



AI 학습용 데이터세트의 송신에 대한 특허법상 보호방안

2020. 10. 23. 오후 4시
인터넷법제도 포럼

이규호 (중앙대학교 법학전문대학원 교수)



차례

- 인공지능 학습용 데이터세트의 정의
- 인공지능 학습용 데이터세트의 필요성
- 인공지능 관련 개념
- 기계학습의 흐름
- 현행 IP법상 보호방안 검토
- 원시데이터의 수집과 AI 학습용 데이터세트 생성 등에 대한 저작권제한사유 (예: 일본)
- AI 학습용 데이터세트 관련 특허의 예
- AI 학습용 데이터세트 청구유형별 보호범위
- AI 학습용 데이터세트의 유통과정과 침해 여부
- AI 학습용 데이터세트의 특허법상 보호의 필요성
- AI 학습용 데이터세트의 특허법상 보호방안: 전제적 검토사항
- AI 학습용 데이터세트의 특허법상 보호방안: 해석론 및 입법론

인공지능 학습용 데이터세트의 정의

- 이를 AI 교육 데이터세트로 부를 수도 있다. 일본에서는 이를 'AI 교사데이터' 또는 'AI 학습데이터'로 부른다. 하지만, AI을 학습시키는 데이터세트를 AI 교사데이터로 칭하는 것은 데이터 세트 자체가 교사가 된다는 의미가 되어 의미상 어색하다. 또한, AI가 생성한 학습데이터와 혼동의 여지가 있다는 점에서 AI을 학습시키는 데이터세트를 AI 학습데이터로 칭하는 것은 문제가 있다.
- 그리고 AI를 학습시키는 데이터세트를 흔히 AI 학습용 데이터라고 칭하기도 한다. 하지만, 이 경우에도 AI 학습용 데이터라는 용어 보다는 AI 학습용 데이터세트가 보다 정확한 표현이다. 왜냐 하면 AI를 학습시키기 위해서는 데이터의 집합물이 필요하기 때문이다.

인공지능 학습용 데이터세트의 필요성

- 학습가능한 양질의 데이터세트(학습용 데이터세트)는 인공지능 기술의 경쟁력을 좌우한다. 즉, 인공지능의 성능은 인공지능 프로그램의 정확성뿐만 아니라 학습용 데이터세트의 내용에 크게 의존한다.
- 원시 데이터를 인공지능 학습용 데이터세트로 가공하여 제공하는 비즈니스가 점차 활성화되고 있다.
- Fintech도 그 중 하나의 예

인공지능 학습용 데이터세트의 필요성

- 최근 9년간(2010~2018) AI 핵심 기술 분야* 특허 출원 동향을 살펴보면, IP5(한국, 미국, 중국, 일본, 유럽) 특허청의 AI 핵심 기술 분야 특허 출원은 연평균 40% 이상 증가하고 있다.
- 국내 인공지능 핵심기술 출원은 2015년 기점으로 연평균 54.3%의 높은 증가율을 보이고 있으며, 출원량으로는 2016년부터 일본과 유럽을 넘어섰다. 2018년 연간 출원량(2,506건)은 중국(11,640건), 미국(6,279건) 다음으로 세계 3위를 유지하고 있다.
- * 이는 인공지능 학습 모델, 인공 신경망 설계, 인공지능 칩(Chip) 등을 구현하는 AI 기술(특허분류 코드: G06N)을 말한다.

인공지능 학습용 데이터세트의 필요성

- 국내 인공지능 핵심기술 출원의 높은 증가세와 더불어 인공지능 학습용 데이터 생성·가공 기술에 관한 특허출원도 함께 증가하고 있다. 인공지능 관련 발명에서 학습용 데이터세트에 대한 권리 보호는 생성단계, 학습용 모델 작성단계, 학습완료 모델의 이용단계에서 매우 중요한 역할을 수행함
- 가칭 '인공지능 산업 발전을 위한 세계 공급사슬과 가치망(global supply chain and value network for AI industry)' 에서도 중요한 부분을 차지함
- => 인공지능 기술뿐만 아니라 이를 위한 학습용 데이터 산업의 육성은 우리나라 미래 산업의 먹거리를 위해서도 매우 중요한 과제라고 생각함

인공지능 관련 개념

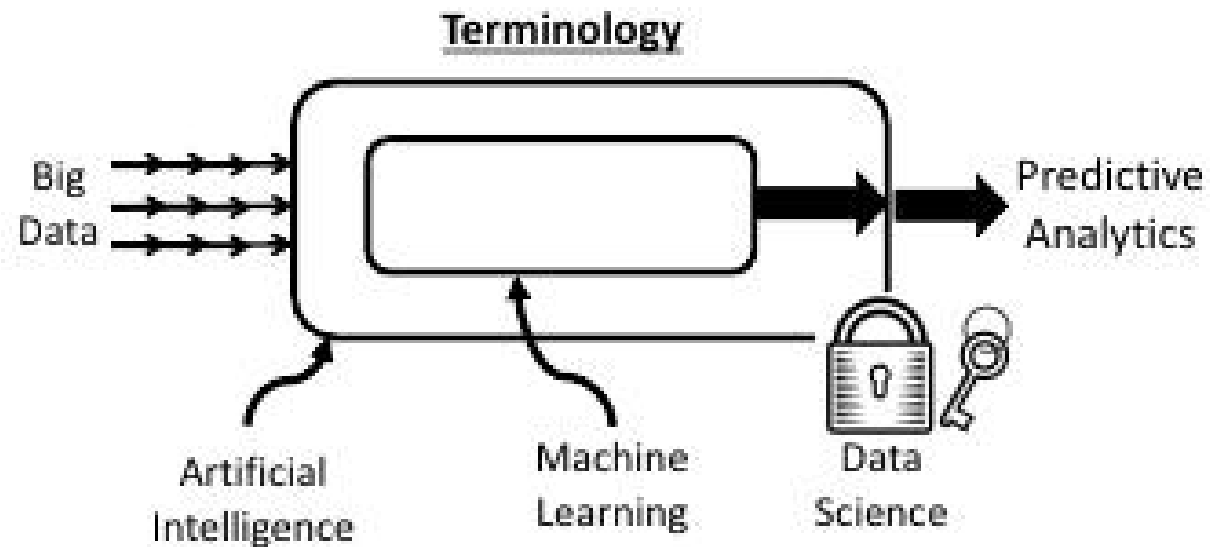
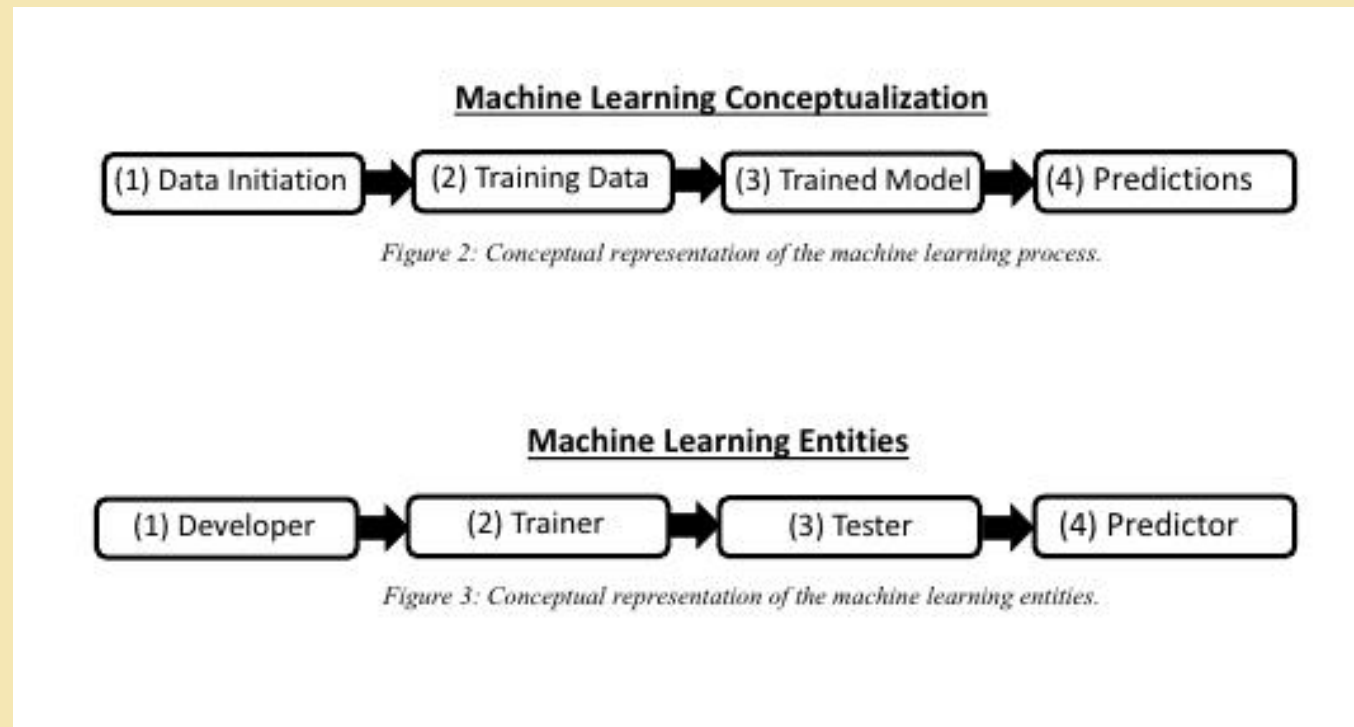


Figure 1: Conceptual representation of artificial intelligence terminology.

Source: Tabrez Y. Ebrahim, Artificial Intelligence Patent Infringement, p. 11 available at https://robots.law.miami.edu/2019/wp-content/uploads/2019/03/Ebrahim_Patent-Infringement.pdf (last visit on July 17, 2020).

기계학습의 흐름



Source: Tabrez Y. Ebrahim, Artificial Intelligence Patent Infringement, p. 11, available at https://robots.law.miami.edu/2019/wp-content/uploads/2019/03/Ebrahim_Patent-Infringement.pdf (last visit on July 17, 2020).

현행 IP법상 보호방안 검토

- 창작성이 있는 데이터세트는 창작성이 있는 저작물과 마찬가지로 저작권법에 의하여 보호된다.
- 또한 AI 학습용 데이터세트의 구성, 선택 또는 배열에 창작성이 있다면 편집저작물로서 보호받을 수 있다. 편집물이 저작물성 있는 지적 창작물에 해당하여 편집저작물로서 저작권 보호를 받게 되더라도 그 저작권 보호는 편집저작물에 포함된 데이터 자체에는 미치지 아니한다.
- 이와 관련하여 데이터세트의 구성, 선택 또는 배열에 창작성이 없다면 편집저작물로 보호받을 수 없을 뿐만 아니라 특허법과는 달리 출원공개가 요건이 아니므로 실제로는 편집저작물과 침해 저작물 사이의 의거관계와 실질적 유사성을 입증하는 것이 용이하지 않을 것이다.

현행 IP법상 보호방안 검토

- "데이터베이스제작자"는 데이터베이스의 제작 또는 그 소재의 갱신·검증 또는 보충(이하 "갱신등"이라 한다)에 인적 또는 물적으로 상당한 투자를 한 자를 말한다. 데이터베이스제작자는 그의 데이터베이스의 전부 또는 상당한 부분을 복제·배포·방송 또는 전송(이하 이 조에서 "복제등"이라 한다)할 권리를 가진다.
- 데이터베이스의 개별 소재는 제1항의 규정에 따른 당해 데이터베이스의 상당한 부분으로 간주되지 아니한다. 다만, 데이터베이스의 개별 소재 또는 그 상당한 부분에 이르지 못하는 부분의 복제등이라 하더라도 반복적이거나 특정한 목적을 위하여 체계적으로 함으로써 당해 데이터베이스의 통상적인 이용과 충돌하거나 데이터베이스제작자의 이익을 부당하게 해치는 경우에는 당해 데이터베이스의 상당한 부분의 복제등으로 본다.

현행 IP법상 보호방안 검토

- 그리고 데이터베이스제작자의 권리제한이 적용된다. 데이터베이스제작자의 권리는 데이터베이스의 제작을 완료한 때부터 발생하며, 그 다음 해부터 기산하여 5년간 존속한다.
- 데이터베이스의 갱신등을 위하여 인적 또는 물적으로 상당한 투자가 이루어진 경우에 당해 부분에 대한 데이터베이스제작자의 권리는 그 갱신등을 한 때부터 발생하며, 그 다음 해부터 기산하여 5년간 존속한다.
- => 인공지능 학습용 데이터세트를 저작권법상 데이터베이스제작자의 권리로서 보호하기 위해서는 침해성립요건, 권리제한, 보호기간 등 여러 제약이 존재한다.

현행 IP법상 보호방안 검토

- 인공지능 학습용 데이터세트의 무단 배포 내지 무단 송신은 부정경쟁방지 및 영업비밀보호에 관한 법률상 타인성과 도용행위로 볼 수도 있다.
- 이 조문은 “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 일 유형으로 규정하고 있다. 이 행위에 대해서는 형사상 처벌규정이 없다는 단점이 있다. 부정경쟁방지법 제2조 제1호 카목은 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.

현행 IP법상 보호방안 검토

- 그리고 같은 법 제2조 제1호 (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 같은 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(公共領域, public domain)에 속하지 않는다고 평가할 수 있어야 한다.

현행 IP법상 보호방안 검토

- 아울러 같은 법 제2조 제1호 (카)목의 '공정한 상거래 관행이나 경쟁질서에 반하는 방법으로 자신의 영업을 위하여 무단으로 사용'한 경우에 해당하기 위해서는 권리자와 침해자가 경쟁 관계에 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.
- 이 요건을 갖추지 못한 자에 대해서는 이 조문이 적용되지 않을 수 있다.

현행 IP법상 보호방안 검토

- 부정경쟁방지법상 인공지능 학습용 데이터셋을 영업비밀로 보호받는 방법이 있을 수 있다. 기밀인 데이터셋 또는 영업상 가치 또는 기술적 가치가 있고 비밀 보유자가 비밀로 관리하는 데이터셋은 특정 사람의 특정 행위, 예를 들어 직원 또는 연구 계약자에 의한 무단 공개 또는 사이버 침입을 통한 도난으로부터 보호된다.
- 부정경쟁방지법상 "영업비밀"이란 공공연히 알려져 있지 아니하고 독립된 경제적 가치를 가지는 것으로서, 비밀로 관리된 생산방법, 판매방법, 그 밖에 영업활동에 유용한 기술상 또는 경영상의 정보를 말한다. 영업비밀에 해당하기 위해서는 비공지성, 비밀관리성, 경제적 유용성 등의 요건을 충족하여야 한다.

현행 IP법상 보호방안 검토

- 인공지능 학습용 데이터세트는 이미 공지된 경우도 많으므로 영업 비밀에 해당하기 곤란한 경우도 상정할 수 있고, 비밀관리성 요건을 주장하여 증명하는 것도 용이치 않은 경우가 있을 수 있다.
- 이들 규정이 적용되지 않은 법적 공백을 고려하고, 역분석 내지 독자적인 개발 등에 따른 인공지능 학습용 데이터세트에 대한 권리자의 피해를 방지하고 사후 구제를 보장하기 위하여는 특허법상 보호받을 수 있는 방안도 마련할 필요가 있다.

현행 IP법상 보호방안 검토

- 이러한 연유로 학습용 데이터세트 시장의 활성화와 학습용 데이터세트의 이용 활성화를 통한 AI 학습용 데이터세트 시장의 확대를 위해서도 특허법상 학습용 데이터세트의 보호 여부를 검토할 필요가 있다.

원시데이터의 수집과 AI 학습용 데이터세트 생성 등에 대한 저작권제한사유 (예: 일본)

II. 일본

학습된 모델 생성을 위해서는 대량의 원시 데이터와 원시 데이터를 기초로 생성된 학습 데이터세트가 필요한데, 이 경우 저작물인 원시 데이터 (텍스트, 사진, 이미지, 동영상 등)를 이용하는 경우도 많음

원시데이터의 수집과 AI 학습용 데이터세트 생성 등에 대한 저작권제한사유 (예: 일본)

일본 저작권법상 저작물은 저작권자의 허락없이 이용 (다운로드 또는 변경 등)하는 것은 불가능하지만, 사실 일본의 2019년 개정 이전의 저작권법 제47조의7에 의거하여 인공지능 개발 목적이라면 일정 한도에서 저작권자의 허락없이 저작물을 이용할 수 있다.

원시데이터의 수집과 AI 학습용 데이터세트 생성 등에 대한 저작권제한사유 (예: 일본)

- 2019년 개정 이전의 일본 저작권법 제47조의7 (정보분석을 위한 복제 등)에서는 “저작물은 전자계산기에 의한 정보분석 (다수의 저작물, 그 밖의 정보로부터 해당 정보를 구성하는 언어, 음, 영상 그 밖의 요소에 관한 정보를 추출, 비교, 분류 그 밖의 통계적인 분석을 행하는 것을 말한다. 이하 이 조문에서는 같다.)을 행하는 것을 목적으로 하는 경우에는 필요하다고 인정되는 한도에서 기록매체에의 기록 또는 번안(이것에 의하여 창작한 2차적저작물의 기록을 포함한다.)을 행할 수 있다. 다만, 정보분석을 행하는 자의 용도에 제공하기 위하여 작성된 데이터베이스의 저작물에 대해서는 그러하지 아니하다.”라고 규정하고 있었다.

원시데이터의 수집과 AI 학습용 데이터세트 생성 등에 대한 저작권제한사유 (예: 일본)

- 다만, 이 조문에도 한계는 존재하였다. 데이터 수집, 데이터 처리, 기계학습, 심층학습이라는 것은 구체적으로는 데이터의 복제물, 정형 데이터, 데이터세트를 이용한 기계학습과 심층학습인데, 이러한 행위는 '복제'나 '번안'에 해당하기 때문에 저작권자의 동의 없이는 원칙적으로 저작권 침해가 되었다.

원시데이터의 수집과 AI 학습용 데이터세트 생성 등에 대한 저작권제한사유 (예: 일본)

- 2019년 1월 1일 일본의 개정 저작권법이 시행되었다. 이 개정 저작권법이 시행되면 종전의 제47조의7이 폐지되고 새로운 조문인 제30조의4와 제47조의5가 효력을 발생하게 되었다. 이에 따라 인공지능 개발을 위해 가능한 행위가 더 확대되어 인공지능 개발이 가속화될 것으로 예상하고 있다.

원시데이터의 수집과 AI 학습용 데이터세트 생성 등에 대한 저작권제한사유 (예: 일본)

일본의 현행 저작권법 제30조의4 (저작물에 표현된 사상 또는 감정의 향유를 목적으로 하지 않는 이용)에서는 “저작물은 다음에 열거한 경우, 그 밖에 저작물에 표현된 사상 또는 감정을 스스로 즐기고 또는 타인에게 누리도록 하는 것을 목적으로 하지 않는 경우에는 필요하다고 인정되는 한도에서 어떤 방법이든지 불문하고 사용할 수 있다. 다만, 해당 저작물의 종류 및 용도 및 해당 이용 양태에 비추어 저작권자의 이익을 부당하게 침해하게 되는 경우에는 그러하지 아니하다.

원시데이터의 수집과 AI 학습용 데이터세트 생성 등에 대한 저작권제한사유 (예: 일본)

1. 저작물의 녹음, 녹화 기타 이용에 관한 기술의 개발 또는 실용화를 위한 시험의 용도로 제공하는 경우
2. 정보분석 (다수의 저작물, 그 밖에 대량의 정보에서 해당 정보를 구성하는 언어, 음, 영상 그 밖의 요소에 관한 정보를 추출, 비교, 분류, 그 밖의 분석을 실시하는 것을 말한다. 제47조의5 제1항 제2호에서 같다.)의 용도로 제공하는 경우
3. 제1호와 제2호의 경우 이외에 저작물의 표현에 대한 사람의 지각에 의한 인식을 수반하지 않고 저작물을 전자계산기에 의하여 정보처리과정에서 이용, 그 밖의 이용(프로그램저작물에 있어서는 저작물의 전자계산기의 실행을 제외한다)에 제공하는 경우”라고 규정하고 있고,

원시데이터의 수집과 AI 학습용 데이터세트 생성 등에 대한 저작권제한사유 (예: 일본)

- 제47조의5(전자계산기에 의한 정보처리 및 그 결과의 제공에 수반되는 경미한 이용 등)에서는 "① 전자계산기를 이용한 정보처리에 의해 새로운 지식이나 정보를 창출하는 것에 의하여 저작물의 이용의 촉진에 이바지하는 다음 각 호의 행위를 하는 자(해당행위의 일부를 행하는 자를 포함하며, 해당행위를 시행령에서 정하는 기준에 따라 실시하는 자에 한한다.)는 공중에 제공 또는 제시(송신가능화를 포함한다. 이하 이 조에서 같다.)가 행해진 저작물(이하 이 조 및 제47조의6 제2항 제2호에서 "공중제공제시 저작물"이라 한다) (공표된 저작물 또는 송신가능화된 저작물에 한한다.)에 대해 해당 각 호에 열거된 행위의 목적상 필요하다고 인정되는 한도에서 해당행위에 부수하여 어떠한 방법이든지 불문하고 이용(해당 공중제공제시 저작물 중 그 이용에 제공되는 부분이 차지하는 비율, 그 이용에 제공되는 부분의 금액, 그 이용에 제공될 때 표시의 정확성, 그 밖의 요소에 비추어 사소한 것에 한한다. 이하 이 조에서 "경미한 이용"이라 한다)을 할 수 있다.

원시데이터의 수집과 AI 학습용 데이터세트 생성 등에 대한 저작권제한사유 (예: 일본)

- 다만, 해당 공중제공제시 저작물을 공중에게 제공 또는 제시하는 것이 저작권을 침해한다는 것(국외에서 행해진 공중에의 제공 또는 제시에 있어서는 국내에서 행해졌다고 했다면 저작권 침해가 되어야 할 것)을 알면서 해당 경미한 이용을 행한 경우, 그 밖에 해당 공중제공제시 저작물의 종류 및 용도 및 해당 경미한 이용의 양태에 비추어 저작권자의 이익을 부당하게 침해하게 되는 경우에는 그러하지 아니하다.
- 1. 전자계산기를 이용하여 검색을 통해 원하는 정보(이하 이 호에서 "검색정보"라 한다.)가 기록된 저작물의 제호 또는 저작자명, 송신가능화된 검색정보에 관한 송신원 식별부호 (자동공중송신의 송신원을 식별하는 글자, 숫자, 기호 기타의 부호를 말한다), 그 밖의 검색정보의 특정 또는 소재에 대한 정보를 검색하는 것과 그 결과를 제공하는 것

원시데이터의 수집과 AI 학습용 데이터세트 생성 등에 대한 저작권제한사유 (예: 일본)

2. 전자계산기에 의한 정보분석을 실시하는 것과 그 결과를 제공하는 것

3. 제1호와 제2호에서 열거한 것 이외에 전자계산기에 의하여 정보처리하여 새로운 지식이나 정보를 창출하고, 그 결과를 제공하는 행위로서 국민생활의 편의성 향상에 기여할 것으로 정령으로 정하는 것

② 제1항 각 호의 행위를 준비하는 자 (해당 행위의 준비를 위한 정보의 수집, 정리 및 제공을 정령으로 정하는 기준에 따라 행하는 자에 한한다.)는 공중제공제시 저작물에 대하여 이 항의 규정에 의한 경미한 이용의 준비를 위해 필요하다고 인정되는 한도에서 복제 또는 공중송신 (자동공송신의 경우에 있어서는 송신가능화를 포함한다. 이하 이 항 및 제47조의6 제2항 제2호에서 같다.)를 행하거나 또는 그 복제물에 의한 배포를 행할 수 있다. 다만, 해당 공중제공제시 저작물의 종류 및 용도 및 해당 복제 또는 배포의 부수 및 해당 복제, 공중송신 또는 배포의 양태에 비추어 저작권자의 이익을 부당하게 침해하게 되는 경우에는 그러하지 아니하다.”라고 규정하고 있다.

AI 학습용 데이터세트 관련 특허의 예

<표 5> 학습용 데이터세트 생성방법에 대한 특허의 예

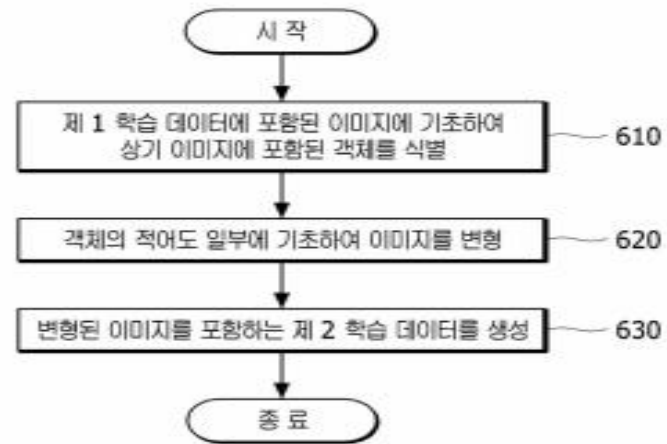
• 주식회사 000 등록특허 제10-*****호 (학습용 데이터세트 생성방법)¹⁴⁷⁾

【청구항 13】

학습 데이터 생성 방법으로서, 컴퓨팅 장치에 포함된 프로세서가 제 1 학습 데이터에 포함된 이미지에 기초하여 상기 이미지에 포함된 객체를 식별하는 단계; 상기 프로세서가 상기 객체의 제 1 부분과 크롭핑된 이미지의 사전 결정된 범위 안의 영역에 포함된 제 2 부분의 적어도 일부가 오버랩되도록 상기 이미지를 크롭핑하는 단계; 및 상기 프로세서가 상기 크롭핑된 이미지를 포함하는 제 2 학습 데이터를 생성하는 단계;를 포함하는, **학습 데이터 생성 방법.**

【청구항 14】

학습 데이터 생성 방법을 제공하기 위한 서버로서, ~ 제 2 학습 데이터를 생성하는 학습 데이터 생성 방법을 제공하기 위한 서버.



출처:
이규호,
152면.

AI 학습용 데이터세트 청구유형별 보호범위

〈표 6〉 학습용 데이터세트 청구유형별 보호범위 점검

청구유형	제3자 실시행위		특허권 보호범위	
(1) 학습용 데이터 세트 생성장치 (물건의 발명)	제2조 제3호 : 정의	모방하여 제작한 학습용 데이터세트 생성장치를 사용하여 학습용 데이터세트 생성	제126조 : 금지 청구권	학습용 데이터세트 생성장치의 생산, 사용, 양도, 대여 등 금지 가능
(2) 학습용 데이터 세트 생성방법 (방법의 발명)		학습용 데이터세트 생성방법을 모방하여 학습용 데이터세트 생성		학습용 데이터세트 생성방법의 사용이나 사용의 청약 금지 가능
(3) 학습용 데이터 세트 구조가 저장된 기록매체 (데이터세트 구조 기록매체 발명)		학습용 데이터세트를 모방제작		학습용 데이터세트 구조 저장된 기록매체의 생산, 사용, 양도, 대여 등 금지 가능

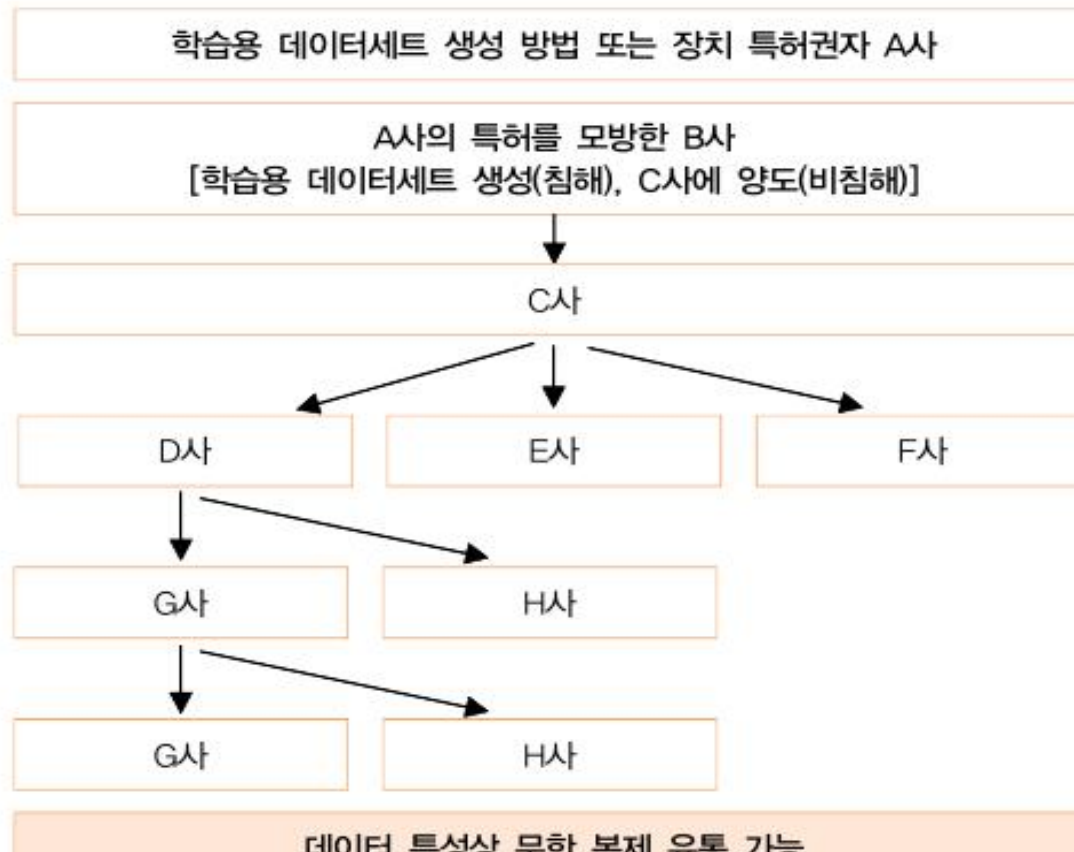
출처:
이규호,
153면

AI 학습용 데이터세트 청구유형별 보호범위

- [표 6]에서 특허발명(1)의 생성물인 학습용 데이터세트의 송신은 특허법상 실시에 해당하지 아니한다.
- 마찬가지로 특허발명(3)의 학습용 데이터세트 구조가 저장된 기록매체 "온라인"상 송신은 특허법상 실시에 해당하지 아니한다. 이 경우 학습용 데이터세트의 구조가 '컴퓨터로 하여금 특정한 기능을 수행'하도록 하는 요건을 만족해야 한다. 이 경우, 발명에 해당하기 어려울 것으로 예상된다.
- 특허발명 (2)의 생성물인 학습용 데이터세트의 송신은 특허법상 사용의 청약에 해당하여 발명의 실시행위에 해당할 수 있음.

AI 학습용 데이터세트의 유통과정과 침해 여부

<표 7> 학습용 데이터세트의 유통과정과 침해 여부



출처: 이규호,
154면.

AI 학습용 데이터세트의 유통과정과 침해 여부

- 학습용 데이터세트 생성방법 또는 학습용 데이터세트를 생성하는 장치에 대한 발명의 특허권자의 학습용 데이터세트는 온라인 송신과 오프라인 배포 양자로 이루어지는 것이 거래현실이다.
- [표 7]에서 B, C사는 모방·생성한 '학습용 데이터세트'를 특허침해 없이 배포하는 것이 가능하다. 이 경우, B, C 등의 학습용 데이터세트 배포는 실시예에 해당하지 아니하여 특허권 침해에 해당하지 아니한다.
-
- 그리고 [표 7]에서 학습용 데이터세트를 생성하는 장치에 대한 발명의 특허권자의 학습용 데이터세트는 특허침해 없이 송신하는 것이 가능하다. 이 경우 B, C 등의 학습용 데이터세트 송신은 실시예에 해당하지 아니하여 특허권 침해에 해당하지 아니한다.
- [표 7]에서 학습용 데이터세트 생성방법에 대한 발명의 특허권자의 학습용 데이터세트는 특허침해 없이 송신하는 것이 가능하다. 이 경우 B, C 등의 학습용 데이터세트 송신은 사용의 청약에 해당할 수 있음.

AI 학습용 데이터세트의 유통과정과 침해 여부

- AI 학습용 데이터세트 시장의 급속한 성장을 위해서는 AI 학습용 데이터세트의 공개가 매우 중요하므로 AI 학습용 데이터세트 관련 특허출원의 증가는 출원공개를 통한 이 시장의 확대를 선도할 것이다.
- 따라서 AI 학습용 데이터세트 관련 발명의 특허법상 보호방안 연구는 시급한 것으로 생각한다.

AI 학습용 데이터세트의 특허법상 보호의 필요성

- 물건의 정의 확대,
- 발명 실시유형의 확대,
- 특허권의 간접침해행위의 확대 등 다양한 선택지를
놓고 분석

AI 학습용 데이터세트의 특허법상 보호방안: 전제적 검토사항

- 특허발명이 방법의 발명인 경우, 발명의 실시행위에 사용의 청약을 포섭하는지:
- 특허법

[시행 2019. 12. 10.] [법률 제16804호, 2019. 12. 10., 일부개정]

구 특허법은 소프트웨어 등과 같은 방법의 발명인 경우 그 방법을 사용하는 행위만을 특허를 받은 발명의 실시로 규정하고 있어 소프트웨어 등을 정보통신망을 통하여 전송하는 행위가 특허를 받은 발명의 실시에 해당하는지 불분명하여 보호하기 어려운 측면이 있는바, 방법의 발명인 경우에 그 방법의 사용을 청약하는 행위를 특허를 받은 발명의 실시에 포함되도록 하되, 이로 인한 소프트웨어 산업의 위축을 방지하기 위하여 특허를 받은 발명의 실시가 방법의 사용을 청약하는 행위인 경우 특허권의 효력은 그 방법의 사용이 특허권 또는 전용실시권을 침해한다는 것을 알면서 그 방법의 사용을 청약하는 행위에만 미치도록 개정함.

AI 학습용 데이터세트의 특허법상 보호방안: 전제적 검토사항

특허법

[시행 2019. 7. 9] [법률 제16208호, 2019. 1. 8, 일부개정]

제2조(정의) 이 법에서 사용하는 용어의 뜻은 다음과 같다.

3. “실시”란 다음 각 목의 구분에 따른 행위를 말한다.

나. 방법의 발명인 경우: 그 방법을 사용하는 행위

제94조(특허권의 효력) ① (현행과 같음)

② <신설>

특허법

[시행 2020. 3. 11] [법률 제16804호, 2019. 12. 10, 일부개정]

제2조(정의) 이 법에서 사용하는 용어의 뜻은 다음과 같다.

3. “실시”란 다음 각 목의 구분에 따른 행위를 말한다.

나. 방법의 발명인 경우: 그 방법을 사용하는 행위 또는 그 방법의 사용을 청약하는 행위

제94조(특허권의 효력) ① (현행과 같음)

② 특허발명의 실시가 제2조제3호나목에 따른 방법의 사용을 청약하는 행위인 경우 특허권의 효력은 그 방법의 사용이 특허권 또는 전용실시권을 침해한다는 것을 알면서 그 방법의 사용을 청약하는 행위에만 미친다.

AI 학습용 데이터세트의 특허법상 보호방안: 전제적 검토사항

- 특허법 제62조 제1호(거절사유) 및 무효사유(특허법 제133조 제1항 제1호)에 특허법 제2조 제1호의 정의에 어긋나는 경우를 포섭할 것인지 여부:
- 우리 특허법상 발명에 해당하지 않으면 거절사유 또는 무효사유에 해당한다. 그럼에도 불구하고 출원발명 내지 특허발명이 특허법 제2조 제1호의 정의에 어긋나는 경우, 이를 거절사유(특허법 제62조 제1호) 또는 무효사유(특허법 제133조 제1항 제1호)로 규정하지 않고 특허법 제29조 제1항의 산업상 이용할 수 있는 발명에 해당하지 않는다는 것을 이유로 거절사유 또는 무효사유로 보는 것은 타당하지 않다.

AI 학습용 데이터세트의 특허법상 보호방안: 전제적 검토사항

- 특허발명이 물건의 발명의 경우 발명의 실시행위에 수출행위를 포함할 것인지 여부:
인공지능 학습용 데이터세트의 이동성을 감안할 때, 그 국제적 활용을 통해 특허권 침해가 발생할 여지가 많으므로 실시의 유형에 '물건의 수출행위'를 포함시킬 필요가 있다. 해외로의 송신행위를 수출행위로 규율하는 것이 가능할 것이다.
- 이미 일본 특허법 제2조 제3항 제1호에서는 "물건 (프로그램 등을 포함한다. 이하 같다.)의 발명에 있어서 그 물건의 생산, 사용, 양도 등 (양도 및 대여를 말한다. 그 물건이 프로그램 등인 경우에는 전기통신회선을 통한 제공을 포함한다. 이하 같다.), 수출 또는 수입 또는 양도 등의 청약 (양도 등을 위한 전시를 포함한다. 이하 같다.)을 하는 행위"라고 규정하고 있다.

AI 학습용 데이터세트의 특허법상 보호방안: 해석론 및 입법론

1. 제1안

특허법 제2조 제3호를 개정하여 실시유형을 추가하는 방안이다. 즉, 제1안은 학습용 데이터세트 배포를 '학습용 데이터세트 생성방법'의 실시 유형에 추가하는 것이다. 이는 학습용 데이터세트 거래실정이나 출원인이 인식하는 보호범위와 일치한다.

AI 학습용 데이터세트의 특허법상 보호방안: 해석론 및 입법론

제2조 제3호. “실시”란 다음 각 목의 구분에 따른 행위를 말한다.

- 나. **방법의 발명**인 경우: 그 방법을 사용하는 행위 또는 그 방법의 사용을 청약하는 행위
- 다. **물건을 생산하는 방법**의 발명인 경우: 나목의 행위 외에 그 방법에 의하여 생산한 물건을 사용·양도·대여 또는 수입하거나 그 물건의 양도 또는 대여의 청약을 하는 행위
- 라. **데이터 세트(전자적산출물? 가칭)를 생성하는 방법**의 발명의 경우: 나목의 행위 외에 그 방법에 의하여 생성된 데이터 세트의 사용·양도·대여 또는 수입하거나 그 데이터 세트의 양도 또는 대여의 청약을 하는 행위

AI 학습용 데이터세트의 특허법상 보호방안: 해석론 및 입법론

- 이는 현재 학습용 데이터세트 관련 출원의 청구범위 양식에 가장 적절하다는 장점이 있다. 이에 따르면, '학습용 데이터세트를 생성하는 방법'을 특허청구범위로 할 수 있게 된다.
- 하지만, 이 안은 소프트웨어 산출물까지 보호하는 등 보호범위가 과도하게 확장될 우려가 있다. 예를 들어, '블록체인을 생성하는 방법'도 이 범주에 해당하게 될 수 있다.

AI 학습용 데이터세트의 특허법상 보호방안: 해석론 및 입법론

2. 제2안

가. 물건의 정의

- 1958년 제정 이전의 구 의용민법 제85조는 “본법에서 물(物)이라 함은 유체물을 말한다.”라고 규정하여 물건을 유체물에 한정하고 있다. 그리고 독일 민법과 일본 민법은 물건을 유체물에 한정하고 있다.
- 하지만 1958년 제정된 우리 민법은 유체물뿐만 아니라 '전기 기타 관리할 수 있는 자연력'을 물건의 개념을 추가하여 물건의 범위를 무체물로 확대하였다. 이는 “사회경제의 발전과 새로운 과학기술의 출현에 대응하기 위한 것이라고 볼 수 있다.”고 할 수 있다.

AI 학습용 데이터세트의 특허법상 보호방안: 해석론 및 입법론

- 민법 제98조와 관련하여 일반적으로 (i) 유체물 및 관리할 수 있는 자연력, (ii) 관리가능성, (iii) 독립성, (iv) 비인격성이 물건의 개념요소를 거론되고 있다. 이에 대하여 '관리할 수 있는'이란 개념은 (ii) 요건과 중복되고 법문의 취지가 '유체물임과 동시에 자연력'이라는 의미는 아니기 때문에 (i)의 요건은 '유체물 또는 자연력'으로 하는 것이 보다 적절한 것으로 판단된다.
- 여기에서 '유체물'이란 "공간의 일부를 차지하고 사람의 감각에 의하여 지각할 수 있는 '형체'를 가진 것'을 말하고, 이는 고체, 액체뿐만 아니라 기체를 포함한다. 이 조문에서는 '관리할 수 있는'을 '자연력'이란 명사를 수식하는 형용사로 위치시킴으로써 유체물은 관리가능성과 무관한 것처럼 오해할 가능성도 있지만 관리가능성은 유체물에 대하여도 요구되는 요건이라는 데에는 이론이 없다. 무체물은 빛, 열, 전기, 에너지, 음향 등과 같이 공간을 차지하는 형체를 가지지 않는 것인데, 이러한 무체물 중 '자연력'만을 물건으로 파악하고 있다.

AI 학습용 데이터세트의 특허법상 보호방안: 해석론 및 입법론

- 이 개념에서 '자연력'이란 "자연계에 존재하는 힘 또는 이와 유사한 것"을 뜻한다. 형법은 재산죄의 객체인 재물의 개념에 관하여 '관리할 수 있는 동력'은 재물로 간주한다는 명문규정(형법 제346조, 제354조, 제361조, 제372조)을 두고 있어 '관리가능한 동력'이 이론의 여지가 없이 재산죄의 객체가 된다는 점 및 특허법상 발명은 물건 또는 방법을 그 대상으로 하는데, '봉입(봉입)된 전기발광성(전기발광성) 인광체 입자(인광체 입자)와 같이 인공적으로 만들어진 물체는 위 물건에 해당한다는 판시한 점에 비추어 볼 때, 인공적으로 창출된 물질이라도 자연계에 실제로 존재하는 것이면 자연력에 해당한다고 보는 것이 타당할 것이다.
- 따라서 자연력은 인공력에 대치되는 개념이 아니다. 그러한 관계로 인위적으로 만들어낸 에너지(발전소에서 생산된 전기 등)는 자연력에 해당한다.

AI 학습용 데이터세트의 특허법상 보호방안: 해석론 및 입법론

- 특히, 디지털정보와 관련하여 (i) 네트워크를 타고 전달되는 도중의 디지털정보는 물리적으로 전기신호이므로 이 때에는 물건이 된다는 입장과 (ii) 이는 유체물이라고 할 수도 없고 자연력이라고 할 수도 없어 민법상 물건의 개념에 해당한다고 해석하기 어렵다는 견해가 대립되고 있다.
- 컴퓨터프로그램의 경우, 독자적인 거래의 대상이 되고 있는바 민법상 물건으로 보아야 한다는 견해와 유체물이나 관리할 수 있는 자연력이라고 볼 수 없으므로 민법상 물건에 해당하지 않는다는 견해의 대립이 있다. 디지털정보에 대하여 입법론적으로 물건의 개념에 포섭하여야 한다는 견해도 존재한다.

AI 학습용 데이터세트의 특허법상 보호방안: 해석론 및 입법론

- 이렇듯 디지털정보가 물건의 범위에 포함되는지 여부에 대하여 의견의 대립이 있는 바, AI 학습용 데이터가 물건의 범위에 포함되는지 여부가 논란이 될 수 있다. 따라서 입법론적으로 이를 해결하는 방안을 제안할 수도 있을 것이다.

나. 제2안

- 물건의 개념에 대한 논란을 피하기 위한 입법론으로서 특허법 제2조 제3호를 개정하여 물건 범위를 확대하는 방안을 고려할 수 있다. 즉, 제2안은 '학습용 데이터세트'를 물건에 포함하는 방안이다.

AI 학습용 데이터세트의 특허법상 보호방안: 해석론 및 입법론

제2조 제3호. “실시”란 다음 각 목의 구분에 따른 행위를 말한다.

가. 물건(데이터세트를 포함)의 발명인 경우: 그 물건을 생산·사용·양도·대여 또는 수입하거나 그 물건의 양도 또는 대여의 청약(양도 또는 대여를 위한 전시를 포함한다. 이하 같다)을 하는 행위

- 이 안은 학습용 데이터 자체를 청구하므로 특허된 후에는 포괄적 보호가 가능하다는 장점이 있다. 하지만, 학습용 데이터 생성기술에 특징이 있어도 생성된 학습용 데이터 자체는 특징이 없는 경우가 많아 학습용 데이터 자체를 특허받기 어려워 실효성이 떨어진다는 단점이 있다. 그리고 데이터 세트에 대한 정의 규정이 없는 경우에는 용어의 정의와 관련하여 논란이 제기될 수 있다는 우려도 있을 것이다.

AI 학습용 데이터세트의 특허법상 보호방안: 해석론 및 입법론

3. 제3안

특허법 제129조를 개정하여 간접침해 유형을 추가하는 방안이다. 즉, 이 안은 학습용 데이터세트 배포를 인공지능 '학습방법' 특허의 간접침해로 규정하는 방안이다.

전용성 요건 때문에 간접침해 인정 사례가 저조하고, 학습방법 청구가 필수적이라는 점에서 이 안의 약점을 찾을 수 있다. 그리고 데이터 세트에 대한 정의 규정이 없는 경우에는 용어의 정의와 관련하여 논란이 제기될 수 있다는 점도 우려된다.

AI 학습용 데이터세트의 특허법상 보호방안: 해석론 및 입법론

제127조(침해로 보는 행위)

3. 특허가 방법의 발명인 경우: 그 방법의 실시에만 사용하는 데이터 세트(가
창)를 생성·양도·대여 또는 수입하거나 그 데이터 세트의 양도 또는 대여의
청약을 하는 행위

AI 학습용 데이터세트의 특허법상 보호방안: 해석론 및 입법론

4. 제4안: 데이터세트를 물건의 발명에 포섭하는 방안

단순히 제시되는 정보의 내용에만 특징이 있는 것으로 정보의 제시를 주된 목적으로 하는 경우(예: 녹음된 음악에만 특징이 있는 CD, 컴퓨터 프로그램 리스트 자체 등)에는 발명에 해당하지 않는다. 하지만 정보의 제시가 새로운 기술적 특징을 가지고 있으면 그와 같은 정보의 제시 그 자체, 정보제시수단, 정보를 제시하는 방법(예: 문자, 숫자, 기호로 이루어지는 정보를 양각으로 기록한 플라스틱카드(정보의 제시 수단에 기술적 특징이 있는 경우))은 발명에 해당될 수 있다.

AI 학습용 데이터세트의 특허법상 보호방안: 해석론 및 입법론

기술은 지식으로서 기량이나 기능과는 달리 타인에게 전달할 수 있는 객관성이 있는 것이어야 한다. 다시 말하면 해당 기술분야에서 평균적 수준을 가진 제3자가 행하더라도 같은 결과에 도달할 수 있어야 한다. 즉 반복재현성이 존재하여야 한다.

일본 학자들의 견해에 따르면, 기술은 일정한 목적을 달성하기 위하여 구체적 수단으로서 실시가능성 및 반복가능성이 있을 필요가 있다고 한다. 참고로 일본 판례에 따르면, 그 기술내용은 반복가능성, 구체성, 객관성을 필요로 한다고 판시한 바 있다.

AI 학습용 데이터세트의 특허법상 보호방안: 해석론 및 입법론

- 발명의 정의에 포함된 개념인 '기술'은 특허요건인 '산업상 이용가능성'과는 구별할 필요가 있다. 산업상 이용가능성 요건이 요구하는 것은 반복실시성의 존재이다. 특허법의 보호대상인 발명에 해당하지 않는 한, 논리적으로 보아 특허요건인 산업상 이용가능성은 논할 수 없다. 오로지 발명에 해당하는 경우에 산업상 이용가능성, 신규성, 진보성 등의 특허요건을 논할 수 있다.
- 발명의 성립성과 관련하여 언급되는 반복재현성의 존재여부는 산업상 이용가능성과 관련하여 언급되는 반복실시성의 존재여부와는 다른 개념으로 보는 것이 타당함에도 불구하고 판례에서는 불분명한 태도를 취하고 있는 것으로 보인다. 그렇다고 반복재현성이란 용어를 식물발명에 한정하여 설명하는 것도 논리적이라고 할 수는 없을 것이다. 다만, 특허법상 발명은 기술적 사상이므로 반드시 기술 그 자체일 필요는 없다.

AI 학습용 데이터세트의 특허법상 보호방안: 해석론 및 입법론

- 전술한 현행법상 논리를 취하면, 인공지능 학습용 데이터세트를 방법의 발명이 아니라 물의 발명으로 보호할 수 없게 되는 문제점이 발생할 여지가 있다. 4차산업혁명 시대에 인공지능을 학습시키는 학습용 데이터세트의 중요성을 감안할 때, AI 학습용 데이터세트는 발명에 해당할 수 있어야 한다.
- 따라서 학습기반 인공지능과 관련하여 인공지능 학습용 데이터 세트 (training datasets for AI)가 발명의 정의 (자연법칙을 이용한 기술적 사상의 창작으로서 고도한 것)를 충족하는 경우, 특허법상 보호의 대상이 되어야 할 것이다.

AI 학습용 데이터세트의 특허법상 보호방안: 해석론 및 입법론

- “인공지능 (학습, 추론, 판단, 이해, 행동 등 인간의 지적 능력의 일부 또는 전체가 전자적 방법으로 구현된 것을 말한다)이 학습용 데이터 세트(부호, 문자, 음성, 음향 및 영상 등으로 표현된 모든 종류의 자료 또는 지식의 집합을 말한다)를 학습하고 이를 기반으로 스스로 판단하여 결론을 도출하는 경우에는 이를 특허의 보호대상에 포함하는 것이 타당하다.
- 데이터세트 (부호, 문자, 음성, 음향 및 영상 등으로 표현된 모든 종류의 자료 또는 지식의 집합을 말한다)가 인공지능을 학습시킬 목적으로 제공된 경우에 이를 인공지능 학습용 데이터세트로 정의할 필요가 있다.

AI 학습용 데이터세트의 특허법상 보호방안: 해석론 및 입법론

현행 특허법

특허법 제2조(정의) 이 법에서 사용하는 용어의 뜻은 다음과 같다.

2. "특허발명"이란 특허를 받은 발명을 말한다.
3. "실시"란 다음 각 목의 구분에 따른 행위를 말한다.
 - 가. 물건의 발명인 경우: 그 물건을 생산·사용·양도·대여 또는 수입하거나 그 물건의 양도 또는 대여의 청약(양도 또는 대여를 위한 전시를 포함한다. 이하 같다)을 하는 행위
 - 나. 방법의 발명인 경우: 그 방법을 사용하는 행위 또는 그 방법의 사용을 청약하는 행위
 - 다. 물건을 생산하는 방법의 발명인 경우: 나목의 행위 외에 그 방법에 의하여 생산한 물건을 사용·양도·대여 또는 수입하거나 그 물건의 양도 또는 대여의 청약을 하는 행위

개정초안 (1순위안)

특허법 제2조(정의) 이 법에서 사용하는 용어의 뜻은 다음과 같다.

2. "특허발명"이란 특허를 받은 발명을 말한다.
 - 2의1. "인공지능"이란 학습, 추론, 판단, 이해, 행동 등 인간의 지적 능력의 일부 또는 전체가 전자적 방법으로 구현된 것을 말한다.
 - 2의2. "데이터 세트"란 부호, 문자, 음성, 음향 및 영상 등으로 표현된 모든 종류의 자료 또는 지식의 집합을 말한다.
 - 2의3. "인공지능 학습용 데이터 세트"란 인공지능을 학습시키기 위하여 제공된 데이터 세트를 말한다.
3. "실시"란 다음 각 목의 구분에 따른 행위를 말한다.
 - 가. 물건의 발명인 경우: 그 물건을 생산·사용·양도·대여 또는 수입하거나 그 물건의 양도 또는 대여의 청약(양도 또는 대여를 위한 전시를 포함한다. 이하 같다)을 하는 행위
 - 나. 인공지능 학습용 데이터세트의 발명인 경우: 그 인공지능 학습용 데이터세트를 사용·양도·양도의 청약·대여·대여의 청약·수입하거나 전기통신회선을 통하여 송신하는 행위
 - 다. 방법의 발명인 경우: 그 방법을 사용하는 행위 또는 그 방법의 사용을 청약하는 행위
 - 라. 물건을 생산하는 방법의 발명인 경우: 나목의 행위 외에 그 방법에 의하여 생산한 물건을 사용·양도·대여 또는 수입하거나 그 물건의 양도 또는 대여의 청약을 하는 행위

AI 학습용 데이터세트의 특허법상 보호방안: 해석론 및 입법론

- 개정초안 제2조 제3호 나목은 현재 학습용 데이터세트 관련 출원의 청구범위 양식에 가장 적절하다. 또한, 제1안과는 달리 소프트웨어 산출물까지 보호하는 등 보호범위의 과도한 확장의 우려를 불식시킬 수 있다는 장점도 있다.
- 그 밖에 인공지능과 데이터세트에 대한 정의도 명확하다. 입법론적으로 볼 때, 필자는 제4안을 1순위안으로 보고 제5안을 2순위안으로 고려할 수 있다고 생각한다.

AI 학습용 데이터세트의 특허법상 보호방안: 해석론 및 입법론

현행 특허법

제127조(침해로 보는 행위) 다음 각 호의 구분에 따른 행위를 업으로서 하는 경우에는 특허권 또는 전용실시권을 침해한 것으로 본다.

1. 특허가 물건의 발명인 경우: 그 물건의 생산에만 사용하는 물건을 생산·양도·대여 또는 수입하거나 그 물건의 양도 또는 대여의 청약을 하는 행위
2. 특허가 방법의 발명인 경우: 그 방법의 실시에만 사용하는 물건을 생산·양도·대여 또는 수입하거나 그 물건의 양도 또는 대여의 청약을 하는 행위

특허법 개정초안(1순위안)

제127조(침해로 보는 행위) 다음 각 호의 구분에 따른 행위를 업으로서 하는 경우에는 특허권 또는 전용실시권을 침해한 것으로 본다.

1. 특허가 물건의 발명인 경우: 그 물건의 생산에만 사용하는 물건을 생산·양도 등·대여 또는 수입하거나 그 물건의 양도 또는 대여의 청약을 하는 행위
2. 인공지능 학습용 데이터세트의 발명인 경우: 그 인공지능 학습용 데이터세트를 사용·양도·양도의 청약·대여·대여의 청약·수입하거나 전기통신회선을 통하여 송신하는 행위
3. 특허가 방법의 발명인 경우: 그 방법의 실시에만 사용하는 물건을 생산·양도 등·대여 또는 수입하거나 그 물건의 양도 또는 대여의 청약을 하는 행위
4. 특허가 물건을 생산하는 방법의 발명인 경우, 그 방법에 의하여 생산한 물건을 업으로서 양도 등 또는 수출을 위하여 소지하는 행위

AI 학습용 데이터세트의 특허법상 보호방안: 해석론 및 입법론

5. 제5안

- 물건에는 유체물뿐만 아니라 관리가능한 자연력도 포함되므로, 인공지능 교육을 위한 교육 데이터는 물론이고 인공지능 생성물인 학습 데이터뿐만 아니라 3D 프린팅의 3D 파일데이터도 포섭하기 위해서는 물건의 정의를 개정하기 보다는 발명의 실시 유형 중 양도에 '전기통신회선을 통한 송신'을 포섭하는 것을 고려할 필요가 있다. '전송'이나 '공중송신'이란 개념은 저작권법상 공중이 수신하거나 접근하게 할 목적으로 무선 또는 유선통신의 방법에 의하여 송신하거나 이용에 제공하는 것을 의미한다(저작권법 제2조 제7호 및 제10호).
- 여기에서 '공중'이란 "공중"은 불특정 다수인(특정 다수인을 포함한다)을 말한다(저작권법 제2조 제32호). 따라서 거래 상대방이 공중이 아닌 경우에는 '전송'이나 '공중송신'의 개념을 사용할 수 없다.

AI 학습용 데이터세트의 특허법상 보호방안: 해석론 및 입법론

- 그러한 측면에서 '전기통신회선을 통한 송신'이란 용어를 활용하는 것이 적절할 것으로 생각한다. 이 경우, 해석론으로서 AI 학습용 데이터세트와 AI 생성물인 AI 학습 데이터세트는 물건으로 취급할 수 있을 것이다. 따라서 이와 관련하여 AI를 교육하기 위하여 활용하는 학습용 데이터세트(AI 교육데이터)(Trainable data sets for AI)와 AI가 생성한 학습데이터세트 (Data sets generated by AI)를 구별하여 분석할 필요가 있을 것이다.
- 학습용 데이터세트를 생성하는 방법을 출원발명으로 하는 경우에는 AI를 교육시키는 학습용 데이터세트의 입력은 '사용의 청약'으로 볼 수 있을 것이다. 그리고 이러한 출원발명을 통해 생성된 AI 학습용 데이터세트의 온라인 송신은 개정초안 제2조 제3호 다목의 '양도 등'으로 포섭할 수 있을 것이다.

AI 학습용 데이터세트의 특허법상 보호방안: 해석론 및 입법론

현행 특허법	특허법 개정초안(2순위안)
<p>3. "실시"란 다음 각 목의 구분에 따른 행위를 말한다.</p> <p>가. 물건의 발명인 경우: 그 물건을 생산·사용·양도·대여 또는 수입하거나 그 물건의 양도 또는 대여의 청약(양도 또는 대여를 위한 전시를 포함한다. 이하 같다)을 하는 행위</p> <p>나. 방법의 발명인 경우: 그 방법을 사용하는 행위 또는 그 방법의 사용을 청약하는 행위</p> <p>다. 물건을 생산하는 방법의 발명인 경우: 나목의 행위 외에 그 방법에 의하여 생산한 물건을 사용·양도·대여 또는 수입하거나 그 물건의 양도 또는 대여의 청약을 하는 행위</p>	<p>3. "실시"란 다음 각 목의 구분에 따른 행위를 말한다.</p> <p>가. 물건의 발명인 경우: 그 물건을 생산·사용·양도 등 (양도와 대여를 뜻한다. 이는 전기통신회선을 통한 송신을 포함한다.)·수출 또는 수입하거나 그 물건의 양도 등의 청약(양도 등을 위한 전시를 포함한다. 이하 같다)을 하는 행위</p> <p>나. 방법의 발명인 경우: 그 방법을 사용하는 행위 또는 그 방법의 사용을 청약하는 행위</p> <p>다. 물건을 생산하는 방법의 발명인 경우: 나목의 행위 외에 그 방법에 의하여 생산한 물건을 사용·양도 등·수출 또는 수입하거나 그 물건의 양도 등의 청약을 하는 행위</p>

AI 학습용 데이터세트의 특허법상 보호방안: 해석론 및 입법론

- 미국 특허법에서는 포괄적인 형태의 '유도침해'규정이 있어 온라인 전송행위를 규제할 수 있고, 영국 및 독일은 간접침해의 대상을 '발명의 본질적 요소와 관련된 수단'으로 규정하고 있어 물건 이외의 대상에 대한 침해 가능성을 열어 두고 있다.
- 또한 일본은 특허법상 물건의 개념에 프로그램을 포함하여 물건 이외의 침해에 대한 대응이 가능한데 우리 특허법에서는 이를 규제할 수 있는 방법이 없다는 견해가 있다. 우리나라 민법에 따르면, 물건의 정의에서 '관리할 수 있는 자연력'을 넓게 해석할 여지는 있다.

AI 학습용 데이터세트의 특허법상 보호방안: 해석론 및 입법론

- 개정초안에 따르면, AI 학습용 데이터세트가 물건의 발명에 해당하는 경우, 그 AI 학습용 데이터세트를 업으로서 전기통신회선을 통하여 송신하기 위하여 소지하는 행위를 침해행위로 간주하여 AI 학습용 데이터세트에 대한 특허권을 보호받을 수 있다(개정초안 제127조 제2호).
- 또한 개정초안에 따르면, AI 학습용 데이터세트를 생성하는 방법이 특허발명인 경우, AI 학습용 데이터세트를 업으로서 전기통신회선을 통하여 송신하기 위하여 소지하는 행위는 침해행위로 간주된다(개정초안 제127조 제4호). 다만, 이와 관련하여 전용성 요건의 완화 여부는 우리나라 판례에 맡겨 해석할 수 있을 것으로 판단된다.

AI 학습용 데이터세트의 특허법상 보호방안: 해석론 및 입법론

현행 특허법

제127조(침해로 보는 행위) 다음 각 호의 구분에 따른 행위를 업으로서 하는 경우에는 특허권 또는 전용실시권을 침해한 것으로 본다.

1. 특허가 물건의 발명인 경우: 그 물건의 생산에만 사용하는 물건을 생산·양도·대여 또는 수입하거나 그 물건의 양도 또는 대여의 청약을 하는 행위
2. 특허가 방법의 발명인 경우: 그 방법의 실시에만 사용하는 물건을 생산·양도·대여 또는 수입하거나 그 물건의 양도 또는 대여의 청약을 하는 행위

필자의 특허법 개정초안 (2순위안)


제127조(침해로 보는 행위) 다음 각 호의 구분에 따른 행위를 업으로서 하는 경우에는 특허권 또는 전용실시권을 침해한 것으로 본다.

1. 특허가 물건의 발명인 경우: 그 물건의 생산에만 사용하는 물건을 생산·양도 등·대여 또는 수입하거나 그 물건의 양도 또는 대여의 청약을 하는 행위
2. 특허가 물건의 발명인 경우, 그 물건을 업으로서 양도 등 또는 수출을 위하여 소지하는 행위
3. 특허가 방법의 발명인 경우: 그 방법의 실시에만 사용하는 물건을 생산·양도 등·대여 또는 수입하거나 그 물건의 양도 또는 대여의 청약을 하는 행위
4. 특허가 물건을 생산하는 방법의 발명인 경우, 그 방법에 의하여 생산한 물건을 업으로서 양도 등 또는 수출을 위하여 소지하는 행위

출처

인공지능 학습용 데이터세트 보호를 위한
특허법상 주요 쟁점 연구*

이규호 (중앙대학교 법학전문대학원 교수), 인공지능 학습용
데이터세트 보호를 위한 특허법상 주요 쟁점 연구, 산업재산권,
제64호, 89-178면

- 
- 경청해 주셔서 진심으로 감사드립니다.
 - 이규호 (중앙대학교 법학전문대학원 교수)
 - E-mail: cion2004@hanmail.net