사를 진행할 때 해결방안에 대한 기술된 영업규칙 및 방법 특징의 기여도를 고려해야 한다.

라. 명세서 및 청구범위의 작성

(1) 명세서의 작성

알고리즘 특징 또는 영업규칙 및 방법 특징을 포함하는 발명특허출원의 명세서는 발명이 그

기술문제를 해결하기 위해 이용하는 해결방안을 명확하고 완전하게 설명해야 한다. 상기 해결방안은 기본적으로 기술특징을 포함하면서, 기술특징과 기능적으로 서로 상호 지원하고 상호작용관계가 존재하는 알고리즘 특징 혹은 영업규칙 및 방법 특징을 더 포함할 수 있다.

명세서에는 기술특징과 그와 기능적으로 서로 상호 지원하고 상호작용관계가 존재하는 알고리즘 특징 또는 영업규칙 및 방법 특징이 어떻게 공동 작용하여 유익한 효과를 생성시키는지를 명확히 설명해야 한다. 예를 들면, 알고리즘 특징을 포함할 때 추상적인 알고리즘과 구체적인 기술분야를 결합시켜야 하고 최소한 1개의 입력 매개변수 및 그와 관련된 출력결과의 정의가 기술분야 중의 구체적인 데이터와 대응되게 연관시켜야 하며 영업규칙 및 방법 특징을 포함할 경우에는 기술문제를 해결하는 전체 과정에 대한 상세한 기재와 설명이 있어야 하고, 해당 기술분야의 기술자가 명세서에 기재된 내용에 따라 해당 발명의 해결방안을 구현할 수 있어야 한다.

명세서는 발명이 종래기술과 비교했을 때 갖는 유익한 효과, 예를 들면 품질, 정밀도 혹은 효율성 향상, 시스템 내부 성능의 개선 등을 명확하고 객관적으로 명시해야 한다. 사용자의 입장에서 봤을 때, 객관적으로 사용자의 사용성을 향상시킨 경우에는 명세서 중에 기재할 수 있다. 이 경우, 상술한 사용자 사용성의 향상이 어떻게 발명의 기술특징 및 그와 기능적으로 서로 상호 지원하고 상호작용관계가 존재하는 알고리즘 또는 영업규칙 및 방법 특징으로 말미암아 발생되었는지를 함께 설명해야 한다.

(2) 청구범위의 작성

알고리즘 특징 또는 영업규칙 및 방법 특징을 포함한 발명특허출원의 청구범위는 명세서에 근거하여, 명확하고 간결하게 특허권리보호의 범위를 한정해야 한다. 청구범위에는 기술특징 및 기술특징과 기능적으로 서로 상호 지원하고 상호작용관계가 존재하는 알고리즘 특징 또는 영업규칙 및 방법 특징을 기재해야 한다.

마. 인공지능 알고리즘의 특허적격성 판단기준¹⁶²⁾

특허적격성의 문제로서 인공지능 알고리즘의 지적규칙 해당여부 관련, 청구범위가 추상적인 알고리즘에 불과하고, 어떠한 기술적 특징도 없다면, 이는 지적규칙에 해당하므로 특허권이 허여될 수 없다. 예컨대, 추상적 알고리즘 기반 어떠한 기술적 특징도 없는 수학적 모델의 구축방법이 여기에 해당된다. 그렇지만, 결합설에 따라 기술적 특징이 포함된 청구범위일 경우에는 지적규칙에 해당한다는 이유만으로 특허권의 허여 가능성을 배제해서는 안 된다고 한다. 다른 한편, 결합설에 근거하였음에도 청구범위가 전체적으로 지적규칙에 해당하여 특허권이 허여될 수 없다면, 기술적 방안의 해당여부를 고려해야 한다. 즉, 청구범위에 기재된 전부의 특징을 고려하는 것이다. 가령, 해당 청구범위에 기술적 문제를 해결하기 위하여 자연법칙을 이용한 기술적 수단이

162) 임춘광, 중국 특허법에서 인공지능 알고리즘의 특허적격성, 아주법학15(2), 2021, 495~496면.

기재되어 있고, 이에 기반 자연법칙에 부합하는 기술적 효과가 발생하였다면, 해당 청구범위에서 한정된 해결방안은 기술적 방안에 해당한다는 것이다. 예컨대, 청구범위에 기재된 알고리즘의 각 절차가 해결하고자 하는 기술적 문제와 ‘밀접한 관계(密切相関)’를 갖고, 알고리즘이 처리하는 데이터가 기술분야에서 명확한 기술적 함의를 갖는 데이터이며, 알고리즘의 수행이 자연법칙을 이용하여 특정 기술적 과제의 해결과정을 나타내고, 기술적 효과의 발생이 있었다면, 일반적으로 해당 청구범위가 한정된 해결방안은 특허법상 기술적 방안에 해당한다.

바. 인공지능 알고리즘 관련 발명에서 ‘밀접한 관계’의 기준 163)

특허심사지침은 청구범위에 기재된 알고리즘의 각 절차가 해결하고자 하는 기술적 문제와 ‘밀접한 관계’를 가질 것을 요구하고 있다. 특히 청구범위에 기술적 특징 기반 이와 상호지원 및 상호작용의 관계에 있는 알고리즘 특징을 포함할 것을 요구하고 있다. 그러나 ‘밀접한 관계’의 정도, ‘상호지원 및 상호작용의 관계’가 무엇인지에 대해서 해석하거나 구체적인 기준을 제시하지는 않았다. 가장 낮은 수준의 관계를 갖는 경우를 생각해 볼 수 있다. 예컨대, 기술적 방안 중 1개의 입력 매개변수와 1개의 관련 출력결과 정의가 기술분야의 구체적인 데이터와 상응된 경우라면 ‘밀접한 관계’를 이루고 있는 것인지가 의문시될 수 있다. 결합설에 따르면 이러한 경우의 기술적 방안도 특허의 대상으로서는 적격하다. ‘밀접한 관계’를 다음과 같이 해석한다. ①알고리즘이 처리하는 데이터는 반드시 기술적 함의를 갖는 것으로서 추상적인 데이터가 아닐 것, ②처리과정에서 해당 데이터에 대해 자연법칙에 부합하는 처리를 하였을 것, ③해당 알고리즘의 처리를 거친 후 출력 데이터도 물론 명확한 기술적 함의를 가질 것, ④알고리즘의 동작이 일정한 기술적 문제를 해결하고 기술적 효과를 발생하였을 것으로 해석할 수 있다.

제4절. 데이터 공개관련 주요 논의/사례/판례의 분석·검토

1. EPO심판원 심결

가. T 0161/18 사건

(1) 발명의 명칭 및 내용

대상 발명은 미국 특허 8,920,327(이하, ‘327특허라 함)이며, Method for Determining Cardiac Output를 발명의 명칭으로 하고 있다. 동 발명은 PCT/AT2006/000457으로 출원되었다. 문제가 되었던, 독립항 제1항과 제8항은 다음과 같이 기재되어 있다.

163) 임춘광, 중국 특허법에서 인공지능 알고리즘의 특허적격성, 아주법학15(2), 2021, 499~500면.

1. 주변부에서 측정된 동맥혈압곡선으로부터 심박출량을 결정하는 방법으로서,
 말초 영역에서 동맥 혈압 곡선을 측정하는 단계; 측정된 동맥 혈압 곡선을 등가 대동맥압으로
 산술 변환하는 단계; 그리고

등가 대동맥압으로부터 심박출량을 계산하고, 여기서 주변부에서 측정된 동맥혈압곡선을 등가
 대동맥압으로 산술변환하는 것은 학습에 의해 결정되는 가중치를 갖는 인공신경망의 도움으로
 수행된다.

8. 말초에서 측정된 동맥혈압곡선으로부터 심박출량을 결정하는 장치로서, 말초에서 혈압곡선
 을 획득하는 측정장치와 측정된 혈압곡선을 변환하는 연산부를 포함한다. 등가대동맥압과 등가대
 동맥압으로부터 심박출량을 산출하는 산술유닛을 포함하고, 측정된 혈압곡선을 변환하는 연산유
 닛은 인공신경망을 가지는 것을 특징으로 하는 학습에 의해 가중치가 결정된 사람.

‘327특허는 “동맥 혈압 곡선에서 심박출량을 측정하는 방법”에 관한 독립 청구항을 포함한
 다. 이 방법은 그림과 같이 심장 장치를 통해 구현된다. fig.1은 ‘327특허의 발명을 구현하기 위한
 장치(1)을 도시한 것으로, 여기서 측정장치(2)는 말초 혈압 곡선을 측정하고 관련 측정 데이터는
 라인(3)을 통해 장치(1)에 공급되고 거기에 저장되고 평가된다. 장치는 광학 디스플레이 장치(4),
 입력 패널(5) 및 정보를 입력하고 표시하기 위한 키(6)를 포함한다. 청구된 방법은 학습에 의해
 결정되는 가중치를 갖는 인공신경망의 사용을 포함한다.

[그림 3-4] 심박측정 장치

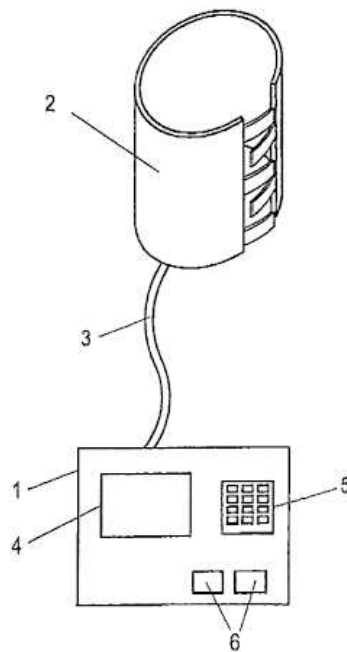


Fig. 1

'327 특허의 fig.3은 청구항 1에 인용된 인공신경망의 개략도이다. '327 특허의 명세서에서 도 3은 신경망의 구조를 도시한 것으로, 신경망이 14, 15, 16의 3개 층으로 구성되어 있다. 이 사양은 지도학습 알고리즘이 모델의 가중치를 학습하는 데 사용되었음을 보여준다.

[그림 3-5] 은닉층 구조도

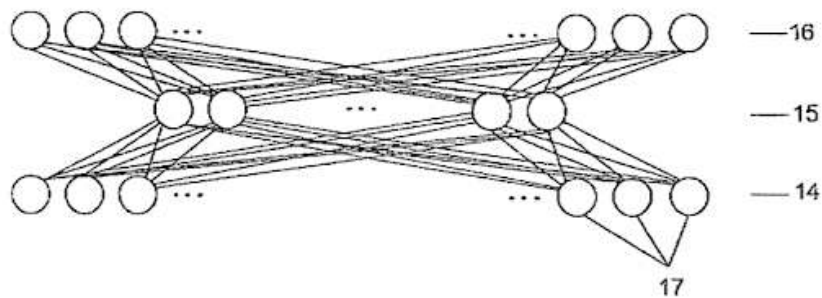


Fig. 3

(2) EPO 심판원(Boards of Appeal)의 심결¹⁶⁴⁾

1) 심결 내용

본 발명에 대하여 심판원은 다음과 같이 심결하였다.

1. 본 발명은 특히 인공신경망과 관련하여 기계학습을 기반으로 하고있으며, 충분한 기술이 공개되지 않은 본 발명에 따라 인공신경망의 학습을 수행할 수 없기 때문에 충분히 공개되지 않은 것이다.

2. 이 사건에서 청구된 발명은 인공신경망에 대한 학습이 구체적으로 공개되지 않은 것만이 종래 기술과 차이가 있으므로, 인공신경망의 사용이 정당화될 수 있는 진보적인 특별한 기술적 효과를 가져오지 않는다.

2) 공개의 충분성(EPC 제83조) 검토

EPC 제83조는 유럽 특허 출원이 평균적 기술자가 수행할 수 있도록 충분히 명확하고 완전한 방식으로 발명을 공개해야 한다고 요구한다. 이를 위해, 출원서의 발명의 개시는 평균적 기술자가 그의 일반적인 전문지식에 기초하여 청구된 발명에 내재된 기술적 사상을 재현할 수 있게 해야 한다. 본 출원은 주변부에서 측정된 혈압 곡선을 등가 대동맥압으로 변환하기 위해 인공신경망을 사용한다. 본 발명에 따른 신경망의 학습과 관련하여, 본 출원은 입력 데이터가 연령, 성별,

164) T 0161/18 (Äquivalenter Aortendruck/ARC SEIBERSDORF) of 12.5.2020.

체질 유형, 건강 상태 등이 다른 광범위한 환자를 포괄해야 함을 공개할 뿐이므로, 네트워크는 전문화되지 않고 있다. 그러나, 본 출원은 어떤 입력 데이터가 본 발명에 따른 인공신경망 학습에 적합한지 또는 당면한 기술적 문제를 해결하는 데 적합한 적어도 하나의 데이터셋을 공개하지 않았다. 따라서 인공신경망의 학습은 평균적 기술자에 의해 재현될 수 없으며, 평균적 기술자는 본 발명을 실시할 수 없다. 따라서, 특히 인공신경망과 관련된 기계학습에 기초한 본 발명은 충분히 공개되지 않았으며, 이는 본 발명에 따른 학습에 상응하는 기술공개의 부족으로 인해 실시될 수 없기 때문이다.

EPO 심판원은 본 출원 공개의 이러한 공백을 고려하여 평균적 기술자가 본 발명을 재현할 수 없을 것이라는 예비 의견으로 항소인에게 통지했다. 이에 청구인은 답변하지 않았다. 따라서 신청은 EPC 제83조의 요구 사항을 충족하지 않는다. 따라서, 본 출원에 대한 항소를 기각한다.

(3) 평가

심결의 요지는 명세서에는 본 발명에 따른 인공신경망 학습에 적합한 입력 데이터 또는 당면한 기술적 문제 해결에 적합한 데이터셋 중 적어도 하나를 공개하지 않았기 때문이라고 판단하였다. 본 발명은 PCT기반 출원건이지만, 미국에서는 충분한 공개에 대한 논란은 없었다. 이는 미국과 유럽의 기술공개에 대한 관행이 다르다는 점을 확인한 것이라고 하겠다.¹⁶⁵⁾

결국, 본 건에 대해서는 실질적으로 데이터 등 기술공개에 대한 명확한 기준을 제시하지 못한다는 점이 반영된 것으로 판단된다. 따라서, AI발명을 포함한 학습데이터와 관련된 가이드라인을 제시하는 것이 필요함을 알 수 있다. 향후, 최소한 IP5에서는 데이터에 관한 구체적인 사실을 파악하여 발명자에게 제시하는 것이 바람직하다고 생각된다.

나. T 1191/19 사건¹⁶⁶⁾

(1) 발명의 명칭

METHOD AND SYSTEM TO SAFELY GUIDE INTERVENTIONS IN PROCEDURES THE SUBSTRATE WHEREOF IS NEURONAL PLASTICITY(신경 가소성인 기질의 절차에 대한 개입을 안전하게 유도하기 위한 방법 및 시스템)

(2) 심결

165) Ryan N. Phelan, A Tale of Two Jurisdictions: Sufficiency of Disclosure for Artificial Intelligence (AI) Patents in the U.S. and the EPO, <https://www.patentnext.com/2021/11/a-tale-of-two-jurisdictions-sufficiency-of-disclosure-for-artificial-intelligence-patents-in-the-u-s-and-the-epo/> <2022.11.22. 접속>

166) ECLI:EP:BA:2022:T119119.20220401.

2. 명확성 (제84조 EPC조)

2.1 결정은 청구항 1167)의 일부 조항이 불명확하다고 하여 이의를 제기하였다. 심사부는 이러한 용어들이 실제로 관련 기술에서 의미를 확립하고 있으며, 이와 관련하여 AX1168)을 일반 상식의 증거로 받아들이지 않는다는 항소인의 주장에 납득하지 못했다. 항소인은 항소 이유를 명시한 진술서에서 AX1, 특히 fig.1(현재 출원의 fig.1과 거의 동일)과 섹션 2.2를 다시 언급하고, 본 발명이 AX1에 공개된 일반적인 메타 학습 체계를 사용했다고 제출했다.

[그림 3-5] 데이터 학습 구조도

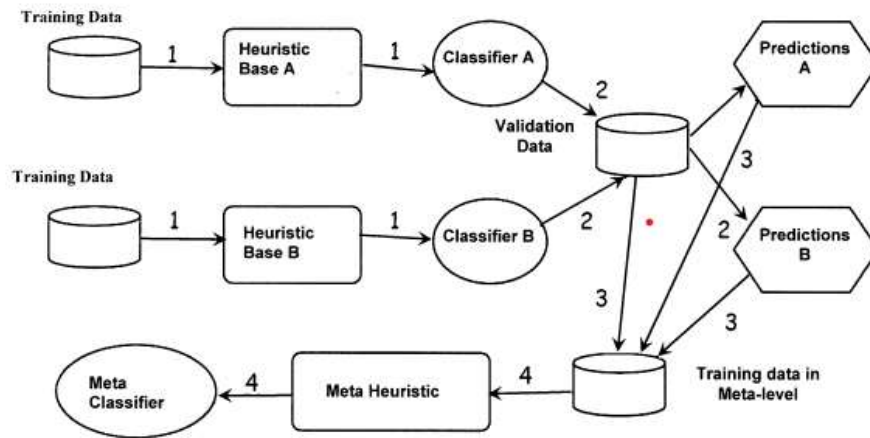


Fig. 1

2.2 청구항 1이 AX1과 동일한 용어를 사용한다는 점을 고려할 때, IASB는 진보성을 평가하기 위한 출발점으로 선호한다. 청구항 1이 EPC 제84조의 요건을 충족하는지 여부는 공개할 수 있다.

3. 진보성 (EPC 제56조)

3.1 이의가 제기된 결정에서 청구항 1은 문서 D1169)과 문서 D11170)의 조합에 관한 진보성을 포함하지 않는 것으로 밝혀졌다. 그러나, 본 발명이 AX1에 개시된 것과 동일한 일반적인 메타

167) 청구항 1항은 “다른 분류기와 다른 각각의 휴리스틱 또는 결정론적 규칙 세트” 및 “적어도 휴리스틱 또는 결정론적 규칙에 기초한 [메타 분류]”를 언급한다.

168) AX1: Prodromidis, et al. "분산 데이터 마이닝 시스템에서의 메타 학습: 문제 및 접근".

169) Gibert et al., "Response to traumatic brain injury neurorehabilitation through an artificial intelligence and statistics hybrid knowledge discovery from databases methodology", Medicinsky Arhiv, Sarajevo, vol. 63, no. 3, June 2008, 132-5.

170) WO 2007/079181.

학습 체계를 사용한다는 상소인의 제출을 고려하여, 위원회는 AX1이 진보성을 평가하는 데 더 적절한 출발점이라고 생각한다.

3.2 항소인은 항소 이유를 제시한 진술서에서 메타 학습을 뇌 가소성과 관련된 과정을 모델링 하고 안내하는 데 적용하는 것은 새롭고 창의적인 전략이라고 주장했다. 그러나 특정 분야의 문제에 알려진 기계 학습 기술을 적용하는 것은 기술의 일반적인 추세이며(이유 T 161/18, 3.6 참조) 그렇게 창의적일 수 없다. 따라서, IASB는 항소인에게 청구항 1의 방법이 당면한 특정 문제에 AX1의 메타 학습 체계를 적용했는지, 즉 기질이 신경 가소성인 과정에서 환자에 대한 개인화된 개입을 예측하는지 여부를 평가해야 한다고 예비 의견에서 통지했다. 위원회는 청구항 1의 방법에서 AX1에 공개된 체계의 추상적 수준에서의 단순한 반복을 넘어 당면한 문제에 대한 AX1의 메타 학습 체계의 적용에 대한 명확하지 않은 세부 사항을 볼 수 없었다.

3.3 항소인이 이사회에 진보적인 반대의견에 응답하지 않았기 때문에, 이사회는 예비의견을 변경할 이유가 없다고 보았다. 따라서, 청구항 1의 주제는 어떠한 진보성도 포함하지 않았다(EPC 제56조).

4. 공개의 충분성(EPC제83조)

4.1 출원은 AX1의 메타 학습 체계가 당업자가 수행할 수 있을 정도로 충분히 명확하고 완전한 방식으로 당면한 문제에 어떻게 적용되었는지를 공개하지 않았다. 그림 1의 용어와 설명을 사용하여, 애플리케이션은 메타 학습 체계가 다음과 같이 요구하는 학습데이터(“청구항 1에서 수행될 개입과 관련하여 적어도 복수의 사용자에게 관한 정보를 가진 데이터베이스) 및 검증데이터(청구항 1에서 “검증 데이터”도 포함)의 어떠한 샘플 세트도 공개하지 않았다. 출원은 의미 있는 예측을 제공하기 위해 학습데이터를 컴파일해야 하는 최소 환자 수와 관련 매개변수 세트를 공개하지 않았다.

분류기 A 및 B를 트레이닝하기 위한 휴리스틱 베이스 A 및 B(청구항 1의 단계 a1에서 “다른 분류기의 규칙과 다른 각각의 휴리스틱 또는 결정론적 규칙 세트”) 및 메타 분류기(청구항 1의 단계 c에 있어서, (적어도 휴리스틱 또는 결정론적 규칙에 기초한) (최소 휴리스틱 또는 결정론적 규칙에 기초한 분류)를 트레이닝하기 위한 메타 휴리스틱이 당면한 문제는 마찬가지로 공개되지 않으며 분류기, 토폴로지, 활성화 기능, 최종 조건 또는 학습 메커니즘으로 사용되는 인공신경망의 구조도 공개되지 않았다(T161/18, 포인트 2 참조). 애플리케이션의 추상화 수준에서, 이용 가능한 공개는 연구 프로그램에 대한 초대에 가깝다.

4.2 이러한 상황에서 숙련된 사람은 신경 가소성을 기반으로 하는 프로세스에서 환자의 개인화된 개입을 예측하는 문제를 해결하기 위해 AX1의 메타 학습 체계의 적용을 과도한 부담 없이 재현할 수 없다.

4.3 IASB는 예비의견에서 이러한 반대의견을 제기하였다. 항소인이 답변서에서 이들을 언급하

지 않았기 때문에, 이사회는 예비 의견을 변경할 이유가 없다고 보았다. 그러므로, 그 신청서는 제83조 EPC의 요건을 충족하지 않았다.

(3) 평가

T 1191/19사건의 대상은 특허출원은 신경 가소성이 기질인 프로세스에서 결정된 사용자에게 대한 개인화된 개입에 대한 예측을 최적화하기 위한 컴퓨터 구현 방법에 관한 것이다. 이 방법발명은 메타 학습 체계를 포함하고 있으며, 이 체계는 학습데이터와 검증 데이터를 입력으로 요구한다. 본 사건은 기술공개라는 측면에서 EPC 제83조에 따른 공개의 충분성이 결여된 AI 관련 특허출원에 관한 것이다.

심판원은 숙련된 사람은 학습데이터, 유효성 검사 데이터 및 최소 환자 수를 공개하지 않고 메타 학습 체계를 포함하는 청구된 방법을 과도한 부담 없이 재현할 수 없으며, 문제 해결을 위한 분류기 A와 B를 학습하기 위한 휴리스틱 기반 A와 B, 분류기로 사용되는 인공 신경망의 구조, 토폴로지, 활성화 기능, 조건 또는 학습 메커니즘 등의 관련 매개 변수 세트 등의 학습데이터는 의미 있는 예측을 제공하기 위해 컴파일되어야 한다고 보았다. 이에 따라, 심판원은 학습데이터 및 검증데이터의 어떠한 샘플 세트도 공개하지 않았다는 점과, 출원은 의미 있는 예측을 제공하기 위해 학습데이터를 컴파일해야 하는 최소 환자 수와 관련 매개변수 세트를 공개하지 않았다는 점에서 충분한 기술공개가 이루어지지 않았다고 판단하였다. 더욱이 인공신경망의 구조도 공개되지 않았다는 점에서 T161/18 사건을 인용하고 있다.

2. 영국 IPO의 데이터 공개에 대한 논의

가. 의의

2021년 3월 23일, 영국 지식재산청(UKIPO)은 ‘인공지능과 지식재산에 관한 대중의견 수렴(Artificial intelligence and Intellectual property: Call for views)’ 결과를 정리하여 공개하였다. 배경은 UKIPO는 2020년 9월 7일부터 2020년 11월 30일까지 인공지능(AI)과 지식재산권(IP)에 관련된 다양한 이슈를 제기하고, 영국의 지식재산 정책 수립을 위해 인공지능의 영향력을 이해하고 중기적 관점에서 인공지능 시대에 IP 분야가 마주할 문제에 대해 모든 이해관계자의 아이디어, 전문성, 통찰력을 수렴하고자 한 것이다. 이러한 의견수렴을 바탕으로, 영국 특허청(IPO)은 인공지능 발명관련 특허출원 심사 안내서¹⁷¹⁾를 통하여 AI발명의 기술공개에 대해서 언급한 바 있다. 심사안내서 12번째 질문 및 그에 대한 답은 아래와 같다.

171)

<https://www.gov.uk/government/consultations/artificial-intelligence-and-intellectual-property-call-for-views/government-response-to-call-for-views-on-artificial-intelligence-and-intellectual-property> <2022.11.30. 접속>

나. 질의 및 응답

(질문 12) 미래에 법이 특허법의 현재 목적을 넘어서는 사회적 이유로 인공지능 발명의 충분한 세부사항을 제공할 이유가 있을 수 있을까요?

(답변) 모든 응답자들은 공개가 특허제도의 기본적인 부분이라는 것에 동의한다. 공개를 위해 테스트를 변경하라는 요청이 없었다. 응답자들은 일반적으로 AI 특허 출원이 숙련된 사람이 발명을 수행하기에 충분한 세부 사항을 포함할 수 있다는 데 동의한다. 그들은 공개는 특허 출원 시 시스템의 스냅샷을 찍는 반면, 그것이 AI의 본질이지만 가능하다고 설명했다. 응답자들은 이것이 법원과 영국 IPO에 도전을 제기할 수 있지만 현재의 영국 특허시스템은 이것을 다룰 수 있다고 말했다. 응답들은 또한 본 발명이 청구범위 전체에 걸쳐 지지될 수 있다는 것을 피드백했다. 예를 들어, 알고리즘이나 학습데이터를 공개해야 하는 요구 사항이 일부 AI 혁신자들의 특허시스템 사용을 저해할 수 있다는 의견이 많았다. 다른 이들은 기밀성을 통한 정보보호와 특허에 의한 정보 보호 사이의 선택은 상업적인 결정이라고 지적한다. 이것은 컴퓨터 구현 발명 분야에서 일반적인 결정이다. 한 응답자는 데이터 보호의 새로운 형태를 제안했다. 그들은 이것이 창조자에게 내재된 가치를 보존하면서 데이터 공개와 공유를 용이하게 할 것이라고 주장했다. 인공지능 특허 출원에서 공개해야 할 정보의 종류와 범위를 알기가 더 어려울 수 있다는 보고도 있다. 다른 사람들은 이것이 다른 복잡한 기술의 상황이기도 하며 AI에만 국한된 것이 아니라고 말한다.

이에 대해 AI 특허 사양의 일부로 대량의 정보를 제출하는 것의 실용성에 대해 우려를 표시했다. 데이터는 학습 세트 또는 기계 연결 가중치의 사양을 포함할 수 있다. 일부에서는 이러한 정보에 대한 기탁(deposit)제도를 만드는 데 효익이 있었다고 제안하였다. 대안으로 영국 IPO가 데이터에 적합한 형식을 개발할 수 있다고 제안했다. 다른 이들은 이 책의 정보가 AI 특허 사양의 공통적인 특징이 될 필요가 없다며 동의하지 않았다.

대부분의 의견제출자들은 특허법 이외의 사회적 감독을 위해 AI 세부사항을 공개하기 위한 향후 요구사항을 유지하는 것이 중요하다고 하였다. 이처럼, 추가 의무공개가 특허제도의 활용을 저해할 수 있다는 우려가 나왔다. 또 다른 사람은 특허 공개를 충족하는 데 필요한 데이터가 어떤 AI 규제 기관이 필요로 하는 정보와 다를 가능성이 높다고 말했다. 일부 의견제출자들은 규제 당국의 접근을 허용하면서도 데이터의 기밀성을 유지하는 새로운 형태의 데이터 보호가 필요할 수 있다고 제안했다.

3. WIPO 차원의 논의

가. WIPO CONVERSATION ON IP AND AI⁷²⁾

(1) 개요

AI는 일과 생활방식을 근본적으로 바꾸어 놓아 세계 공통의 도전 과제를 안기고 있고 관련 정책에 대해서도 의문과 도전을 불러일으키고 있다는 점과, 또한 전자적 데이터는 AI의 머신러닝 뿐만 아니라 디지털 경제에서 지식재산권과 혁신의 핵심 자원이기 때문에 이와 관련한 문제에 대하여 대중의 의견을 수렴하였다. 이를 위하여 WIPO는 2019년 AP와 AI관련 이슈를 정리하는 보고서를 준비하였고, 이를 바탕으로 설문을 진행하였다.¹⁷²⁾ WIPO의 AI에 대한 의견모집은 AI가 지식재산에 어떤 영향을 미치는지에 관한 것으로 주요 질문 내용과 이슈는 다음과 같다.¹⁷⁴⁾

(2) 질문 내용 및 이슈

AI에 관한 이슈가 어떤 영역에서 발생하는지에 대해서는 특허, 저작권, 데이터, 디자인, 기술격차 및 역량 구축, 지식재산 관리 결정에 대한 책임 등을 설문 문항으로 제시하고 있다. 또한 AI에 의해 자동적으로 생성된 발명에 관하여 그 발명자를 누구로 할 것인가에 대한 이슈를 포함하고 있으며, 컴퓨터 기반 발명의 특허 대상 문제 및 특허성에 관한 가이드라인, 비자명성 관련 이슈에 대해 묻고 있다. 이외에도 특허의 공개와 관련하여, AI가 생성한 발명에 공개의 요건을 부과해야 하는지 등의 문제를 다루고 있다. 저작권과 관련 권리에 대해서는 침해 및 예외, 딥 페이크 기술(deep fakes technology), 일반적 정책 이슈 분야의 질문을 하고 있다. 데이터에 대하여는 지식재산 정책이 데이터와 관련한 새로운 권리의 창작을 고려해야만 하는지, 만약 고려해야만 한다면 어떤 형태의 데이터를 보호대상으로 감안해야 하는지 등을 묻고 있다.

(3) 기술공개 관련 이슈

WIPO에서는 2020년 인공지능과 지식재산에 있어서 이슈를 정리하고 이에 대해 의견수렴을 진행한 바 있다. 250건의 의견이 접수되었으며, 이 중 기술공개(disclosure)에 부문에 관하여 AI관련 사업을 영위하는 제안자를 대상으로 하는 유의미한 의견에 대하여 아래와 같이 선별하여 정리하였다.

(Issue 5: Disclosure) 21. 특허 제도의 기본 목표는 기술을 공개하여 시간이 지남에 따라 공개 영역이 풍부해지고 인류 기술의 체계적인 기록이 이용 가능하고 접근 가능하도록 하는 것이다. 특허법은 발명의 공개가 관련 기술 분야의 숙련된 사람이 발명을 복제할 수 있을 만큼 충분한 것을 요구한다.

172) WIPO, REVISED ISSUES PAPER ON INTELLECTUAL PROPERTY POLICY AND ARTIFICIAL INTELLIGENCE, WIPO/IP/AI/2/GE/20/1 REV, 2020.

173) 2019년 초안을 마련하여 의견요청을 하였고, 그 요청에 대한 회신을 정리하여 2020년 이슈페이퍼로 정리한 것이다.

174) https://www.wipo.int/edocs/mdocs/mdocs/en/wipo_ip_ai_ge_20/wipo_ip_ai_2_ge_20_1.pdf

- (i) 공개에 관한 현재 규칙은 AI 생성 및 AI 지원 발명에 어떻게 적용되며 기본 정책 근거를 충족하기에 충분합니까?
- (ii) AI 지원 또는 AI 생성 발명이 공개요건에 대해 제시하는 문제는 무엇입니까?
- (iii) 입력 데이터에 따라 결과가 변경될 수 있고 알고리즘이 뉴런 연결과 관련된 가중치를 조정하여 실제 결과와 예측 결과의 차이를 조정하는 기계학습의 경우 초기 알고리즘을 공개하는 것으로 충분합니까?
- (iv) 미생물의 기탁과 유사한 AI 응용 프로그램이나 학습데이터의 기탁 시스템이 유용할까요?
- (v) 알고리즘 학습에 사용되는 데이터는 공개 목적으로 어떻게 취급되어야 합니까? 알고리즘 학습에 사용되는 데이터를 특허 출원서에 공개하거나 설명해야 합니까?
- (vi) 데이터를 선택하고 알고리즘을 학습하는 데 사용되는 인간의 전문성을 공개해야 합니까?

나. 의견¹⁷⁵⁾

(1) 인텔¹⁷⁶⁾

학습데이터에 대한 지식 없이는 AI 발명을 수행, 사용 또는 최상의 실시로 실행할 수 없는 범위 내에서 인텔은 그러한 데이터가 US § 112(a)의 요구사항을 충족할 수 있도록 상세하게 설명되어야 한다고 제안한다.

(2) American Bar Association¹⁷⁷⁾

- (i) 인공지능 지원 또는 인공지능 생성 발명이 공개요건에 대해 제시하는 문제는 무엇입니까?

(답변) 이 섹션은 블랙박스 설명과 기술 분야에서 일반적인 기술을 가진 사람이 보유한 지식에 대한 허용 가능한 가정을 균형 있게 고려하여 AI 발명에 대한 서면 설명(written description)이 블랙박스로 쉽게 전환될 수 있다고 우려한다. 이 문제는 기존의 다양한 소프트웨어 발명에서 정기적으로 다루어지기 때문에 AI 발명에만 국한된 것은 아니다. 그리고 기존의 소프트웨어 발명에서와 마찬가지로, 잘 알려진 공정을 상세히 설명할 필요가 없는 반면, 보상되어야 할 발명(즉, 청구된 특징)은 소유를 보여주고 당업자에 의한 통상적인 기술이 가능하도록 그러한 상세하게 설

175) 제출된 의견은 아래 사이트에서 확인할 수 있으며, 다양한 분야에서 250건 이상이 제출되었다.
https://www.wipo.int/about-ip/en/artificial_intelligence/submissions-search.jsp?type_id=&territory_id=&issue_id=2421 <2022.12.12. 접속>

176)
https://www.wipo.int/export/sites/www/about-ip/en/artificial_intelligence/call_for_comments/pdf/org_intel.pdf <2022.12.12. 접속>

177)
https://www.wipo.int/export/sites/www/about-ip/en/artificial_intelligence/call_for_comments/pdf/org_american_bar_association.pdf <2022.12.12. 접속>

명되어야 한다. 비표준 AI 애플리케이션이 구성요소로 주장되는 경우, 표준 AI 애플리케이션만 식별되는 경우보다 더 많은 공개가 필요하다. 예를 들어, 일반적인 딥러닝 신경망의 경우, 서면 설명은 불필요하고 기능적으로 참조될 수 있기 때문에 사용된 숨겨진 레이어 또는 뉴런의 수를 공개할 필요가 없다. 반대로, 컨볼루션 신경망과 같은 일부 유형의 신경망의 경우, 특정 인식 문제에 대한 은닉 계층 설계는 상세하게 설명되어야 한다. 인공지능 애플리케이션 자체가 발명의 구성요소가 아니지만 인공지능 애플리케이션의 기여가 비표준인 경우, 청구된 발명이 달리 활성화되지 않은 경우 활성화를 위해 더 많은 공개가 필요할 수 있다.

(ii) 시간이 지남에 따라 데이터에 대한 접근으로 알고리즘이 변경되는 기계학습의 경우, 초기 알고리즘의 공개는 충분한가?

(답변) 청구항에 기재된 발명이 그것이 접근하는 데이터에 기초하여 실질적으로 변경되는 알고리즘을 포함하는 경우, 그러한 데이터에 기초하여 알고리즘의 진화를 설명하고 가능하게 하기 위해, 접근이 제공되는 데이터의 유형에 대응하여 알고리즘이 어떻게 변경될 것으로 예상되는지에 대한 일부 공개가 필요할 수 있다. 그러나 변경된 알고리즘이 예측 가능한 개선이라면 초기 알고리즘과 접근되는 데이터의 유형의 공개가 적절할 수 있다.

(iii) 미생물의 기탁과 유사한 알고리즘을 위한 기탁 시스템이 유용할까요?

(답변) 알고리즘이 특정 학습데이터셋과 함께 개발되거나 진화하고 청구를 충족할 수 있는 적절한 능력이 없는 서면 설명에서 “블랙박스” 알고리즘을 포함하는 청구의 경우, 출원인에게 연구를 촉진하기 위해 알고리즘의 사본을 보관할 수 있는 선택적 옵션을 제공하는 것이 바람직할 수 있다. 기탁된 데이터베이스를 보호하는 데 사이버 보안 문제뿐만 아니라 의미 있는 참여가 있을 것인지에 대한 의문이 있다.

(iv) 알고리즘을 학습하는 데 사용되는 데이터는 공개 목적을 위해 어떻게 처리되어야 하는가?
알고리즘을 학습하는 데 사용되는 데이터는 특허 출원에 공개되어야 하는가 아니면 기술되어야 하는가?

(답변) 공개의 표준은 다양해서는 안 된다. 소유와 활성화를 보여주는 적절한 설명이나 표준 AI를 포함한 표준 도구와 학습데이터에 대한 적절한 설명으로 주장되는 기술 분야의 일반 기술자를 가능하게 하는 데 필요한 모든 것은 기존 특허 표준을 충족해야 한다. 청구된 발명이 좁게 지정된(또는 정밀한) 학습데이터를 켜는 경우, 해당 데이터와 학습된 알고리즘과의 상호작용이 공

개되어야 한다.

(v) 데이터를 선택하고 알고리즘을 학습하는 데 사용되는 인적 전문지식이 공개되어야 하는가?

(답변) 이 단계에서 이 문제를 완전히 고려하지는 않았지만, 청구항이 “정신적 단계“에 대한 자격을 상실하지 않는 한, 그러한 정신적 단계는 발명의 소유를 보여주고 청구된 발명을 가능하게 하기에 충분히 상세하게 설명되어야 한다고 간주할 수 있다.

(3) 국제 지적재산권 보호협회(AIPPI)¹⁷⁸⁾

International Association for Protection of Intellectual Property(AIPPI)는 추가적인 설명을 할 가치가 있다고 믿는다. 이슈페이퍼 초안의 10번 문항은 AI 시스템으로 구성되거나 AI 시스템에 지시된 발명에만 관련되고 AI 지원 또는 AI가 생성한 발명과는 관련이 없는 고려사항을 제시하는 것으로 보인다. 후자의 발명 범주의 경우, 가능한 공시의 문제는 발명이 어떻게 또는 어떤 방법론에 의해 발생하는지에 대한 문제와 완전히 관련이 없다. 특히 명세서에 충분한 정보가 제공되어 특허가 만료되면 당업자가 해당 정보와 일반적인 지식에 기초하여 청구된 장치를 제조하거나 청구된 과정을 수행할 수 있음을 의미한다. 따라서 이러한 맥락에서 이를 만드는 데 사용된 AI 시스템과 독립적인 독립형 발명과 AI 시스템이 분리할 수 없는 부분을 형성하는 발명을 구별하는 것이 중요할 것이다. 발명이 인공지능 시스템의 혁신이나 개선 또는 그 사용에 관한 것이라면, 발명의 공개와 관련하여 제기된 질문들(즉, 질문 10(i)부터 질문 10(iv)은 이러한 맥락에서 관련이 있는 것으로 보인다.

그러나 질문 10(v)는 발명이 인간의 전문지식이 아닌 기술적 장치나 방법에만 있을 수 있기 때문에 발명의 공개와 관련하여 관련성이 적은 것으로 보인다. 인공지능 지원 또는 인공지능 생성 발명과 관련하여, 공개 요구사항은 발명 자체를 포함하며, 발명을 작성하기 위해 사용된 인공지능 시스템 또는 방법은 포함하지 않는다. 예를 들어, 독립형 발명품은 자동차 또는 항공기의 공기역학적 윤곽이나 모양에서 찾을 수 있으며, 여기서 문제의 윤곽이나 모양은 AI 지원 설계 프로세스의 결과이다. 이러한 상황에서 가능한 공개는 인공지능 시스템의 세부사항을 포함할 필요가 없다. 왜냐하면 발명의 윤곽이나 모양을 특징으로 하는 물품의 제조는 그것을 처음 만드는 데 사용된 기본 인공지능 시스템에 의존할 필요가 없기 때문이다. 인공지능의 도움을 받아 만들어진 전술한 예에 따른 독립형 발명은 인공지능 기술에 대한 가능한 공개를 요구하지 않으며, 이러한 맥락에서 인공지능은 발명품이 어떻게 생산되거나 제조되어야 하는지 숙련된 사람에게 가르쳐주지 않는다. 보다 구체적으로, 독립형 발명이 구상된 알고리즘 및/또는 데이터를 공개하는 것은

178)

https://www.wipo.int/export/sites/www/about-ip/en/artificial_intelligence/call_for_comments/pdf/org_aippi.pdf <2022.12.12. 접속>

그러한 경우에 특허법의 의미에서 가능한 공개의 요건을 충족하지 못한다.

(4) Guido Noto La Diega 교수¹⁷⁹⁾

데이터에 접근하여 시간이 지남에 따라 알고리즘이 바뀌는 머신러닝의 경우, 초기 알고리즘을 공개하는 것으로 충분합니까?

(답변) 특허 갱신의 조건은 갱신 요청일 현재의 기계학습 알고리즘을 갱신 공개하는 것이다. 그렇지 않으면 충족 요건 위반이다. 사실, 초기 알고리즘으로 제한되는 공개는 Novartis v Johnson & Johnson¹⁸⁰⁾에서 설명된 바와 같이 당업자가 발명을 수행할 수 있을 만큼 충분히 명확하고 완전한 방식으로 발명을 공개하지 않을 것이다. 미생물 기탁과 유사한 알고리즘 기탁시스템이 유용할 것이다. 또한, 알고리즘 학습에 사용된 데이터는 공개에 포함되어야 한다.

(5) 막스프랑크 연구소¹⁸¹⁾

12. 질문 10(i)은 공개요구사항에 대해 AI 지원 또는 AI 생성 발명이 존재하는 이슈가 무엇인지 묻는다. 원칙적으로 특허는 숙련된 사람이 재현할 수 있도록 기술적 해결책을 충분히 공개하는 것을 의미하는 반면, 발명을 개발하는 과정에서 도움이 될 수 있는 방법이나 도구의 공개는 관련이 없기 때문에 인공지능 지원 발명에 대한 이 질문의 관련성은 간단해 보이지 않는다. 그러나 공개 요건은 AI 기반 시스템의 기능을 통해 발명의 기술적 해결책이 가능한 상황, 즉 청구된 주제 내에 AI가 구성되는 상황, 특히 이러한 AI 도구가 보호 주제를 형성하는 경우에 문제가 있다. 이와 관련하여 먼저 현행 공개규정이 AI를 주장하는 발명에 어떻게 적용되는지, 그리고 이들이 기초적인 정책적 근거를 확보하기에 충분한지에 대한 의문을 제기하는 것이 적절할 것이다.

13. 공개요구사항을 특허성에서 제외하기 위하여(또는 달리 취급하기 위하여) AI가 자율적으로 생성한 발명을 식별하기 위한 수단으로 본다면, 이 요구사항이 그 목적에 적합한 수단인지는 의문이다. 먼저 특허 시스템이 발명이 어떻게 생겨났는지에 대한 공개를 요구한 적이 없는 이유를 고려할 필요가 있다.

(6) 컴퓨터통신산업협회(CCIA)¹⁸²⁾

179)

https://www.wipo.int/export/sites/www/about-ip/en/artificial_intelligence/call_for_comments/pdf/ind_noto_la_diega.pdf <2022.12.12. 접속>

180) Novartis Consumer Health, Inc. v. Johnson & Johnson-Merck Consumer Pharmaceuticals Co., 129 F. Supp. 2d 351 (D.N.J. 2000).

181)

https://www.wipo.int/export/sites/www/about-ip/en/artificial_intelligence/call_for_comments/pdf/org_max_planck_institute_for_innovation_and_competition.pdf <2022.12.12. 접속>

(i) 인공지능 지원 또는 인공지능 생성 발명이 공개요건에 대해 제시하는 문제는 무엇입니까?

(답변) AI 발명에 대한 충분한 공개는 AI가 어떻게 발명에 통합되는지와 해당 AI의 종류에 달려 있다. 현재 AI의 대부분은 기계학습 기술에 기반을 두고 있으며, 이러한 의견은 기계학습과 가장 관련이 있는 공개에 초점을 맞추고 있다. 그러나 다른 형태의 인공지능이 존재하고 개발될 가능성이 높으며, 다른 인공지능 아키텍처에도 뚜렷한 공개 고려사항이 적용될 수 있다. 자율적 인공지능 생성 발명은 발명의 주제에 대한 전형적인 것을 넘어 뚜렷한 공개 고려사항을 만들지 않을 것이다. 예를 들어, 인공지능에 의해 식별된 새로운 화합물은 화학 화합물에 대한 일반적인 서면 설명 제약 조건을 준수해야 한다.

기계학습을 위한 새로운 알고리즘과 같이 AI 발명(AI 개선 발명)의 기능 향상과 관련된 발명은 일반적으로 컴퓨터가 구현한 발명과 동일한 공개 우려를 내포할 것이다. 서면 설명은 AI의 원하는 측면뿐만 아니라 알고리즘을 공개하거나 구조를 구현하여 AI가 어떻게 달성되는지, 그리고 이를 활용하는 방법을 설명해야 한다. 예를 들어, 새로운 딥러닝 구조에 대한 특허 출원은 계층의 수, 계층 당 단위의 수, 단위 또는 계층 간의 데이터 흐름 및 활성화 기능을 설명할 수 있다. 새로운 학습 알고리즘에 대한 특허 출원은 학습될 구조, 학습에 사용되는 시작점과 데이터 및 발명을 재현하는 데 필요한 기타 관련 정보를 설명할 것이다.

AI 머신 비전 또는 AI 약물 발견과 같이 AI 영역 밖의 문제에 AI를 적용하는 발명(AI 응용 발명)은 공개와 관련하여 추가적인 고유 고려사항을 제시한다. 대부분의 인공지능 애플리케이션 발명에 적용되는 인공지능 발명과 관련하여 위에서 설명한 컴퓨터 구현 우려를 넘어, 인공지능 애플리케이션 발명은 특히 기능적인 용어로 설명될 가능성이 높으며 무효에 대한 우려를 피하기 위해 적절한 인공지능 구조의 공개를 요구할 것이다.

명세서의 그러한 구조적 설명은 발명자가 발명을 복제하기 위한 충분한 정보에 대한 접근을 제공하도록 요구함으로써 독자에 의한 복제성을 보장할 필요가 있을 것이다. 그것은 학습데이터의 유형과 학습 방법론을 설명하는 형태일 수도 있고, 학습 후 숨겨진 계층 가중치와 같은 학습된 모델 자체의 형태일 수도 있다.

최신 AI 기술에 의존하는 애플리케이션은 구조가 특정 문제를 해결하기 위해 어떻게 조정되는지에서 오는 참신함으로 잘 알려진 구조(예: 딥 러닝 네트워크)를 구현할 수 있다. 필요한 공개는 사용된 학습데이터의 유형, 사용된 데이터의 전처리 또는 문제를 해결하기 위해 출력을 분석하고 적용하는 방법에 대한 설명일 수 있다.

두 가지 상황에서 AI 시스템의 특성은 공개 및 활성화에 대한 추가적인 우려를 제기할 수 있

182)

https://www.wipo.int/export/sites/www/about-ip/en/artificial_intelligence/call_for_comments/pdf/org_ccia.pdf <2022.12.12. 접속>

다. 특히, AI 시스템은 예측할 수 없을 수 있다. 학습데이터 또는 알고리즘의 결보기에 사소한 변경은 매우 다양한 출력 모델을 생성할 수 있다. 이를 고려할 때, 특히 AI 애플리케이션 발명의 경우, 명세서가 청구의 전체 범위를 가능하게 하는 것이 중요하다. 이는 단일 종의 공개가 일반적인 주장을 뒷받침하기에 충분하지 않다는 것을 의미한다. 예를 들어, 개의 이미지를 인식하는 딥 러닝 모델의 공개는 m을 제공하기 위한 명세서의 추가적인 지원 없이는 “딥 러닝 모델을 사용하여 유형별 이미지를 인식” 한다는 주장을 뒷받침하기에 충분하지 않다. AI로 기능을 한다는 주장은 단순히 AI의 특정 모델이 아니라 어떤 형태의 AI에 대해서도 기능을 하는 방법을 가르쳐야 가능하다고 봐야 한다. 그러한 주장을 다른 방식으로 처리하는 것은 특허가 새로운 AI 아키텍처를 사용하여 주장된 기능을 수행할 수 있도록 예측 가능하게 할 수 있다는 것을 의미한다.

(ii) 머신러닝의 경우, 데이터에 접근하여 시간이 지남에 따라 알고리즘이 변경되는데, 초기 알고리즘을 공개하는 것으로 충분합니까?

(답변) 위에서 설명한 것처럼 알고리즘의 공개만으로는 충분하지 않은 경우가 많다. 모델만으로는 발명을 구현하지 않는다. 다른 사람이 특허를 복제할 수 있도록 하려면 학습데이터 및 방법론 또는 학습된 모델 자체에 대한 액세스가 필요하다.

(iii) 미생물 기탁과 유사한 알고리즘 기탁 시스템이 유용할까요?

(답변) 기탁 시스템이 유용한 이유는 확실하지 않습니다. 기존 AI 도구를 활성화하고 설명하는 데 유용한 알고리즘, 교육 방법론 및 모델 세부 정보는 일반적으로 텍스트 또는 디지털 설명이 가능하며 특허출원의 일부로 제출해야 한다. 특허 문서에 포함하기에 적절한 것보다 더 방대한 발명을 가능하게 하고 공개하는 데 필요한 추가 데이터는 컴퓨터 프로그램 목록에 대해 허용된 부록과 유사하게 응용 프로그램의 부록으로 제출될 수 있다. AI 아키텍처는 기탁 시스템에 적합하며, 그러한 시스템은 당시에 생성될 수 있다.

(iv) 알고리즘 학습에 사용되는 데이터는 공개 목적으로 어떻게 취급되어야 합니까? 알고리즘 학습에 사용되는 데이터를 특허 출원서에 공개하거나 설명해야 합니까?

(답변) 숙련된 사람이 과도한 실험 없이 발명을 복제할 수 있도록 충분한 정보가 특허 출원의 일부로 제공되어야 한다. 이것은 다양한 방법으로 달성될 수 있다. AI 개선 또는 AI 응용 발명에 대해 서면 설명이 제공되는 정확한 방법은 발명가가 청구된 AI를 복제할 수 있도록 충분한 설명을 제공하는 한 발명가의 재량에 맡길 수 있다. 이것은 학습된 모델 자체와 이를 적용할 알고리

즘에 대한 설명, 적절한 학습데이터 및 학습 방법론에 대한 설명, 실제 학습데이터 및 학습 방법론의 제공 또는 기타 데이터 조합의 형식일 수 있다. 제공된 정보의 전체가 복제를 가능하게 하기에 충분한 경우에 한한다.

복제 불가능성은 불충분한 공개의 명백한 증거이며, AI 개선 또는 AI 응용 발명의 복제 가능성은 본질적으로 모델을 생성하기에 충분한 학습데이터 또는 학습된 모델 자체의 공개를 요구한다. 그러한 데이터가 복제에 불필요하다는 발명가의 주장은 승인되기 전에 추가 증거로 뒷받침되어야 한다. 반대로, 학습된 모델의 제공은 기본적으로 항상 복제를 허용하여 활성화 및 설명을 위한 안전한 피난처(safe harbor) 역할을 한다. 따라서, 필수는 아니지만 충분하다.

(v) 데이터를 선택하고 알고리즘을 학습하는 데 사용되는 인간의 전문성을 공개해야 합니까?

(답변) CCIA는 일반적으로 데이터를 선택하고 알고리즘을 학습하는 데 사용되는 인간의 전문성이 공개되어야 한다고 생각한다. 학습된 모델 자체가 제공되고 학습된 모델이 청구항의 전체 범위를 실행하기에 충분하도록 청구항이 충분히 좁은 경우에는 인간의 전문지식은 생략될 수 있다. 그러나 AI 기술의 성공적인 적용을 위한 데이터 선택 및 학습의 중요성을 감안할 때 일반적으로 AI 학습에 필요한 인간의 전문성은 신청자가 불필요하다는 것을 보여주지 않는 한 공개되어야 한다고 가정해야 한다.

(7) 에릭슨

특허제도의 근본적인 목표는 기술을 공개함으로써 시간이 지남에 따라 공공영역이 풍부해지고 인류의 기술에 대한 체계적인 기록을 이용하고 접근할 수 있도록 하는 것이다. 특허법은 발명의 개시가 당해 기술분야에서 숙련된 자가 그 발명을 재현할 수 있도록 하기에 충분하도록 요구한다.

(i) 인공지능 지원 또는 인공지능 생성 발명이 공개요건에 대해 제시하는 문제는 무엇입니까?

(답변) AI 발명에 고유한 특정 공개 관련 고려사항이 있다고 생각하지 않는다.

(ii) 시간이 지남에 따라 데이터에 대한 접근으로 알고리즘이 변경되는 기계학습의 경우, 초기 알고리즘의 공개는 충분한가?

(답변) 초기 알고리즘의 공개를 넘어, 본 발명을 실행하는 데 필요한 발명과 관련되는 알고리즘의 진화가 어떻게 구현되는지도 설명되어야 한다.

(iii) 미생물의 기탁과 유사한 알고리즘을 위한 기탁 시스템이 유용할까요?

(답변) 발명을 실행하기 위한 충분한 공개이며, 이는 반드시 충족되어야 한다. 따라서 어떤 경우에는 알고리즘 또는 그 일부의 공개가 필요하거나 도움이 될 수 있으며, 어떤 경우에는 특허 명세서에 넣기에는 너무 길 수 있다. 이러한 경우, 알고리즘에 대한 기탁 시스템이 도움이 될 수 있다. 그러나 우리는 이것이 어느 정도까지 필요할지 예상할 수 없다.

(iv) 알고리즘을 학습하는 데 사용되는 데이터는 공개 목적을 위해 어떻게 처리되어야 하는가? 알고리즘을 학습하는 데 사용되는 데이터는 특허출원에 공개되어야 하는가 아니면 기술되어야 하는가?

(답변) 이것은 공개의 충분성에 대한 문제이다. 이와 관련하여 학습데이터에 대한 정보가 필요한 경우 이를 포함해야 한다. 그러나 이는 데이터와 관련된 다른 요구사항(예: 개인 데이터의 기밀성 또는 보호)과 균형을 이루어야 한다. 대부분의 경우 데이터가 필요한 것이 아니라 데이터의 유형과 잠재적인 양이 필요하다.

4. 학술연구 부문에서의 논의

NeurIPS(Neural Information Processing Systems Conference)¹⁸³⁾에서의 논의는 특허법적 측면에서 논의가 아닌 학술적인 부분에서 재현가능성을 확보하기 위한 차원에서 논의된 것이다. 다만, 재현가능성은 특허법에서 발명의 실시에서 필수적인 사항이라는 점에서 논의의 필요성이나 방향에서는 다르지 않고, 실질적인 개발자나 연구자집단에서 활용되는 사항이라는 점에서 살펴볼 필요성이 크다고 하겠다.

가. NeurIPS의 기계학습 재현성 체크리스트(v2.0, 2020년 4월 7일)¹⁸⁴⁾

NeurIPS 2018의 The Posner Lecture에서 Joelle Pineau는 이 분야의 학술 논문에 대한 재현성 체크리스트를 제안하였다. 이 재현성 체크리스트는 2019년 4월 2일부터 NeurIPS 2019에 제출되는 논문의 필수 요구사항으로 채택되었다.¹⁸⁵⁾

183) 신경정보처리시스템학회(Neural Information Processing Systems)로, 뉴립스는 ICML(국제 머신러닝 학회), ICLR(표현학습 국제학회) 등과 함께 세계적인 머신러닝(ML)·AI 학회로 꼽힌다. 연구자들이 제출한 논문을 학회가 승인할 확률이 약 20%로 낮다. 그만큼 채택 논문들의 품질이 높다고 평가된다. 특히 뉴립스에서는 중장기적 관점의 기술 성과가 많이 공유된다. AI타임스, 2022.04.27.일자.

184) <https://www.cs.mcgill.ca/~jpineau/ReproducibilityChecklist.pdf>. <2022.12.12. 접속>

185) 기계학습의 재현성에 대해서는 Joelle Pineau 외, Improving Reproducibility in Machine Learning Research, Journal of Machine Learning Research 22 (2021) pp. 1-20에서 구체적으로 살펴볼 수 있다.

(1) 표시된 모든 모델 및 알고리즘에 대해 다음을 포함하는지 확인

- 수학적 설정, 알고리즘 및/또는 모델에 대한 명확한 설명
- 모든 가정에 대한 명확한 설명
- 알고리즘의 복잡성(시간, 공간, 표본 크기)에 대한 분석

(2) 이론적 주장에 대해서는 다음 사항을 포함하는지 확인

- 클레임에 대한 명확한 진술
- 클레임의 완전한 증거

(3) 사용되는 모든 데이터셋에 대해 다음을 포함하는지 확인

- 예제 수와 같은 관련 통계량
- 열차/유효성 검사/테스트 분할에 대한 세부 정보
- 제외된 모든 데이터와 모든 전처리 단계에 대한 설명
- 데이터셋 또는 시뮬레이션 환경의 다운로드 가능한 버전에 대한 링크
- 수집된 새 데이터의 경우 데이터 수집 프로세스에 대한 전체 설명
- 주석자에 대한 지침 및 품질 관리 방법

(4) 이 작업과 관련된 모든 공유 코드에 대해 다음을 포함하는지 확인

- 학습 코드,
- 평가 코드
- (사전) 학습된 모델
- README 파일 : 결과 테이블이 포함되어 있으며, 실행하여 결과를 생성할 수 있는 정확한 명령이 포함되어 있음

(5) 보고된 모든 실험 결과에 대해 다음을 포함하는지 확인

- 고려되는 하이퍼 파라미터¹⁸⁶의 범위, 최적의 하이퍼 파라미터를 선택하는 방법
- 결과를 생성하는 데 사용되는 모든 하이퍼 파라미터의 구성 및 사양
- 학습 및 평가 실행의 정확한 수

186) 하이퍼 파라미터는 학습 과정을 제어하는 ANN 모델의 구성 설정이다. 하이퍼 매개 변수는 모델 아키텍처(예: 숨겨진 노드 및 계층의 수), 손실 함수, 학습률, 정규화 등)에 제공되며, 모델 매개 변수는 데이터(가중치 및 편향)에서 추론된다. Kim, Daria and Alber, Maximilian and Kwok, Man Wai and Mitrovic, Jelena and Ramirez-Atencia, Cristian and Rodriguez Perez, Jesus and Zille, Heiner, Ten Assumptions About Artificial Intelligence That Can Mislead Patent Law Analysis (August 1, 2021), p. 57.

- 결과를 보고하는 데 사용되는 특정 측도 또는 통계에 대한 명확한 정의
- 중심 경향(예: 평균) 및 변동(예: 오류 막대)이 있는 결과에 대한 설명
- 각 결과의 평균 런타임 또는 예상 에너지 비용
- 사용된 컴퓨팅 인프라에 대한 설명

나. NeurIPS의 기계학습 재현성 체크리스트(2019 버전)

소스 코드의 다운로드 가능한 버전과 데이터셋 또는 시뮬레이션 환경에 대한 링크가 논쟁의 여지가 있을 수 있지만, 체크리스트의 나머지 요구 사항은 현재 모범 사례가 아니더라도 최소한 미래의 모범 사례로 고려해야 하는 사항을 효과적으로 요약하여야 한다.

(1) 제시된 모든 알고리즘에 대해 다음을 포함하는지 확인

- 알고리즘에 대한 명확한 설명
- 알고리즘의 복잡성(시간, 공간, 샘플 크기) 분석
- 모든 종속성을 포함하여 다운로드 가능한 소스 코드에 대한 링크

(2) 이론적 주장의 경우 다음을 포함하는지 확인

- 결과에 대한 진술
- 모든 가정에 대한 명확한 설명
- 주장에 대한 완전한 증거

(3) 경험적 결과를 제시하는 모든 그림과 표에 대해 다음을 포함하는지 확인

- 샘플 크기를 포함한 데이터 수집 프로세스에 대한 완전한 설명
- 데이터셋 또는 시뮬레이션 환경의 다운로드 가능한 버전에 대한 링크
- 학습/검증/테스트를 위해 샘플이 할당된 방식에 대한 설명
- 제외된 데이터에 대한 설명
- 고려되는 하이퍼 매개변수의 범위, 최상의 하이퍼 생성기 구성을 선택하는 방법 및 결과를 생성하는 데 사용되는 모든 하이퍼 매개변수의 사양
- 정확한 평가 실행 횟수
- 실험이 실행된 방법에 대한 설명
- 결과를 보고하는 데 사용되는 특정 측정 또는 통계에 대한 명확한 정의
- 명확하게 정의된 오차 막대
- 중심 경향(예: 평균) 및 변동(표준 편차)을 포함한 결과 설명
- 사용된 컴퓨팅 인프라

제5절. AI 발명의 데이터 공개 범위 및 방안

1. AI발명에 있어서 기술공개

가. 공개의 수준에 대한 문제 제기

AI발명의 공개를 어느 수준으로 해야할 것인지는 이미 살펴본 바와 같이 일반적인 발명의 경우도 마찬가지다. 발명의 기술공개 목적이나 취지는 기술의 공개를 통하여 다양한 후속 발명을 유도함으로써 기술발전과 이를 통하여 산업 및 국민경제의 발전을 목적으로 하는 특허법의 이념과 같다.

인공지능 발명에서의 평균적 기술자 수준에 대해서 보면, 인공지능 기술 분야에서 통상의 기술자는 인공지능 기술 분야의 기술 상식을 보유하고 있고, 출원발명의 과제와 관련되는 출원전의 기술수준에 있는 모든 것을 입수하여 자신의 지식으로 할 수 있는 자로서, 실험, 분석, 제조 등을 포함하는 연구 또는 개발을 위하여 통상의 수단을 이용할 수 있으며, 설계변경을 포함한 통상의 창작능력을 발휘할 수 있는 특허법상 상상의 인물이다.¹⁸⁷⁾ 그렇지만, 발명이라는 것은 추상적인 기술적 사상을 명세서로 통해 구체화하는 과정을 거치기 때문에 추상화를 구체화하는 것은 문언에 따른 내용이기 때문에 어느 정도의 공개 내지 발명의 기술이 이루어져야 할 것인지는 통상의 기술자라는 가상의 인물을 통하여 판단할 수 밖에 없다. 그렇지만, 이러한 가정은 특허 심사관이라는 실체가 존재함에도 불구하고 가상의 인물을 내세우는 것의 타당성은 별론으로 하고, 통상의 지식인이 해당 발명의 명세서를 통하여 쉽게 실시할 수 있을 정도여야 할 것이라는 점이다.

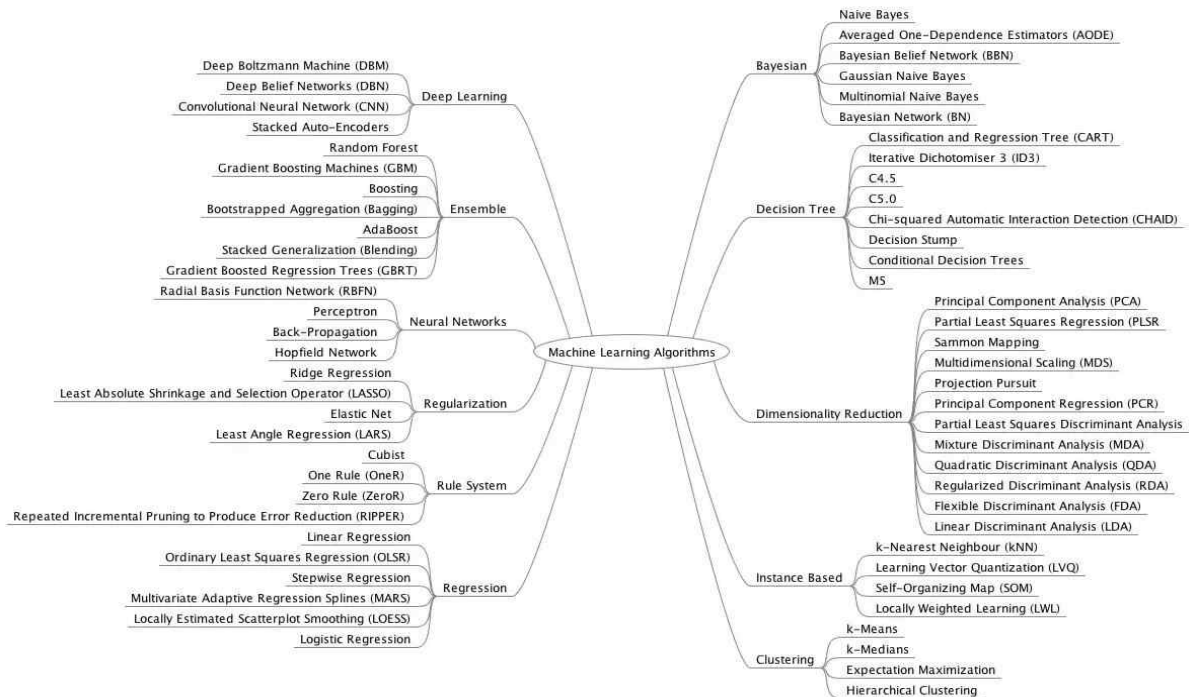
나. 구성요소로서 일반적인 알고리즘을 사용한 경우

현재 공개된 인공지능 알고리즘의 일반적인 유형은 다음과 같다. 발명에 있어서 공개된 알고리즘을 이용할 경우에는 해당 알고리즘에 대해 구체적인 기술을 필요로 하지는 않는다. 이미 평균적 기술자는 해당 알고리즘에 대한 이해를 전제로 기술공개를 할 것이기 때문에 세부적인 기술내용에 대한 공개는 사실상 해당 명세서에서는 필요한 요건이라고 보기는 어렵다. 따라서, 명세서의 발명의 상세한 설명은 해당 알고리즘의 기재가 아닌 실질적으로 출원된 발명의 성립성을 위해 뒷받침되어야할 해당 발명의 독창적인 기술적 사상을 명확하게 기재하는 것이 필요하다.¹⁸⁸⁾

187) 특허청, 기술분야별 심사실무가이드, 2020, 1311~1312면.

188) 심사지침에서도 청구항에 기재된 발명이 선행기술 또는 주지관용 기술을 인공지능 기술로 구현한 경우에는 청구항에 해당 인공지능 기술이 특정의 과제를 수행하도록 학습된 모델에서 얻어지는 특유의 정보처리에 관하여 특정하고 있고, 발명의 설명 및 기술상식을 참작하여 볼 때 선행기술 또는 주지관용 기술에 비하여 더 나은 효과가 있는 경우에는 그 발명의 진보성이 인정된다. 그러나 발명에서 채택하고 있는 인공지능 기술이 학습모델에서 얻어지는 특유의 정보처리에 관하여 특정하지 않은 채 단순히 선행기술 또는 주지관용 기술을 인공지능 기술로 구현한 것에 불과하고, 선행기술 또는 주지관용 기술에 비하여 더 나은 효과가 있는 것으로 인정되지 않는 경우에는 청구항에 기재된 발명의 진보성이 있는 것으로 보기 어렵다고 한다. 기술분야별 심사실무가이드, 특허청, 2020.

[그림 3-4] 기계학습 알고리즘 유형



<출처 : 구글검색, 2022>

위의 그림 이외에도 다양한 알고리즘이 공개되어 있다. 공지공용의 기술로서 공개된 경우에는 해당 알고리즘이 주된 경우에는 해당 알고리즘이 특성을 기재하는 것도 필요하다 할 것이다. 구체적인 방식에 대해서는 IP5의 사례를 분석함으로써 개선할 수 있는 방안을 찾고자 한다. 아울러, 데이터가 구조화된 경우를 포함하여 데이터가 사용되는 발명에 있어서도 해당 데이터의 처리 방식에 대한 구체성이 있어야 한다. 구체적인 발명의 설명이 아니라면 해당 발명은 법률적 요건을 모두 갖추었다고 보기 어렵기 때문이다.

다. AI발명에서의 공개 사항

미생물과 유사하게, 우리는 설명할 수 없는 것과 같은 어려움에 직면한 AI 시스템에 대한 기탁 규칙을 촉구한다. 기탁하려는 AI 모델에는 코드, 데이터, 출력 결과 등 AI 시스템의 핵심 구성 요소가 특허청에 디지털 형태로 제출될 수 있다. 기탁된 AI 모델은 AI발명을 설명하고, AI 시스템이 어떻게 작동하는지, 특허 출원에서 주장한 대로 AI 도구를 어떻게 만들고 사용하는지, AI 도구의 구성 요소가 무엇을 의미하는지 등을 답변하는 실시예로 활용될 수 있다. AI 모델의 시연은 서면 기재, 실시가능, 신규성 및 효용성 등 요구사항을 포함한 많은 특허성 기준을 포함하는 불충분한

설명의 문제를 해결할 것이다. 또한 기탁은 나중의 침해 사례에서 잠재적인 증거로 작용할 수 있다.¹⁸⁹⁾

(1) 일반적인 SW의 경우

일반적인 SW라면, 알고리즘을 공개함으로써 용이하게 실시할 수 있거나 과도한 실험이 없는 경우라면 공개의 수준에 대한 별다른 문제는 없을 것이라고 생각된다. 다만, 알고리즘이 무엇인지에 대해서는 명확하게 정리할 필요가 있다. 알고리즘이란 기본적으로 문제를 해결하기 위한 기법이나 방법을 의미하며, 발명의 개념에서 기술적 창작을 의미한다고 볼 수 있다. 따라서, SW발명의 명세서에 기재된 설명(description)은 알고리즘이거나 알고리즘에 관한 내용이라고 할 것이다.

(2) AI인 경우

AI는 내부적인 데이터 학습과정을 외부에서 확인할 수 없다는 점에서 공개사항을 확대할 필요가 있다. 알고리즘으로 그 내용을 확인하기도 어렵기 때문이다. 따라서, AI발명에 있어서 보다 구체적인 사항에 대한 논의가 필요하고, SW발명보다는 보다 높은 수준의 공개요건을 구성할 수 있을 것이다. 특히, 발명의 주된 요소는 데이터 기반에 의한 정보처리나 데이터셋을 생성하는 경우 등 다양하다는 점에서 데이터의 공개에 대한 구체적인 사항을 살펴보는 것은 본 연구의 주된 내용이기도 하다.

1) AI모델의 기본구조(알고리즘)

AI발명에 있어서 공개할 사항은 데이터를 기반으로 학습된 CNN등 다양한 AI모델을 들 수 있다. 표준이거나 공개된 알고리즘일 경우에는 해당 알고리즘을 기재하면 될 것이다. 만약, AI 모델이 표준이 아닌 구성 요소(예: 사용자 지정 활성화 계층)를 사용하는 경우 이 새로운 구성 요소는 수학적 형식, 의사 코드 또는 실제 컴퓨터 코드로 정확하게 설명되어야 한다. 이것 역시 큰 도전이 되지 않아야 한다. 새로운 최적화 계획이나 비표준 피드백 루프에 대해서도 마찬가지이다.¹⁹⁰⁾

2) 학습 및 학습된 계수(Training and trained coefficients)

기본 모델이 적절하게 기술된 후에도 숙련된 사람은 여전히 모델을 구현하기 위한 충분한 정

189) Hagen, Gregory R., AI and Patents and Trade Secrets (February 1, 2021). in Florian Martin-Bariteau & Teresa Scassa, eds., Artificial Intelligence and the Law in Canada (Toronto: LexisNexis Canada, 2021), Available at SSRN:<https://ssrn.com/abstract=3734654>

190)

<https://www.iam-media.com/global-guide/iam-yearbook/2020/article/ai-inventions-and-sufficiency-of-disclosure-when-enough-enough> <2022.11.26. 방문>

보를 가지고 있지 않다. 즉, 본 발명은 충분히 공개되지 않았다. 여전히 필요한 것은 학습 데이터에 대한 참조를 포함하여 모델이 학습되는 방식에 대한 설명 또는 모델의 모든 학습된 계수 또는 가중치 중 적어도 하나이다. 이것의 중요성은 아무리 강조해도 지나치지 않는다. 딥러닝 모델의 구조에서 미리 결정된 유일한 부분은 시작 부분의 입력 이미지와 끝부분의 출력값이다. 입력과 출력 사이의 모든 레이어를 숨겨진 레이어라고 하며, 이는 학습 중에 형성된 후 이러한 레이어에서 정확히 어떤 일이 발생하는지는 잘 알려져 있지 않다.

AI 시스템의 구성 요소로서의 모델은 별도로 주장될 수도 있고 시스템의 다른 구성 요소와 관련하여 주장될 수도 있다. 예를 들어 심층 신경망 학습 알고리즘에서 모델이나 구조 개선은 학습 알고리즘으로 주장될 수 있으며 그 구조, 층 및 뉴런의 수, 그리고 그에 따라 정의될 수 있다. 그들 사이의 내부 연결 또는 모델은 알려진 모든 조정된 가중치를 가진 실행 단계에서 학습되고 배치된 모델로 주장될 수 있다. 알려진 모델의 구조를 공개하는 것만으로는 당업자가 제안된 모델을 실제로 사용할 수 없을 것이다. 따라서 발명자는 네트워크의 가중치와 계수를 공개해야 하며, 그렇지 않으면 학습데이터셋을 공개하고 숙련된 사람이 학습 알고리즘을 학습하고 모델을 배포할 수 있도록 학습 방법을 제공할 수 있다.¹⁹¹⁾

3) 학습데이터

비용이 많이 드는 데이터 수집 프로세스를 통해 얻은 학습데이터셋을 공개하는 것은 혁신적인 주체에게 심각한 문제가 될 수 있다.¹⁹²⁾ 출원인은 경쟁업체가 다른 AI 모델(신청자의 주장을 침해하지 않도록 신중하게 선택)을 빠르게 학습하는 데 사용할 수 있고, 따라서 불공정한 경쟁우위를 얻을 수 있다고 보아 이 데이터셋을 대중에게 공개하는 것을 꺼릴 수 있다. 귀중한 데이터를 청소하고 추출하는 과정은 데이터 수집 비용을 증가시킬 수 있는 특정 지식과 기술을 필요로 한다.¹⁹³⁾ 따라서 AI 기반 기업은 이 귀중한 데이터를 공개하지 않고 영업 비밀로 유지하는 것을 선호한다.¹⁹⁴⁾ 학습데이터가 생략된 상태에서 설명에 포함된 학습 방법에 대한 설명만 있는 경우가 있는 이유 중 하나이다.

2. 데이터 공개의 수준 및 범위

데이터를 공개하는 것은 앞서 살펴본 T/0161/18 사건처럼, 공개되지 아니할 경우에 발생할 수 있는 발명의 거절이나 무효 등에 대해 적절하게 대처할 수 있는 것임을 확인하였다. 따라서, 공

191) Medhi Poursoltani, Disclosing AI Inventions, TEXAS INTELLECTUAL PROPERTY LAW JOURNAL Vol.29, 2021, p. 56.

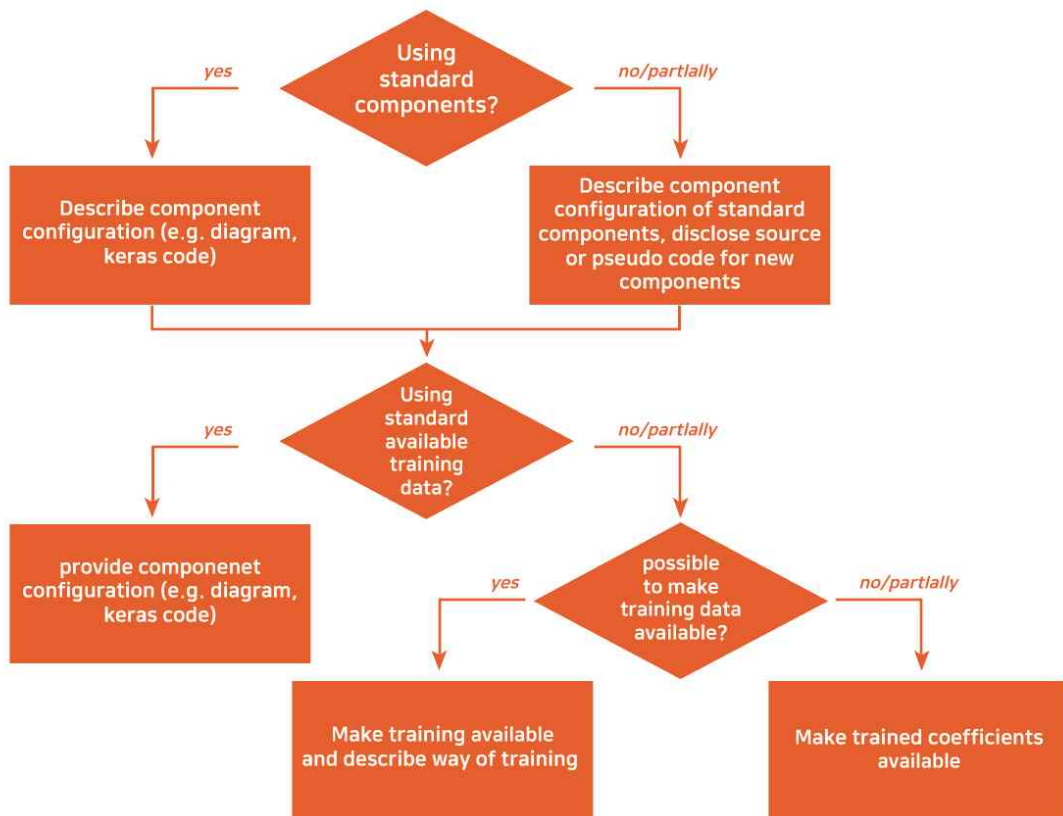
192) Volha Litskevich, How Much Does Artificial Intelligence (AI) Cost in 2019, AZATI (Jan. 23, 2019)

193) Volha Litskevich, How Much Does Artificial Intelligence (AI) Cost in 2019, AZATI (Jan. 23, 2019)

194) Medhi Poursoltani, Disclosing AI Inventions, TEXAS INTELLECTUAL PROPERTY LAW JOURNAL Vol.29, 2021, p. 57.

개의 필요성이 인정되며, 데이터를 공개해야할 경우에 대해 판단을 내릴 수 있어야할 것이다. 만약, 공개를 할 경우에는 데이터의 어떤 정보를 포함해야할 것인지는 발명을 보호하고자 하는 입장에서 최소화하거나 요부가 아닌 것으로 선택하려고 할 것이다. 아래 그림은 기본적으로 공개된 데이터셋을 이용한 경우인지, 아니면 직접 제작하거나 구매하여 사용한 경우인지를 구별할 수 있도록 구성된 일종의 판별식이다.

[그림 3-5] 데이터 공개방식에 따른 판별식



<출처 : Law Business Research, 2022>

가. 공개 대상

데이터셋을 공개하는 경우에는 몇 가지 사항을 고려할 수 있을 것이다. 우선, 공개하는 경우는 유형화할 필요가 있다. 모든 데이터를 공개할 필요가 있는 경우는 드물 것으로 보이고, 외부에 공개된 표준화된 데이터셋을 이용한 경우라면 굳이 공개하고자 하는 대상에 포함할 필요가 없는 경우이다. 깃허브와 같이 공개된 사이트나 AI허브와 같이 독자적인 구축형태로 공개된 데이터셋인지 여부를 표시함으로써 판단기준을 제시할 수 있을 것이다. 독자적인 구축이라면, 어느 정도를 공개할 것인지에 대한 기준이 제시될 필요가 있다.

(1) 구체적인 설명으로 대체 가능한 경우

인터넷에 공개된 정보를 바탕으로 생성할 수 있는 데이터셋으로 용이하게 취득될 수 있는 데이터셋이라면 어떻게 취득할 수 있는지에 대해 기록하면 될 것이다. 대표적으로 플랫폼사업자들이 자사 회원들의 SNS 등을 이용한 경우이다.

(2) 외부 공개데이터를 이용한 경우

이미지넷(ImageNet)이나 깃허브 등 표준화된 데이터셋과 같이 외부에 데이터셋이 공개된 경우라면, 해당 데이터셋이 공개된 사이트나 플랫폼에 대한 출처표시를 명확히 하는 것이 필요하다. 물론, 이러한 경우라도 데이터셋에 대한 상세한 설명이 필요한 경우라면 그렇게 하는 것이 바람직하다. 다만, 외부 데이터셋이 소멸되거나 변경될 경우에는 발명의 동일성이나 재현가능성이 담보될 수 없다는 한계를 갖는다.

(3) 자체 제작 등의 데이터셋을 이용한 경우

데이터셋을 자체적으로 제작한 경우라면, 제한없이 필요한 사항을 기재하거나 공개할 수 있을 것이다. 다만, 이용허락을 받거나 구매한 경우라면 계약조건에 따라 다를 수 있을 것이다. 무엇보다, 자체 제작의 경우에는 데이터셋의 전체를 공개할 것인지, 일부를 공개할 것인지, 샘플링할 것인지 선택할 수 있도록 할 필요가 있다.

청구항에 기재된 발명에 대하여 발명가는 추가적인 데이터셋을 공개하고 숙련된 사람이 모델을 수정하고 조정하는 방법을 설명하고 관련 작업에 사용할 새로운 모델을 개발할 수 있다. 그럼에도 불구하고, 학습데이터셋의 공개는 청구항에 구체적으로 공개된 데이터셋으로 제한할 수 있다. 또한 유사한 유형의 데이터셋을 공개하면 생성된 결과에 사소한 차이가 있더라도 완전히 다른 모델을 생성할 수 있기 때문에 보호를 못할 수 있다. 따라서, 발명자가 서로 다른 공개된 데이터셋이 실질적으로 유사한 결과를 산출하는 모델을 생성할 것이라는 것을 보여줄 수 없다면, 청구범위는 공개된 데이터셋으로 제한되어야 한다.¹⁹⁵⁾

세부적인 내역에 대해서는 위에서 살펴보는 공개내용에 따르면 될 것이다. 외부의 데이터를 이용한 경우는 계약 범위 내에서 자체 제작의 경우에 준하여 공개 수준을 선택할 수 있을 것이다.

나. 공개 수준

공개 수준은 데이터셋을 공개할 경우에 어느 수준에서 정할 것인지 여부이다. AI 응용 발명에

195) Mehdi Poursoltani, Disclosing AI Inventions, TEXAS INTELLECTUAL PROPERTY LAW JOURNAL Vol.29, 2021, p. 58; Generally Artificial Intelligence (AI) Patents, PATENT ATTORNEYS <https://www.techlaw.attorney/artificial-intelligence-ai-patents> <2022.11.22. 접속>.

대한 학습데이터의 공개는 전체 데이터셋의 명시적 공개에서 데이터셋의 특성 설명에 이르기까지 다양할 수 있다. 전체 데이터셋의 명시적 공개는 특히 초안 작성자에게 법적 및 실질적인 문제를 제시할 수 있다. 예를 들어, 개인 정보 보호법 또는 이러한 데이터의 독점 특성으로 인해 특허 사양에서 트레이닝 데이터를 제외해야 하는 경우가 있다. 전체 데이터셋을 공개하는 것이 허용되더라도 최신 딥러닝 네트워크를 학습하는 데 사용되는 방대한 양의 데이터를 고려할 때 실용적이지 않은 경우가 많다. 따라서 학습 데이터의 출처와 양, 학습 데이터의 전처리나 정리, 학습 데이터의 세부적인 특성에 대한 설명이 포함되어야 응용 발명을 충분히 공개할 수 있다.¹⁹⁶⁾

다음 사항을 고려하여 공개 수준을 정하는 것이 필요하다. 다만, 기밀데이터 또는 독점데이터와 같이 데이터셋의 공개를 원하지 않을 경우에는 데이터 획득방법에 대한 완전한 세부적인 정보가 포함될 필요가 있다. 이러한 방식으로 숙련된 사람은 데이터셋을 재생산할 수 있기 때문이다.¹⁹⁷⁾

[표 3-1] 공개 수준

데이터셋의 공개 수준	비고
<ul style="list-style-type: none"> • 전체 범위를 공개할 것인지 • 일부를 공개할 것인지 • 획득 방법 등의 세부적인 정보 • 일정 제한 요건을 갖출 것인지 • 오픈소스 라이선스 방식으로 할 것인지 	<p>공개수준에 대해서는 아래 사항을 고려하여 결정</p> <ul style="list-style-type: none"> • 지역적 한계를 설정할 것인지 여부 • 제3자 제공금지 등에 대한 사항

다. 공개 내용(항목)

데이터셋을 공개하는 구체적인 항목은 데이터셋이 갖고 있는 속성들에 따라 다를 수 있을 것이다. 다만, 데이터셋을 공개하는 경우라면 몇 가지 고려할 사항도 있다. 반드시 필요한 항목과 데이터셋이 공개된 경우에 실제 심사관이 해당 내용을 확인할 수 있는 항목들을 지정할 필요가 있는지에 관한 것이다. 아울러, 데이터셋의 표준화의 여부이다. 향후, 표준화는 데이터셋을 기탁하고 기탁된 데이터셋이 외부에 공개될 경우에 활용가능성을 높일 수 있는 방법이기 때문이다.

사전 학습된 모델의 가중치 초기화는 심층 신경망 학습 알고리즘에서 중요한 단계이다. 초기화 단계에서 가중치의 초기값을 무작위화로 지정할 수 있으며, 가중치의 최종값은 학습 과정 후

196) Edwin D. Garlepp, Disclosing AI Inventions - Part II: Describing and Enabling AI Inventions. 2021. <https://www.theipatentblog.com/disclosing-ai-inventions-part-ii-describing-and-enabling-ai-inventions> <2022.11.22. 접속>

197) <https://www.appleyardlees.com/artificial-intelligence-and-machine-learning-sufficiency-and-plausibility/> <2022.11.22. 접속>

반에 조정될 것이다. 초기값은 사전에 결정되지 않으며 랜덤화를 포함하여 많은 방식으로 정의될 수 있다. 동일한 구조 또는 아키텍처와 동일한 학습데이터셋을 가진 동일한 모델을 사용하더라도, 서로 다른 가중치 초기화는 서로 다른 후보 모델로 이어질 수 있다. 비록 후보 모델들의 결과들의 차이가 미미할 수 있지만, 그럼에도 불구하고, 이러한 최소한의 차이는 발명의 범위를 정의하고, 발명자가 청구된 모델의 소유권과 유효 출원일 현재 그것의 최소 변형을 가지고 있는지를 결정하는 데 문제가 될 수 있다. 이러한 AI 모델의 비결정론적 특성은 요구되는 공개 기준을 변경할 수 있으며, 모델 관련 AI 발명에 더 높은 공개 기준을 부과할 수 있다.¹⁹⁸⁾

또한, 서로 다른 가중치 초기화 방법, 서로 다른 모델 하이퍼 파라미터, 데이터셋의 전처리, 각 반복에서 데이터셋의 무작위화 및 학습 방법은 서로 다른 후보 모델로 이어질 수 있다.¹⁹⁹⁾

생명공학 분야와 마찬가지로, AI 모델의 비결정론적이고 블랙박스과 같은 특성은 더 높은 수준의 불확실성과 예측 불가능성을 초래할 수 있다. 따라서, 인공지능 분야의 발명자들은 청구된 모델의 다수의 실시예, 모델의 정확한 구조, 샘플 데이터셋, 중량 초기화 방법 및 샘플 초기 중량 값을 포함하여 청구된 모델의 전체 범위를 보유하고 있음을 보여주는 보다 상세한 정보를 공개해야 할 수 있다. 발효일 당시 발명 및 당업자가 과도한 실험 없이 청구의 전체 범위를 작성하고 사용할 수 있도록 하는 것이다.²⁰⁰⁾

위와 같은 논의에 따라, 공개할 수 있는 항목에 대해 다음과 같이 정리하고자 한다. 항목들은 공개된 데이터셋을 통하여 실질적으로 확인할 수 있는 사항들을 기재하였다. 물론, 데이터셋이 자체적으로 담겨있는 정보를 통하여 확인할 수도 있겠지만, 데이터셋에 표시할 수 있는 항목들을 제시하는 것도 공개제도가 갖는 특성이기 때문이기도 하다.

[표 3-2] 데이터 항목

데이터셋의 공개 항목	비고
<ul style="list-style-type: none"> • 데이터셋 • 데이터의 구조 • 파라미터 초기화 등 작업 • 가중치 • 데이터의 유형 • 데이터 가공방식 • 파일 형식 : 이미지, 동영상, 텍스트, CVS²⁰¹⁾, PDF, XML, 	<p>공개 항목은 데이터 표준화 논의를 통해서 구체화하는 방안도 고려할 수 있을 것임</p>

198) Mehdi Poursoltani, Disclosing AI Inventions, TEXAS INTELLECTUAL PROPERTY LAW JOURNAL Vol.29, 2021, p. 59.

199) Mehdi Poursoltani, Disclosing AI Inventions, TEXAS INTELLECTUAL PROPERTY LAW JOURNAL Vol.29, 2021, p. 61.

200) Medhi Poursoltani, Disclosing AI Inventions, TEXAS INTELLECTUAL PROPERTY LAW JOURNAL Vol.29, 2021, p. 62.

<p>JASON²⁰²⁾ 방식 등</p> <ul style="list-style-type: none"> • 합성데이터인지 여부 • 저작권 등 권리 획득 여부 • 수집 및 취득 방법 등 데이터의 출처 • 전처리 방식 • 데이터 학습에 사용된 알고리즘 모델 	
--	--

라. 공개 방법

(1) NFT의 신기술 활용

1) 스마트계약과 NFT

스마트 계약(smart contract) 또는 스마트 컨트랙트란 블록체인 기반으로 금융거래, 부동산 계약, 공증 등 다양한 형태의 계약을 체결하고 이행하는 것을 말한다. 엄밀히 말하자면, 스마트계약은 블록체인 기술을 기반으로 계약 협상과 이행을 자동으로 실행되는 컴퓨터 언어로 작성된 계약 프로그램이다. 이는 우리가 일상적으로 사용하는 전통적 의미의 계약이 아니라 블록체인 기술에 의하여 일정한 조건이 성취되면 계약이 자동적으로 실행되는 컴퓨터 프로그램으로 볼 수 있다. 스마트계약은 일종의 자동화 계약 시스템으로 기존에는 계약이 체결되고 이행되기까지 수많은 문서가 필요했다면 스마트 계약은 계약조건을 컴퓨터 코드로 지정해 두고 조건이 맞으면 계약을 이행된다.²⁰³⁾

스마트계약은 2013년 비탈릭 부테린(Vitalik Buterin)이 비트코인의 블록체인 기술을 이용하여 대금결제, 송금 등 금융거래뿐 아니라 모든 종류의 계약을 처리할 수 있도록 기능을 확장하면서 널리 확산되었다. 부테린은 기존 비트코인의 소스 코드를 일부 수정하여 스마트 계약 기능을 구현하고자 하였으나, 비트코인 커뮤니티에서 자신의 요구가 받아들여지지 않자, 비트코인을 포크(fork)하여 새로 이더리움(Ethereum)이라는 가상화폐를 만들고 스마트 계약 기능을 구현하였다. 이 기능을 사용하면, 개발자가 직접 계약 조건과 내용을 코딩할 수 있기 때문에, 원칙적으로 인간이 상상할 수 있는 모든 종류의 계약을 이더리움 플랫폼을 이용해 구현할 수 있다. 다만, 솔리

201) CSV(영어: comma-separated values)는 몇 가지 필드를 쉼표(,)로 구분한 텍스트 데이터 및 텍스트 파일이다. 확장자는 .csv이며 MIME 형식은 text/csv이다. comma-separated variables라고도 한다. 오래전부터 스프레드시트나 데이터베이스 소프트웨어에서 많이 쓰였으나 세부적인 구현은 소프트웨어에 따라 다르다. (위키백과 2022.11.26.일 방문)

202) JSON이란 JavaScript Object Notation의 축약어로 데이터를 저장하거나 전송할 때 많이 사용되는 경량의 DATA 교환 형식이다. 즉, Javascript에서 객체를 만들 때 사용하는 표현식을 의미한다. JSON 표현식은 사람과 기계 모두 이해하기 쉬우며 용량이 작아서, 최근에는 JSON이 XML을 대체해서 데이터 전송 등에 많이 사용한다.

203) 손경한 편저, 블록체인과 법, 박영사, 2019, 147면.

더티(Solidity)라는 자바 기반의 독립적인 프로그래밍 언어를 알아야 하기 때문에, 프로그래머가 아닌 일반인들이 직접 스마트 계약의 조건과 내용을 코딩하기는 어렵다.

2) 증명서로서 NFT

NFT(Non-Fungible Token)는 블록체인에 저장된 데이터 단위로, 특정 기초자산의 소유 및 거래 기록을 블록체인에 영구적으로 저장하여 디지털 자산화를 지원하는 기술을 의미한다. NFT와 관련하여 볼 수 있는 법률인 특정 금융거래정보의 보고 및 이용 등에 관한 법률(특정금융정보법)은 가상자산을 “경제적 가치를 지닌 것으로서 전자적으로 거래 또는 이전될 수 있는 전자적 증표”로 포괄적으로 정의하고 있다. 따라서, NFT는 그 자체로서 독립적인 경제적 가치를 지닌다기 보다는 경제적 가치를 지닌 기초자산에 대한 일종의 등기권리증 역할을 수행한다고 볼 수 있어 과연 NFT가 특정금융정보법상 가상자산에 포섭되는지 여부는 분명하지 않다. 다만, 국제자금세탁방지기구인 FATF는 2021. 10. 28. 개정 발표한 ‘가상자산 및 가상자산사업자 위험기반 접근법 지침서’를 통해 NFT는 일반적으로 가상자산의 범위에 포함되지는 않으나, NFT가 가상자산의 일반적인 성격인 결제 및 투자 용도로 사용된다면 가상자산의 범위에 포함될 수도 있다고 판단된다.

NFT는 코드 내지 링크로서 증명서 역할을 한다. NFT 생성을 위한 요소로서 간단한 이미지, 이름, 외부 링크 정도로 구성된다. 기초 자산에 해당하는 디지털 작품의 원본파일이 NFT안에 저장되기보다는 외부 링크로 연결되는 것이다.²⁰⁴⁾

NFT는 자산의 희소성 보장 및 원본성 증빙, 소유자 정보 및 거래이력 등을 영구적으로 저장할 수 있는 기술로 평가되어, 현재 스포츠, 엔터테인먼트, 게임, 패션 등 다양한 분야에서 활용된다. 특히, 디지털 수집품(Digital Collectible) 시장에서 많이 사용되고 있다. 다만, NFT는 원본 인증서로서 역할을 하지만 토큰은 기초자산이 원본이라는 것을 쉽게 확인하는 것이지 이 토큰이 기초자산의 원본임을 검증하는 것은 아니다.²⁰⁵⁾

204) 장세형, NFT 실체와 가치, 위키북스, 2022, 165면.

205) 장세형, NFT 실체와 가치, 위키북스, 2022, 169면.

[그림 3-6] NFT 생성 정보

Create New Item

* Required fields

Image, Video, Audio, or 3D Model *
File types supported: JPG, PNG, GIF, SVG, MP4, WEBM, MP3, WAV, OGG, GLB, GLTF. Max size: 100 MB



Name *

External link
OpenSea will include a link to this URL on this item's detail page, so that users can click to learn more about it. You are welcome to link to your own webpage with more details.

<출처 : Opensea 블로그>

3) 데이터 공개 방식으로써 NFT의 활용

이러한 NFT나 블록체인 등의 기술을 활용할 수 있을 것으로 생각된다. 기술적으로 위변조 등이 어렵다는 점에서 데이터의 변개가 어려울 수 있기 때문에 상대적으로 데이터의 신뢰성을 확보할 수 있을 것이다. 다만, 데이터가 블록체인에 공개되더라도 활용이 용이하지 않다는 점에서 실효성은 떨어질 수 있다. 따라서, 블록체인에 데이터를 탑재하기 어렵기 때문에 외부의 기관에 데이터를 저장하는 방식으로 운용될 수 있을 것이다. 이 경우에는 전반적인 관리체계를 갖추어야 할 것이기 때문에 전문기관을 지정하는 방안을 고려할 수 있으며, 제3의 기관에 공개하는 방식과 유사하며, 다만 방식에서의 차이라고 할 것이다.

NFT 등의 신기술을 활용할 경우에는 별도의 민팅(minting)²⁰⁶⁾ 과정을 거치기 때문에 일명 가스비라고 하는 민팅 비용이 발생한다. 이러한 점은 대규모의 데이터에 적용하는 경우라면 발생하는

206) 민팅이란 그림이나 영상 등 디지털 자산의 대체불가능토큰(NFT)을 생성하는 것을 일컫는 용어다. 민팅은 화폐를 주조한다는 의미에서 유래한 말이다.

비용이나 관리비용 측면에서 실효성이 떨어질 수 있기 때문에 기술적인 사항과 더불어 경제적인 면까지 고려하여야 할 것이다.

(2) 공개사이트를 이용한 경우

깃허브, AI-허브 등에서 공개된 데이터를 이용하였다면 학습데이터를 발명자가 자체적으로 제작한 것이 아닌 공개된 데이터를 이용한 경우이므로, 해당 사이트를 표기하면 될 것이다. 오픈소스 방식이나 별도의 비용 없이 이용할 수 있는 경우에는 별다른 공개 절차가 필요하지 않은 경우라고 할 것이다.

(3) 제3의 기관에 공개하는 방안

학습데이터를 공개할 경우, 깃허브나 AI허브 등 기존의 사이트를 이용하는 방법도 고려할 수 있다. 그렇지만, 공개된 데이터는 영구적인 보관이나 운영이 이루어져야 한다. 무엇보다, 특허요건으로 데이터를 기탁하는 경우라면 특허제도가 유지되는 동안에는 기탁된 데이터를 보유하고 있어야 하기 때문이다. 다만, 기술적인 방식이나 공개사이트를 통해서 공개하는 것은 여러 가지 운영상의 리스크가 있음을 확인하였다. 따라서, 이를 위해서 제3의 기관에서 운영하는 것이 바람직하다. 만약 제3의 기관이 운영하게 된다면, 해당 기관은 미생물 기탁기관과 같이 여러 가지 요건을 갖추어야 할 것이며 미생물기탁제도와 같이 특허청이 지정하는 곳으로 하는 것이 바람직하다.

무엇보다 제3의 기관을 지정할 경우, 고려할 수 있는 요건으로는 특허와 관련 전문성을 갖출 것, 대용량 데이터를 처리할 수 있는 기술력을 갖추어야 할 것, 향후 다른 데이터와 연계하여 확장할 수 있는 사업능력을 갖출 것, 다양한 특허정보를 데이터셋으로 제작하여 연계할 수 있는 능력을 갖추어야 할 것이다.

3. 인센티브의 부여 방안

인공지능 분야에 종사하는 사람들을 포함한 발명가들이 자신들의 발명에 대한 데이터나 다른 정보를 공개하는 것을 주저하는 것은 드문 일이 아니다. 이는 다른 사람들이 정보를 사용하거나 오용하거나 시장에서 발명자와 경쟁할 가능성에 대한 우려 때문일 수 있다. 또한, 데이터의 잠재적인 도난이나 오용에 대한 우려 또는 경쟁자들이 공개된 정보에 기초하여 발명을 리버스 엔지니어링할 수 있는 가능성 때문이다.²⁰⁷⁾ 이러한 점을 우려하여 영업비밀 제도에 대한 검토 및 인센티브 방안을 제안할 수 있을 것이다. 인센티브 방안에 대해서는 자발적으로 유도할 것인지, 강제적인 방식으로 접근할 것인지에 따라 여러 가지 사항을 고려할 수 있을 것이다.

207) Hagen, Gregory R., AI and Patents and Trade Secrets (February 1, 2021). in Florian Martin-Bariteau & Teresa Scassa, eds., Artificial Intelligence and the Law in Canada (Toronto: LexisNexis Canada, 2021), Available at SSRN:<https://ssrn.com/abstract=3734654>

가. 영업비밀과 데이터 공개의 충돌

(1) 관계

새롭게 창작하거나 구조·성능 등을 개량한 기술정보에 대하여, 현행법상 법적 장치는 특허제도(실용신안제도)가 대표적이다. 특허법은 일정한 요건을 갖춘 발명에 대해 이를 공개하는 조건으로 그 발명자에게 일정기간 동안 독점권을 부여하고 있으므로, 특허권자는 특허제도를 이용하여 자신의 발명을 특허로 출원하여 등록을 받음으로써 특허법이 보장하는 강력한 독점권을 행사할 수 있다. 그러나 실제로 특허권의 보호대상이 되기 위해서는 산업상 이용가능성, 신규성, 진보성 등의 까다로운 요건을 갖추어야 하므로, 특허권으로 보호될 수 있는 기술정보는 매우 제한적이다. 더구나 발명자는 기술정보의 공개를 꺼려하여 특허권에 의한 보호를 원하지 않을 수도 있다. 이러한 경우 가능한 법적 보호 중의 하나로서 대표적인 것이 바로 부정경쟁방지법상 영업비밀에 의한 보호이다.²⁰⁸⁾ 두 법이 추구하는 목적이 상이하다는 점에서 어느 일방에 의한 적용이 배제되는 것은 아니다. 데이터셋에 대해서는 특허를 통한 공개와 영업비밀로서의 보호가 가능하다.

(2) 데이터의 영업비밀성 요건

영업비밀은 공연히 알려져 있지 아니하고 독립된 경제적 가치를 가지는 것으로서, 상당한 노력에 의하여 비밀로 유지된 생산방법, 판매방법 그 밖에 영업활동에 유용한 기술상 또는 경영상의 정보를 말하는 것인데, 여기서 ‘공연히 알려져 있지 아니하다’는 것은 정보가 간행물 등의 매체에 실리는 등 불특정 다수인에게 알려져 있지 않기 때문에 보유자를 통하지 아니하고는 정보를 통상 입수할 수 없는 것을 말하고, ‘독립된 경제적 가치를 가진다’는 것은 정보 보유자가 정보의 사용을 통해 경쟁자에 대하여 경쟁상 이익을 얻을 수 있거나 또는 정보의 취득이나 개발을 위해 상당한 비용이나 노력이 필요하다는 것을 말하며, ‘상당한 노력에 의하여 비밀로 유지된다’는 것은 정보가 비밀이라고 인식될 수 있는 표시를 하거나 고지를 하고, 정보에 접근할 수 있는 대상자나 접근 방법을 제한하거나 정보에 접근한 자에게 비밀준수의무를 부과하는 등 객관적으로 정보가 비밀로 유지·관리되고 있다는 사실이 인식 가능한 상태인 것을 말한다.²⁰⁹⁾

데이터의 경우도 관리가능성, 독립된 경제성, 비밀유지성이 있는 상태라면 영업비밀로서 보호받을 수 있다.²¹⁰⁾ 따라서, 이러한 요건하에 있는 데이터셋이 공개될 경우에는 영업비밀성이 해소된다고 할 것이다. 그렇다면, 명세서에 데이터를 공개하지 않은 상태로 출원하는 경우에는 영업

208) 한국특허법학회편, 특허판례연구, 박영사, 2017, 36면.

209) 대법원 2011. 7. 14. 선고 2009다12528 판결 [손해배상(기)]

210) USPTO, Public Views on Artificial Intelligence and Intellectual Property Policy, October, 2020, p. 36.

비밀성에는 하등 문제될 것은 아니라고 하겠다. 다만, 그러할 경우에는 충분한 설명이 이루어진 것으로 보지 않을 가능성도 있다. 즉, 데이터에 대한 2차적 정보가 기재된 경우라는 점에서 발명을 용이하게 실현할 수 없는 상태라고 판단될 경우라면 발명이 완성되지 못한 미완성발명으로 무효사유에 해당할 수 있다는 우려도 있기 때문이다.

(3) 중복 보호의 가능성

특허출원을 하기 위한 특허출원서에는 발명의 명세서와 필요한 도면 및 요약서를 첨부하여야 하고, 발명의 상세한 설명에는 그 발명이 속하는 기술분야에서 통상의 지식을 가진 자가 용이하게 실시할 수 있을 정도로 그 발명의 목적·구성 및 효과를 기재하여야 하며, 특허청구범위에는 발명이 명확하고 간결하게 그 구성에 없어서는 아니되는 사항을 기재하여야 하므로(특허법 제42조 제2항, 제3항, 제4항 참조), 그 기술분야에서 통상의 지식을 가진 자라면 누구든지 공개된 자료를 보고 실시할 수 있다 할 것이니, 특허출원된 발명에 대하여 영업비밀을 주장하는 자로서는 그 특허출원된 내용 이외의 어떠한 정보가 영업비밀로 관리되고 있으며 어떤 면에서 경제성을 갖고 있는지를 구체적으로 특정하여 주장·입증하여야 할 것이다.²¹¹⁾

특허제도는 기술을 공개한 대가로 일정기간 동안 독점적, 배타적 권리를 갖는 제도인 반면, 영업비밀은 어떤 정보를 비밀로 유지하려는 노력에 대해 보호가 주어진다는 점을 고려할 때, 출원 공개되지 않은 기술이라면 몰라도, 공개된 특허 기술에 대해 다시 이를 영업비밀에 의해 보호된다고 보는 것은 영업비밀제도의 본질상 문제가 있다고 생각된다. 특허로 공개된 기술 그 자체는 특허로써 보호하면 충분하고 그와 별도로 영업비밀 보호제도에 의해 별도로 보호할 필요는 없을 뿐만 아니라, 현행법상의 영업비밀 보호제도에 비추어 볼 때 영업비밀로서 중복하여 보호하는 것이 가능하지도 않을 것이다. 다만, 특허로 공개된 기술 이외에 어떤 '특별한 기술정보'가 더 존재한다면, 이는 동일한 정보에 대한 중복 보호가 아니라 별도로 영업비밀로서의 보호가 가능하다.²¹²⁾ 대법원도 “특허출원된 발명에 대하여 영업비밀을 주장하는 자로서는 그 특허출원된 내용 이외의 어떠한 정보가 영업비밀로 관리되고 있으며 어떤 면에서 경제성을 갖고 있는지를 구체적으로 특정하여 주장, 입증하여야 한다”²¹³⁾고 판시함으로써, 이를 뒷받침하고 있다.

나. 데이터 공개에 따른 인센티브 방안

데이터 공개에 대해서는 아직은 생소하기 때문에 데이터를 공개를 강제하거나 강행요건으로 하는 것은 지양할 필요가 있다. 일정한 요건에 따라 데이터 공개가 필요한 경우를 구체적으로 제시하는 것도 하나의 방안이 될 수 있을 것이다. 이러한 과정에서 인센티브를 부여할 수 있는 방안도 같이 고려된다면 발명자의 입장에서 부담없이 데이터를 공개할 수 있을 것이다.

211) 대법원 2004. 9. 23. 선고 2002다60610 판결.

212) 한국특허법학회편, 특허판례연구, 박영사, 2017, 39면.

213) 대법원 1998. 11. 10. 선고 98다45751 판결.

(1) 자발적 공개 방식

발명가에 의해 자발적으로 이루어지는 경우에는 인센티브를 부여하도록 하는 방안으로, 데이터 공개가 영업비밀을 공개하는 것이 아니라는 점을 명시적으로 제시할 필요가 있다. 자발적 공개의 경우에 가능한 인센티브를 확대할 필요가 있다.

(2) 강제적 공개 방식

쉽게 실시의 요건에 따른 공개로써, 실질적인 효과를 뒷받침할 수 있는 데이터셋의 공개를 유도하는 것으로, 강제적 공개라도 하더라도, 전면적인 공개는 지양될 필요가 있다. 아울러, 학습데이터가 필요한 발명을 유형화하여 심사지침의 개정도 이루어져야할 것이다. 예를 들면, 화학발명이나 미생물발명과 같이 쉽게 실시가 가능하도록 한 것처럼 일정한 요건의 경우에는 데이터를 공개하도록 하는 것이다.

다. 보편적인 인센티브

데이터를 공개하도록 제도화함으로써 얻을 수 있는 보편적인 인센티브와 제도적으로 설정할 수 있는 인센티브를 나누어 살펴볼 수 있다. 공개제도가 갖는 장점으로써 발명의 상세한 설명의 요건을 충족할 수 있다는 점이다. 무엇보다, 데이터가 포함된 발명에서 데이터가 공개되지 않음으로써 나타날 수 있는 충분한 설명인지에 대한 다툼을 해소할 수 있기 때문이다. 이로써, 심사과정이나 발명의 실시에서 예측가능성을 높일 수 있으며, 사업자에게는 법적안정성을 담보할 수 있을 것이다. 발명의 공개가 갖는 기본적인 혜택이라고 할 수 있는 기술공개로써 발명의 진흥이나 기술 및 산업발전을 이끌어내면서 얻을 수 있는 사회적 후생을 높일 수 있다는 점도 인센티브가 될 수 있다. 또한, 데이터공개를 통하여 중복투자를 방지하거나 또는 다양한 데이터를 확보할 수 있다면 다른 부문에 투자를 확대할 수 있다는 점도 인센티브가 될 수 있다.

라. 제공가능한 인센티브

앞에서 살펴본 보편적인 인센티브와는 별개로 산업정책적으로 공개자에게 부여할 수 있는 인센티브는 다음과 같다.

(1) 제공가능한 인센티브

우선심사의 확대, 출원 관련 비용지원, 발명의 유지비용의 지원, 특허기간의 연장 등을 고려할 수 있을 것이다.²¹⁴⁾ AI 기반 엔티티들은 그들의 값비싸고 가치있는 수집된 빅데이터를 공개하는

214) 자발적 공개를 시작으로 공개 인센티브를 확대하는 등 특허정책의 개정 통해 AI 공개 강화가 이뤄질 수 있다고 주장한다. 이러한 인센티브의 범위는 (1)우선심사, (2)관납료 인하, (3)더 긴 특허보호를 가능하게 하는 조건, (4)완전한 특허 출원의 전제 조건으로서의 작업모델의 요구에 이르기까지 다양하다. Tabrez Ebrahim, Artificial

대가로 특허를 얻고 싶어하지 않는다. 그러나 이러한 결함은 AI 특허와 관련하여 높아진 공개 기준의 균형을 맞추기 위한 보완적 인센티브로서 정부 기금, 세액 공제, 판매세 감면, 보조금 및 보충금과 같은 다른 인센티브로 해결될 수 있다.²¹⁵⁾ 심사과정에서 데이터가 공개된 경우에는 기재 요건 등을 갖추었다고 판단할 수 있도록 한다. 또한, 발명가에게는 다음과 같은 안전장치가 있다는 점을 제시할 필요가 있다.

[표 3-4] 인센티브 현황

제도적 인센티브
<ul style="list-style-type: none"> • 공개시에 해당 데이터를 안전하게 보관 • 기술적, 관리적, 물리적인 보안시스템을 구축하도록 하는 방안 • 전문성이 있는 제3의 기관을 지정하는 방안 • 영구적인 보관이 가능하도록 하는 방안 • 영업비밀로 보호되는 데이터도 보호가능
경제적 인센티브
<ul style="list-style-type: none"> • 우선심사의 확대 • 관납료 등 유지비용의 지원 • 특허기간의 연장 • 보조금 • 세액 공제

(2) 인센티브 제공시 예상되는 비판

기술공개에 따른 독점권 부여를 넘어선 보호기간 등의 연장을 추가로 하는 것에 대한 비판도 가능하나, 행정절차 등에 따라 특허를 행사할 수 없었던 경우에는 연장하는 제도가 있기 때문에 이와 유사하게 설정할 수 있다.

4. 공개시의 기대효과

학습데이터의 공개시 기대할 수 있는 효과는 기술공개에 따른 효과에 더하여, AI발명이 갖는 문제점을 극복할 수 있을 것으로 기대한다. 다만, 기술공개가 완전하다거나 그 기준이 완비된 것은 아니라는 점에서 위에서 살펴보았던 공개요건 등을 보다 구체화함으로써 기대효과를 높일 수 있을 것이다. 물론, 이를 반대하는 입장에서 제시하는 부정적인 면에 대해서도 인센티브를 확대

Intelligence Inventions & Patent Disclosure, Iowa Legal Studies Research Paper No. 2021-48, November 2021, p. 162.

215) Medhi Poursoltani, Disclosing AI Inventions, TEXAS INTELLECTUAL PROPERTY LAW JOURNAL Vol.29, 2021, p. 63.; Ted M. Sichelman, Patents, Prizes, and Property, 30 Harv. J.L. & Tech. 279, 280-81 (2017).

하는 것과 데이터의 공개를 강제하는 것이 아닌 자발적으로 이루어낼 수 있도록 하는 것이 필요하다.

가. 긍정적 효과

데이터 공개에 따른 긍정적인 효과는 무엇보다도, 기술공개제도가 추구하는 발명의 확산을 기할 수 있다는 점이다. 즉, 데이터 공개에 따른 쉽게 실시가 가능해지기 때문에 기술공개에 대한 논란을 줄일 수 있다. 이에 더하여, 기재불비 내지 미완성 발명에 대한 리스크를 감소할 수 있을 것이다. 아울러, 데이터가 공개됨으로써 해당 발명에 대한 재현가능성을 높일 수 있다는 점을 긍정적인 효과로 정리할 수 있다. 또한, 공개된 데이터를 제3자가 활용할 수 있도록 한다면 다른 기계학습을 위한 데이터셋으로 산업적 활용이 가능해질 수 있다. 다만, 이는 요건을 완전하게 공개할 것인지에 따라 다를 수 있다.

또한, 일부 인공지능 발명가들은 인공지능 연구개발 분야에 기여하거나 해당 분야의 최첨단 기술을 발전시키기 위해 자신들의 발명에 대한 데이터나 다른 정보를 공개하는 것을 선택할 수 있다. 이러한 경우에, 발명자는 자신의 데이터나 그 밖의 정보를 다른 사람들이 이용할 수 있도록 하는 데 더 적극적일 수 있다.²¹⁶⁾

나. 부정적 효과

긍정적인 효과와 달리, 부정적인 결과가 나올 가능성도 배제하기 어렵다. 가장 큰 것은 발명자가 데이터를 영업비밀로 유지하려고 할 경우이다. 이러한 경우에는 특허출원이 줄어들 가능성도 예상된다. 물론, 앞에서 검토한 바와 같이 발명을 통해 공개하더라도 영업비밀제도로써도 학습데이터를 보호할 수 있는 방안이 충분히 있다는 점을 통해서 이러한 우려를 불식시킬 수 있을 것이다.

행정적인 면에서의 부정적인 효과는 데이터 공개에 따른 출원절차가 늦어질 가능성에 관한 것이다. 데이터에 대한 형식적인 심사 내지 실질적인 심사가 이루어질 경우에는 그만큼 심사기간이 걸릴 수 있다는 점에서 발명자나 심사관의 입장에서 부담이 될 수 있다. 이에 따라, 특허정책적인 측면에서 인센티브를 부여하거나 형식적인 요건을 완비할 경우에는 별도 심사를 거치지 않도록 하는 것도 방안이 될 수 있다.

또한, 발명자는 학습데이터의 공개에 따른 시간과 비용에 대한 추가적인 부담이 발생할 수 있다. 기존의 출원절차에 없던 항목이 추가된 것이라는 점에서 부정적인 효과를 가져올 가능성도 있다. 다만, 화학발명에서 약리데이터가 없을 경우에는 미완성 발명으로 볼 수 있다는 점이나 이미 미생물발명에서도 기탁제도가 이루어지고 있다는 점에서 설득할 수 있을 것이다.

216) Hagen, Gregory R., AI and Patents and Trade Secrets (February 1, 2021). in Florian Martin-Bariteau & Teresa Scassa, eds., Artificial Intelligence and the Law in Canada (Toronto: LexisNexis Canada, 2021), Available at SSRN:<https://ssrn.com/abstract=3734654>

제6절. 특허제도 개선의 필요성 검토

1. 데이터 공개의 필요성에 대한 의견 수렴

가. 개요

학습데이터를 공개하는 것에 대해 다음과 같이 실무자나 전문가의 의견을 수렴하였다. SW관련 사업자 단체나 관련 사업자, 데이터셋에 대한 다양한 투자를 하고있는 포털사업자와 리걸테크(legal tech) 기업 이나 통신사업자 등 플랫폼사업자의 관계자, 제조기업의 관계자, 변리사, 법학 교수 및 국책연구원의 원구원을 대상으로 데이터 공개에 대한 의견을 수렴하였다.²¹⁷⁾

나. 의견 요청 내용

데이터 공개에 대해서는 데이터셋의 성격에 따라 다양한 의견이 제시될 수 있을 것으로 보았으며, 실제 데이터를 기반으로 하는 사업자의 입장과 전통적인 제조업자들의 차이가 있음을 확인하였다. 질의 항목은 다음과 같았다.

- 데이터셋을 공개하는 것에 대한 의견
- 데이터를 공개하는 경우라면 어떤 문제가 예상되는지에 대한 의견
- 데이터(구조) 발명에 대한 인정 여부에 대한 의견

다. 의견 내용

(1) 부정적인 의견

AI를 서비스의 영역에서 구체적으로 다루는 기업에서는 데이터셋을 공개하는 것에 대해 부정적인 의견을 보였다. 제도화에 대해서는 정부 정책일 경우에는 따라야 한다는 입장이었으나, 그러할 경우에는 데이터의 공개가 오히려 발명의 출원 경향성을 낮출 수도 있다는 의견이었다. 대체적인 의견은 다음과 같다.

- 데이터의 보호범위와 데이터 공개에 따른 이익형량이 필요할 것으로 보임
- 데이터의 성질에 따른 대응이 달라질 수 있을 것이나, 개인정보, 저작권, 영업비밀, 기밀정보, 금지되는 데이터가 활용되는 경우에는 공개는 어려울 수밖에 없음

217) 다양한 전문가나 실무자의 의견을 듣는 것은 회사나 소속된 곳의 공식적인 의견이 아닌 개인적인 의견을 제시한 것이라는 점에서 실명이나 회사명을 공개하는 것에 부정적이었기 때문에 가명화하여 내용을 정리하였음을 밝히며, 의견내용을 임의로 수정하거나 변경하지 않았음을 밝힌다.

- 데이터의 공개는 데이터셋이라는 회사의 핵심적인 자산이 경쟁사에게 넘어갈 수 있다는 우려가 가장 크다고 할 것임

(2) 긍정적인 의견

데이터를 활용하는 경우가 적은 회사의 경우라면, 공개의 필요성을 강조하였다. 정확히는 제3자적 입장이라고 볼 수 있으며, 데이터 공개에 대한 가이드라인을 제시할 경우에는 발명자는 예측가능성을 높일 수 있기 때문에 긍정적이라는 의견을 제시하였다.

- 구체적인 예측가능성과 투자에 따른 법적 안정성을 높일 수 있음
- 실질적인 데이터의 보호 가능성을 높일 수 있을 것임
- 데이터의 가치를 인정받을 수 있다는 점에서 투자유인이 될 수 있음
- 미생물발명과 같이 기탁제도는 의미있는 제도로 활용될 수 있을 것임

(3) 특허제도에 대한 논거

현행 특허제도하에서 데이터 공개에 대해서는 견해 차이가 있다. 다만, 데이터 공개에 대해서 그동안 고민해본 사항은 아니기 때문에 이에 대한 구체적인 의견을 제시한 것은 아니란 점을 밝힌다.

적극적인 해석론의 입장에서, 실무적으로는 데이터 자체에 대한 특허가 인정되고 있으므로, 데이터를 물건으로 충분히 볼 수 있다는 견해이다. 특허법 제94조²¹⁸⁾에 따라 특허발명의 실시 규정을 통해서도 데이터 자체의 보호가능성도 있다. 즉, 특허권의 효력규정에 따라 데이터가 물건의 발명으로 인정될 경우에는 청약하는 경우를 실시로 볼 수 있다는 것이다. 참고로, 특허발명의 실시가 제2조 제3호 나목에 따른 방법의 사용을 청약하는 행위인 경우에는 특허권의 효력은 그 방법의 사용이 특허권 또는 전용실시권을 침해한다는 것을 알면서 그 방법의 사용을 청약하는 행위에만 미친다.

소극적인 입장에서는 현행 특허법상 데이터는 자연법칙을 이용한 사상이 아니라는 점에서 보호범위에 포함되지 않는다는 입장이다. 따라서, 특허법의 개정을 통하여 데이터발명에 대한 구체적인 사항을 규정할 필요가 있으며, 이 과정에서 데이터 공개를 구체화하는 방안이 필요하다는 것이다. 공개에 대해서는 조금 더 고려해볼 사안이라는 점에서 소극적인 입장이라고 할 것이다. 또한, 데이터공개에 대해서는 영업비밀 원본증명제도와 같은 성격을 지닐 수 있다는 의견을 주기

218) 제94조(특허권의 효력) ① 특허권자는 업으로서 특허발명을 실시할 권리를 독점한다. 다만, 그 특허권에 관하여 전용실시권을 설정하였을 때에는 제100조 제2항에 따라 전용실시권자가 그 특허발명을 실시할 권리를 독점하는 범위에서는 그러하지 아니하다. <개정 2019. 12. 10.>

② 특허발명의 실시가 제2조 제3호 나목에 따른 방법의 사용을 청약하는 행위인 경우 특허권의 효력은 그 방법의 사용이 특허권 또는 전용실시권을 침해한다는 것을 알면서 그 방법의 사용을 청약하는 행위에만 미친다. <신설 2019. 12. 10.>

도 했다.

라. 정리

자문의견을 통하여 데이터공개에 대해서는 기술공개라는 점에서 긍정하지만, 데이터공개가 자사의 경쟁우위에 해당하는 경우에는 공개가 가져올 수 있는 제도적인 우려에 대한 의견을 주었다. 데이터가 갖는 성격에 따른 것으로 볼 수 있으며, 데이터의 전처리에도 불구하고 현행법에서 금지하고 있는 다양한 이슈를 담고있기 때문인 것으로 판단된다. 따라서, 데이터가 갖는 속성이나 문제점들에 대한 가이드라인을 제시할 필요가 있다고 본다. 또한, 영업비밀로 보호되는 데이터를 공개하는 것은 공개제도가 갖는 입법목적까지도 오해할 수 있는 부분이라는 점에서 앞서 인센티브로서 살펴본 바와 같이, 특허제도와 공존할 수 있다는 점도 제시하는 것이 필요하다고 보았다.

2. 심사지침 개정에 대한 검토

가. 현재 지침의 한계

데이터 구조, 구조화된 데이터의 발명에 대해서는 명확한 가이드를 제공하고 있다. 다만, 데이터 자체에 대해서는 발명의 성립가능성이 부재한 것으로 판단할 수 있다. 현행 심사지침은 SW나 AI발명이나 데이터 구조 발명은 알고리즘성이 인정된다는 점에서 발명의 성립성이 부정되지는 않고 있다. 그렇지만, SW를 포함하여 데이터 관련 발명은 물건발명으로 인정되고 있다. 자연법칙을 이용하는 알고리즘의 성격에 따른 것이다.

불명확하다는 점이 지적되지만 사실상 데이터의 구조화된 집합체인 데이터셋의 보호가능성도 충분하다. 이미 데이터셋을 청구항으로 하는 발명이 등록되고있다는 점에서 사실상 데이터 자체의 보호가능성이 부정된다고 보기는 어렵다. 다만, 심사지침에서 명시적으로 가이드가 제시되지 않고있을 뿐이다.

나. 개정 방향

일반적으로 기계장치 등에 관한 발명에 있어서는 특허출원의 명세서에 실시예가 기재되지 않더라도 평균적 기술자가 발명의 구성으로부터 그 작용과 효과를 명확하게 이해하고 용이하게 재현할 수 있는 경우가 많으나, 이와는 달리 이른바 실험의 과학이라고 하는 화학발명의 경우에는 당해 발명의 내용과 기술수준에 따라 차이가 있을 수는 있지만 예측가능성 내지 실현가능성이 현저히 부족하여 실험데이터가 제시된 실험예가 기재되지 않으면 평균적 기술자가 그 발명의 효과를 명확하게 이해하고 용이하게 재현할 수 있다고 보기 어려워 완성된 발명으로 보기 어려운 경우가 많고, 특히 약리효과의 기재가 요구되는 의약의 용도발명에 있어서는 그 출원 전에 명세

서 기재의 약리효과를 나타내는 약리기전이 명확히 밝혀진 경우와 같은 특별한 사정이 있지 않은 이상 특정 물질에 그와 같은 약리효과가 있다는 것을 약리데이터 등이 나타난 시험예로 기재하거나 또는 이에 대신할 수 있을 정도로 구체적으로 기재하여야만 비로소 발명이 완성되었다고 볼 수 있는 동시에 명세서의 기재요건을 충족하였다고 볼 수 있을 것이며, 이와 같이 시험예의 기재가 필요함에도 불구하고 최초 명세서에 그 기재가 없던 것을 추후 보정에 의하여 보완하는 것은 명세서에 기재된 사항의 범위를 벗어난 것으로서 명세서의 요지를 변경한 것이다.²¹⁹⁾

따라서, AI발명이나 데이터 발명에 있어서는 발명의 내용과 기술수준에 따라 차이가 있을 수는 있지만 예측가능성 내지 실현가능성이 현저히 부족하여 데이터가 제시된 실험예가 기재되지 않으면 평균적 기술자가 그 발명의 효과를 명확하게 이해하고 용이하게 재현할 수 있다고 보기 어려운 경우라고 할 것이다. 이러한 경우라면, 심사지침에 데이터셋이 알고리즘성을 갖는 경우로 한정하여 데이터발명으로 기재할 수 있도록 한다면 발명자 입장에서 법적 안정성과 예측가능성을 높일 수 있을 것이다. 또한, 데이터발명의 경우에는 데이터셋을 공개하도록 하는 내용도 포함될 필요가 있다. 데이터셋의 공개방법, 공개범위 및 수준 등에 대해서는 앞서 살펴본 바를 참조할 수 있을 것이다.

또한, 데이터를 기재하지 않아도 이를 대신할 수 있는 정도라면 학습데이터셋을 대신할 수 있을 만한 AI모델에 입력한 데이터의 속성과 전처리 및 효과 등을 기재하도록 해야할 것이다.

3. 특허법제의 개정 사항

가. 현재 특허법의 한계

특허법은 SW발명의 청약에 대해서는 보호범위에 포함하고 있다. 다만, 명시적으로 SW발명이나 온라인 실시를 규정하고 있는 것은 아니지만 심사지침을 통하여 사실상 SW발명이 인정되고 있다. 더욱이, 실시에 대해서는 명확한 것은 아니지만 제94조에 있어서 해석 가능성이 존재한다. 이미 심사지침을 통하여 물건발명이 인정되고 있다는 점에서도 데이터 발명의 온라인 실시에 대해서 보호되지 못한다고 볼 수 있을지 의문이다.

또한, SW의 보호가 매체에 따라 보호범위가 달라지는 것이 합리적인 것인지는 의문이다. SW는 그 자체로서 독립된 가치재이다. 그렇다면 매체에 저장되거나 그렇지 않거나 상관이 있다고 보기 어렵다. 둘을 구분하여 다루어야할 실익이 있다고도 보기 어렵다. 민법상 물건성 논의에서 유체물성과 무체물성을 기준으로 하나, 전통적인 민법 체계에서는 정보나 데이터라는 개념을 입법과정에서 고려한 적이 없기 때문에 당연하다. 법이 보수적이며, 특히 사인간의 거래관계를 명확히 하는 민법은 더욱 그러하다. 그렇지만, 민법의 특별법적 관계에 있는 다양한 법률에서 SW나 정보 등에 대해 규율하고 있으며, EU의 제조물책임 지침 개정 논의에서 전통적인 제조물의

219) 대법원 2001. 11. 30. 선고 2001후65 판결 [보정각해(특)].

개념을 넘어서,²²⁰⁾ 제조물의 범위에 SW나 디지털 제품에 대해서도 포함하고 있다는 점은 시사하는 바가 크다.²²¹⁾ 즉, SW의 제조물성을 논의해야할 실익이 없다는 점이다. SW발명에 있어서도 매체성에 한정된 것이 SW성격에 부합한 것인지 의문이다. 특히, SW의 거래구조는 패키지 형태의 사용이 아닌 앱방식이나 SaaS나 클라우드 방식으로 전환되고 있다는 점에서 매체성을 고집할 필요가 있는 것인지는 의문이다.

나. 데이터 발명의 간접침해 적용 가능성

일본은 데이터가 청구항에 기재되어 등록된 특허는 우리나라와 같이 데이터의 생성·수집 방법특허 및 그 물건(시스템, 장치), 물건발명으로 “데이터구조” 또는 “데이터 매체 청구항”을 인정하고 있다. 다만 우리나라와 달리 “데이터 매체 청구항” 이외에도 “구조를 가진 데이터”를 청구항의 말미에 기재하면, 그 데이터는 발명해당성이 있는 것으로 간주하여 특허대상에 포함시키고 있다. 이러한 청구항의 구성요소 일부에 해당하는 데이터만을 실시하는 경우, 데이터를 물건으로 간주하여 일본특허법 제107조에서 규정하는 간접침해 규정을 적용할 수 있는 것인지에 대해서는 쟁점이 되고 있다. 즉 특허발명의 구성요소 일부인 데이터(「특정 구조를 가진 데이터」)만을 사용하거나 다른 사람에게 네트워크를 통하여 전송 또는 제공되는 경우, 그 행위가 일본특허법상 간접침해가 성립하는 것인지가 쟁점이 되고 있다. 이러한 구성요소로서 데이터의 사용행위 또는 전송행위에 대하여 간접침해가 성립하기 위해서는 데이터의 물건성이 인정되어야 하는데 현행 일본특허법상 데이터를 물건으로 간주할 수 없기 때문에 야기되는 문제로 인식되고 있다. 따라서 AI와 관련한 데이터산업의 촉진을 유도하기 위하여 데이터를 물건으로 간주하여 데이터의 실효적 특허보호를 강화하려는 움직임이 진행되고 있다.²²²⁾

우리나라는 데이터 발명이 간접침해를 구성하는 물건발명으로 볼 수 없다는 견해에 따라 현행 특허법에서는 어렵다는 견해가 다수적이다. 그렇지만, 이러한 견해는 데이터 구조를 데이터발명으로 인정하고 있으며 이는 매체에 저장되는 형식의 물건발명이라는 점을 간과하고 있다. 따라서 데이터 발명은 물건발명으로 구성한 경우에는 제127조의 적용가능성을 배제하기는 어렵다고 판단된다.

다. 데이터공개에 대한 특허법 개정 방향

현행 특허법의 해석 방식에서는 SW특허의 온라인 실시 등에 대해서는 크게 이슈가 되지 않을 것으로 보인다. 물론, 데이터 발명에 대해서 특허법을 개정하여 명확히 하자는 주장도 있는 것이 사실이다. 별도의 데이터 기탁 요건을 가진 특허보호를 위한 법령을 제정하자는 주장도 있

220) EU제조물지침 개정안에서는 제조물의 개념에 동산을 포함하여 SW, 디지털제조파일이 명시적으로 규정되어있다. 또한, AI는 SW로 간주되기 때문에 AI도 제조물책임법상 제조물에 해당하여 그 책임을 지게된다.

221) SW에 대한 제조물책임 논의에 대해서는 권오승 외, 제조물책임법, 법문사, 2002 참조.

222) 권지현, AI발명과 데이터의 실효적 특허보호방안, 法學研究 第24輯 第1號, 2020, 12~13면.

다.²²³⁾ 그 주장도 타당하다고 보나 현실적으로 현행 법체계 내에서 적극적으로 해석하거나 가이드를 제시할 수 있다고 판단된다. 특허법 개정을 통하여 방안을 제시하는 것도 의미있다고 판단되나 SW발명에 대한 관계부처의 다툼은 지난한 것이라는 점에서 현행 특허법의 적극적인 해석을 통해서 SW를 포함한 AI, 데이터 발명에 대한 보호범위를 명확히 하는 것이 우선이라고 판단된다.

참고로, 미생물발명의 경우에는 기탁서류를 첨부하도록 되어있다. 이러한 경우는 기탁사실을 기재한 서류를 첨부함으로써, 공개요건을 갖춘 것으로 보고 있다. 미생물발명의 기탁에 대해서는 특허법 시행령에서 규정하고 있는 사항으로, 서면주의를 취하고 있는 우리 특허제도의 변형적인 방법이라고 하겠다. 미생물 자체를 서류에 포함시킬 수 없기 때문에 미생물의 기탁과 관련된 사실관계를 기재한 서류를 제출토록 하고있는 것이다. 데이터의 경우도 유사하게 고려할 수 있을 것이다. 이러한 경우에는 해당 사이트에서 다운로드 받은 데이터의 유형이나 URL 등을 기재하는 방식으로 명세서에 나타내는 것이면 족할 것으로 판단된다. 구체적인 기재요건은 위에서 살펴본 바와 같다. 이와 같이, 기재하거나 공개하는 내용을 특허명세서에 첨부함으로써 가름할 수 있을 것이다. 특허법의 개정사항이라기 보다는, 특허법 시행령을 통하여 데이터를 공개하는 방식이 입법정책적으로 타당하다고 판단된다.

이처럼, 신경망은 생물학적 미생물 등으로 모델링되기 때문에 신경망에 기반한 인공지능 발명이 궁극적으로 생물학적 물질을 통하여 재현되는 것과 유사하게 학습데이터를 기탁함으로써 미생물 특허와 유사한 발명의 특성을 갖는다는 점에서 데이터 기탁을 위한 제도설계에서 미생물 기탁제도는 중요한 가이드가 될 수 있을 것이다.

223) Tabrez Ebrahim, Artificial Intelligence Inventions & Patent Disclosure, Iowa Legal Studies Research Paper No. 2021-48, p. 215.

제4장. AI 학습데이터의 보호·활용을 위한 제도개선 기초 연구

제1절. 논의의 필요성

1. 발명의 투명성 확보

가. AI발명의 불투명성에 따른 문제제기

AI발명에 있어서 중요한 것은 데이터이지만, 인공지능이 데이터를 처리하는 과정의 불투명성은 알고리즘의 문제라는 점과 더불어 데이터셋이 공개되지 않는다는 점 때문이기도 하다. 데이터에 포함된 다양한 개인정보의 처리를 위하여 거치는 데이터의 가명화 또는 익명화 과정에서 외부의 제3자는 어떻게 처리되어 학습데이터로 가공되는지 알 수 없다. 딥러닝 알고리즘이 어떻게 데이터를 처리하는지도 명확하지 않다. 실제, 개인정보 수집과정에서 동의받는 목적에 부합하게 데이터가 처리되는지 알 수 없다. 데이터 양의 방대함²²⁴⁾, 데이터 처리의 불명확성 및 데이터 학습 알고리즘의 불명확성으로 인하여, 개발자는 물론 이용자도 그 내용을 알 수 없는 블랙박스 현상이 발생한다. 수집하거나 제공되는 데이터를 분석하여, 결과물을 만들어냈지만 데이터와 그 결과의 상관관계를 확인할 수 있을 뿐, 인과관계를 확인하기 어렵다. 법적으로 원인에 따른 결과를 인과관계라고 하지만, 이를 인공지능은 명확하게 설명하기 어렵다는 문제가 있다.²²⁵⁾ 이러한 기계학습 전반의 불투명성은 AI발명은 물론 AI 자체의 신뢰성을 의심받게되는 것이다. 이러한 신뢰성을 얻고 AI발명의 투명성을 확보하기 위한 방안으로써 데이터를 공개하는 방안을 고려할 수 있을 것이다.

나. 데이터를 통한 투명성 확보

데이터 관련 발명을 실시하기 위해서는 구성요건이 완비되어야 한다. 이렇게 완비된 후에야 용이하게 실시할 수 있을 것이다. 따라서 발명의 결과에 대해서 신뢰하기 위해서는 발명 과정에서 소요된 데이터 등이 왜곡되거나 확보할 수 없는 방법으로 이루어진 것인지는 확인이 필요하다.

공서양속에 위배되거나,²²⁶⁾ 데이터 자체의 원본성이나 저작권 등의 권리침해가 이루어지지 않았다는 점을 증명하기는 어렵다. 데이터의 경우도 다르지 않다. 물론, 저작권을 침해하는 데이터를 바탕으로 발명이 이루어졌다고 하더라도, 발명의 성립성이 부정될 수는 없다. 다만, 해당 데이

224) 빅데이터라는 표현은 데이터를 통계적으로 표본을 샘플링해서 분석하는 것이 아니라, 모든 데이터(n=all)를 사용하여 분석할 수 있는 기술적인 배경이 자리하고 있다.

225) 김윤명, 블랙박스를 열기위한 인공지능법, 박영사, 2022, 466면.

226) 제32조(특허를 받을 수 없는 발명) 공공의 질서 또는 선량한 풍속에 어긋나거나 공중의 위생을 해칠 우려가 있는 발명에 대해서는 제29조제1항에도 불구하고 특허를 받을 수 없다.

터를 확보할 수 없는 경우에도 이를 확보할 수 있는 것으로 기재하여 발명이 성립된 경우는 무효화될 수 있을 것이다.

신뢰성 확보를 위한 데이터의 확보나 제3자가 이를 확인해줄 수 있는 방법을 고려할 필요가 있는 이유이다. 특히 화학발명의 경우에는 데이터가 필요한 경우가 있다.²²⁷⁾ 실제 복잡한 구조에 대한 이해를 위해서는 실험 데이터가 필요하기 때문이다. 데이터 발명의 데이터와 화학발명이나 의약의 용도발명에서 약리데이터는 각각의 성격의 차이가 있으나, 작용에 있어서 데이터의 역할을 발명의 신뢰성을 검증할 수 있다는 점에서 의미가 있다. 따라서, 데이터의 기탁 등의 제도를 고려할 실익은 있다고 판단된다.

2. 발명의 반복재현성 확보

가. 실시가능성으로서 재현

발명의 공개는 쉽게 실시할 수 있도록 함으로써 반복재현성²²⁸⁾을 위한 것이다. 기본적으로 재현이란, 물건발명에서는 그 물건을 생산, 사용 등을 할 수 있고, 방법발명에서는 그 방법을 사용할 수 있고, 물건을 생산하는 방법의 발명에서는 그 방법에 의하여 그 물건을 생산할 수 있을 정도를 말한다.²²⁹⁾ 따라서, 반복재현이란 특허법상 실시와 다름이 없다. 실질적으로 실시권을 하여 받아 해당 발명을 실시한다는 것은 반복재현을 한다는 것과 다르지 않기 때문이다. 실시권을 하여받은 경우라면 영업비밀이나 노하우까지도 이전하기 때문에 재현가능성을 높일 수 있겠지만, 사실상 특허권을 부여하면서 기술공개를 요건으로 하는 입법취지에 비추어볼 때, 실시가능성과 실시권의 허여를 통해 재현하는 것이 달라지는 것이라면 문제라고 하겠다.

특허법은 실시요건으로서 명세서기재에 대해 “그 발명이 속하는 기술분야에서 통상의 지식을 가진 사람이 그 발명을 쉽게 실시할 수 있도록 명확하고 상세하게 적을 것”을 요구하고 있다. 경우에 따라서는 실시례를 기재할 수도 있을 것이다. 발명의 성격이나 기술내용 등에 따라서는 명세서에 실시례가 기재되어 있지 않다고 하더라도 통상의 기술자가 발명을 정확하게 이해하고

227) 특허출원서에 첨부하는 명세서에 기재될 '발명의 상세한 설명'에는 그 발명이 속하는 기술분야에서 통상의 지식을 가진 자가 당해 발명을 명세서 기재에 의하여 출원시의 기술 수준으로 보아 특수한 지식을 부가하지 않고서도 정확하게 이해할 수 있고 동시에 재현할 수 있도록 그 목적·구성·작용 및 효과를 기재하여야 하고, 특히 약리효과와 기제가 요구되는 의약의 용도발명에 있어서는 그 출원 전에 명세서 기재의 약리효과를 나타내는 약리기전이 명확히 밝혀진 경우와 같은 특별한 사정이 있지 않은 이상 특정 물질에 그와 같은 약리효과가 있다는 것을 약리데이터 등이 나타난 시험 예로 기재하거나 또는 이에 대신할 수 있을 정도로 구체적으로 기재하여야만 비로소 발명이 완성되었다고 볼 수 있는 동시에 명세서의 기재요건을 충족하였다고 볼 수 있다. 대법원 2004. 12. 23. 선고 판결 [거절결정(특)].

228) 발명의 목적을 달성하기 위한 수단이 형식적으로 제시되어 있으나 그 제시한 수단에 의하여 발명자가 얻은 성과와 객관적으로 동일한 결과를 얻을 수 없는 경우, 즉 반복하여 실시할 수 없는 것은 발명에 해당될 수 없다. 여기에서 출원발명의 반복재현성은 반드시 100%의 확률로 효과를 얻을 수 있는 것만을 의미하는 것이 아니고, 100%보다 적은 확률이라도 효과를 얻을 수 있는 것이 확실하다면 반복재현성이 있다고 본다. 특허청, 특허·실용신안 심사기준, 2021, 3106면.

229) 윤선희, 특허법, 법문사, 2019, 382면.

재현하는 것이 용이한 경우도 있으므로 구 특허법 제42조 제3항이 정한 명세서 기재요건을 충족하기 위해서 항상 실시례가 기재되어야만 하는 것은 아니다. 또한 구성요소의 범위를 수치로써 한정하여 표현한 발명의 경우, 그러한 수치한정이 단순히 발명의 적당한 실시 범위나 형태 등을 제시하기 위한 것으로서 그 자체에 별다른 기술적 특징이 없어 통상의 기술자가 적절히 선택하여 실시할 수 있는 정도의 단순한 수치한정에 불과하다면, 그러한 수치한정에 대한 이유나 효과의 기재가 없어도 통상의 기술자로서는 과도한 실험이나 특수한 지식의 부가 없이 그 의미를 정확하게 이해하고 이를 재현할 수 있을 것이므로, 이런 경우에는 명세서에 수치한정의 이유나 효과가 기재되어 있지 않더라도 구 특허법 제42조 제3항에 위배된다고 할 수 없다.²³⁰⁾

무엇보다, 특허법은 특허청구범위에 보호받고자 하는 사항을 기재한 항(청구항)은 발명의 상세한 설명에 의하여 뒷받침될 것을 규정하고 있는데, 그 취지는 특허출원서에 첨부된 명세서의 발명의 상세한 설명에 기재되지 아니한 사항이 청구항에 기재됨으로써 출원자가 공개하지 아니한 발명에 대하여 특허권이 부여되는 부당한 결과를 막기 위한 것으로서, 청구항이 발명의 상세한 설명에 의하여 뒷받침되고 있는지는 특허출원 당시의 기술수준을 기준으로 하여 통상의 기술자의 입장에서 특허청구범위에 기재된 사항과 대응되는 사항이 발명의 상세한 설명에 기재되어 있는지에 의하여 판단하여야 한다.²³¹⁾

나. 재현가능성 확보와 데이터의 관계

SW발명의 경우에는 데이터가 필요로 하지 않는 것으로 코딩을 통한 알고리즘단에서 의도하는 발명의 목적을 달성할 수 있었다. 그렇지만, 기계학습 기반의 인공지능은 데이터 학습을 통하여 발명이 완성된다는 점에서 데이터가 중요한 구성요소가 되고 있다. 이러한 점에서 발명의 재현가능성을 확보하는 것은 방법의 용이한 실시를 위해서 필수적인 요건이 될 수 있다. 반면, AI발명에서 고려할 사항은 일반적인 SW발명과는 차이가 있다. 발명을 실시로서 재현한다는 것의 의미는 명세서에 담긴 기술적 사상으로서 발명이 그대로 완벽하게 재현된다는 것을 의미하는 것인지 여부이다. 기존에는 재현한다는 것은 동일 발명으로서 의미를 가진 것이었다면 이제는 동일 발명으로 보기에 어려움이 발생한다. 왜냐하면, 인공지능 발명은 데이터가 달라지면 발명의 결과도 달라질 수 있기 때문이다. 용이하게 실시할 수는 있겠지만, 데이터가 달라질 경우에는 AI모델에서 처리하는 값이 달라질 수도 있기 때문이다.

물건의 발명의 경우 그 발명의 실시라고 함은 그 물건을 생산, 사용하는 등의 행위를 말하므로, 그 발명의 특허청구범위에 특정된 물건 전체의 생산, 사용 등에 관하여 위와 같은 정도의 명세서 기재가 없는 경우에는 위 조항에서 정한 기재요건을 충족한다고 볼 수 없다. 따라서 구성요소의 범위를 수치로써 한정하여 표현한 물건의 발명에서도 그 특허청구범위에 한정된 수치범위

230) 대법원 2011. 10. 13. 선고 2010후2582 판결

231) 대법원 2011. 10. 13. 선고 2010후2582 판결

전체를 보여주는 실시 예까지 요구되는 것은 아니지만, 통상의 기술자가 출원 시의 기술 수준으로 보아 과도한 실험이나 특수한 지식을 부가하지 않고서는 명세서의 기재만으로 위 수치범위 전체에 걸쳐 그 물건을 생산하거나 사용할 수 없는 경우에는, 위 조항에서 정한 기재요건을 충족하지 못한다고 보아야 한다. 데이터 발명을 물건의 방법으로 본다는 점에서, 실시하여 재현할 수 있는 수준이 아니라면 기재불비나 미완성발명으로 볼 수 있을 것이다. 또한, 출원시의 발명과 등록시의 발명이 달라질 수 있다. 이러한 발명의 재현은 동일한 발명이어야 하는가? 그렇다면, 발명은 평균적 기술자에 의하여 용이하게 실시될 수 있어야 한다는 점에서 구성요소인 데이터가 요구된다는 점은 부인하기 어렵다. 물론, 등록시의 발명은 데이터의 차이로 인한 것이기 때문에 동일할 필요는 없다. 다만, AI발명에 있어서 재현가능성을 높이기 위한 방안으로써 미생물 기탁 제도를 고려할 수 있을 것이다. 제3의 기관에 데이터를 기탁하고, 해당 데이터를 통하여 재현가능성을 확인할 수 있도록 하자는 것이다.

제2절. 데이터 관련 제도 방안 검토

신뢰성 확보와 재현가능성을 담보할 수 있는 방안으로써 여러 가지 기술적, 관리적 방안에 대해 살펴보고자 한다. 발명에 부속되어 사용되었다는 사실관계의 인증으로 활용할 수 있을 것이다.

1. 실무적인 검토사항

가. AI 허브(데이터 레이크)

우리나라에서 정책적으로 데이터셋을 구축하고 있는 운영하는 공개 데이터셋이다. 현재, 9억 건이 구축되어있으며, 그 이용의 제한은 없다. 누구라도 AI기술 개발을 위해서 사용할 수 있는 무상의 데이터집합이다. 이용은 AI허브 사이트에 회원가입 절차를 거치고나서 필요로 하는 데이터셋을 다운로드 받아서 활용하면 된다. 다만, 외국인의 사용은 제한적이라고 하겠다.

기본적으로 정책적인 데이터셋을 구축하여 운영하는 것을 목적으로 하기 때문에 영속적이거나, 외부의 데이터를 수용할 수 있는 체계는 갖추지 못하고 있다. 따라서, 학습데이터를 기탁할 수 있는 시스템으로써 활용하기에는 부적합한 면이 있다는 점에서 실무적인 활용가능성은 높지 않다고 판단된다.

나. 깃허브 데이터 공개

깃허브는 MS에서 운용하고 있는 공개 사이트로, 오픈소스 방식으로 프로그램 소스코드와 필요한 데이터셋을 공개하는 방식으로 이루어진다. 해당 사이트에 가입하여 누구라도 이용할 수 있다. 다만, 깃허브는 오픈소스 방식을 따르기 때문에 공개된 방식에 따라 다양한 오픈소스 라이선

스 정책을 따라야 한다. 만약, 해당 정책을 따르지 않을 경우에는 라이선스를 받을 수 없기 때문에 계약위반에 따른 책임과 특허와 관련된 경우에는 특허권 침해에 따른 리스크가 예상된다.

깃허브는 오픈소스를 바탕으로 이루어지기 때문에 누구나 자유롭게 이용할 수 있다는 장점이 있으나 운영주체가 MS라는 점, 서버의 소재가 국외에 있다는 점 등에 비추어볼 때 기탁방식의 운용시스템으로 활용할 수 있는 가능성은 크지 않다고 생각된다.

다. NFT 등 신기술 활용 방안

NFT의 효용성은 다양하다. 신뢰성을 확보하기 위한 다양한 방안 중 하나로 고려할 수 있을 것이다. 스마트 계약을 통하여 이용할 수 있을 것이나 데이터셋을 어디에 저장할 것인지, 스스로 관리하기 어려운 구조라는 점, 데이터 자체의 진정성을 확보하기 어렵다는 점에서 데이터를 암호화하는 방식으로서 이용통제나 누가 이용하는 것인지는 확인할 수 있을 것이나 증명방식으로서 NFT가 갖는 특성상 활용 가능성은 높지 않을 것으로 판단된다.

다만, 데이터 기탁제도를 운용하게 될 경우에는 관리기관을 지정할 것이기 때문에 기술적인 방식을 도입할 필요는 없을 것으로 생각된다. 물론, 관리기관에서는 데이터에 대해서 물리적, 관리적, 기술적 방안을 세우고 기탁자의 기대를 저버리지 않도록 관리해야할 것이다.

2. 유사 제도 검토

가. 저작권법상 에스크로우제도

소프트웨어 거래 시 저작권자가 사용을 허락받은 자(이하, 사용권자라 함)를 위하여 소스코드 및 기술정보 등을 신뢰성 있는 제3의 기관에 임치해 두고, 저작권자의 폐업 또는 자연재해로 인한 소스코드의 멸실 등으로 유지보수를 계속할 수 없게 되는 조건이 발생하는 경우 수치기관이 해당 원시코드(이하, 소스코드라 함) 및 기술정보를 사용권자에게 교부함으로써 사용권자가 안정적·계속적으로 당해 소프트웨어를 사용할 수 있도록 하는 제도이다. 임치에 대한 법적 근거는 저작권법 제101조의7(프로그램의 임치), 동법 시행령 제39조의2(임치기관)에 따른다. 임치 대상은 아래 표와 같다.

[표 4-1] 임치 대상

SW분야	IT분야
<ul style="list-style-type: none"> • 소스프로그램 • 오브젝트프로그램 • 실행프로그램 • 설계서 · 사양서 	<ul style="list-style-type: none"> • 설계도, 회로도 • 반도체 칩 • 데이터베이스 • 디지털콘텐츠

<ul style="list-style-type: none"> • 플로차트 • 매뉴얼 • 유지보수 자료 • 개발기술자 정보 	<ul style="list-style-type: none"> • 회계 관련 문서 • 이외 계약 당사자 간 협의된 기술정보 등 핵심 기술자료를 담고 있는 SW 및 문서
--	---

<출처 : 저작권위원회(2022)>

임치제도가 갖는 장점은 개발자 측면과 사용자 측면으로 나누어 살펴볼 수 있다. 먼저, 개발자 측면에서는 소프트웨어 개발업체는 저작권을 사용권자에게 양도하지 않고 원천기술을 계속 소유함으로써 기술 유출의 위험을 줄이고, 해당 기술에 대한 지속적인 연구·개발과 원활한 영업활동을 할 수 있다는 점이다.

다음으로, 사용자 측면에서는 중소 소프트웨어 개발업체의 폐업·파산이 빈번한 상황에서 유보수를 위한 소스코드 등을 확보함으로써 안정적인 사업수행이 가능하고, 저작권을 제외한 사용권만의 이관을 통해 예산절감의 효과, 국내 소프트웨어 산업의 보호를 통한 국가경쟁력 강화, 국내 소프트웨어에 대한 신뢰를 강화함으로써 수출 증대에 기여, 국내 우수 소프트웨어 기술의 유출방지 및 원천기술 확보를 통한 소프트웨어 산업의 경쟁력 강화에 기여할 수 있다는 점이다.

다만, 저작권법에 근거하여 저작물 위주로 이루어지는 시스템이라는 점에서 학습데이터를 기탁하기에는 부적합한 면이 있으며 현재로서는 오프라인 방식의 매체에 담겨진 것만을 대상으로 한다는 점에서 온라인으로의 접근가능성이 낮은 상황에서 적용하기에는 부적합한 면이 있다.

나. 기술자료임치제도

거래관계에 있는 대기업과 중소기업이 일정한 조건하에 서로 합의하여 핵심 기술자료를 신뢰성 있고 임치설비를 갖춘 대·중소기업·농어업협력재단에 안전하게 보관해 둬으로써 중소기업은 기술유출 위험을 줄일 수 있고 대기업은 해당 중소기업의 파산·폐업시 해당 임치물을 이용하여 관련 기술을 안전하게 활용할 수 있도록 하는 제도이다. 법적 근거는 대·중소기업 상생협력 촉진에 관한 법률 제24조의 2(기술자료 임치제도)²³²⁾에 따른다.

232) 제24조의2(기술자료 임치제도) ① 수탁·위탁기업(수탁·위탁기업 외에 단독 또는 공동으로 기술자료를 임치(任置)하고자 하는 기업을 포함한다)은 전문인력과 설비 등을 갖춘 기관으로서 대통령령으로 정하는 기관(이하 “수치인”(受置人)이라 한다)과 서로 합의하여 기술자료를 임치하고자 하는 기업(이하 “임치기업”이라 한다)의 기술자료를 임치할 수 있다. <개정 2010. 12. 7.>

② 위탁기업은 다음 각 호의 어느 하나에 해당하는 경우에는 수치인에게 수탁기업이 임치한 기술자료를 내줄 것을 요청할 수 있다.

1. 수탁기업이 동의한 경우
2. 수탁기업이 파산선고 또는 해산결의로 그 권리가 소멸되거나 사업장을 폐쇄하여 사업을 할 수 없는 경우 등 위탁기업과 수탁기업이 협의하여 정한 기술자료 교부조건에 부합하는 경우

③ 수치인은 중소벤처기업부장관이 정하는 기술자료 교부조건에 부합하는 경우에 임치기업의 기술자료를 요청한 자에게 이를 교부한다. <신설 2010. 12. 7., 2017. 7. 26.>

④ 정부는 수치인에게 예산의 범위에서 필요한 지원을 할 수 있다. <개정 2010. 12. 7.>

⑤ 그 밖에 기술자료의 임치 등에 필요한 사항은 대통령령으로 정한다. <개정 2010. 12. 7.>

[그림 4-1] 기술임치계약 프로세스



<출처 : 대중소협력재단, 2022>

기술자료임치제도에 따른 이용효과로는 개발기술의 효율적인 보호, 개발기업의 기술자료가 유출되었을 경우, 기술자료 임치물을 통해 개발기업의 기술 개발사실 입증, 내부직원 및 산업스파이 등에 의해 기술자료가 유출되더라도 임치물을 통해 개발기업은 해당 기술의 개발사실 입증 및 기술 경쟁력 유지라는 점이다.

다. 미생물발명의 기탁제도

(1) 기탁제도의 의의

미생물은 구조가 복잡하고 살아있는 것이어서 미생물에 관계되는 발명을 특허출원하는 경우, 명세서에 제3자가 반복 재현할 수 있도록 기재하는 것이 곤란하기 때문에 해당 미생물을 공인된 기탁기관에 기탁함으로써 발명의 재현성을 뒷받침하고 기탁된 미생물을 제3자가 분양받아 해당 발명을 재현할 수 있도록 함으로써 특허출원명세서의 기재사항을 보완하기 위한 제도이다. 즉, 특허명세서에 기재된 발명은 당업 자가 명세서의 기재에서 용이하게 실시할 수 있어야 하지만, 미생물에 관련한 발명 가운데에는 아무리 상세히 기재되어 있다고 하더라도, 당해 미생물의 입수 없이는 실시 불가능한 것도 많다 그와 같은 실험결과를 그대로 따라해 보는 것이 곤란한 것은 서면만으로는 미완성으로 취급되고 말기 때문에 서면주의의 예외로서 당해 미생물의 기탁에 의해 발명이 완성된 것으로서 다루어지게 된다.²³³⁾ 이처럼 미생물기탁제도를 둔 것은 발명의 실시

233) 中山信弘, 特許法, 法文社, 2001, 117면.

가능성의 확보와 발명공개를 통하여 후속 기술개발을 촉진하기 위한 것이다.²³⁴⁾

미생물기탁제도에 대한 법적 근거로는 특허법·시행령 및 시행규칙에서는 전문기관의 등록 등(특허법 제58조), 전문기관 등록의 취소 등(특허법 제58조의2), 미생물의 기탁(시행령 제2조), 미생물에 관계되는 발명의 특허출원명세서 기재(시행령 제3조), 미생의 분양(시행령 제4조), 미생물 기탁·분양에 관한 전문기관의 등록기준 등(시행령 제8조의4), 미생물의 수탁번호변경신고(시행규칙 제22조), 미생물의 분양절차(시행규칙 제23조), 전문기관의 등록(시행규칙 제36조의2)에 관해 규정하고 있다.

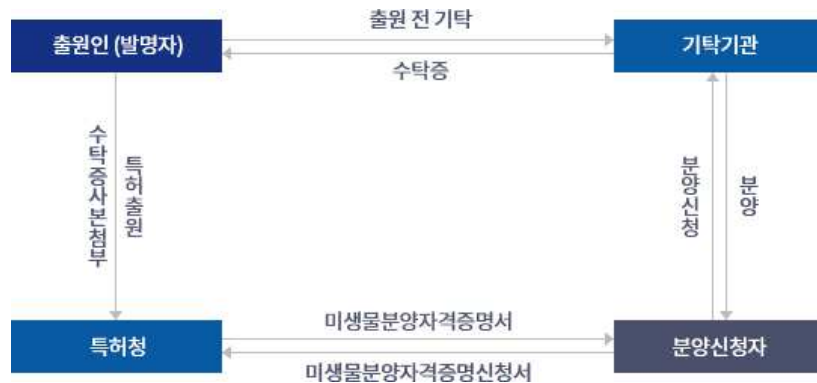
(2) 대상이 되는 미생물

기탁대상이 되는 미생물이라 함은 유전자, 세균, 바이러스, 곰팡이, 효모, 조류, 동물세포, 식물세포, 수정란, 종자 등 일체의 생물학적 물질(biological material)을 의미한다.

(3) 기탁 및 분양 프로세스

미생물 기탁이나 기탁된 미생물을 분양받기 위해서는 기탁기관을 통하여 이루어지게 된다. 다만, 2022년 기탁제도가 개선된 바 있다. 특허청이 지정한 4개 기탁기관과의 정보공유체계를 구축하여 기탁정보의 실시간 공유가 가능하게 되어, 미생물 관련 특허를 출원할 때 별도의 미생물기탁증 사본을 제출하지 않아도 된다.

[그림 4-2] 미생물기탁제도



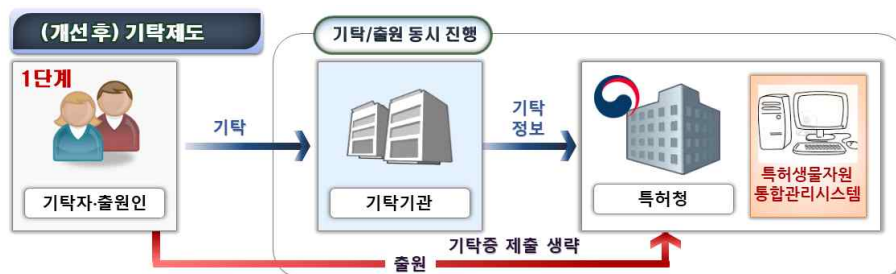
<출처 : 특허청(2022)>

그동안은 특허청과 기탁기관 간에 기탁정보가 실시간으로 공유되지 않아, 미생물 관련 특허를 출원하려면 출원인이 기탁기관에 해당 미생물을 기탁하고, 직접 기탁기관에 가서 미생물기탁증을

234) 심미랑 외, 특허절차상 생물학적 물질 기탁제도에 관한 연구, 산업재산권 제37호, 2012, 266면.

발급받아 그 사본을 특허청에 제출해야 하는 불편함이 있었다. 무엇보다, 기탁서류가 첨부되지 않는 경우에는 미완성발명에 해당거나 기재요건불비로 출원인에게는 부담이 되기도 하였다.

[그림 4-3] 변경 기탁제도



<출처 : 특허청(2022)>

특허청의 ‘특허생물자원 통합관리 시스템’의 개통으로, 기탁기관에 미생물을 기탁한 경우에, 특허 출원 시 ‘미생물 기탁 사실 증명 서류’ 제출을 생략할 수 있어 미생물 기탁 후에 바로 출원을 할 수 있게 되었다.²³⁵⁾

(4) 특허미생물 기탁기관 현황

우리나라에서 특허미생물 기탁이 가능한 기관은 한국생명공학연구원 생물자원센터(KCTC: Korean Collection for Type Cultures), 한국미생물보존센터(KCCM: Korean Culture Center of Microorganisms), 한국세포주연구재단(KCLRF: Korean Cell Line Research Foundation), 농촌진흥청 국립농업과학원 미생물은행(KACC: Korean Agricultural Culture Collection) 등 4곳이 특허청으로부터 지정되어있다. 특히, 미생물에 관계되는 발명을 국내 출원하려는 경우에는 KCTC, KCCM, KACC 중 한 곳에 미생물을 기탁하면 되고, 국제 출원하려는 경우에는 국제기탁기관인 KCTC, KCCM, KCLRF, KACC 중 한 곳에 미생물을 기탁할 수 있다.

(5) 기탁제도의 효과

미생물기탁제도의 취지는 극미의 세계에 존재하는 미생물의 성질상 그 미생물의 현실적 존재가 확인되고 이를 재차 입수할 수 있다는 보장이 없는 한 그 발명을 재현하여 산업상 이용할 수 없기 때문이다. 또한, 미생물 기탁제도는 미생물 자체가 실제로 존재한다는 것을 밝힘과 동시에 평균적 기술자가 그 발명의 실시를 위하여 필요한 때에는 그 미생물의 시료를 받을 수 있게 하기 위한 것으로, 미생물에 관계되는 발명의 완성을 입증하거나 재현성을 확인하는 데에 필요한 경우에만 적용되는 제도²³⁶⁾라는 점에서 시료를 기탁함으로써 발명의 재현성을 확인할 수 있다는

235) 특허청, 미생물 발명, 특허출원·분양신청 쉽고 빠르게, 보도자료, 2022.6.27.일자.

236) 특허법원 2002. 10. 10. 선고 2001허4722 판결 : 확정.

점을 제도의 효과로 들 수 있다.

기탁기관을 지정하는 것은 미생물 시료를 기탁하여 보관하거나 분양할 수 있는 전문적인 설비와 인력이 필요한 제도의 효율적인 운영을 위한 것이다. 특허청이 자체적으로 운용하기 보다는 전문기관을 통하여 운용함으로써 발명의 재현가능성을 높이고, 시료의 보관 등에 대한 안전성 등을 확보할 수 있을 것이다.

라. 영업비밀의 원본증명제도 검토

(1) 국회 논의

영업비밀 원본증명제도에 대해 국회심사보고서에서는 다음과 같이 평가하고 있다.²³⁷⁾ 영업비밀 원본증명제도는 영업비밀을 포함하고 있는 전자문서의 원본 여부를 증명하기 위하여 그 전자문서로부터 고유의 식별값인 전자지문²³⁸⁾을 추출하여 원본증명기관에 등록하고, 필요한 경우 원본증명기관이 전자지문을 이용하여 그 전자문서가 원본임을 증명하는 제도로써, 기업 등이 영업비밀을 외부로 반출하지 못하도록 할 수 있다. 영업비밀 유출 소송에서 피해 기업 등은 유출된 영업비밀을 특정하여야 하나, 특허 등 산업재산권이나 저작권과 달리 별도의 등록공시(公示) 제도가 없는 영업비밀의 경우 피해 기업 스스로 어떤 영업비밀을, 어느 시점에 보유하고 있었는지를 위변조에 대한 의심의 여지 없이 명확히 입증하는 것은 어려움이 있다. 이에 특허청은 이와 같은 입증 부담을 완화하기 위해 2010년 영업비밀 원본증명제도를 도입하였으며, 미국·일본·독일 등 선진국에서도 전자지문을 활용한 원본증명제도가 민간기업을 통해 시행되고 있다. 영업비밀 원본증명제도에 대한 법적 근거를 마련함으로써, 제도의 공공성을 강화하고 신뢰성을 제고하기 위해 필요한 것으로 보이며, 유사한 인증 제도인 공인인증기관, 공인전자문서센터, 기술자료 임치제도도 각각 「전자서명법」, 「전자문서 및 전자거래기본법」, 「대중소기업 상생협력 촉진에 관한 법률」에 근거를 두고 있는 바, 이러한 취지의 개정안은 타당한 입법이다.

(2) 영업비밀 원본증명제도의 의의

영업비밀 원본증명제도는 전자문서로 보관중인 영업비밀이 도용이나 유출 등으로 영업비밀 보유자가 해당 영업비밀 보유에 대한 입증이 필요한 경우, 영업비밀의 원본 존재와 보유시점의 입증에 도와주는 제도이다.²³⁹⁾ 영업비밀 원본증명제도의 법적 근거는 영업비밀보호법 제9조의2이

237) 이동근, 부정경쟁방지 및 영업비밀보호에 관한 법률 일부개정법률안 검토보고서, 국회산업통상자원위원회, 2013.

238) 전자지문(hash code) : 전자문서로부터 난수를 생성하는 수법에 의해 생성된 값. 전자문서가 수정되면 다른 코드가 생성됨. 해쉬코드를 원래 전자문서로 되돌리는 것은 불가능함. 전자지문(hash code)을 이용하여 데이터의 수정 여부가 검증되므로, 전자지문을 생성·등록하게 되면 해당 전자문서가 특정 시각에 존재하고 있었다는 것과 그 시각 이후 데이터가 변경되지 않았음을 증명할 수 있음

239) 최정렬·이규호, 부정경쟁방지법, 진원사, 2022, 412면.

다.240)

영업비밀 원본증명제도에 따라 영업비밀 보유자는 영업비밀이 포함된 전자문서의 원본 여부를 증명하기 위해서 지정받은 영업비밀 원본증명기관²⁴¹⁾에 그 전자문서로부터 추출된 고유의 식별값을 등록할 수 있는데, 그 전자문서로부터 추출한 고유의 식별값을 법에서는 ‘전자지문’ 이라고 한다.

[그림 4-4] 원본증명제도 프로세스



<출처 : 지식재산보호원(2022)>

영업비밀 원본증명의 절차는 예를 들어, A라는 파일에 제품 설계도가 담겨 있는 경우 이 A라는 파일의 고유 식별값을 추출하면 이를 전자지문이라 하고, 나중에 영업비밀 보유자가 보관하고 있는 A라는 파일로부터 추출한 고유식별값이 전자지문과 동일하다면 이 영업비밀 보유자가 A라

240) 영업비밀보호법 제9조의2(영업비밀 원본 증명) ① 영업비밀 보유자는 영업비밀이 포함된 전자문서의 원본 여부를 증명받기 위하여 제9조의3에 따른 영업비밀 원본증명기관에 그 전자문서로부터 추출된 고유의 식별값[이하 “전자지문”(電子指紋)이라 한다]을 등록할 수 있다.

② 제9조의3에 따른 영업비밀 원본증명기관은 제1항에 따라 등록된 전자지문과 영업비밀 보유자가 보관하고 있는 전자문서로부터 추출된 전자지문이 같은 경우에는 그 전자문서가 전자지문으로 등록된 원본임을 증명하는 증명서(이하 “원본증명서”라 한다)를 발급할 수 있다.

③ 제2항에 따라 원본증명서를 발급받은 자는 제1항에 따른 전자지문의 등록 당시에 해당 전자문서의 기재 내용대로 정보를 보유한 것으로 추정한다.

241) 영업비밀보호법 제9조의3(원본증명기관의 지정 등) ① 특허청장은 전자지문을 이용하여 영업비밀이 포함된 전자문서의 원본 여부를 증명하는 업무(이하 “원본증명업무”라 한다)에 관하여 전문성이 있는 자를 중소벤처기업부장관과 협의하여 영업비밀 원본증명기관(이하 “원본증명기관”이라 한다)으로 지정할 수 있다. <개정 2017. 7. 26.>

② 원본증명기관으로 지정을 받으려는 자는 대통령령으로 정하는 전문인력과 설비 등의 요건을 갖추어 특허청장에게 지정을 신청하여야 한다.

③ 특허청장은 원본증명기관에 대하여 원본증명업무를 수행하는 데 필요한 비용의 전부 또는 일부를 보조할 수 있다.

④ 원본증명기관은 원본증명업무의 안전성과 신뢰성을 확보하기 위하여 다음 각 호에 관하여 대통령령으로 정하는 사항을 지켜야 한다.

1. 전자지문의 추출·등록 및 보관
2. 영업비밀 원본 증명 및 원본증명서의 발급
3. 원본증명업무에 필요한 전문인력의 관리 및 설비의 보호
4. 그 밖에 원본증명업무의 운영·관리 등

⑤ 원본증명기관 지정의 기준 및 절차에 필요한 사항은 대통령령으로 정한다.

는 파일의 보유자라고 볼 수 있다. 이러한 영업비밀 원본증명제도는 영업비밀침해 사건(형사사건, 민사사건 등)에서 영업비밀 보유자의 입증부담을 완화시키고, 고유의 식별값으로 입증하므로 자료의 외부 유출 없이도 영업비밀의 입증을 도울 수 있는 제도라 할 수 있다.

영업비밀 원본증명제도의 법적 효력은 실제 사건에서 법원이나 수사기관에 어느 정도 효력을 미치는지가 제일 중요한데, 이에 대하여 영업비밀보호법 제9조의2 제3항은 원본증명을 발급받은 자는 전자지문의 등록 당시에 해당 전자문서의 기재 내용대로 정보를 보유한 것으로 추정토록 함으로써, 입증책임이 곤란한 점을 완화하게 되었다.

(3) 학습데이터의 적용가능성 검토

데이터 공개제도를 고려함에 있어서, 영업비밀 원본증명제도와 연계할 수 있는 방법은 데이터 자체의 원소유권을 확인하거나 또는 정당한 이용권을 확보하고 있다는 점을 입증할 수 있을 것이다. 다만, 이는 문서상의 증명이라는 점에서 진정한 저작권자인지, 이용허락을 받은 것인지를 증명해주는 것은 아니라는 점에도 본래의 제도적 취지에 부합할 수 있는 것인지는 검토가 필요하다. 증명시점부터 보유한 것을 증명하는 것이기 때문에 데이터 자체에 대한 증명이나 보유관계를 확인할 수 없다는 점에서 차이가 있다. 다만, 제3의 기관을 통하여 기탁하는 경우에 있어서 이를 확인해주는 증빙으로서 활용할 수도 있을 것이나 그 때에는 별도의 제도가 설계되어야 할 것이다. NFT화 하는 것에 대해 살펴본 바와 같이, 증명제도는 제3의 기관이 설립되는 경우에는 해당 기관에서 이루어지는 것이기 때문에 영업비밀보호법 상의 증명제도와는 다른 방식이어야 할 것으로 판단된다.

3. 데이터 기탁제도에 대한 비교·분석

가. 미생물발명과 AI발명

미생물의 배양자체는 유전자 방식에 따라 알 수 없으나, 시간이 흐름에 따라 그 결과가 확인되는 과정을 거친다. 이 제도는 출원인으로 하여금 해당 미생물을 공인기관에 기탁하도록 하여 제3자가 일정조건(기탁자 또는 특허청 허락)하에 분양을 받아 당해 발명을 쉽게 실시할 수 있도록 하는 제도로써, 미생물발명의 일정한 확실성과 반복생산 가능성의 문제를 해결하고 있다. 이러한 미생물발명처럼, 인공지능 발명도 인공지능경망을 통하여 구현된 구조는 블랙박스화되기 때문에 인간이 이해하기는 어려움이 있다. 이러한 점에서 유전자방식과 인공지능경망 방식의 유사성에 따라 발명을 쉽게 실시할 수 있는 방안으로써 기탁제도를 고려할 수 있을 것이다.

나. 미생물기탁제도를 통해 본 데이터기탁제도

(1) 기탁 요건

미생물은 특허법 시행령에 따라 기탁을 요건으로 하고 있다.²⁴²⁾ 미생물을 이용한 발명에 대하여 특허출원을 하고자 하는 자는 특허청장이 지정하는 기탁기관에 그 미생물을 기탁하고 그 기탁사실을 증명하는 서면을 출원서에 첨부하여야 하며, 다만 그 미생물이 그 발명이 속하는 기술 분야에서 통상의 지식을 가진 자(평균적 기술자)가 용이하게 얻을 수 있는 때에는 기탁을 하지 아니할 수 있다고 규정하고 있는바, 이 규정의 취지는 극미의 세계에 존재하는 미생물의 성질상 그 미생물의 현실적 존재가 확인되고 이를 재차 입수할 수 있다는 보장이 없는 한 그 발명을 재현하여 산업상 이용할 수 없기 때문이라 할 것이고, 다만 최종 생성물이나 중간 생성물은 비록 그 자체가 기탁되어 있지 아니하더라도 이를 생성하는 과정에 필요한 출발 미생물들이 평균적 기술자가 용이하게 얻을 수 있는 것이고, 또 명세서에 이를 이용하여 중간 생성물이나 최종 생성물을 제조하는 과정이 평균적 기술자가 용이하게 재현할 수 있도록 기재되어 있는 경우라면 그 최종 생성물이나 중간 생성물 자체의 기탁을 요구할 것은 아니다.²⁴³⁾

(2) 서면주의 방식의 보충적 방식으로 기탁

미생물에 관계되는 발명에 대하여 특허출원을 하려는 자는 법 제42조 제2항에 따른 명세서(특허출원서에 최초로 첨부한 명세서를 말한다)를 적을 때 제2조 제1항 본문에 따라 미생물을 기탁한 경우에는 국내기탁기관, 국제기탁기관 또는 지정기탁기관에서 부여받은 수탁번호를, 같은 항 단서에 따라 그 미생물을 기탁하지 않은 경우에는 그 미생물의 입수방법을 적어야 한다.²⁴⁴⁾

구 특허법 시행령 제1조 제2항이 정한 '기탁사실을 증명하는 서면'은 미생물의 수탁기관이 발행하는 미생물수탁번호통지서나 수탁증 등과 같이 당해 미생물의 기탁사실을 객관적으로 증명하

242) 특허법 시행령 제2조(미생물의 기탁) ① 미생물에 관계되는 발명에 대하여 특허출원을 하려는 자는 특허출원 전에 다음 각 호의 어느 하나에 해당하는 기관에 특허청장이 정하여 고시하는 방법에 따라 해당 미생물을 기탁해야 한다. 다만, 해당 발명이 속하는 기술 분야에서 통상의 지식을 가진 자가 그 미생물을 쉽게 입수할 수 있는 경우에는 기탁하지 않을 수 있다.

1. 「특허법」(이하 “법”이라 한다) 제58조제2항에 따라 미생물 기탁 및 분양에 관한 업무를 담당하는 전문기관으로 등록된 기관(이하 “국내기탁기관”이라 한다)
2. 「특허절차상 미생물기탁의 국제적 승인에 관한 부다페스트조약」 제7조에 따라 국제기탁기관으로서의 지위를 취득한 기관(이하 “국제기탁기관”이라 한다)
3. 다음 각 목의 요건을 모두 충족하는 국가에서 미생물 기탁 및 분양에 관한 업무를 담당하는 전문기관으로 지정한 기관(이하 “지정기탁기관”이라 한다)

가. 「특허절차상 미생물기탁의 국제적 승인에 관한 부다페스트 조약」의 당사국이 아닐 것
나. 해당 국가의 특허청장이 대한민국 국민에게 특허절차상 미생물기탁에 대해 대한민국과 동일한 조건의 절차를 인정하기로 특허청장과 합의한 국가일 것

② 제1항에 따라 미생물을 기탁한 자는 특허출원서에 산업통상자원부령으로 정하는 방법에 따라 그 취지를 적고, 미생물의 기탁 사실을 증명하는 서류(국제기탁기관에 기탁한 경우에는 「특허절차상 미생물기탁의 국제적 승인에 관한 부다페스트조약 규칙」 제7규칙에 따른 수탁증 중 최신의 수탁증 사본을 말한다)를 첨부하여야 한다. 다만, 국내에 소재지가 있는 국내기탁기관 또는 국제기탁기관에 해당 미생물을 기탁한 경우에는 미생물의 기탁 사실을 증명하는 서류를 첨부하지 않을 수 있다.

③ 특허출원인 또는 특허권자는 제1항의 미생물의 기탁에 대하여 특허출원후 새로운 수탁번호가 부여된 때에는 지체없이 그 사실을 특허청장에게 신고하여야 한다.

243) 대법원 2002. 11. 8. 선고 2001후2238 판결 [거절사정(특)].

244) 특허법 시행령 제3조(미생물에 관계되는 발명의 특허출원명세서 기재)

는 서면을 말하는 것이므로, 특허발명의 출원시에 제출된 명세서에 당해 미생물의 기탁번호·기탁기관의 명칭 및 기탁연월일을 기재하였다고 하더라도, 이는 구 특허법 시행규칙 제31조의2 제1항의 명세서 기재요건을 충족한 것으로 볼 수 있을 뿐, 이러한 출원서의 제출을 들어 위 시행령 제1조 제2항의 '기탁사실을 증명하는 서면'이 제출되었다고 할 수는 없다. 이 사건 미생물이 미국의 ATCC에 기탁되어 있다는 사실만으로는 이 사건 특허발명의 우선일인 1984. 12. 4경 이 발명이 속하는 기술분야에서 통상의 지식을 가진 자가 이 사건 미생물을 용이하게 얻을 수 있는 것이라고 할 수도 없으며, 특허발명의 심사단계에서 미생물 기탁증명서의 미제출을 간과한 채 특허가 되었다고 하여 그 출원절차상의 하자가 치유된다거나 출원에 있어서의 하자를 들어 특허의 효력을 부인하는 것이 금지된다고 보아야 할 아무런 근거가 없다. 따라서 미생물의 기탁에 관한 요건을 충족하지 못한 이 사건 특허발명은 미완성 발명에 해당하고, 미완성 발명의 경우는 특허무효심결의 확정 전이라도 그 권리범위를 인정할 수 없는 법리이므로, 원고의 확인대상발명이 이 사건 특허발명과 대비할 것도 없이 이 사건 특허발명의 권리범위에 속하지 않는다고 한 원심의 판단은 정당하다고 보았다.²⁴⁵⁾ 기탁이 이루어지지 않는 상태로 용이하게 구현할 수 있는 수준으로 명세서에 기재하면 미완성발명에 대한 논란은 해소될 것이다. 데이터의 경우도 서면주의에 따른 보충적 방식으로 기탁할 수 있을 것이다. 다양한 데이터셋의 정보를 기재하는 경우라면 서면주의 방식으로 가능할 것으로 생각된다. 물론, 데이터셋의 공개가 필요한 경우에는 정보의 기술이 아닌 구체적인 데이터셋의 기탁을 통하여 재현가능성을 확보할 수 있도록 해야 할 것이다.

(3) 데이터의 추가 또는 변경의 경우

데이터를 이용한 발명에서 데이터가 변경되거나 추가되는 경우에는 출원발명과 다를 수 있다. 예를 들면, 출원시의 발명과 심사시의 발명이 달라질 수 있기 때문에 동일한 발명인지에 대한 논란이 가능하다. 물론, 출원시의 데이터를 통하여 발명을 실시할 수 있다면 다른 발명이라고 판단할 필요는 없을 것이다. 실제, 미생물의 기탁과 관련해서도 유사한 사례가 발견다. 즉, 발명에 이용된 미생물이 신규의 미생물로 인정되는 경우에는 미생물기탁제도의 취지에 비추어 보아 일용이를 기탁할 필요성이 있다고 할 것이나, 최종생성물이나 중간생성물은 비록 그 자체가 기탁되어 있지 아니하더라도 이를 생성하는 과정에 필요한 출발미생물들이 공지된 균주이거나 그 발명이 속하는 기술분야에서 통상의 지식을 가진 자가 용이하게 얻을 수 있는 것일 경우에는 그 최종미생물이나 중간생성물은 자체의 기탁을 요구할 것은 아니라 할 것이다.²⁴⁶⁾

다. 기탁기관의 지정

기탁기관은 별도의 자격을 갖추도록 필요한 요건을 지정할 필요가 있다. 참고할 수 있는 사례

245) 대법원 2005. 9. 28. 선고 2003후2003 판결 [권리범위확인(특)].

246) 대법원 1997. 3. 25. 선고 96후658 판결.

는 미생물기탁기관의 등록 및 운영에 관한 고시²⁴⁷⁾이다. 향후, 데이터 기탁기관의 등록 및 운영에 관한 고시를 제정하여 운영하여야 할 것으로 보인다.

(1) 지정 요건

기탁기관은 엄격한 요건하에 지정이 이루어져야 할 것이다. 무엇보다도 국내에서 사업을 영위하여야 하며, 물리적인 서버를 국내에 두어야 할 것이다. 또한, 전문적인 인력과 시설을 갖추어야 할 것이다. 미생물의 기탁보다는 용이하게 운영될 수 있기 때문에 미생물과 같은 요건보다는 외부적인 해킹이나 사이버보안과 관련된 이슈에 대해서 대응할 수 있는 물리적, 기술적 관리체계를 갖추는 것이 필요하다고 하겠다.²⁴⁸⁾

(2) 기탁기관의 의무²⁴⁹⁾

247) [시행 2017. 8. 1.] [특허청고시 제2017-16호, 2017. 8. 1., 일부개정].

248) 미생물기탁기관의 등록 및 운영에 관한 고시 제4조(기탁기관의 등록 또는 지위취득 보증의 요건)에 따르면, 특허청장으로부터 국내기탁기관으로 등록을 받거나 국제기탁기관으로 지위취득 보증을 받고자 하는 기관은 다음 각 호의 요건을 충족하여야 한다.

1. 대한민국 영토 내에 위치하고 계속적으로 존재할 것
2. 다음 각 목의 요건을 충족하는 인력을 갖추는 것
 - 가. 미생물 분야와 관련된 학사이상의 학위 보유자 또는 미생물의 수탁, 분양 등과 관련된 업무를 2년 이상 수행한 경험이 있는 자
 - 나. 기탁기관의 등록 또는 지위취득 보증의 계획공고에서 정하는 미생물의 종류에 따른 업무분야 별로 1명 이상일 것
3. 다음 각 목의 요건을 충족하는 시설을 갖추는 것
 - 가. 수탁 미생물을 오염시키지 않고 생존상태로 보관할 수 있는 전문장비를 확보할 것
 - 나. 화재, 정전 등 비상시를 대비한 화재대비안전관리체계 및 비상전원공급체계를 확보할 것
 - 다. 관계자 이외의 출입을 통제할 수 있는 보안체계를 확보할 것
4. 다음 각 목의 미생물 관련 업무를 수행할 수 있을 것
 - 가. 미생물 수탁 및 수탁증 발행 업무
 - 나. 미생물 생존시험 및 생존증명서 발행 업무
 - 다. 미생물 보관 업무
 - 라. 미생물 분양 업무
5. 수탁 미생물에 대한 비밀을 유지할 수 있을 것
6. 이 고시에 따른 기탁기관의 의무를 성실히 이행하고 관련 업무를 공정하고 객관적으로 수행할 수 있을 것

249) 제8조(기탁기관의 의무) ① 기탁기관은 특허청장의 요청이 있을 시 다음 각 호의 사항을 기재한 서면을 특허청장에게 제출하여야 한다.

1. 미생물의 수탁 및 분양 현황
2. 생존시험 및 생존증명서 발급 현황
3. 인력 및 시설에 관한 사항
4. 기타 특허청장이 필요하다고 인정하는 사항

② 기탁기관은 특허청장이 제1항 각 호에 관한 실태 조사를 요구하는 경우에는 이에 성실히 응해야 한다.

③ 기탁기관은 제12조 제1항의 업무처리규정에 따라서 미생물 수탁 및 분양 업무를 성실히 수행하여야 한다.

④ 기탁기관은 천재지변 등에 의하여 미생물이 멸실 또는 훼손된 경우에는 즉시 특허청장에게 보고한 후 제12조 제2항의 안전관리규정에 따라 신속하게 조치하여야 하고 그 피해 및 복구 상황은 피해발생 후 15일 이내에 특허청장에게 보고하여야 한다.

⑤ 기탁기관은 미생물의 멸실 또는 훼손에 대비하여 미생물의 복제본을 제4조제3호 각 목과 동일한 요건을 갖춘 별도의 보관시설에 보관하여 신속한 복구가 가능하도록 하여야 한다.

수탁 및 분양에 대한 현황을 특허청장에게 제출하여야 한다. 특허청은 해당 기관에 대해 실태 점검을 진행함으로써 운영상의 문제는 없는지 수시로 파악할 수 있어야 할 것이다. 또한, 기탁받은 데이터가 해킹을 통하여 유출되거나 파손되는 등의 사이버보안 이슈가 발생하지 않도록 하여야 할 것이다.

(3) 점검

특허청은 제3자가 연구 또는 시험을 위해 특허관련 미생물의 분양을 신청할 경우 미생물분양 자격증명서를 발급하여 기탁미생물의 분양을 허락하고 있으며, 기탁된 미생물의 관리실태를 분기별로 점검하고 있다.

(4) 비상대책 등의 마련

사이버 보안이나 데이터센터의 화재와 같은 물리적인 문제를 포함하여 지진이나 전쟁 등의 비상사태시에 기탁된 미생물을 보관할 수 있는 체계나 시스템의 구축은 무엇보다 중요하기 때문에 이에 대한 비상계획을 수립운영할 수 있어야 할 것이다.

라. 데이터기탁제도의 한계

데이터기탁제도 갖는 긍정적인 면에도 불구하고 현실적으로 생각할 수 있는 기술적인 한계는 데이터나 파라미터가 적지 않다는 점을 들고 있다. 즉, “인공지능 마다 사용 플랫폼이 상이할 수 있고, 초대형언어모델의 경우(구글의 Switch Transformer는 1조 6,000억개, 네이버의 HyperClova는 2,040억개, GPT-3는 1,750억개의 파라미터를 가짐) 기탁 및 실행이 용이하지 않을 것으로 생각되며, 이처럼 운용 및 유지에 비용 및 어려움이 있는데 반해, 실질적으로 이용이 활성화 되기는 어려울 수 있다. 또한 기탁제도는 공개를 전제로 하기 때문에 인공지능 개발자 입장에서는 영업비밀 등의 유출 문제로 제출을 꺼려할 수 있을 것”²⁵⁰⁾이라는 주장이다. 그렇지만 이러한 주장 등에 대해서는 앞에서 살펴본 바와같이, 영업비밀제도와 특허제도가 공존할 수 있는 여지가 충분하다는 점에서 이러한 주장을 보완할 수 있을 것이다. 다만, 기술적인 면에서 수많은 파라미터를 가진 AI모델 발명의 경우에는 해당 데이터를 공개하는 것이 실질적으로 기술공개에 대한 기여를 할 수 있을 것인지에 대한 의문으로 보인다. 발명자 입장에서는 제기할 수 있는 의문이라고 볼 수 있으나, 데이터의 공개는 발명의 재현가능성에 대한 부분이고 이를 발명의 상세한 설명으로 입증하기 어려운 경우에 데이터기탁으로 해결할 수 있는 일종의 인센티브로서 역할을 할 수 있다는 점에서 제도의 특징이 있다고 하겠다.

250) 한국지식재산연구원·AI-IP특별위원회, 인공지능(AI)-창작물의 권리보호 방안 수립 등 AI-지식재산 이슈 대응 연구, 국가지식재산위원회, 2022, 65면.

마. 정리

데이터 기탁에 대한 기술적인 솔루션이 필요할 것으로 보이며, 이는 제3의 기관에 진행하는 것으로 고려할 수 있을 것이다. 물론, 데이터 기탁 자체를 강제적으로 시행하기 보다는 여러 가지 파일럿 형태로 진행하는 것도 방안이라고 본다. 아울러, 미생물 발명과 같이 특허법 시행령을 통하여 법적인 근거를 마련할 수도 있을 것이다. 보다 구체적인 것은 심사지침에서 구체화하는 것이라고 하겠다. 다만, 제도화될 경우에는 데이터셋에 있어서 공개할 수 있는 내용을 가능한한 최대한 공개하는 것이 필요할 것으로 보인다.

제3절. 데이터의 보호·활용을 위한 제도개선

1. 데이터의 신뢰성 확보 방안 검토

가. 신뢰성 확보의 의의

신뢰성은 발명의 요소로서 데이터가 활용되었다는 사실관계의 확인이며, 그 과정에서 사용되었던 데이터라는 점을 확인하는 것이다. 신뢰성 확보를 위해서는 필요에 따라 미생물을 분양받는 것처럼 데이터도 제공받아서 이용할 수 있는 체계가 고려될 수 있다.

나. 데이터 권리관계를 증명할 수 있는지 여부

신뢰성은 데이터가 구성요소로 사용되었고 그러한 결과가 방명으로 성립되었다는 점을 확인하는 것이다. 다만, 데이터 자체에 대해서는 증명할 수 있는 방법이 없다. 특허심사처럼 공지공용의 기술인지를 심사과정에서 확인할 수 있는 방법도 없다. 따라서, 데이터에 대해서는 요건사실만 확인할 수 밖에 없다는 한계를 갖는다. 결국, 데이터를 제3자가 이용하는 과정에서 발생하는 문제를 당사간에 약정으로 정할 수밖에 없는 것이다. 그렇지만, 민사적인 계약관계이외에 위법한 경우에 대해서 까지 면책할 수 없다는 점에서 권리관계의 증명성은 사실상 어려운 일이다.

다. 데이터 증명에 대한 법적 책임

데이터의 저작권이나 타인의 데이터자산을 임의로 사용한 것이 아니라는 점은 데이터를 제공한 자가 져야 한다. 이러한 점에서 데이터 자산에 대한 증명책임은 데이터를 제공한 자가 지기 때문에 데이터의 신뢰성을 확보하려는 시도와 실제 데이터를 제공하는 것의 패리가 생길 수밖에 없다. 따라서, 신뢰성 확보의 취지가 필요하다고 하더라도 그러한 취지에 반하는 현실적인 이유 때문에 제도의 효용성은 찾기 어려울 것으로 생각된다.

2. 데이터 재현가능성 확보 방안 검토

가. 재현가능성 확보의 의의

데이터를 이용한 발명에서 재현가능성은 동일한 발명의 실시가능성과 같다. 즉 발명의 실시와 동일한 개념으로 재현을 이해할 수 있는 것이다. 따라서, 발명의 구성요소가 완비되어야 동일한 발명을 실시할 수 있기 때문에 데이터는 재현가능성을 확보하기 위해 필요하다. 따라서 데이터 관련 발명에서 재현가능성은 데이터의 완비가 뒷받침되어야 한다는 점에서 필수적인 것이다. AI 발명에서 데이터가 구성요건으로 포함될 경우에는 데이터를 통하여 실시가 가능하게 됨으로써 재현가능성 확보할 수 있을 것이라 짐이다.

나. 데이터기탁제도의 타당성 검토

데이터기탁제도를 도입할 수 있을 것인지는 필요성에 동의할 수 있으나 실현가능성에는 의문을 표하기도 한다. 신뢰성 확보나 재현가능성을 확보하기 위하여 기탁 등의 방식을 취한다고 하더라도 데이터를 제3자가 활용할 수 있다는 점에서 발명자는 소극적일 수 밖에 없다. 경쟁가능성이 있는 기업에게 발명을 공개하는 것이 될 수 있다는 점이다.

발명제도의 목적은 기술공개이지만 사실상 기술공개는 경쟁업자의 기술우회를 가능하게 할 수 있기 때문이다. 이러한 점에서 AI발명의 핵심적인 구성요소인 데이터를 공개하는 것은 기업의 입장에서 부정적일 수밖에 없을 것이다. 다만, 앞서 살펴본 바와 같이 특허출원이 되더라도 구체적인 기술적 사상은 공개하지 않을 수도 있기 때문에 영업비밀을 모두 공개할 필요는 없을 것이다. 이러한 점을 발명자에게 제시함으로써 기술우회 가능성에 대한 우려를 불식시키는 것도 필요하다고 하겠다.

다. 데이터원본증명 제도의 연계 가능성 검토

데이터 자체를 공개하는 것은 기업 입장에서는 리스크가 있다고 판단하는 경향이 있음을 알 수 있다. 데이터를 제3의 기관에 저장하는 방식은 활용가능성을 높일 수 있다고 보지만, 당사자 입장에서는 거부감도 예상되는 방식이라고 하겠다. 따라서, 데이터 자체를 공개하는 방식보다는 제3의 기관을 통하여 데이터의 권리관계를 확인하거나 발명에 사용되었다는 사실관계를 확인하는 수준의 증명방식을 고려할 수 있을 것이다.

이러한 증명방식은 데이터의 원본성이나 저작권 침해 여부등에 구애됨이 없이 사용되었다는 점만을 확인시켜준다는 점에서 의의가 있다. 따라서, NFT와 같이 기술적인 증명방식을 활용할 수도 있을 것이다. 다만, NFT는 제3의 관리기관이 지정되어 운용될 경우에는 사용할 실익은 크지 않다고 하겠다.

3. 기탁제도 도입시에 고려되어야할 법률 문제

가. 데이터 내용과 관련된 문제

제3의 기관에 학습데이터를 기탁하는 경우에 해당 데이터셋에 대한 실질심사는 지양될 필요가 있다. 무엇보다, 심사기간이 지연될 수 있기 때문이다. 아울러, 데이터의 진성정, 저작물성, 공서양속의 위배가능성 등에 대해서는 발명자가 제시하는 내용에 대해 특허심사 절차처럼 진행할 수도 없는 상황이다. 데이터셋의 저작권 침해, 개인정보 이슈, 영업비밀의 침해나 계약위반 등에 대해서는 기탁자의 책임으로 두어야 할 것이다.

나. 데이터 관리와 관련된 문제

앞서 기탁관의 지정요건에서 제시한 사항에 대한 책임문제에 관한 것이다. 기탁된 데이터에 대해 해킹이나 침해사고 등 사이버 보안 사고에 대해 누가 책임질 것인지 여부이다. 기본적으로는 기탁기관의 관리책임이기 때문에 사이버 보안 이슈가 발생되지 않도록 하여야 할 것이다. 특히, 영업비밀성이 있는 데이터가 공개될 경우에는 확대손해가 발생할 수 있기 때문이다. 이러한 침해사고에 대해 물리적, 기술적 관리가 필요하다.

제4절. 데이터에 대한 독자적 권리화 방안

1. 독자적 권리화의 필요성

데이터의 공개 및 이를 관리하기 위한 기탁제도는 입법적인 추진이 이루어질 필요가 있다는 점에서 대안으로써 독자적 권리화 방안에 대해 고려할 필요가 있다. 특허권이나 저작권의 대상이 되기 어려운 경우에는 제3의 방식으로써 데이터에 대한 권리를 부여하자는 의미이다. 물론, 데이터 관련 법제에서 데이터자산이라는 보호법익을 부여하고 있으나, 이는 특허제도 내에서 필요한 경우와는 차이가 있다.

따라서, 독자적 권리화는 기존 저작권이나 특허권과는 다른 보호체계와 기간 등을 부여하는 것이 필요하고, 이는 이용활성화라는 측면에서 채권적 권리화 내지는 형사처벌을 배제하는 방향으로 설계할 필요가 있다.

2. 데이터 특허화 방안

데이터 자체가 구조화되지 않는 경우에는 특허법적인 보호가능성은 낮다. 다만, 데이터셋은 그 자체가 알고리즘으로서 소프트웨어적 성격을 갖는 경우라면 특허법의 대상이 될 수 있다. 다만, 이러한 범위에 포함되지 않는 경우에는 별도의 권리화하는 방안이다.

AI 학습용 데이터를 특허권 유사의 독자적인 방법으로 보호해야 할 필요성은 AI 관련 발명에 있어 ‘데이터’가 갖는 중요성을 전제로 하여, 데이터 수집 및 분석한 자에 대한 인센티브를 부

여하는 방안이다.²⁵¹⁾ AI 학습용 데이터는 AI 성능에 직결되고, AI 사회에 있어서 기본적인 인프라가 될 수 있으므로, ① 될 수 있는 한 많이 양질의 학습용 데이터를 수집하고, ② 이를 광범위한 사람들이 이용할 수 있도록 하는 것이 중요한 요소가 될 것이다. AI 학습용 데이터는 비밀로 관리하지 않아도, 공개되어 공지가 되어도 AI 학습용 데이터를 보호할 수 있도록 영업비밀에 의한 보호가 아니라, 지식재산권 유사 보호가 필요하다. 이를 위해 특허시스템을 차용하면, 데이터 특허출원은 ① 데이터 클레임, ② 데이터 명세서, ③ 실제 AI 학습용 데이터(기탁·공개)를 제출하는 형식을 고려할 수 있을 것이다.²⁵²⁾

데이터 특허에 대해서는 출원만을 하여 출원 장려금을 받으려고 하는 경우도 있지만, 심사청구를 하여 특허청에 의한 심사를 받을 수도 있다. 데이터 특허의 심사에 있어서는 진보성은 필요 없고, 예를 들면, ① 신규성, ② 학습가능 요건, ③ 학습 유용성 요건, ④ 기재요건 등의 심사방법을 생각할 수 있다. ① 신규성은 클레임 및 출원시에 기탁된 데이터가 종래 데이터베이스에 존재하지 않는지 여부를 심사한다. 다만, 타인이 작성한 데이터를 무단으로 출원하면 모인출원이 된다. ② 학습가능 요건은 기존 특허의 실시가능 요건에 대응하는 것이며, 데이터를 실제로 AI에 학습시킬 수 있는지 여부를 심사한다. 기탁된 데이터에 누락이 있는 등 학습하지 못할 경우에는 이 요건을 만족시키지 않는다. 심사 시에는 특허청에 구비된 복수 종류의 표준 뉴럴 네트워크 등을 학습에 이용하여 학습가능 요건을 판정한다. 표준 뉴럴 네트워크는 일반적으로 공개하여, 출원인이 학습가능 요건을 사전에 체크할 수 있게 하는 것을 생각할 수 있다. 또한 출원인이 특수한 뉴럴 네트워크를 이용한 경우에는 해당 뉴럴 네트워크도 기탁하면 학습가능 요건을 충족한 것으로 본다. 다만, 데이터 특허에 대해서는 데이터의 실질적인 내용을 심사할 수 있을지는 의문이다.

3. 데이터 등록제도의 도입

앞서 살펴본, 데이터 특허화와는 다른 개념으로서 데이터 등록제도를 제안한다. 기본적인 골격은 데이터베이스 제작자의 권리 또는 반도체집적회로의 배치설계에 관한 법률(이하, 반도체설계법이라 함)에 따른 반도체배치설계권과 유사한 개념으로 이해할 수 있다. 이 또한 별도의 독자적인 권리화 방안이다. 학습데이터의 기술적인 진보성 보다는 데이터의 활용가능성을 목적으로 하는 것이다. 디자인 무심사제도 내지는 반도체배치설계권과 같이 형식적인 등록체계를 갖추으로써 권리화하는 방안이다. 등록받은 데이터는 특허청 산하기관 중에서 전문성이 있는 기관을 지정하여 기탁 내지는 입치 등을 할 수 있을 것이다. 이를 위하여 본 연구에서 기탁기관으로써 제3의

251) 岡本 義則, “人工知能(AI)の学習用データに関する知的財産の保護”, *パテント*(Vol.70 No. 10), 2017, 91-92頁; 전성태, AI 학습용 데이터의 보호에 관한 소고 - 지식재산법상의 보호를 중심으로 -, *법학논총* vol.28, no.1, 2021, 95면.

252) 岡本 義則, “人工知能(AI)の学習用データに関する知的財産の保護”, *パテント*(Vol.70 No. 10), 2017, 93-94頁; 전성태, AI 학습용 데이터의 보호에 관한 소고 - 지식재산법상의 보호를 중심으로 -, *법학논총* vol.28, no.1, 2021, 96~97면.

전문기관에 관한 사항을 참조할 수 있을 것이다.

권리의 내용으로는 설정등록이 이루어진 경우에는 독점권을 부여하며, 이용권의 설정을 가능하도록 한다. 다만, 이미 공개된 데이터를 활용한 경우에는 별도의 등록사항이 될 수 없다는 점을 명확히 고지하여야 하며, 고의로 등록을 할 경우에 한하여 형사처벌 등의 벌칙규정을 두는 방안도 고려할 수 있을 것이다.

기본적으로 형사처벌 보다는 민사적인 구제를 원칙으로 한다. 다만, 앞서 살펴본 바와 같이 허위등록을 할 경우에는 이는 사회적 법익의 침입이기 때문에 형사처벌의 대상으로 두는 것이 바람직하다고 본다. 또한, 데이터의 등록이나 이용과정에서 발생할 수 있는 분쟁에 대해서는 소송으로 해결할 경우에는 시간과 비용이 소요될 수 있기 때문에 분쟁조정제도를 도입하여 간이하게 해결할 수 있도록 한다.²⁵³⁾

제5절. 소결

데이터의 기탁이나 등록 등의 제도는 데이터가 갖고있는 다양한 특성에 따른 고민의 결과이다. 데이터를 활용한 발명에서는 데이터가 구성요소를 이루기 때문에 용이실시를 위해서 필요한 요소이다. 이러한 요소가 부재할 경우에는 발명의 실시가 제대로 이루어지기 어렵다. 그러할 경우, 발명의 충분한 공개가 이루어졌다고 보기 어려울 수 있다. 실상 심사과정에서 그러한 사실을 확인한다면 거절결정이 내려질 수 있으며 등록이 이루어진 시점에서는 무효화될 가능성이 높다. 그렇기 때문에 특허법에서 요구하는 사항을 충족시키기 위해서도 데이터에 대한 확인가능성을 높일 수밖에 없을 것이다.

이러한 필요성에도 불구하고, 실제 산업계에서는 데이터를 공개하는 것에 대해 부정적인 입장이다. 특히, AI기업의 경우는 데이터가 기본적인 자산이기 때문에 이를 공개하기 어렵다는 입장이다. 영업비밀성을 유지하는 것이 경쟁력을 유지할 수 있다는 것이다. 다만, 선택적으로 발명자가 데이터를 공개할 수 있는 방안을 제시하는 것은 필요해 보인다. 해당 기업이 가지고 있는 자산을 공개하여 다른 발명을 활성화한다는 특허제도의 취지에 맞게 운영할 수 있다고 보기 때문이다. 따라서, 시범사업 형태로 데이터 기탁사업을 추진하는 것도 필요해 보인다. 예를 들면, 다양한 특허정보를 구축하고 있는 특허정보원의 인프라를 활용한다면 시스템의 구축 및 운영은 크게 어려움이 없을 것으로 생각된다.

시범사업 형태로 진행하고 필요할 경우, 실질적인 심사지침 등을 개정하여 선택적으로 기탁을 받을 수 있도록 하는 것이다. 다만, 시스템을 운영하기 위해서는 기존의 특허정보 제공과는 다른 차원의 운영지침이나 방식이 요구될 것으로 보인다. 이러한 점은 법률문제 부분에서 검토한 바와 같다.

253) 강영주, 인공지능을 활용한 민사분쟁 해결에 관한 고찰, 안암법학 no.64, 2022 참조.

제5장. AI 학습데이터를 활용한 데이터산업 활성화 방안 연구

제1절. 데이터산업의 활성화

1. 데이터산업 현황

가. 데이터 시장 현황²⁵⁴⁾

국내 데이터산업의 성장세는 두드러진다. 2016년부터 2021년까지 데이터산업의 성장률을 10% 이상을 이끌어내고 있다. 다만, 데이터 판매 증가는 27% 이상이라는 점에서 데이터 자체의 거래나 시장형성을 이끌어가고 있음을 알 수 있다.

[표 5-1] 2021년 데이터산업 시장규모-매출(단위: 억원)

	연도별 시장규모			증감율		연평균 증감율		
	2019년	2020년	2021년(E)	2019-2020	2020-2021(E)	2019-2021(E)	2016-2021(E)	
데이터 처리 및 관리 솔루션 개발·공급업	데이터 수집·연계 솔루션 개발·공급업	1,871	2,122	2,787	13.4%	31.3%	22.0%	15.7%
	데이터베이스 관리 시스템 솔루션 개발·공급업	7,510	8,979	11,062	19.6%	23.2%	21.4%	12.5%
	데이터 분석 솔루션 개발·공급업	2,014	2,586	2,951	28.4%	14.1%	21.0%	18.8%
	데이터 관리 솔루션 개발·공급업	5,203	6,022	7,151	15.7%	18.7%	17.2%	5.9%
	데이터 보안 솔루션 개발·공급업	1,975	2,558	3,146	29.5%	23.0%	26.2%	-
	빅데이터 통합 플랫폼 솔루션 개발·공급업	2,231	2,866	3,468	28.4%	21.0%	24.7%	16.6%
소계	20,805	25,133	30,566	20.8%	21.6%	21.2%	14.2%	
데이터 구축 및 컨설팅 서비스업	데이터 구축/가공 서비스업	62,223	72,805	81,188	17.0%	11.5%	14.2%	8.3%
	데이터 관련 컨설팅 서비스업	3,189	4,194	5,147	31.5%	22.7%	27.0%	32.1%
	소계	65,412	76,999	86,335	17.7%	12.1%	14.9%	9.1%
데이터 판매 및 제공 서비스업	데이터 판매/증개 서비스업	11,332	16,054	20,781	41.7%	29.4%	35.4%	27.7%
	정보제공 서비스업	71,033	81,838	93,289	15.2%	14.0%	14.6%	9.3%
	소계	82,364	97,891	114,071	18.9%	16.5%	17.7%	11.6%
데이터산업 전체	168,582	200,024	230,972	18.7%	15.5%	17.1%	10.9%	

<출처 : 데이터산업현황조사(2022)>

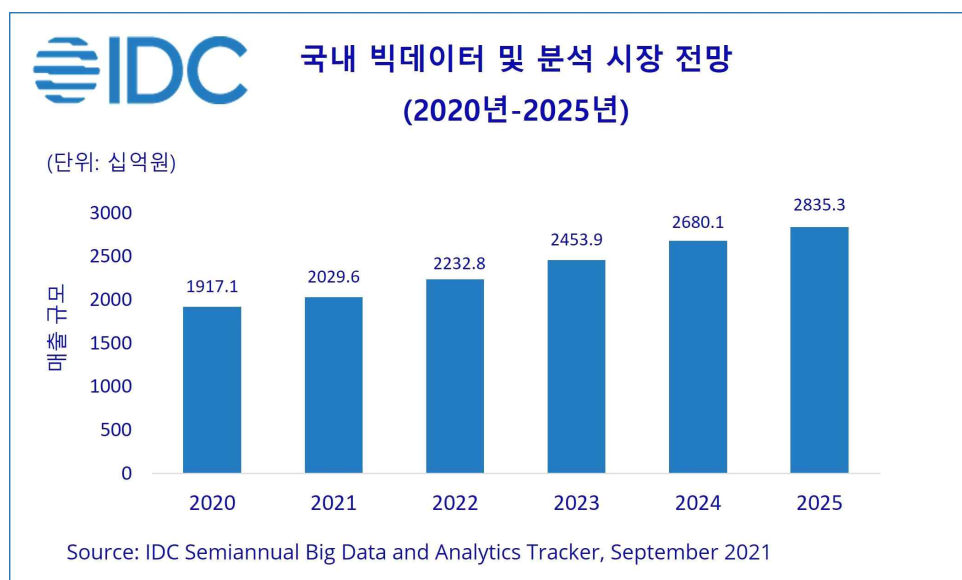
나. 데이터 시장 전망

IDC의 연구 보고서에 따르면, 국내 빅데이터 및 분석 시장은 2021년 전년 대비 5.5% 성장하여

254) 과학기술정보통신부, 2021 데이터산업현황조사, 한국데이터진흥원, 2022.

2조 296억원의 매출 규모를 형성할 전망이다이라고 한다. 해당 시장은 향후 5년간 연평균 성장률 6.9% 을 기록하며 2025년까지 2조 8,353억원 규모에 이를 전망이다. 다양한 산업에서 이전보다 더 많은 데이터를 확보하고 이를 활용하기 위한 수요가 높아지며 자체 데이터 플랫폼 구축 및 관련 시스템 도입이 적극 이루어지는 추세다.²⁵⁵⁾

[그림 5-1] 데이터 시장 전망



<출처 : IDC, 2022>

국내 빅데이터 및 분석 시장은 코로나 팬데믹으로 인한 시장 변동성에 대응하기 위해 다양한 산업에서 디지털 혁신을 가속화하며 전반적인 성장세를 보인 것으로 나타났다. 이는 국내 기업을 중심으로 높아진 데이터 가치에 따라 이를 효율적으로 활용하기 위한 비즈니스 워크플로우 재정립 및 다양한 업무 프로세스를 데이터화하는 움직임이 유효하게 작용한 것으로 분석됐다. 일반 기업의 데이터 기반 디지털 전환을 이루는 움직임과 더불어 공공기관의 빅데이터 플랫폼 구축 사업 지원과 금융산업의 마이데이터(본인신용정보관리업) 사업의 확산과 같이 다양한 산업에서 데이터를 중심으로 디지털 현대화가 가속화 될 전망이다. 보고서에 따르면 로우코드·노코드 플랫폼의 등장 및 데이터 시각화 분석 도구와 같은 솔루션의 성장으로 이전에는 데이터 과학자 또는 일부 개발자들의 영역이었던 전문화된 데이터 분석을 마케팅 부서 또는 사업 담당자와 같은 현업에서도 사용되고 있는 것으로 나타났다. 또한 기업에서도 데이터 활용 기술의 저변화를 통해 기업의 데이터 문맹률 문제를 개선하고, 비즈니스 민첩성을 확보하여 데이터 기반의 조직으로 거듭나기 위한 움직임을 보인다. 하지만 데이터를 다루는 인원이 확장됨에 따라 이에 상응하는 기술적 책임 또한 전보다 요구되고 있다.²⁵⁶⁾

255) IDC, 국내 빅데이터 및 분석 시장 전망 , 2021-2025, 2022.

2. 데이터산업 활성화 방안

가. 데이터산업의 정의

데이터산업이란 데이터와 관련된 경제활동이 이루어지는 산업이라고 할 수 있다. 데이터산업 법에서는 데이터산업을 경제적 부가가치를 창출하기 위하여 데이터의 생산·유통·거래·활용 등 일련의 과정과 관련된 행위와 이와 관련되는 서비스를 제공하는 산업으로 정의하고 있다.

크롤링이나 스크레이핑하여 데이터를 수집하는 경우, 비정형 데이터를 정형 데이터로 가공하거나 데이터에 레이블링하는 것을 데이터의 생산이라고 할 수 있다. 또한, 데이터거래소 등이나 개별적인 협의를 통하여 데이터를 판매하거나 이용에 제공하는 것은 유통이나 이용허락을 받거나 구매하는 것을 데이터 거래라고 할 수 있다. 이렇게 확보된 데이터를 기계학습이나 서비스에 이용하는 것을 활용이라고 할 수 있을 것이다. 따라서, 실제 데이터의 생애주기에 따른 모든 활동이 데이터산업에 포함된다고해도 과언이 아니다. 데이터산업을 영위하는 자를 데이터사업자라고 정의하고 있으므로, 데이터거래사업자, 분석제공자 등도 데이터사업자에 포함된다.

나. 데이터산업의 활성화 방안

(1) 공개된 데이터의 활용

데이터와 관련된 경제활동을 데이터산업이라고 정의하고 있기 때문에 데이터산업의 활성화는 데이터의 생성, 유통, 거래 및 활용을 위해 필요한 모든 방안을 강구하는 것이 정책적인 목표라고 하겠다. 본 연구와 관련하여 데이터산업의 활성화는 데이터 발명의 보호를 통한 산업활성화나 데이터 공개를 통하여 누구나 데이터를 이용하여 발명을 실시하고 그 발명을 통하여 AI서비스를 활용하거나 자신의 사업에 데이터를 활용하도록 하는 것이 포함된다. 따라서, 데이터를 공개하는 경우에는 해당 데이터를 분양받아서 활용할 수 있다면 데이터 생성이나 유통에 대한 비용을 줄일 수 있다는 점에서 투자를 집중할 수 있을 것이다.

(2) 데이터셋의 표준화

데이터의 궁극적인 목적으로 이용하는 것이라는 점에서 데이터셋의 품질관리가 필요하다. 이는 앞서 살펴본 데이터 공개의 항목과도 연계되는 사항이다. 실질적으로 데이터를 구축하는 과정에서 표준화가 되지 않을 경우에는 데이터의 전처리과정에서 시간과 비용이 소요될 수 있으며, 동일한 데이터임에도 서로 다른 결과가 나타날 수 있기 때문에 발명의 실시가 제대로 이루어지지 못할 수도 있기 때문이다. 따라서, 데이터의 표준화는 데이터산업 활성화는 물론 데이터 공개

256) <https://www.idc.com/getdoc.jsp?containerId=prAP48914822>. <2022.11.26. 접속>

의 실효성을 확보할 수 있는 방안이라고 하겠다.

AI-Hub를 구축하는 과정에서 정리된 인공지능 학습용 데이터 품질관리 가이드라인 v1.0에서는 구축데이터 품질관리에 대해 가이드라인을 제시하고 있다.²⁵⁷⁾ 인공지능 학습용 데이터의 구축 시 생산되는 원시데이터, 원천데이터, 라벨링 데이터의 품질을 확보하기 위한 제반 활동을 의미하며, 원시데이터와 라벨링 데이터의 품질관리가 학습데이터의 품질 측면에서 중요하다. 단, 구축하는 데이터가 인공지능 학습목적에 따라 텍스트, 음성, 이미지, 영상 등 다양한 유형의 데이터로 구분될 수 있고, 학습목적에 따라서는 분류, 인식, 검색, 식별, 예측 등 매우 다양한 목적으로 다시 세분화될 수 있어, 데이터 유형이나 학습목적별로 품질 확보 방안을 상세화하여 제시하는 것은 현실적으로 어려운 측면이 있다. 따라서 원시데이터와 라벨링 데이터가 갖는 공통의 데이터품질요건을 제시하며, 수행기관은 제시된 품질관리 요건을 참조하여 해당 구축데이터의 학습목적에 부합하는 데이터를 확보할 수 있도록 품질관리 방안을 구체화해야 한다.

제2절. 데이터 거래 체계 및 거래 현황

1. 데이터 거래 체계

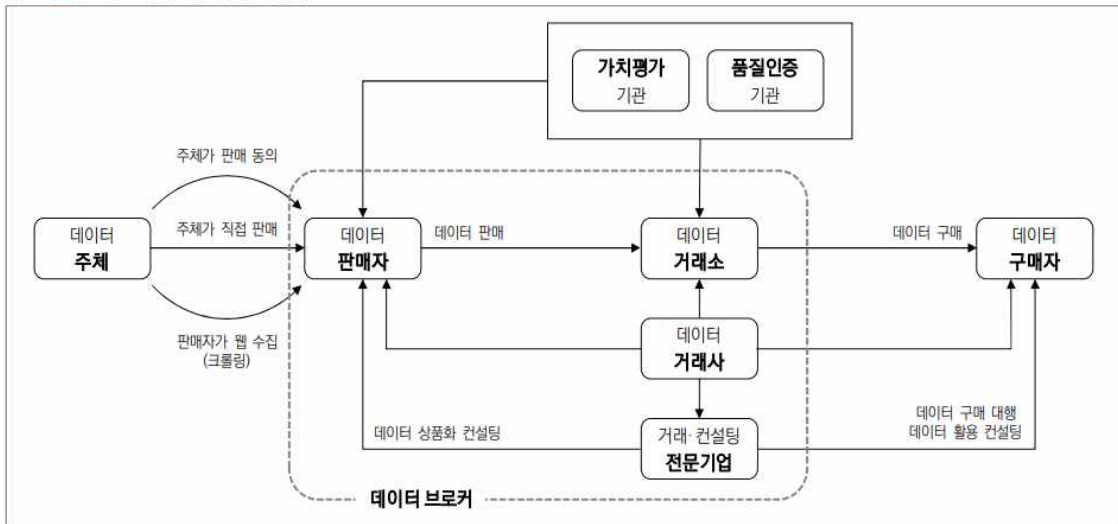
가. 데이터 거래의 의의

데이터 거래 체계를 단순화하여 나타내면 아래 그림의 주체-판매자-거래소(거래사)-구매자 순서와 같다. ‘데이터 주체’는 데이터를 최초로 만든 사람·기관·단체 등을 말하며 개인정보 보호법 제2조 제3호의 정보주체, 공공데이터를 생성하는 공공기관, SNS에 게시물을 올리는 개인, 사물데이터를 만들어내는 사업자 등이 포함된다. ‘데이터 판매자’는 데이터 주체로부터 데이터를 모아서 거래가능한 규모로 키우고 상품성을 높여서 시장에 상품으로 판매하는 자를 말한다. 데이터 판매자는 데이터 주체로부터 동의를 받아 데이터를 가져오거나, 금전적 대가를 주고 구입하거나, 웹사이트에 공개되어있는 데이터를 수집하기도 한다.

‘데이터 거래소’는 하나의 판매자가 보유한 데이터만 판매하는 형태도 있고, 여러 판매자가 자유롭게 거래에 참여할 수 있도록 하는 형태도 있다. ‘데이터 거래사’는 개인사업자 또는 데이터 거래소에 소속된 직원 신분 등으로 데이터 거래를 지원하는 사람이다.

257) 한국지능정보화진흥원, 인공지능 학습용 데이터 품질관리 가이드라인 v1.0, 2021, 30면.

[그림 5-2] 데이터 거래 체계도



<출처 : 입법조사처(2022)>

데이터 판매자·거래소·거래사는 각각 독립적으로 운영될 수도 있고, 서로 결합되어 하나의 사업체 안에서 운영될 수도 있다. 우리나라 데이터 산업법은 각각을 독립적인 것으로 전제한다. 이와 달리 세계 데이터 거래에서 큰 비중을 차지하는 미국의 ‘데이터 브로커(data broker)’는 판매자·거래소·거래사가 결합되어 있는 형태로 운영된다. 데이터 브로커는 여러 기업·기관으로부터 데이터를 수집하여 분석·가공한 다음 구매자에게 판매하는 역할을 하나의 사업체 안에서 모두 수행한다.

나. 데이터 거래소 현황

국내 데이터거래소가 다양하게 형성되고 있음을 알 수 있다. 공공영역에서 이루어지거나 민간과 공공이 협력하여 운영되기도 한다. 데이터댐이나 데이터레이크 사업에 따라 구축된 데이터의 거래 및 활용을 위한 거래소가 형성되고 있기 때문인 것으로 분석된다.

[표 5-2] 데이터거래소 현황

운영주체	데이터 거래소 사례
공공	<ul style="list-style-type: none"> • 공공데이터포털(행정안전부/한국지능정보사회진흥원 운영, data.go.kr) • AI허브(과학기술정보통신부/한국지능정보사회진흥원 운영, aihub.or.kr) • 데이터스토어(과학기술정보통신부/한국데이터산업진흥원 운영, datastore.or.kr) • 금융데이터거래소(금융위원회/금융보안원 운영, findatamall.or.kr) 등 다양함
민관합동 (정부지원 민간운영)	<ul style="list-style-type: none"> • 문화 빅데이터 플랫폼(주관: 한국문화정보원) • 지역경제 빅데이터 플랫폼(주관: 경기도청) • 통신 빅데이터 플랫폼(주관: 케이티) • 산림 빅데이터 플랫폼(주관: 한국임업진흥원) • 유통 빅데이터 플랫폼(주관: 매일방송) • 소방안전 빅데이터 플랫폼(주관: 소방청) • 헬스케어 빅데이터 플랫폼(주관: 국립암센터) • 스마트치안 빅데이터 플랫폼(주관: 경찰대학 치안정책연구소) • 교통 빅데이터 플랫폼(주관: 한국교통연구원) • 해양수산 빅데이터 플랫폼(주관: 한국해양수산개발원) • 환경 빅데이터 플랫폼(주관: 한국수자원공사) • 농식품 빅데이터 플랫폼(주관: 한국농수산식품유통공사) • 금융 빅데이터 플랫폼(주관: 비씨카드) • 라이프로그 빅데이터 플랫폼(주관: 원우연세의료원) • 중소기업 빅데이터 플랫폼(주관: 더존비즈온) • 디지털산업혁신 빅데이터 플랫폼(주관: 한국산업기술시험원)
민간	<ul style="list-style-type: none"> • SK C&C 데이터 큐레이션 플랫폼 AccuRator(아큐레이터) (skdt.co.kr/accurator) • KT 빅사이트 (bigsight.kt.com) • 빅데이터 마트 (bigdatamart.co.kr) 등 다양함

<출처 : 입법조사처(2022)>

2. 데이터의 거래

가. 거래 현황

전 세계에서 생성·복사·소비되는 데이터의 총량은 2010년 기준 2 제타바이트(zetabite,ZB)⁴에서 2020년 64.2 ZB로 10년간 32배 증가했다. 앞으로 데이터 총량은 더욱 빠르게 증가하여 2024년에는 181 ZB가 될 전망이다. 데이터 총량의 증가에 따라 데이터 거래도 빠르게 증가하고 있다. 외국의 데이터 브로커 시장 규모는 2019년 기준 2,326억 달러에서 연평균(CAGR) 5.8%씩 성장하여 2026년에는 3,452억 달러에 이를 전망이다. 특히, 세계 최대 규모의 데이터 브로커시장을 형성하고 있는 미국 시장 규모는 2018년 기준 약 1,832억 달러(약 220조 원)이다. 미국은 일찍부터 사후 거부권(opt-out) 방식을 채택했기 때문에 자유롭게 개인정보를 수집하여 분석하는 데이터 브로커 중심의 데이터 유통 시장이 성장할 수 있었다. 현재 미국에는 2,500여개 이상의 민간 데이터 브로커 기업이 운영 중이다.²⁵⁸⁾

국내의 데이터 거래 규모는 ‘데이터 판매 및 중개 서비스업’을 통해 추정할 수 있으며, 다음의 표와 같이 2020년 기준 1조 6,054억 원이다. 데이터산업의 전체 시장 규모는 20조 24억원인데, 이 중에서 데이터 거래가 차지하는 비중은 약 8%이다. 앞서 살펴본 미국의 2018년 데이터 브로커 시장 규모와 비교하면 약 0.7% 수준에 불과하다. 그러나 시장 규모는 작지만 증가율은 전년 대비 41.7%로, 데이터산업의 다른 세부 업종과 비교했을 때 상당히 높은 수준이다.

258) 박소영 외, 데이터 거래 활성화를 위한거래소·거래사·크롤링의 현황과 개선과제, 입법조사처, 2022.

나. 데이터의 확보

데이터를 확보하는 것은 플랫폼사업자 입장에서는 자사의 내부적인 데이터로 활용하는 것으로 보인다. 국내의 경우와 달리, 해외는 주로 데이터 거래소나 공개된 데이터를 활용하는 경우가 많은 것으로 보인다. 따라서, 공공 영역에서 데이터를 공개를 확대하는 것도 의미있다고 판단된다. 다만, 공공 부문의 데이터는 공공데이터 포털을 통해서 공개되기 때문에 어느정도 비중이 있는 것으로 보인다. 또한, 국가데이터정책위원회의 발족 등으로 데이터 정책이 정합성을 이루게 되면 데이터 거래 및 확보 등에서도 활성화를 기대할 수 있을 것으로 기대된다.

제3절. 기탁된 데이터의 활용 방안

1. 데이터의 제공

가. 제공 방식

데이터의 활용방안은 데이터 거래소를 통하여 활용되기도 하며, 다양한 공개 데이터셋을 이용하기도 한다. 이를 위해서는 직접적으로 데이터를 다운로드 하여 이용하는 방식과 플랫폼 사업자들이 공개하는 방식이다. 다운로드 방식은 특정 사이트에 데이터와 관련된 소스코드 등이 오픈소스 방식으로 공개되어있기 때문에 이를 무상으로 활용할 수 있는 것이다. 다만, 오픈소스 방식은 다양한 라이선스 체계를 갖추고 있기 때문에 라이선스를 위반할 경우에는 해당 데이터를 포함한 소스코드 등을 활용할 수 없으며, 저작권 침해 내지 특허권 침해 분쟁에 휘말릴 수 있다.

기탁을 받은 데이터에 대해서는 분양과 유사한 절차를 수립해야할 것이며, 이는 데이터를 복제하여 제공하는 방식이 될 것이다. 다만, 실질적으로 매체에 저장하여 제공하는 방식은 클라우드 환경에 맞지 않기 때문에 온라인 방식으로 제공하는 것이 행정 편의에 있어서 부합한 방식이라고 본다.

나. 이용 방식

깃허브나 AI허브와 같이 운영기관이 데이터 자체를 일반인에게 공개하는 경우에는 데이터를 다운로드 받아 해당 플랫폼에서 제시하는 계약조건에 따라 이용하면 된다. 기업에서 공개하는 경우에는 다르게 운영될 가능성도 있으나, API를 해당 기업에 요청하고 그 요청에 따라 이루어지는 방식이라면 데이터의 제공자가 요구하는 방식에 따라 진행하면 될 것이다. 대체적으로 요청방식은 데이터 다운로드 방식이나 API 연계 방식이다. API란, Application Programming Interface의 약어로 특정 프로그램의 기능이나 데이터를 다른 프로그램이 접근할 수 있도록 미리 정한 통신규약으로 네트워크상으로 서로 다른 프로그램간 기능·데이터를 연결하는 매개체의 역할의 역할을

한다. API는 해당 API에 접속할 수 있는 권한을 가진 자의 범위에 따라 ‘폐쇄형(Closed API)’ 과 ‘공개형(Open API)’ 으로 구분할 수 있다.²⁵⁹⁾

[표 5-3] API의 유형

폐쇄형 (Closed API)	· 해당 회사·기관 내부에서만 API를 통한 프로그램 접근 가능
공개형 (Open API)	· 회사·기관 외의 제3자에게도 API를 통한 프로그램 접근 허용 - 제3자의 범위에 따라 Partner (상호협약에 따른 파트너만 접근 가능), Member (자격 요건 등을 정하고 있는 커뮤니티에 속하는 멤버만 접근 가능), Acquaintance (일정한 요건을 갖춘 경우 누구나 접근 가능)로 구분하기도 함

<출처 : 금융위원회(2019)>

데이터가 API형태로 제공되면, 각 서비스 사업자가 개별적으로 데이터를 획득·가공하는 것이 아니라, API를 통하여 데이터 보유기업의 DB서버에 직접 접속하고, 데이터를 실시간 조회할 수 있게 된다. 이에 따라서 서비스 사업자는 데이터의 획득·가공 측면에 대한 고민을 덜고, 자신의 고객에 대한 데이터의 표시 측면만을 고려할 수 있게 되는 편리함을 갖게 된다.²⁶⁰⁾

이와 달리, 기탁기관을 통하여 데이터를 제공받는 경우에는 기탁기관 체결하는 계약조건에 따라 해당 데이터를 이용하게 될 것이다. 이용방식은 보안절차를 따라야 하겠지만, 암호화된 방식을 통하여 이루어도록 해야할 것이다.

2. 기탁된 데이터의 활용

AI나 데이터 발명에 있어서 미생물발명과 유사한 점을 들어, 재현가능성을 높일 수 있는 방법으로서 데이터를 공개하는 경우에는 해당 공개된 데이터를 제3의 기관에 기탁하거나 보관하는 방안을 제시하였다. 기탁된 경우라면 API방식으로 분양의 방식처럼 제공할 수 있을 것이다. 참고로, 현재 기탁된 미생물에 대해서는 공공데이터포털을 통하여 기탁된 목록에 대해서 API방식으로 공개하고 있는 것으로 확인된다.

259) 금융위원회, 금융권 Open API 보도자료, 2019.2.21.

260) 김보현 외, 데이터경제와 디지털금융, 법문사, 2021, 137면.

[그림 5-3] 특허기탁미생물 공보 API

DATA .GOV.KR 데이터찾기 국가데이터맵 데이터요청 데이터활용 정보공유 이용안내

오픈API 상세 f URL 복사

XML 특허청_특허기탁미생물_REST API ↗ 바로가기

기탁미생물 제도는 특허 출원시 명세서에 타인이 반복재현할 수 있도록 기재하여야 하나 미생물은 구조가 복잡하고 살아있는 것이어서 미생물에 관한 발명을 특허청원하는 경우 특허 명세서에 타인이 반복재현할 수 있도록 기재가 불가능하여 출원된 미생물을 공인된 기관에 기탁하고 공개 후 제3자가 분양받을 수 있도록 함으로써 명세서 기재사항을 보완하기 위한 제도

하지만 기탁 기관에 대한 정보를 별도로 기재하지 않아 검색으로 찾아야 하여 불편함을 초래하는 바, 명세서와 서지 내에 기탁코드를 포함한 정육이 있는 공보를 리스트화하여 제공

다른 사용자들이 활용한 데이터

로그인하셔서 다른 사용자들이 활용한 데이터를 추천받아 보세요

OpenAPI 정보

분류체계	과학기술 - 과학기술진흥	제공기관	특허청
관리부서명	한국특허청보철	관리부서 전화번호	02-6915-1446
API 유형	LINK	데이터포맷	XML
활용신청	29	키워드	특허 미생물, 특허기탁미생물
등록	2021-01-13	수정	2022-06-07
URL	http://eplus.kipris.or.kr/portal/data/uttl/DBII_000000000000224/view.do?menuNo=210007		
비용부과유무	무로	신청가능 트래픽	해당 기관의 정책에 따라 트래픽 수는 상이 할 수 있음
이용허락범위	이용허락범위 제한 없음		
참고문서			

<출처 : 공공데이터 포털(2022)>

제3의 기관에 기탁된 데이터는 AI 허브나 깃허브 등의 방식과 같이 누구나 자유롭게 이용할 수 있도록 데이터 플랫폼을 구축하는 것이 필요하다고 하겠다.

3. 법적 근거의 마련

가. 특허법상 데이터 기탁제도

다양한 부처에서 데이터 플랫폼을 구축할 것으로 보이며 관련 법제를 정비하고 있다는 점에서 기탁된 데이터를 활용하거나 이용에 제공할 수 있는 법적인 근거를 마련하는 것도 필요하다고 판단된다. 현재, 특허출원되는 다양한 명세서는 특허데이터로서 역할을 할 수 있을 것이며, 이를

활용하는 것은 지식재산에 대한 국가경쟁력 확보에도 의미있는 작업이라고 생각된다.

현행 미생물기탁제도의 경우에는 특허법이 아닌 특허법 시행령에서 근거하고 있다. 실무적인 측면이 강하기 때문에 특허법 보다는 동법 시행령에 근거토록 한 것으로 판단된다. 따라서, 데이터 기탁제도를 구체화하게된다면 미생물기탁제도에 준하여 시행령에 근거토록 하는 것도 무방하다고 본다.

나. 기탁기관의 운영

특허법에 근거하여 데이터기탁제도의 법적 근거가 마련될 경우에 특허청은 고시를 통하여 데이터기탁 운영기관에 관한 가이드라인을 제시할 필요가 있다. 다만, 제3의 전문기관을 지정하는 것에 대해서는 고려될 사항이 있음을 앞서 살펴보았다. 즉, 전문성과 공공성 및 지속적인 운영가능성이 담보되어야한다는 점이다. 미생물기탁기관처럼, 민간의 역량있는 기관도 고려할 수 있겠지만 데이터기탁기관은 제도 자체가 구체화된 것으로 보기 어렵기 때문에 시범사업 형태로 운영하는 것이 필요하다고 본다. 따라서, 특허청 산하기관과 업무협약을 통하여 데이터 기탁을 시범적으로 추진하는 것을 고려할 수 있을 것이다.

다음은 전문기관으로 지정할 수 있는 특허청 산하기관의 현황을 정리한 것이다. 다만, 설립목적과 주요사업 현황에 따라 전문기관을 지정할 수 있을 것으로 판단된다.

[표 5-4] 전문기관 지정 대상 기관

기관명	법적 근거	설립목적	주요 사업
한국지식재산연구원	발명진흥법 제51조	지식재산권 관련 국내외 분쟁 대응방안 수립, 지식재산권 동향 분석 및 신지식재산권 분야 연구	- 국내외 지식재산 조사·연구 - 국내외 지식재산 관련 국제교류 협력, 인식제고, 정보수집, IP전문 도서관 운영 등
한국발명진흥회	발명진흥법 제52조	발명진흥사업을 체계적·효율적 추진 및 발명가의 이익증진 도모	- 발명진흥에 대한 조사·연구 - 인재 양성, 교육시설 운영 - 발명교육·연구, 발명교원 육성 - 발명진흥을 위한 전시·행사 및 국제 교류·협력 등
한국지식재산보호원	발명진흥법 제55조의2	지식재산 보호에 관한 지원사업 추진	- 국내외 지식재산 보호에 관한 조사·연구, 기반 조성, 교육·홍보 및 국제협력 - 분쟁예방 및 대응지원 등
한국특허전략개발원	발명진흥법 제55조의5	산업재산전략 수립 및 연구개발 수행에 관한 사업을 효율적으로 지원	- 산업재산 정보 조사·분석 지원 - R&D 기획단계에서의 산업재산 정보의 동향 조사 - 연구개발과정에서의 산업재산 창출 전략 지원 등
한국특허정보원	발명진흥법	산업재산권 정보화 및	- 산업재산권 정보 DB 구축관리,

기관명	법적 근거	설립목적	주요 사업
	제20조의3 ('22.1.11. 본회의 통과)	산업재산권 정보의 활용기반 구축에 관한 사업을 효율적으로 지원	시스템 구축운영 및 연계 지원 - 산업재산권 정보의 가공·보급 지 원 등

물론, 전문기관으로 지정받을 수 있는 기관에 대해서는 발명진흥법이나 국회에 발의중인 「산업재산 정보의 관리 및 활용 촉진에 관한 법률」(이하, 산업재산정보 활용촉진법(안)이라 함)에 규정하는 것이 필요하다고 본다. 아직은 발의중이지만 산업재산정보 활용촉진법(안)에 근거하는 것이 기탁제도의 성격상 부합하다고 할 것이다.²⁶¹⁾

다. 기탁시스템의 구축 및 운영

산업재산정보 활용촉진법(안)에서는 산업재산 정보를 산업재산의 창출·보호 및 활용 과정에서 수집·생성되거나 이를 조사·분석·가공·연계하는 등의 방법으로 처리한 모든 종류의 지식 또는 자료로 정의하고 있다. 실질적으로 데이터를 포함한 다양한 지식재산 관련된 2차적인 정보까지도 포함하는 개념이다. 따라서, 학습데이터도 대상이 될 수 있다. 또한, 동법안에서는 산업재산 정보 데이터베이스²⁶²⁾와 산업재산 정보 시스템²⁶³⁾을 구축 및 운영토록 하고 있다. 즉, 법안 제9조에 따르면 특허청장은 업무 수행과정에서 수집·생성된 산업재산 정보를 체계적으로 관리하기 위하여 산업재산 정보 데이터베이스를 구축할 수 있다. 또한, 산업재산 정보 데이터베이스의 구축·관리 등을 위하여 관계 중앙행정기관의 장, 지방자치단체의 장, 공공기관의 장 및 관련 기업·법인 또는 단체 등에게 필요한 자료의 제출 또는 협조를 요청할 수 있다. 산업재산 정보를 체계적으로 관리한다는 측면에서 본다면, 기탁시스템의 구축 및 운영에 관한 사항을 포괄적으로 규정할 수 있을 것이다. 따라서, 데이터 기탁에 관한 사항을 해당 데이터베이스 내지 정보시스템과

261) 산업재산정보 활용촉진법(안)의 입법취지는 “산업재산 정보는 기업 등이 각고의 노력과 비용을 들여 만든 연구 결과의 핵심결정체로 구체적인 기술내용 뿐만 아니라 기업 등의 기술·시장전략, 발명자 등 핵심인력을 파악할 수 있는 다양한 부가정보를 담고 있음. 또한, 산업·기업별 기술경쟁력에 대한 객관적 비교분석을 통해 효율적인 연구개발(R&D)을 가능케 하는 등 경제적 부가가치 창출을 위한 기업의 경영자원으로서 산업재산 정보의 활용가치는 매우 높다고 할 수 있음. 최근 들어 글로벌 기술패권 경쟁이 심화되고 디지털 경제로의 전환이 가속화되면서 전 세계 산업·기술 변화의 흐름을 신속·정확히 파악하여 핵심기술을 확보하기 위한 객관적 의사결정의 도구로서 산업재산 정보의 중요성은 더욱 증대되고 있음. 그러나, 우리나라의 경우 「발명진흥법」의 일부 조항에서 산업재산 정보의 제공 및 활용에 대하여 간략히 규율하고 있을 뿐, 산업재산 정보의 수집·가공 등 관리, 제공 및 활용 전반을 체계적으로 뒷받침하기 위한 법적 기반은 미흡한 실정임. 이러한 입법공백은 국가와 기업의 산업·기술 전략 수립에 필수적인 산업재산 정보의 전략적 활용을 저해하여 주요국과의 첨단기술 선점 경쟁에서 뒤처지는 결과를 초래할 우려가 큼. 이에 기존 「발명진흥법」, 「특허법」 등의 산업재산 정보 관련 일부 조항을 이관하는 한편, 산업재산 정보의 체계적 관리 및 효과적 활용을 촉진하기 위한 근거조항들을 신설하여 「산업재산 정보의 관리 및 활용 촉진에 관한 법률」을 제정하고자 함. 이를 통해 범국가적 차원에서 산업재산 정보의 활용을 촉진함으로써 국가와 기업의 효율적인 연구개발 수행 및 산업·기술 전략 수립을 위한 기반을 마련하고, 나아가 국가 기술혁신 역량 및 산업경쟁력을 높이는 데 기여하려는 것임”이라고 밝히고 있다.

262) 산업재산정보 활용촉진법(안) 제2조 4. “산업재산 정보 데이터베이스”란 산업재산 정보를 체계적으로 정리하여 사용자가 검색하고 활용할 수 있도록 가공한 정보의 집합체를 말한다.

263) 산업재산정보 활용촉진법(안) 제2조 5. “산업재산 정보 시스템”이란 산업재산 정보의 수집·생성·가공·저장·관리·검색·송신·수신 및 그 활용과 관련되는 기기와 소프트웨어의 조직화된 체계를 말한다.

연계함으로써 가능하다고 할 것이다.

또한, 기탁된 데이터의 분양과 관련해서는 동법안 제10조에 따라, 특허청장은 산업재산 정보의 수집·검색·가공 및 분석 등의 업무를 효율적으로 수행하고 산업재산 정보 이용자에게 산업재산 정보를 원활하게 제공하기 위하여 산업재산 정보 시스템을 구축·운영할 수 있다. 또한, 산업재산 정보 시스템의 구축·운영을 위하여 필요한 경우 관계 중앙행정기관의 장, 지방자치단체의 장 및 공공기관의 장에게 해당 기관이 운영하는 정보 시스템과의 연계를 요청할 수 있다. 따라서, 동 규정에 근거하여 분양을 필요로 하는 사람에게는 해당 데이터를 분양할 수 있을 것이다. 다만, 구체적으로 데이터기탁 및 분양에 대한 구체적인 규정을 두는 것이 보다 보다 명확하다고 생각된다.

제4절. 데이터 산업법제와의 정합성 검토

1. 데이터보호 체계는 정합성을 갖는가?

데이터를 보호하는 법제유형은 앞서 살펴본 바와 같다. 저작권법 체계는 인접권으로서 데이터베이스제작자에게 배타적 권리를 부여하고 있다. 반면, 부정경쟁법리를 차용하고 있는 부정경쟁방지법, 콘텐츠산업 진흥법 및 산업 디지털전환 촉진법 등에서는 부정경쟁법리에 따른 채권적 권리를 부여하고 있다. 보호요건으로는 상당한 인적, 물적 투자를 한 경우를 상정하고 있으나, 결론적으로 보호방법은 다르다. 데이터베이스의 경우나 데이터의 경우나 투자의 정도는 다르지 않다. 그럼에도 불구하고, 보호방법이나 수준은 상이하다. 이는 데이터 또는 데이터베이스 산업에 대한 전반적인 상황을 염두에 둔 것이 아닌 개별 법제가 요구하는 입법에 치중한 결과이다.

다양한 법제에서 데이터에 관한 규정을 두면서, 데이터의 보호와 이용의 관계를 명확하게 설정하지 못한 면이 있다. 예를 들면, 저작권법 개정안에서는 저작재산권을 제한하여 정보분석이 가능하도록 하고있으나, 부정경쟁방지법 등에서는 별도 데이터자산을 권리화하고 있다는 점에서 데이터 정책에 대한 일관성이 문제될 수 있다. 권리자에게는 권리를 제한함으로써 인공지능 산업의 발전을 이끌어내고 있으나, 반면 산업데이터에 대해서는 보호범위에 포함함으로써 오히려 투자목적으로 만들어낸 것이 아님에도 불구하고 그 이용을 제한고있다는 점에서 데이터 정책의 정합성이 떨어질 수 있기 때문이다.

2. 개선 방안

다양한 데이터 법제가 제정되거나 개정되면서 우선순위나 보호법리의 적용 등이 상이하기 때문에 전체적인 체계정합을 위해서라도 데이터에 관한 기본법 체계를 수립하여야 한다. 모든 부처에서 데이터 법제를 제정하고자하는 의지가 강하기 때문에 데이터에 대한 기본법제를 두는 것이 필요하다. 데이터 산업법과 데이터 관련 법제의 체계를 제시하는 것도 의미가 있다. 또한, 현행

데이터 산업법상 국가데이터정책위원회 위원장을 국무총리로 하고 있으나, 민간 전문가를 공동위원장으로 하여 국가 CDO(chief data officer)를 지정할 수 있는 근거를 마련하는 것이 필요하다. 다양한 부처에서 데이터를 중요한 정책적 과제로 두면서 정책의 정합성이 떨어질 수 있기 때문이다.

제5절. 데이터 이용활성화를 위한 제도적 가이드라인의 제시

1. 기탁 데이터의 분양

기탁된 데이터의 활용 과정에서의 문제는 분양받은 데이터의 복제를 할 수 있을 것인지 여부이다. 해당 데이터는 공공의 영역에 있는 것이기 때문에 자유롭게 이용할 수 있으며, 선행기술로서 인정받을 수 있을지 여부이다. 발명자가 원치 않을 수도 있다는 점에서 기탁된 데이터를 활용하는 방안에 대해서는 선택할 수 있도록 절차를 마련하는 것이 필요하다. 따라서, 데이터를 활용하는 것이 정책적으로 부합하더라도 제도 초기에는 강제하는 것보다는 발명자가 선택할 수 있도록 하는 것이 바람직하다고 본다. 따라서, 운영과정에서 데이터 공개나 분양에 대해 의무나 강제요건으로 두기 보다는 선택적 사항으로 두는 것이 제도 정착에 효과적이라고 생각된다.

2. 산업재산정보 활용촉진법(안)과의 연계

가. 기탁기관에의 위임

특허제도는 특허법 등의 개정 사항으로 데이터공개에 관한 구체적인 규정을 두기는 입법정책적 차원에서 쉽지 않다. 따라서, 발명진흥법에 구체적인 지정요건 등을 규정하는 것도 하나의 방안이라고 하겠다. 다만, 발명진흥법의 분법 및 산업재산정보 활용촉진법(안)이 발의되어있기 때문에 데이터 기탁시스템 구축 및 운영에 관한 근거 법률로서 이용하는 것도 고려할 필요가 있다. 데이터 기탁에 대한 법적인 근거는 미생물기탁제도와 같이 특허법 시행령에 두고, 지정에 대한 사항은 고시를 통하여 산업재산정보 활용촉진법(안) 상의 특허정보원 등의 전문기관을 활용하는 것이다. 그 필요성은 앞서 살펴본 바와 같으며, 해당 기관은 공공기관으로써 운영에 대한 지속가능성을 담보할 수 있다는 점에서 기탁된 데이터의 유지관리 등에 있어서 안정적이라고 보기 때문이다. 산업재산정보 활용촉진법(안)에서도 업무를 위탁할 수 있는 규정을 두고 있기 때문에 충분히 가능하다고 생각된다.²⁶⁴⁾

나. 미국의 특허정보 데이터셋의 구축

264) 산업재산정보 활용촉진법(안) 제27조(업무의 위탁) 특허청장은 이 법에 따른 업무의 일부를 대통령령으로 정하는 바에 따라 문서전자화기관, 진단기관, 정보원, 전략원 또는 그 밖의 관련 기관·법인 또는 단체에 위탁할 수 있다.

특허정보를 활용한 데이터셋의 구축에 있어서도 의미있는 역할을 할 것으로 기대된다. 참고로, 2021년 7월 1일, 미국 특허상표청(USPTO)은 1,320만 개의 미국 특허 및 특허공보 문서 중에서 어떤 것이 인공지능(AI)을 포함하는지를 식별하는 신규 인공지능 특허 데이터셋(Artificial Intelligence Patent Dataset, AIPD)²⁶⁵⁾을 발표하였다. 주요 내용은 USPTO는 이번 신규 인공지능 특허 데이터셋을 통해 연구자, 정책입안자 및 일반 대중이 발명에 대한 인공지능에 관한 영향을 확인할 수 있도록 지원할 것이라고 밝혔다. AIPD는 인공지능 구성요소 기술(AI component technologies)에 대한 기계학습 모델을 사용하여 구축되었으며, 데이터셋을 생성하는데 사용된 기계학습 접근방식은 기존 방식과 비교하여 우수한 성능을 수행한다. 이번 데이터셋은 총 2개의 파일로 구성되어 있는데, 첫 번째 데이터 파일은 1976년부터 2020년까지 등록된 미국 특허와 2020년까지의 공개특허공보(pre-grant publications)를 식별하고 여기에는 기계학습, 자연언어 처리, 컴퓨터 비전, 음성, 지식 처리, AI 하드웨어, 진화 알고리즘(evolutionary computation) 등이 포함된다. 두 번째 데이터 파일은 기계학습 모델을 학습하는데 사용되는 특허 문서가 포함되어 있다.²⁶⁶⁾

다. 산업재산 데이터셋의 구축

산업재산정보 활용촉진법(안)이 제정될 경우, 우리도 특허청에 등록되어있는 명세서 등의 산업재산 정보를 적극적으로 활용할 수 있도록 한다면 기술 발전에 있어서도 유의미한 정책이 될 것이다.

국회 통과를 조건부로 하지만, 산업재산정보 활용촉진법(안)에서는 산업재산 정보를 산업재산의 창출·보호 및 활용 과정에서 수집·생성되거나 이를 조사·분석·가공·연계하는 등의 방법으로 처리한 모든 종류의 지식 또는 자료로 정의하고 있다. 산업재산정보는 특허청에 출원된 정보를 기반으로 하기 때문에 특허청이 데이터셋을 구축하여 이용에 제공하는 것이 가능하다. 동법안에서는 특허청장은 국가 및 민간 연구개발의 효율성을 높이고 기술·산업 관련 전략의 수립·추진 및 평가 등을 효과적으로 지원하기 위하여 「특허법」 등 관련 법령에 따라 공개된 산업재산 정보를 수집·가공하여 이용하거나, 수집·가공된 정보를 제공할 수 있다(안 제14조).

3. 기계학습에서의 공정이용 가이드라인

기계학습 과정에 대한 공정이용 여부에 대한 판단기준은 이용의 목적이나 성격과 시장대체성이다. 저작물의 종류나 용도는 그 활용에 따른 문화적 향유를 위한 수단이 아니라는 점에서 크게 다루어질 것은 아니기 때문이다. 또한, 저작물의 전체에 대해서 보더라도, TDM을 통하여 확보되

265) 자세한 내용은 Artificial Intelligence Patent Dataset 페이지에서 확인이 가능하다. (<https://www.uspto.gov/ip-policy/economic-research/research-datasets/artificial-intelligence-patent-dataset>)

266) Alexander V. Giczy et. al., Identifying artificial intelligence (AI) invention: A novel AI patent dataset, USPTO Economic Working Paper No. 2021-2 August 2021.

거나 사실 정보에 불과한 경우라면 비중은 큰 의미가 없다.

가. 개정안 검토

유럽연합이나 일본 등 여러 나라에서 데이터를 확보하려는 목적으로 저작권법을 개정해, 저작권자의 허락 없이 저작물을 활용해 데이터 마이닝이 가능하게 하고 있다. 우리나라도 정보분석을 할 때는 저작권자의 허락 없이 저작물을 이용할 수 있도록 관련 법조문을 거의 전부 고치는 수준의 저작권법 전부 개정안이 발의되어 있다.

[표 5-5] 저작권법 개정안

제43조(정보분석을 위한 복제·전송) ① 컴퓨터를 이용한 자동화 분석기술을 통해 다수의 저작물을 포함한 대량의 정보를 분석(규칙, 구조, 경향, 상관관계 등의 정보를 추출하는 것)하여 추가적인 정보 또는 가치를 생성하기 위한 것으로 저작물에 표현된 사상이나 감정을 향유하지 아니하는 경우에는 필요한 한도 안에서 저작물을 복제·전송할 수 있다. 다만 해당 저작물에 적법하게 접근할 수 있는 경우에 한정한다.

② 제1항에 따라 만들어진 복제물은 정보분석을 위하여 필요한 한도에서 보관할 수 있다.

저작권법 개정안이 의미하는 저작물을 향유한다는 것은 인간이 교육이나 학습 또는 일상생활에서 읽고, 보고, 듣고 또는 느끼는 감정 활동을 말한다. 하지만 기계는 이런 활동 자체를 할 수 없으므로 기계가 활용한다는 것은 저작물에 투여된 저작자, 즉 창작자의 의도를 알 수 없이 그 특징값만 분석해 이용한다는 것이다. 이는 저작물의 표현에 대한 사람의 인식을 동반하지 않고, 해당 저작물을 컴퓨터에 의한 정보처리 과정에서 이용하거나 기타의 이용이라면 침해를 인정하지 않겠다는 의미이다. 사실 저작권법은 인간이 저작물에 담겨진 저작자의 의도와 작품의 가치를 이해할 수 있도록 보고나 듣는 것을 전제하고 있다. 이로써, 데이터셋을 구축하는 과정이 보다 용이해질 것으로 기대한다.²⁶⁷⁾

나. 공정이용 규정 검토

(1) 데이터 이용으로서 기계학습

인공지능의 신경망이 인간의 것과 큰 차이가 없는 것을 본다면 인공지능의 학습방식도 인간의 학습방법과 크게 다르지 않다. 기계학습 과정에서 인공지능은 저작물을 분석하여 특징들을 수치화한다. 수치화한다는 것은 인간의 뇌에 저작물을 복제하는 것이 아닌 인간이 이해할 수 있는 특

267) 김윤명·이민영, 인공지능의 생각, 이다북스, 2022, 119면.

정만을 기억하는 것을 의미한다. 이미지의 경우, 해당 이미지의 특징을 분석하여 수치화하며, 텍스트의 경우는 말뭉치(corpus)를 인덱싱하여 데이터 값을 부여한다. 이 과정은 저작물의 의미를 이해하거나 활용하는 것이 아닌 단어나 문장의 구성을 분석하는 것이다.

이러한 이용은 저작권법이 의도하는 인간의 저작물 이용방식과는 차이가 있다. 분석된 결과물은 저작물 그 자체가 아닌 저작물에 담겨있는 특성, 패턴, 스토리, 구조 등의 것이다. 따라서, 인공지능이 학습하는 것은 메모리에 복제하는 것이 아닌, 데이터를 이해하는 상태로 분석하고 추상화하는 상태이기 때문에 복제가 일어나는 것으로 보기 어렵다.²⁶⁸⁾ 인공지능의 학습과 관련되어 적용할 수 있는 저작권법상 제한규정은 공정이용에 관한 일반조항이 최선이다. 즉, 저작권법상 저작물의 공정한 이용에 관한 규정이다. 공정이용에 대한 고려에서 필요한 것은 헌법 제22조에 따른 창작자의 권리를 보호하는 것의 해석이다. 공정이용은 헌법상의 창작자 보호, 저작권법상의 제한규정이외에도, 저작물의 통상적인 이용 방법과 충돌하지 아니하고 저작자의 정당한 이익을 부당하게 해치지 아니하는 경우에는 저작물을 이용할 수 있기 때문이다. 창작자의 권리를 저작권법으로 보호하지만, 저작권법은 공정한 이용도 도모하고 있다. 이러한 점에서 볼 때, 공정이용은 권리보호와 함께 중요한 법정책적 임무를 부여받은 것으로 이해된다.

(2) 통상적인 이용인지 여부

먼저, 공정이용 제1요건인 통상적인 이용 여부에 대해 본다. 인공지능의 학습은 정보를 분석하고, 그 패턴이나 특징 값을 분석하여 이용하는 것이기 때문에 일반적인 저작물의 이용 형태와는 다름을 알 수 있다. 즉, 인간의 저작물 이용과 달리, “저작물 그 자체를 향유하는 것이 아니라 단지 정보를 습득하고자 그 저작물을 구성하는 언어나 기호 등을 통계적으로 분석하는 경우에는 그 저작물 등을 복제하거나 번역 등 필요한 형태로 변환할 수 있다.”²⁶⁹⁾ 기계학습은 저작물 등의 복제나 단순한 2차적 저작물의 작성이 아닌 창작적 이용(creative use)²⁷⁰⁾이라는 점에서 보면 공정이용에 해당할 가능성이 높다. 저작권법의 목적은 문화의 창달이며, 기존의 저작물을 향유하는 과정에서 새로운 창작적 표현을 만들어낼 수 있는 동인으로 작용하게 된다. 창작적 이용을 인정하는 수단으로써 공정이용은 이러한 저작권법의 일반적인 목적규정을 통하여 확인할 수 있는 것이다. 미연방대법원은 변형적 이용을 “새로운 표현, 의미 또는 메시지를 가지고 원저작물을 변형해, 다른 목적 또는 다른 성질을 가지고 원저작물의 표현에 무언가 새로운 것을 추가한 경우”²⁷¹⁾라고 판시하였다. 이처럼 변형적 이용에 해당하는 경우라면 “원저작물과는 다른 목적의

268) 만약, 기계학습 과정이 저작권 침해행위로 본다면 인공지능의 학습은 불가능할 수 있다. 아니면 빅 데이터를 확보할 수 있는 인터넷기업들만이 경쟁에서 살아남을 수 있는 것이다.

269) 임원선, 실무자를 위한 저작권법, 저작권위원회, 2014, 231~232면.

270) 공정이용인지 여부에서 저작물의 변형적 이용(transformative use)에 대해 판단한다. 여기서는 이를 창작적 이용이라고 표현하나 그 의미는 변형적 사용과 다르지 않다. 정상조, 딥러닝에서의 학습데이터와 공정이용, LAW&TECHNOLOGY 제16권 제1호 (통권 제85호) 2020, 12면.

271) Campbell v. Acuff-Rose Music, Inc., 510U.S.569(1994).

이용이고 원작의 성질에 대한 새로운 표현을 부가해 변화를 준 것” 272)에 해당한다. 새로운 가치를 부여함으로써 저작자가 의도했던 가치 이상을 더해주는 경우라면, 이는 공정이용으로 판단될 가능성을 높이는 것이다. 이러한 맥락에서 인공지능은 저작물을 학습하고, 그로써 새로운 가치를 만들어낼 수 있는 상태를 구축한 것으로 볼 수 있다.

(3) 종류 및 용도의 적합성

기계학습은 인공지능의 성능을 향상시키기 위한 것이라는 점이 주된 용도이다. 이때 학습 주체는 인공지능 알고리즘이다. 알고리즘이 프로그래밍 된 바대로 데이터의 패턴이나 특징을 인식하고 분석하여, 의도한 결과를 만들어 내거나 또는 의도성을 가지고 학습하는 것이다. 따라서, 기계학습 과정은 용도라는 것이 저작물을 향유하는 과정과는 다른 용도라는 점에서 본질적인 차이가 있다. 공정이용 규정에서의 용도와 기계학습에서의 용도는 다른 기준점에서 봐야하며 저작물의 유형에 따라 달리 봐야하는 것은 맞다. 즉, 기능적, 사실적 저작물과는 차이(差異)가 있다.

인간이 학습하는 것은 다양성 확보를 위한 것인 것처럼, 기계학습도 인공지능의 다양한 기능의 향상을 위한 것으로 궁극적으로는 인간의 사고와 유사한 범용 인공지능을 개발하기 위한 것으로 볼 수 있다. 이 과정에서 끊임없는 학습이 이루어질 것이며, 다른 차원의 기술개발도 병행될 것이다. 궁극적으로 데이터 없이도 학습할 수 있는 강화학습 형태로 발전해 갈 것이다.

정리하자면, 인공지능의 기계학습과 인간의 문화향유는 기본적인 용도나 목적이 상이하다. 따라서, 용도의 차이라는 점에서 본다면 기계학습은 저작물을 향유하는 것이 아닌 데이터에 담겨진 패턴과 특징값을 찾는 것이기 때문에 공정이용에 해당한다. 다만, 학습용으로 제작된 데이터베이스라면 예외로 두는 것이 바람직하다.²⁷³⁾

(4) 비중 및 중요성²⁷⁴⁾

딥러닝에 의한 학습데이터의 이용의 경우에는 사실저작물이나 기능저작물뿐만 아니라 비록 예술저작물이라고 하더라도 저작물의 문예적, 심미적 가치를 이용하는 것이 아니라 그 속의 데이터로서의 비표현적 가치를 이용하는 변형적 이용이 많기 때문에 저작물의 종류 및 용도가 커다란 영향을 미치지 않을 것이다.

딥러닝에서의 학습데이터 이용에 있어서 투입단계와 산출단계의 원저작물 이용분량이 서로 다

272) 한국정보법학회 지음, 인터넷, 그 길을 묻다, 중앙Books, 2012, 544면.

273) 우리 저작권법도 데이터베이스 제작자의 권리에 대한 제한규정을 두고 있다. 다만, “제한사유가 무제한적으로 활용된다면 데이터베이스제작자의 경제적 권익에 큰 위협이 될 수도 있을 정도이다. 그러한 점을 감안하여 저작권법은 “당해 데이터베이스의 통상적인 이용과 저촉되는 경우에는 그러하지 아니하다”고 규정하고 있다. 따라서 교육, 학술, 연구에 종사하는 기관을 고객층으로 하여 제작된 데이터베이스가 있을 경우에 이를 유상으로 이용하지 않고 무단 복제하여 이용한다면 그것은 법이 허용하는 범위를 넘어선 것이라고 보아야 할 것”이라고 한다. 이해완, 저작권법, 박영사, 2019, 1021면.

274) 정상조, 딥러닝에서의 학습데이터와 공정이용, LAW&TECHNOLOGY 제16권 제1호 (통권 제85호) 2020, 18~19면.

를 수 있기 때문에, 산출단계의 경미한 이용이 공정이용으로 허용된다는 점을 명시적으로 규정하는 저작권법 개정도 검토를 해볼 만한 가치가 있다.

일본 개정 저작권법에 의하면 컴퓨터를 이용한 정보처리에 의한 새로운 지식 또는 정보를 창출함으로써 저작물이용의 촉진에 이바지하는 정보검색서비스라거나 정보해석서비스는 타인의 공개된 저작물의 이용용도 및 분량에 비추어 저작권자의 이익을 부당하게 해치지 않는 소위 경미한 이용으로서 적법한 이용에 해당된다.

(5) 시장대체성

다음으로, 시장대체성 여부에 대해 본다.²⁷⁵⁾ 시장대체성의 범위는 “원저작물 자체뿐만 아니라 2차적 저작물의 시장이나 가치도 포함된다.”²⁷⁶⁾고 한다. 기계학습은 인간의 이용이 아닌 정보 내용이나 표현의 특성을 학습하기 때문에 일반적인 이용형태와 다를뿐더러, 일반 소비자에게 제공되는 것과는 다른 시장을 형성하게 될 것이다. 기계학습은 인공지능의 지능 수준을 높이기 위한 것에 불과할 수 있기 때문이다.²⁷⁷⁾ 다만, 기계학습을 위한 별도의 정보를 구축하여 제공한다면 이는 시장대체성을 인정받을 것이다.

일본은 이러한 상황을 입법론으로 정리하고 있다. 즉, 정보 분석을 위한 데이터 등의 이용을 공정이용으로 규정하고 있는 것이다.²⁷⁸⁾ 따라서, 인공지능을 학습하는 과정에서 이루어지는 저작물의 습득 자체는 학습 매커니즘이지 저작물을 복제하여 배포하는 것으로 보기 어렵다. 또한, 인공지능의 학습 형태에서 데이터 등의 정보를 이용하는 것은 “저작물 등을 구성하는 언어나 기호 등의 요소들 또는 그들이 관계 등을 분석하려는 것일 뿐 그 저작물 등 자체를 이용하고자 하는 것이 아니고, 그 분석의 결과물을 그 저작물 등과는 전혀 별개로서 그에 원저작물이 드러나지 않으므로 그 저작물 등의 통상적인 이용과 충돌하거나 저작자의 정당한 이익을 부당하게 저해할 우려가 적다.”²⁷⁹⁾고 생각한다. 인공지능의 학습과 유사하게 적용할 수 있는 기존 사례는 썸네일 검색으로 볼 수 있다. 인터넷상에 공개된 정보를 크롤링하여, 이를 데이터베이스화하고 검색어가 입력되면 해당 정보를 제공하는 것은 기계학습 매커니즘과 유사기 때문이다. 썸네일 검색은 그 결과를 보여주는 것이지만, 크롤링은 기계학습을 위한 데이터 수집 내지 수집된 데이터를 인덱싱

275) Robert Merges, Peter Menell, Mark Lemley, Intellectual Property in the New Technological Age, Wolters Kluwer, 2012, p.646.

276) 최호진, “썸네일 이미지와 공정이용”, LAW&TECHNOLOGY 제8권 제3호, 2012, 70면.

277) 물론, 수많은 인공지능에 탑재할 목적으로 이용했다면 시장대체성을 상실할 가능성도 부인하기 어렵다.

278) 일본 저작권법 제47조의7(정보해석을 위한 복제 등) 저작물은, 전자계산기에 의한 정보해석(다수의 저작물 기타의 대량의 정보로부터, 당해 정보를 구성하는 언어, 음, 영상 기타의 요소와 관련된 정보를 추출, 비교, 분류 기타의 통계적인 해석을 행하는 것을 말한다. 이하 이 조에서 같다)을 하는 것을 목적으로 하는 경우에는, 필요하다고 인정되는 한도에서 기록매체에의 기록 또는 번안(이에 의하여 창작한 2차적저작물의 기록을 포함한다)을 할 수 있다. 다만, 정보해석을 하는 자의 이용에 제공하기 위해 작성된 데이터베이스저작물에 대하여는 그러하지 아니하다.

279) 임원선, 실무자를 위한 저작권법, 저작권위원회, 2014, 232면.

하여 관리 값을 부여하는 것과 같기 때문이다. 기계학습은 특정 값을 분류해내는 과정이라는 점에서 차이가 있지만, 정보를 분석하여 분류하는 과정에서는 다르지 않다. 물론, 정보를 분석하여 이용 가능한 상태에 놓인 것은 공개되거나 출시된 것이 아니기 때문에 시장대체를 논하는 것이 타당한 것인 지하는 지적도 가능하다. 그렇지만, 시장대체성을 해당 저작물의 이용과정에서 고려하는 예측에 대한 판단이기 때문에 이를 부인할 필요는 없다. 이상과 같이, 썸네일 형태로 검색 결과에 노출되는 것도 정보의 위치를 알려주는 것으로 공익적 성격으로써 공정이용이 인정되고 있고,²⁸⁰⁾ 대법원도 같은 취지로 저작권 침해를 부인한 바 있다.²⁸¹⁾

다. 소결

기계학습에서 저작물이 이용되는 매커니즘은 원저작물의 시장을 대체할 수 있는 형태로 보기 어렵다. 기계학습의 공정이용을 고려할 수 있는 사례로는 검색엔진의 크롤링과 검색결과의 현시를 들 수 있다. 크롤링 과정에서 많은 데이터를 수집하지만, 그 자체는 정보검색의 용이성을 위한 것이기 때문에 공정이용으로 보는 것이다. 물론, 시장대체성을 넓게 보아 인공지능을 통하여 형성할 수 있는 시장까지 볼 가능성도 부인하기 어렵다. 그렇지만, 저작물의 이용이라는 것은 미래의 특정 시점에 도래하는 기술적 수준에 의한 것을 대상으로 제한하기 어렵다. 또한, 공정이용 규정 자체가 기술적 발전에 대응하기 위한 것이며, 그 요건에 해당하는 경우라면 면책을 부여하는 것이 타당하다. 만약, 기계학습이 저작자의 권리를 심대하게 침해하는 경우가 발생한다면 판례 또는 입법론적으로 대응하는 것이 타당하다.

280) 구글검색엔진의 높은 수준의 변형적 이용과 사회적 편익을 제공한다는 점에서 공정이용에 해당한다고 판시한 바 있다(Perfect10, Inc. v. Amazon, Inc., 508 F.3d 1146(9th Cir, 2007)).

281) 피고인 회사가 썸네일 이미지를 제공한 주요한 목적은 보다 나은 검색서비스의 제공을 위해 검색어와 관련된 이미지를 축소된 형태로 목록화하여 검색서비스를 이용하는 사람들에게 그 이미지의 위치 정보를 제공하는 데 있는 것이지 피고인들이 공소외인의 사진을 예술작품으로서 전시하거나 판매하기 위하여 이를 수집하여 자신의 사이트에 게시한 것이 아닌 만큼 그 상업적인 성격은 간접적이고 부차적인 것에 불과한 점, 공소외인의 사진작품은 심미적이고 예술적인 목적을 가지고 있다고 할 수 있는 반면 피고인 회사의 사이트에 이미지화된 공소외인의 사진작품의 크기는 원본에 비해 훨씬 작은 가로 3cm, 세로 2.5cm 정도이고, 이를 클릭하는 경우 독립된 창으로 뜬다고 하더라도 가로 4cm, 세로 3cm 정도로 확대될 뿐 원본 사진과 같은 크기로 보여지지 아니할 뿐만 아니라 포토샵 프로그램을 이용하여 원본 사진과 같은 크기로 확대한 후 보정작업을 거친다 하더라도 열화현상으로 작품으로서의 사진을 감상하기는 어려운 만큼 피고인 회사 등이 저작물인 공소외인의 사진을 그 본질적인 면에서 사용한 것으로는 보기 어려운 점, 피고인 회사의 검색사이트의 이 사건 썸네일 이미지에 기재된 주소를 통하여 박범용의 홈페이지를 거쳐 공소외인의 홈페이지로 순차 링크됨으로써 이용자들을 결국 공소외인의 홈페이지로 끌어들이게 되는 만큼 피고인 회사가 공소외인의 사진을 이미지검색에 제공하기 위하여 압축된 크기의 이미지로 게시한 것이 공소외인의 작품사진에 대한 수요를 대체한다거나 공소외인의 사진 저작물에 대한 저작권침해의 가능성을 높이는 것으로 보기는 어려운 점, 이미지 검색을 이용하는 사용자들도 썸네일 이미지를 작품사진으로 감상하기보다는 이미지와 관련된 사이트를 찾아가는 통로로 인식할 가능성이 높은 점 및 썸네일 이미지의 사용은 검색사이트를 이용하는 사용자들에게 보다 완결된 정보를 제공하기 위한 공익적 측면이 강한 점 등 판시와 같은 사정 등을 종합하여 보면, 피고인 회사가 공소외인의 허락을 받지 아니하고 공소외인의 사진작품을 이미지검색의 이미지로 사용하였다고 하더라도 이러한 사용은 정당한 범위 안에서 공정한 관행에 합치되게 사용한 것으로 봄이 상당하다. 대법원 2006.02.09. 선고 2005도7793 판결.

제6장. 결 론

본 연구는 AI발명에서 필수적인 구성요소인 데이터에 관한 것으로, 구체적으로는 기술공개로서 데이터의 공개 및 기탁 등에 관한 것이다. 특허권이라는 독점권을 부여함에 따라 이루어지는 기술공개는 특허청과 발명자와의 공적 계약이다. 계약조건으로 쉽게 실시할 수 있을 정도의 기술공개를 요구하고 있으며, 해당 조건이 만족되지 아닐 경우 해당 특허권은 무효화될 수 있다. 그렇기 때문에 기술공개에 대해서는 엄격하게 할 경우에는 특허권자의 특허권이 제한될 수 있으며, 반대의 경우에는 충분한 기술공개가 이루어지지 않음으로써 발명의 진흥이라는 원래의 취지를 살릴 수 없게 된다. 이러한 점에서 AI발명의 기술공개는 쉽지 않다.

특허법에서 기술공개의 충분성(sufficiency of disclosure)은 특허출원이 해당 분야에 숙련된 자가 과도한 실험 없이 발명을 제조하고 사용할 수 있도록 충분한 정보를 제공해야 한다는 요건을 말한다. 전통적인 발명은 기술공개를 통해서 다양한 기술발전을 가져왔지만, AI발명은 기술적인 사상이 블랙박스화 되는 경우가 있다는 점에서 전통적인 발명의 기술공개와는 다른 방식이 필요할 수 있다. 다만, 미생물발명 등과 같이 전통적인 발명과도 유사한 면이 있다는 점은 상정적이다. 실제, 인공지능경망을 통하여 구현된다는 점과 미생물 발명의 내부적인 구조에 따라 구현된다는 점에서 유사성이 있기 때문이다.

데이터 기탁제도는 블랙박스화되고 있는 AI발명의 특성에 따라 유사한 점이 발견되었고, 다른 유사제도와 비교했을 때 이를 뒷받침할 수 있는 제도로서 합리적이라고 판단되었다. 무엇보다, AI발명에서 사용되는 데이터셋은 재현가능성과 투명성 및 신뢰성을 담보할 수 있기 때문에 이를 공개하는 것은 의미있는 일이다. 이를 위하여, 기술공개의 충분성을 입증하는 방법 중 하나는 발명을 위해 사용되는 데이터를 공개적으로 접근가능한 저장소에 저장하는 것이다. 기술공개의 충분성을 위한 방안으로써 데이터를 기탁하는 것에는 몇 가지 이점이 있다. 첫째, 인공지능발명에 대한 투명성을 향상시킬 수 있다는 점이다. 즉, 데이터를 보관하면 다른 사람이 발명의 결과와 결론을 검토하고 검증할 수 있어 특허출원의 투명성과 신뢰성을 높일 수 있다. 둘째, 인공지능 발명의 향상된 재현성을 가져올 수 있다는 점이다. 즉, 데이터를 이용할 수 있게 함으로써, 다른 사람들은 결과를 재현하고 발명의 타당성을 확인할 수 있으며, 이는 특허출원을 강화할 수 있다. 셋째, 관련 분야에서의 향상된 협업을 가능하게 할 수 있다는 점이다. 즉, 공개적으로 액세스할 수 있는 기탁기관에 데이터를 보관하면 다른 사용자가 자신의 연구개발 노력을 위해 데이터에 액세스하고 사용할 수 있으므로 연구자와 조직 간의 협업을 촉진할 수 있다. 이로써, 협업을 촉진하고 과학과 기술의 발전을 지원하는 데 도움이 될 수 있다.

이와 같이, 기술공개의 충분성을 위해 데이터를 기탁하는 것은 특허출원이 정확하고 신뢰할 수 있으며 투명성을 보장하는 데 도움이 될 수 있다. 또한, 데이터 기탁은 데이터에 대한 접근성을 개선하고 과학기술의 발전을 촉진하는 것과 같은 더 광범위한 이점을 가질 수 있다. 데이터를

다른 사람이 이용할 수 있게 함으로써, 그것은 혁신과 분야의 발전을 지원할 수 있을 것이다.

마지막으로 공개되거나 기탁된 데이터를 활용하여 산업을 육성할 수 있는 방안을 살펴보았다. 실질적으로 데이터에 따라 인공지능의 능력과 실제 서비스되는 인공지능의 품질이 달라질 수 있기 때문에 많은 데이터를 공개하여 이용할 수 있도록 뒷받침하는 것은 무엇보다 중요한 산업정책이다. 따라서, 이를 뒷받침할 수 있도록 전문기관을 두거나 또는 전문기관을 통하여 기탁 내지 관련 업무를 수행토록 하는 것을 제안하였다. 전문기관은 특허관련 전문성, 특허데이터 이용의 확장가능성 등 다양한 분야에서 전문성과 공공성을 담보할 필요가 있다. 이러한 점을 염두에 두고 제3의 기관에 대한 필요한 조건들을 제안하였다.

참고문헌

■ 국내문헌

- 강영주, 인공지능을 활용한 민사분쟁 해결에 관한 고찰, 안암법학 no.64, 2022
- 권오승 외, 제조물책임법, 법문사, 2002
- 권지현, AI발명과 데이터의 실효적 특허보호방안, 法學研究 第24輯 第1號, 2020
- 권지현, AI발명에 있어서 데이터의 물건특허 인정방안, 서울法學 제28권 제4호, 2021
- 권지현, 비기술적 요소가 포함된 AI발명의 진보성 판단기준, 법학논총 제48집, 2020
- 권태복, 제4차 산업혁명과 특허전략, 한국지식재산연구원, 2019
- 금융위원회, 금융권 Open API 보도자료, 2019.2.21.
- 김 영, 특허의 적극적 요건 및 명세서 기재요건으로서의 발명의 효과, 특허소송연구 제2집, 특허법원, 2001
- 김관식, 컴퓨터프로그램의 특허법상 보호에 관한 특허법 제2조 개정시안의 법적 의의 및 과제, 법학연구 제23권 제1호, 2012
- 김보현 외, 데이터경제와 디지털금융, 법문사, 2021
- 김원오, 데이터 산업법 제정안에 관한 소고, 산업재산권 제68호, 2020
- 김윤명, SW특허는 기술발전을 이끄는가?. 홍익법학 제15권 제4호, 2014
- 김윤명, 발명의 컴퓨터 구현 보호체계 합리화를 위한 특허제도 개선방안 연구, 특허청, 2014
- 김윤명, 블랙박스를 열기위한 인공지능법, 박영사, 2022
- 김윤명·이민영, 인공지능의 생각, 이다북스, 2022
- 김인중, 기계학습의 발전 동향, 산업화 사례 및 활성화 정책 방향, 이슈리포트 2015-017, 소프트웨어정책연구소, 2015
- 김진형, AI취강의 수업, 매일경제신문사, 2020
- 김태수 외, 중국특허법, 한빛지적소유권센터, 2015
- 김태훈 외, 마이데이터의 시대가 온다, 지식공감, 2022
- 박동식, 유럽특허법, 세창출판사, 2009
- 박소영 외, 데이터 거래 활성화를 위한거래소·거래사·크롤링의 현황과 개선과제, 입법조사처, 2022
- 박준석, 우리 특허법상 ‘발명’의 개념에 관한 고찰, 서울대학교 법학 제54권 제3호, 2013.9
- 반병현, 142가지 데이터셋, 생능북스, 2022
- 반용성, 화학발명에서의 진보성 판단기준에 관한 연구, 충남대학교 석사학위논문, 2002
- 샌딕 우탐찬다니, 데이터 민주화와 셀프서비스 데이터, 에이콘, 2022

소프트웨어정책연구소, 2021 인공지능산업 실태조사, 과학기술정보통신부, 2022

손경한 편저, 블록체인과 법, 박영사, 2019

손경한 편저, 특허허법론, 법영사, 2005

심미랑 외, 특허절차상 생물학적 물질 기탁제도에 관한 연구, 산업재산권 제37호, 2012

안재현, XAI 설명가능한 인공지능을 해부하다, 위키북스, 2020

알렉스 고렐릭, 엔터프라이즈 빅데이터 레이크, 에이콘, 2020

양종모, “인공지능 이용 범죄예측 기법과 불심검문 등에서의 적용에 관한 고찰”, 형사법의 신 동향 통권 제51호, 2016.6

오요한 외, 인공지능 알고리즘은 사람을 차별하는가?, 과학기술학연구 제18권 제3호, 2018

윤선희, 특허법, 법문사, 2019

이규호 외, 부정경쟁방지법, 진원사, 2022

이규호, 인공지능 학습용 데이터세트 보호를 위한 특허법상 주요 쟁점 연구, 산업재산권 제64호, 2020

이수호, AI인사이트, 한빛미디어, 2022

이해완, 저작권법, 박영사, 2019

이혜진, 특허와 상표, 2021.03.22.

임 호, 특허법, 법문사, 2003

임원선, 실무자를 위한 저작권법, 저작권위원회, 2014

임춘광, 중국 특허법에서 인공지능 알고리즘의 특허적격성, 아주법학15(2), 2021

장세형, NFT 실체와 가치, 위키북스, 2022

전성태, AI 학습용 데이터의 보호에 관한 소고, 法學論叢 제28집 제1호, 2021

정상조 박성수 공편, 특허법 주해 I, 박영사, 2010

정상조, 딥러닝에서의 학습데이터와 공정이용, LAW&TECHNOLOGY 제16권 제1호 (통권 제85호) 2020

정상조·박준석, 지식재산권법, 홍문사, 2013.

정우성, 특허문서론, 에이콘, 2017

정원준 외, AI 학습데이터의 특허법적 보호방안 및 산업계 영향 분석, 특허청, 2021

정준화, 빅데이터 플랫폼의 운영 실태와 개선방안, 입법조사처, 2020

조상준 외, 빅데이터 커리어 가이드북, 길벗, 2022

조영선, 명세서 기재요건으로서의 발명의 효과, 인권과정의 Vol.427, 2012

조영선, 특허법 2.0, 박영사, 2018

차상욱, 데이터법, 세창출판사, 2022.

채수근, 산업디지털 전환 촉진법안, 산업의 디지털 전환 및 지능화 촉진에 관한 법안 검토보고

서, 국회산업통상자원중소벤처기업위원회, 2020.

최승재 외, 신미국특허법, 법문사, 2020

최호진, 썸네일 이미지와 공정이용, LAW&TECHNOLOGY 제8권 제3호, 2012

특허청, 기술분야별 심사실무가이드, 2020

특허청, 특허·실용신안 심사기준, 2021

특허청, 미생물 발명, 특허출원·분양신청 쉽고 빠르게, 보도자료, 2022.6.27.

한국데이터진흥원, 2021 데이터산업현황조사, 과학기술정보통신부, 2022.

한국정보법학회 지음, 인터넷, 그 길을 묻다, 중앙Books, 2012

한국지능정보화진흥원, 인공지능 학습용 데이터 품질관리 가이드라인 v1.0, 2021

한국지식재산연구원·AI-IP특별위원회, 인공지능(AI)-창작물의 권리보호 방안 수립 등 AI-지식 재산 이슈 대응 연구, 국가지식재산위원회, 2022

한국특허법학회편, 특허판례연구, 박영사, 2017

함영욱, AI 관련 발명의 성립성 판단기준에 관한 연구, 한양대학교 박사학위논문, 2022

IDC, 국내 빅데이터 및 분석 시장 전망, 2021-2025

Superb AI, 한국기업들은 Data Centric AI에 얼마나 준비되어 있을까?, 2021.

■ 국외문헌

Alexander V. Giczy et. al., Identifying artificial intelligence (AI) invention: A novel AI patent dataset, USPTO Economic Working Paper No. 2021-2 August 2021.

David Leslie, Understanding artificial intelligence ethics and safety, Alan Turing Institute, 2019

Fromer, Jeanne C., Patent Disclosure. Iowa Law Review, Vol. 94, p. 539, 2009.

Joelle Pineau et. al, Improving Reproducibility in Machine Learning Research, Journal of Machine Learning Research 22, 2021

Hagen, Gregory R., AI and Patents and Trade Secrets (February 1, 2021). in Florian Martin-Bariteau & Teresa Scassa, eds., Artificial Intelligence and the Law in Canada (Toronto: LexisNexis Canada, 2021), Available at SSRN:<https://ssrn.com/abstract=3734654>

Kim, Daria and Alber, Maximilian and Kwok, Man Wai and Mitrovic, Jelena and Ramirez-Atencia, Cristian and Rodriguez Perez, Jesus and Zille, Heiner, Ten Assumptions About Artificial Intelligence That Can Mislead Patent Law Analysis (August 1, 2021).

Liva Rudzite, Algorithmic Explainability and the Sufficient-Disclosure Requirement under the European Patent Convention, Juridica International 31, 2022.

Medhi Poursoltani, Disclosing AI Inventions, TEXAS INTELLECTUAL PROPERTY LAW

JOURNAL Vol.29, 2021.

Robert Merges, Peter Menell, Mark Lemley, Intellectual Property in the New Technological Age, Wolters Kluwer, 2012

Romer, P. “Endogenous technological change.” Journal of Political Economy 98, no. 5, Part 2(1990): S71-S102.

Tabrez Ebrahim, Artificial Intelligence Inventions & Patent Disclosure, Iowa Legal Studies Research Paper No. 2021-48, November 2021

Ted M. Sichelman, Patents, Prizes, and Property, 30 Harv. J.L. & Tech. 279, 280-81 (2017).

Travis A. Dyer, et. al. The Effect of Patent Disclosure Quality on Innovation, September 14, 2020. Available at SSRN: <https://ssrn.com/abstract=3711128>

USPTO, Public Views on Artificial Intelligence and Intellectual Property Policy, October, 2020.

Valinasab, Omid, Big Data Analytics to Automate Patent Disclosure of Artificial Intelligence’ s Inventions (April 19, 2022).

Vladimir Fedak, Top 10 Most Popular AI Models, DZone (Nov. 8, 2018), <https://dzone.com/articles/top-10-most-popular-ai-models>.

Vijay/Devesh et al., AI and Indian Patent Law- Sufficiency of Disclosure for Artificial Intelligence-based Patents, International Journal of Mechanical Engineering, Vol. 7 (Special Issue, Jan.-Feb. 2022)

Volha Litskevich, How Much Does Artificial Intelligence (AI) Cost in 2019, AZATI (Jan. 23, 2019).

WIPO, FURTHER STUDY ON THE SUFFICIENCY OF DISCLOSURE (PART I), SCP/34/5, SEPTEMBER 1, 2022.

WIPO, REVISED ISSUES PAPER ON INTELLECTUAL PROPERTY POLICY AND ARTIFICIAL INTELLIGENCE, WIPO/IP/AI/2/GE/20/1 REV, 2020.

日本特許庁, 「特許・実用新案審査ハンドブック 附属書B 第1章 コンピュータソフトウェア関連発明」, SW審査基準, 2021. 3.

中山信弘, 特許法, 法文社, 2001

岡本 義則, “人工知能 (AI) の学習用データに関する知的財産の保護”, パテント(Vol.70 No. 10), 2017.

Yutaka Matsuo, 인공지능과 딥러닝, 동아엠앤비, 2015

■ 웹사이트

Best Public Datasets for Machine Learning and Data Science,
<https://pub.towardsai.net/best-datasets-for-machine-learning-data-science-computer-vision-nlp-ai-c9541058cf4f>

Edwin D. Garlepp, Disclosing AI Inventions - Part II: Describing and Enabling AI Inventions. 2021.
<https://www.theaipatentblog.com/disclosing-ai-inventions-part-ii-describing-and-enabling-ai-inventions>

Generally Artificial Intelligence (AI) Patents, PATENT ATTORNEYS.
<https://www.techlaw.attorney/artificial-intelligence-ai-patents>

<http://kpaanews.or.kr/news/view.html?section=91&category=120&no=4842>

<https://hiseon.me/data-analytics/tensorflow/tensorflow-dataset/>

<https://learn.microsoft.com/ko-kr/windows/ai/windows-ml/what-is-a-machine-learning-model>.

<https://www.cs.mcgill.ca/~jpineau/ReproducibilityChecklist.pdf>.

<https://www.gov.uk/government/consultations/artificial-intelligence-and-intellectual-property-call-for-views/government-response-to-call-for-views-on-artificial-intelligence-and-intellectual-property>

<https://www.iam-media.com/global-guide/iam-yearbook/2020/article/ai-inventions-and-sufficiency-of-disclosure-when-enough-enough>

<https://www.ibm.com/kr-ko/topics/data-pipeline>.

<https://www.idc.com/getdoc.jsp?containerId=prAP48914822>.

<https://www.integrate.io/ko/blog/structured-data-vs-unstructured-data-ko/>

https://www.jpo.go.jp/system/laws/rule/guideline/patent/document/cs_shinsa/cs_shinsa.pdf

[https://www.uspto.gov/ip-policy/economic-research/research-datasets/artificial-intelligence-patent-dataset\)](https://www.uspto.gov/ip-policy/economic-research/research-datasets/artificial-intelligence-patent-dataset)

https://www.wipo.int/about-ip/en/artificial_intelligence/submissions-search.jsp?type_id=&territory_id=&issue_id=2421

https://www.wipo.int/edocs/mdocs/mdocs/en/wipo_ip_ai_ge_20/wipo_ip_ai_2_ge_20_1.pdf

https://www.wipo.int/export/sites/www/about-ip/en/artificial_intelligence/call_for_comments/pdf/ind_noto_la_diega.pdf

https://www.wipo.int/export/sites/www/about-ip/en/artificial_intelligence/call_for_comments/pdf/org_aippi.pdf

https://www.wipo.int/export/sites/www/about-ip/en/artificial_intelligence/call_for_comments/pdf/org

_american_bar_association.pdf

https://www.wipo.int/export/sites/www/about-ip/en/artificial_intelligence/call_for_comments/pdf/org_ccia.pdf

https://www.wipo.int/export/sites/www/about-ip/en/artificial_intelligence/call_for_comments/pdf/org_intel.pdf

https://www.wipo.int/export/sites/www/about-ip/en/artificial_intelligence/call_for_comments/pdf/org_max_planck_institute_for_innovation_and_competition.pdf

<https://www.patentnext.com/2021/11/a-tale-of-two-jurisdictions-sufficiency-of-disclosure-for-artificial-intelligence-patents-in-the-u-s-and-the-epo/>

<https://www.appleyardlees.com/artificial-intelligence-and-machine-learning-sufficiency-and-plausibility/>

https://www.epo.org/modules/epoweb/accdocument/epoweb2/468/en/CA-PL_5-20_en.pdf

<https://www.superb-ai.com/ko-blog/everything-about-computer-vision-data>

AI 학습데이터의 활성화 방안 연구

- AI 학습데이터 공개·보호·활용을 중심으로 -

(A Study on the Utilization of AI Learning Data

- Focusing on the disclosure, protection, and utilization of AI learning data -)

발행일 : 2022년 12월 21일

발행인 : 특허청장 이인실

발행처 : 특허청 등록과(www.kipo.go.kr)

대전광역시 서구 청사로 189 대전정부청사 4동

TEL. (042) 481-5236

FAX. (042) 472-3467

이용허락 유형	표시 마크	이용허락범위
[제4유형] 제1유형+상업적 이용금지+변경금지		- 출처 표시 - 비상업적 이용만 가능 - 변형 등 2차적 저작물 작성 금지

AI 학습데이터의 활성화 방안 연구

- AI 학습데이터 공개 · 보호 · 활용을 중심으로 -

(A Study on the Utilization of AI Learning Data

- Focusing on the disclosure, protection, and utilization of AI learning data -)



대전광역시 서구 청사로 189 정부대전청사 4동

Tel. 042-481-5225 <http://www.kipo.go.kr>

ISBN : 979-11-6884-089-8 13500

DOI : 10.8080/P9791168840898