

# 학술정보 라이선스 계약서 내 TDM 및 AI 조항 분석 연구\*

## An Analysis of TDM and AI Clauses in Scholarly Information License Agreements

안 병 군 (Byoung-Goon An)\*\*  
정 진 우 (Ginu Chung)\*\*\*  
이 혜 진 (Hyejin Lee)\*\*\*\*

### 목 차

- |            |             |
|------------|-------------|
| 1. 서론      | 4. 분석 결과    |
| 2. 선행연구 분석 | 5. 결론 및 시사점 |
| 3. 연구 방법   |             |

### 초 록

TDM 및 AI 활용이 확대됨에 따라 학술정보 라이선스 계약서에서 콘텐츠의 TDM 및 AI 활용을 제한하는 조항이 포함되고 있다. 이와 같은 조항은 학술 연구와 도서관 운영 측면에서 중대한 문제를 야기할 가능성이 있으나, 국내에서는 아직 관련 논의가 미비하다. 이에 따라 본 연구에서는 계약서에 포함된 TDM 및 AI 활용 제한 조건을 분석하는 실증적 연구를 수행하였다. 귀납적 방법을 통해 실제 계약서에서 나타나는 TDM 및 AI 조항을 유형화하는 기준을 수립하였다. TDM 및 AI 조항을 TDM 입력, TDM 출력, AI 도구 활용, AI 학습 활용으로 구분하고, TDM 입력, AI 도구 활용, AI 학습 활용을 위한 분류 기준과 TDM 출력을 위한 분류 기준을 각각 수립하였다. 이후 연구를 통해 수립된 분석 기준을 적용하여 각 조항의 현황을 분석하였다. 모든 유형에서 가장 높은 빈도를 차지한 것은 '조건부 허용(공개)'였으나, '허용'의 정도나 세부적인 구분에서는 차이를 보였다. 연구를 통해 개발한 TDM 및 AI 조항의 유형 분류 기준은 향후 관련 조항이 확대됨에 따라 변화하는 양상을 일관성있게 관찰하는 기준이 될 수 있을 것으로 기대된다. 또한 국내에서는 아직 논의가 미비한 라이선스 내 TDM 및 AI 활용 제한 조항에 대한 문제를 제기하였다는 측면에서도 의미가 있다.

### ABSTRACT

As the use of TDM and AI expands, clauses restricting the use of content for TDM and AI are increasingly being included in license agreements. While such clauses have the potential to raise significant issues for academic research and library operations, related discussions remain limited in South Korea. In response, this study conducted an empirical analysis of the restrictions on TDM and AI usage in license agreements. Through an inductive approach, the study developed a classification framework for TDM and AI clauses based on their actual expressions in licensing documents. The clauses were categorized into four domains: TDM input, TDM output, use of AI tools, and use for training of AI. Separate classification criteria were established for (1) TDM input, use of AI tools, and AI training, and (2) TDM output. The resulting framework was applied to examine the current state of each clause type. Across all types, the most frequently observed category was "conditionally permitted (public)," though differences were observed in the levels of permission granted. The classification criteria for TDM and AI clauses are expected to serve as a consistent analytical tool for tracking changes as such clauses continue to evolve.

키워드: 학술정보 라이선스, 라이선스 계약서, AI 조항, TDM 조항, AI/TDM 제한

Scholarly information license, License agreement, AI clause, TDM clause, Restrictive clauses on AI/TDM usage

\* 이 논문은 2025년도 한국과학기술정보연구원(KISTI)의 기본사업으로 수행된 연구입니다.

(과제번호: (KISTI) K25L3M1C1)

\*\* 한국과학기술정보연구원 선임연구원(rns0305@kisti.re.kr / ISNI 0000 0005 0283 2705) (제1저자)

\*\*\* 한국과학기술정보연구원 선임기술원(ginu@kisti.re.kr / ISNI 0000 0004 7383 6618) (공동저자)

\*\*\*\* 한국과학기술정보연구원 책임연구원(hyejin@kisti.re.kr / ISNI 0000 0004 6490 0147) (교신저자)

논문접수일자: 2025년 7월 21일 최초심사일자: 2025년 7월 30일 게재확정일자: 2025년 8월 11일

한국문헌정보학회지, 59(3): 181-201, 2025. <http://dx.doi.org/10.4275/KSLIS.2025.59.3.181>

※ Copyright © 2025 Korean Society for Library and Information Science

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (<https://creativecommons.org/licenses/by-nc-nd/4.0/>) which permits use, distribution and reproduction in any medium, provided that the article is properly cited, the use is non-commercial and no modifications or adaptations are made.

## 1. 서론

오늘날 텍스트·데이터 마이닝(Text Data Mining, 이하 TDM) 및 인공지능(Artificial Intelligence, 이하 AI) 기술이 발전함에 따라 논문 전체를 요약, 번역하거나 논문을 짧은 영상으로 생성하는 등 AI 도구를 활용하는 사례가 증가하고 있다. 또한 TDM 및 AI를 위한 양질의 학습 데이터로서 논문의 중요성도 높아졌다. 이와 같은 상황에서 학술정보 콘텐츠 공급사는 학술정보 라이선스 계약서(이하 라이선스 계약서)에 콘텐츠의 TDM 및 AI 활용을 제한하는 조항을 포함하기 시작했다(ICOLC, 2024). TDM 및 AI 활용 제한 조항은 장기적으로 연구 발전과 도서관 운영에 부정적인 영향을 미치게 될 것이다. Sandiumenge(2023)는 공개된 저작물로만 AI를 학습시킬 경우 AI 성능 저하 및 산출물의 편향 가능성이 존재함을 지적하였다. 또한 AI 학습 대상의 제한은 AI를 활용한 연구 질문의 제한으로 이어지고, 연구 주제와 연구 방법에도 편향성이 발생할 수 있다는 우려도 제기되었다(CNI, 2023). 도서관 운영 측면에서 발생하는 문제도 존재한다. Samberg(2023)는 공급사가 학술정보 라이선스 계약서에 TDM 및 AI 활용을 금지한 후, TDM 및 AI 활용을 허가하는 라이선스를 다시 판매할 가능성이 존재한다고 주장한다. 이 경우 도서관의 재정적 부담이 가중될 뿐만 아니라, 자금이 부족한 분야, 기관, 국가에서는 연구를 위한 콘텐츠 확보가 어려워질 수 있다. 이러한 문제 상황에 선제적으로 대응하기 위해 여러 사서가 문제를 제기하고 해결 방안을 모색하고 있으며(Taplin, 2024; Nobre & Hackett, 2025), EU,

미국, 일본 등 여러 나라에서 TDM 및 AI 활용과 관련한 제도와 법률을 마련하고 있다(차상욱, 2021; 김경숙, 2025).

국내에서도 TDM 및 AI를 고려하여 저작권법을 개정하고자 하였으나 아직 법률이나 정책이 마련되지 못한 상태(류시원, 2024)에서 최근 국내에서 체결된 학술정보 라이선스 계약에도 TDM 및 AI를 금지하는 조항이 포함되기 시작했다. 공적인 법률과 규제가 정착되기 이전에는 도서관과 학자가 TDM 및 AI 활용 권리를 지키기 위해 적극적으로 행동할 필요가 있으나(Samberg et al., 2024a), 국내에서는 관련 논의가 충분히 이루어지지 않고 있다.

본 연구는 라이선스 계약서에서 TDM 및 AI 제한을 최소화하고 이용자의 권익을 보장하는 방안을 모색하는 연구의 일환으로, 라이선스 계약서에 명시된 TDM 및 AI 제한 조건에 주목하였다. 구체적으로, 현재 라이선스 계약서에서 TDM 및 AI 조항이 명시되는 양상을 파악하고 어떠한 방식으로 TDM 및 AI 활용을 제한하고 있는지에 초점을 맞추어 분석하였다. 이를 위하여 실제 라이선스 계약서 내 명시된 TDM 조항과 AI 조항을 조사하고 제한 혹은 허용 정도를 유형화하였으며, 각 유형에서 드러나는 빈도와 양상을 분석하였다.

## 2. 선행연구 분석

### 2.1 TDM 및 AI 이용 규정 관련 연구

TDM 및 AI의 이용이 활발해지면서 관련된 규정의 필요성이 대두되었다. 이에 따라 여러

분야에서 다양한 관점으로 TDM 및 AI의 이용 규정에 대한 연구를 진행하고 있다. Vrakas (2024)는 근거 이론 접근법을 토대로 TDM에 적용되는 규제 프레임워크를 검토한 88편의 문헌을 조사하여 합법적인 TDM에 대한 미국, EU, 영국에 소속된 학계 및 실무자의 견해를 종합적으로 분석하였다. TDM에 적용되는 규제를 다루는 문헌은 7개 범주를 가진다는 점을 밝혔으며(Taxonomy, Copyright and Related Rights, Data Protection, Fundamental Rights, Contract Law, Competition Law, Generalist Approach), 각 분류에 해당하는 문헌의 구체적인 내용을 기술하였다. 연구에서는 TDM과 더불어 생성형 AI에 대한 논의를 함께 다루면서 TDM에 대한 논의의 연장선에서 생성형 AI에 대한 논의가 이루어지고 있음을 드러냈다.

Kim(2024)은 한국의 저작권법 개정안을 분석하고 세계적인 동향과 어떠한 차이를 보이는지 연구하였다. 연구에서는 TDM과 생성형 AI는 방대한 학습 데이터를 활용하므로 학습 단계에서부터 저작권 침해가 발생할 수 있다고 밝혔다. 또한 현행 국내 공정 이용 조항만으로는 적용 범위가 제한적이므로 TDM 예외 및 제한에 대한 규정을 제정함으로써 문제의 해결책을 마련할 수 있다고 주장했다. 연구에서는 국내에서 저작권법의 개정 방안으로 제안된 TDM 예외 조항과 EU의 디지털 단일시장 저작권 지침(Directive on Copyright in the Digital Single Market, 이하 DSM), 일본 저작권법의 TDM 예외 조항을 비교하였다. 이와 더불어, TDM 예외 조항이 포함되지 않을 경우 현행 한국 저작권법의 공정 이용(Fair Use) 조항(제35조의5)을 TDM 예외에 적용할 수 있다고 제시하였

다. 이를 위하여 TDM과 AI에 대한 판례를 추적해 온 미국의 공정 이용 원칙을 분석하였다.

Kumar et al.(2023)은 AI 기술 중에서도 자연어 처리 모델(Natural Language Processing Model, NLP 모델)에 주목하여 AI 규제와 법안에 대해 분석하였다. 이를 위하여 연구에서는 우선 OECD, UNESCO 등의 AI 규제 이니셔티브를 조사하였으며, 이후 AI 사용에 적용될 수 있는 GDPR(General Data Protection Regulation) 조항들을 제시하였다. 연구에서는 EU에서 진행되는 AI 법안에서 채택한 “AI의 위험 기반 접근 방식”과 “고위험 인공지능 모델”을 논의하면서, 특정 조건을 충족할 경우 AI 금지조항으로 인해 연구가 제한되어서는 안 된다는 점을 언급했다. 마지막으로 대규모 언어 모델(Large Language Model, 이하 LLM)이 유발할 수 있을 것으로 추정되는 위험과 그에 대한 논박을 제시하였다.

Xalabarder(2025)에서는 TDM 및 AI 학습과 AI 결과물이 저작권 침해에 해당하는지를 중심으로 미국과 EU의 AI 규정이 어떠한 차이를 보이는지 분석하였다. 연구에 따르면 EU의 DSM 지침은 연구 목적의 TDM은 무보상이 허용된다고 명시하고 있다. 또한 연구에서는 EU가 AI와 관련된 정책 문서를 발의하고 AI 시스템을 4단계 위험 수준으로 구분한 AI법을 제정하였다고 설명하였다. 연구는 미국의 경우 공정 이용 조항이 과학과 예술의 진보를 위한 TDM에도 적용될 수 있다는 관점인 것을 설명하면서 관련 판례를 제시했다. 유럽의 DSM 지침과 미국의 공정 이용 조항 모두 산출물을 고려한 조항이 아니므로, 생성형 AI의 산출물과 관련해서는 앞으로의 판단이 중요해질 것이라고 언급하였다.

## 2.2 학술정보 라이선스 계약에서 TDM 및 AI 이용 조항 분석 연구

라이선스 계약에서 TDM 및 AI 이용 관련 조항이 빈번하게 나타나기 시작한 것이 비교적 최근의 사안이므로, 아직 관련 연구가 다양하지 않다. 그러나 실제 계약 과정에서 불공정한 이용 제한을 경험한 사서들이 문제를 제기하고 있으며, 도서관과 관련한 여러 기구에서 이러한 문제의 해결 방안을 모색하고 있다.

Samberg et al.(2024b)은 최근 공급사와의 라이선스 계약에 연구자의 콘텐츠 활용과 공정 이용을 제한하는 조항이 포함되고 있다고 지적하였다. 이러한 조항은 콘텐츠 활용을 위하여 별도의 라이선스 비용을 지불해야 하는 환경을 조성하고 있으며, TDM 및 AI 관련 연구에 제한점으로 작용될 수 있다고 주장하였다. TDM과 AI를 위한 학습 대상으로 저작권이 있는 콘텐츠를 활용하는 것은 공정 이용에 해당하고 공급사와의 계약에서 이러한 권리를 지켜야 한다고 이야기하는 한편, AI 학습과 AI 출력의 차이점을 밝히면서 생성형 AI의 출력이 공정 이용에 해당하는지는 예측하기 어렵다고 언급하였다.

Lazarova(2024)는 저작권 콘텐츠의 이용에 대한 라이선스 조건을 평가하는 연구의 일환으로 100건의 계약서에 포함된 TDM 제한 조항을 분석하였다. 연구에서는 저작권 콘텐츠 이용이라는 측면에서 계약서를 '허용', '금지', '제한', '비언급', '불명확'으로 구분하였으며, TDM에 대한 직접적인 금지 언급이 없더라도 간접적인 금지 혹은 제한을 명시한 경우까지 모두 포함하였다. 그 결과 TDM을 허용하는 계약서는 9개인 반면 직간접적으로 제한하는 경우가

28건, 금지가 31건으로 높게 나타났다.

Nobre와 Hackett(2025)에서는 사서가 공급사와 라이선스 계약을 체결할 때 발생할 수 있는 불공정한 관행을 다루면서, TDM 및 AI 금지 조항을 주요하게 언급했다. 연구에서는 실제 계약서 조항을 제시하며 2023년부터 AI 불허 조항이 나타나기 시작했으며, AI 학습뿐만 아니라 AI 도구 사용까지 제한하고 있다고 밝혔다. 나아가 개별 도서관의 협상력이 공급사보다 약하다는 점을 이용하여 공급사가 AI에 이어 TDM도 제한하는 조항을 포함하려 한다고 언급했다. 연구에서는 이러한 제한으로 인해 공급사가 지나치게 과도한 통제권을 가지게 될 수 있다고 주장하였다.

Taplin(2024)은 JISC가 인공지능 시스템에서 콘텐츠 사용을 제한하는 조항을 제외하기 위해 노력해 왔으며, JISC의 모델 라이선스에서는 머신러닝이나 AI를 목적으로 라이선스된 콘텐츠를 사용하는 것을 제한하지 않는다고 밝혔다. 연구에서는 ICOLC의 AI 성명을 지지하면서 협상 시 반대 입장을 표해야 하는 TDM 및 AI 제한 조항을 구체적으로 나열하고 대체 조항을 제안하였다. 연구에서는 TDM 및 AI 제한 조항을 제외하는 주장을 강화하기 위해서는 공동의 협력을 통해 강한 협상력을 발휘할 수 있어야 한다고 강조하였다.

## 3. 연구 방법

### 3.1 분석 대상 선정

본 연구에서는 실제 계약서 조항을 분석하고

자 ESAC Registry를 통해 공개된 OA 전환계약서를 분석의 대상으로 선정하였다. 일반적인 구독 계약서는 기밀 조항으로 인해 공개되지 않는 반면, OA 전환계약은 투명성을 지향하며 계약서를 공개하는 경우가 다수 존재한다. 선행연구에서도 OA 전환계약서에 대한 분석을 토대로 TDM 및 AI 제한 조항을 분석한 사례가 존재한다(Nobre & Hackett, 2025). 선행연구와 충분한 계약서 입수 측면을 고려하여 공개된 OA 전환계약서를 연구의 분석 대상으로 결정하였다.

ESAC Registry에 등록된 1,355개 OA 전환계약 중 계약서가 공개됨을 명시한 사례는 540건이었다(2025.06.08. 기준). 이들 중 실질적으로 계약서에 접근 가능한 494건을 분석 대상으로 하되, Temple와 Young(2004) 연구를 준용하여 영어가 아닌 외국어의 경우 번역된 문서를 검토하는 과정에서의 제한점을 고려하여 제외하였다.

### 3.2 분석 절차 및 방법

계약서 내 TDM 및 AI 조항을 분석하기 위하여 접근 가능한 계약서 494건을 전수조사하였다. 이들 중 영어가 아닌 외국어로 작성된 23건을 제외한 471건에 대하여 TDM 혹은 AI 관련 조항의 존재 여부와 그 내용을 파악하였다. 자동 다운로드 혹은 색인 프로그램(로봇, 스파이더, 크롤러 등)의 사용을 금지하는 조항을 TDM 금지로 평가하는 선행연구도 있으나(Lazarova, 2024), 본 연구에서는 TDM 혹은 AI에 대한 직접적인 언급이 있는 조항만을 분석 대상으로 포함하였다.

이후 개별 조항의 TDM 및 AI 허용 정도를 검토하였다. TDM 조항은 TDM을 개발하기 위한 자료의 입력(input)과 개발한 TDM 생성물의 이용과 관련한 출력(output)으로 조항의 내용이 구성되는 점을 파악하고, TDM 입력과 출력을 구분하여 허용 정도를 조사하였다. AI 조항의 경우 라이선스된 콘텐츠를 이용하는 과정에서 제3자가 개발한 AI 도구를 사용하는 것과 콘텐츠를 AI 학습 데이터로 활용하는 것은 성격이 다르므로(Nobre & Hackett, 2025), AI 조항을 AI 도구 활용과 AI 학습 활용으로 나누고 각각의 허용 정도를 조사하였다. TDM 입력과 AI 도구 활용, AI 학습 활용의 경우 모든 조항의 허용 정도를 '허용', '조건부 허용', '불허', '비언급'으로 구분하고 조건부 허용을 구체적으로 세분하기 위한 기준을 수립하였다. 하나의 세부 조건이 전체 계약서 중 9% 이상의 계약서에서 동일하거나 유사한 내용으로 기술된 경우 세분화된 조건으로서 추가하는 귀납적 방식으로 조건부 허용의 세분화 기준을 수립하였다(Fung et al., 2003). TDM 출력 조항은 생성물의 공개와 관련된 것이므로, TDM 입력과 동일한 기준을 사용하기 어렵다고 판단하여 별도의 공개 허용 정도 기준을 수립하였다.

TDM 입력, TDM 출력, AI 학습, AI 도구 관련 조항 존재 여부와 허용 정도를 조사한 내용을 토대로 분석을 진행하였다. 우선 허용 정도 구분에 따라서 계약서상 어떠한 방식으로 기술되는지를 확인하고 대표적인 기술 방식을 정리하였다. 이후 계약서상 TDM 및 AI 관련 조항이 나타나는 빈도와 변화 양상을 조사하였으며, 조건부 허용의 세분화된 구분에 따른 기술 분석과 시계열 분석을 수행하였다. 라이선



〈그림 1〉 연구 절차 도식

스 계약 기간은 1년부터 3년까지 상이하지만, 해당 계약을 기반으로 서비스가 제공되기 시작하는 연도를 기준연도로 정하고 분석을 진행하였다. 〈그림 1〉은 본 연구의 절차를 도식화한 것이다.

## 4. 분석 결과

### 4.1 조항 분류 기준 수립

#### 4.1.1 조항 세분화 기준 수립

저작권과 관련하여 TDM은 데이터 수집, 데이터 처리, 데이터 처리 후 과정으로 구분된다(김용주, 2020). 라이선스 계약서에서는 계약

된 콘텐츠를 TDM에 활용하기 위한 수집과 처리 과정을 하나의 입력 과정으로 간주하고 있으며, 처리 후 과정을 생성물의 활용으로 명시한다. AI의 경우 AI 학습, AI 생성물 활용, AI 플랫폼 활용 과정으로 논의된다(김경숙, 2025). 라이선스 계약서에서도 라이선스된 콘텐츠를 AI 학습 자료로 활용하는 것과 콘텐츠를 이용하는 과정에서 제삼자가 개발한 AI 도구를 활용하는 것을 구분하고 있다. 다만 생성형 AI를 이용하여 만들어진 산출물과 관련한 내용은 라이선스 계약서에서 주요하게 다루이지 않는다.

본 연구에서는 TDM 및 AI의 저작권과 관련된 선행연구를 토대로 계약서 조항을 TDM 입력과 TDM 출력, AI 도구 활용, AI 학습 활용으로 구분하였다(김용주, 2020; Vrakas, 2024:

김경숙, 2025; Nobre & Hackett, 2025). 이후 네 가지 구분과 관련된 개별 조항의 양상을 분석하였으며, 이를 기반으로 분류 기준을 수립하고 대표적인 표현을 정리하였다. TDM 입력과 AI 도구 활용, AI 학습 활용은 라이선스 콘텐츠의 활용과 관련된 부분인 반면, TDM 출력은 라이선스 콘텐츠를 토대로 만들어진 TDM의 생성물과 관련된 부분이므로 조항의 성격이 달라 동일한 기준을 적용하기 어렵다고 판단하였다. 이에 따라 TDM 출력을 위한 구분 기준을 별도로 마련하고 대표적인 표현을 정리하였다. 본 연구의 조항 구분은 <표 1>과 같다.

#### 4.1.2 TDM 입력, AI 활용, AI 학습의 허용 정도 분류 기준 수립

라이선스 계약서 내에서 TDM 및 AI에 대한 허용 정도는 크게 '허용', '조건부 허용', '불허', '비언급'으로 구분하였다. 라이선스 계약서에서 TDM 및 AI 조항을 포함하지 않는다면 TDM 및 AI에 대한 명시적인 제약은 없는 한편, 문제 발생 시 관할법원 조항, 수집 금지 조항 등 다른 관계 조항을 토대로 기관(이용자)과 공급사가 법률적 분쟁을 해야 한다(Association of Research Libraries, 2024). 따라서 TDM 및 AI 조항에 대한 언급이 있는지 자체가 중요한 의미를 지니므로 '비언급'을 구분에 포함하였다.

'조건부 허용'의 경우 구체적으로 어떤 조건을 제시하며 TDM 및 AI 사용을 허가 혹은 제한하는지를 확인할 필요가 있으나, 관련 구분 기준이나 전반적인 사례를 제시한 선행연구가 존재하지 않는다. 따라서 대상 계약서를 검토하며 다수의 계약서에서 빈번하게 나타나는 유형들을 정리하는 귀납적 방식을 채택했다. 이용 대상자를 권한이 있는 이용자로 제한하는 것과 같이 비단 TDM 및 AI 조항이 아닌 계약서 전반에서 전제조건이 되는 제한사항에 대해서는 세분화 과정에서 세부 조건으로 포함하지 않았다. 서로 다른 계약서에서 다소간에 표현적 차이가 있더라도 의미적으로 동일하다면 동일한 제한 조건으로 간주하였으며, 한 계약서에서 여러 조건을 제시하는 경우 중복으로 분류하였다. 그 결과, '조건부 허용'은 '서면 허용', '목적', '환경', '방식', '사전 보고', '사후 삭제' 6가지로 구분되었다.

'서면 허용' 조건은 TDM 및 AI 관련 조항의 시작 부분에서 주로 명시되며, TDM 및 AI를 수행하기 이전에 별도의 서면 허가를 받은 경우가 아니라면 이용을 금지한다는 내용을 포함하는 것이 일반적이다. 다만 단서 조항을 통해 '서면 허용' 조건이 있더라도 서면 허용 없이 TDM 및 AI 활용을 허용하는 다른 조건을 제시하기도 한다. '목적' 조건은 특정한 목적으로

<표 1> 분석 대상 조항 분류

구분		설명
TDM	TDM 입력	콘텐츠를 TDM 학습 및 분석에 활용하는 것과 관련한 내용이 명시된 조항
	TDM 출력	콘텐츠를 기반으로 개발된 TDM의 생성물에 대한 권리 및 활용 사항을 명시한 조항
AI	AI 도구	콘텐츠 활용 시 AI 도구를 사용하는 내용이 명시된 조항
	AI 학습	콘텐츠를 AI 학습에 활용하는 내용이 명시된 조항

사용할 수 있거나, 특정 목적으로는 사용할 수 없다는 조항이다. 일반적으로 학술, 연구, 교육 목적의 사용만 허용하거나 상업적 목적의 사용을 금지하는 내용이 포함된다. '환경' 조건은 TDM 및 AI를 수행하는 컴퓨팅 인프라와 관련된 내용으로, 계약서에서 요구하는 보안이 갖춰진 서버나 내부 서버에서만 사용할 것을 요구하는 조항이 이에 해당한다. '방식' 조건은 TDM 및 AI 활용을 위한 데이터 수집 방식에 제한을 두는 조건이다. '방식' 조건에 해당하는 대부분의 조항에서는 TDM 및 AI를 위한 데이터 수집 과정에서 크롤링, 스크래핑 등의 방법을 제한하고 특정 서비스나 API를 통해서만 데이터를 수집할 수 있음을 명시하고 있다. '사전 보고' 조건은 TDM 및 AI 프로젝트를 진행하기 이전, 공급사에 별도의 보고 절차를 요구하는 경우이다. 이메일을 통한 통지나 플랫폼을 통한 등록 등 방식에 따라서 통지(notify), 허가(authorization, permission), 동의(consent), 승인(approval), 등록(register)과 같은 다양한 표현이 사용되고 있으나, 공급사가 TDM 및 AI 작업을 사전에 인지하고자 요구하는 의도는 동일하다. '방식' 조건과 '사전 보고' 조건은 모두 공급사 콘텐츠에 대한 TDM 및 AI를 전체적으로 통제·관리하고 모니터링하고자 하는 것이다. 이를 통해 허가된 수집 활동과 허가되지 않은 크롤링 및 스크래핑을 구분하고 서버에 과부하 문제가 생기는 것을 방지하고자 하는 것으로 추정된다. '사후 삭제' 조건은 TDM 및 AI 활용이 종료된 이후 혹은 라이선스 권한이 종료된 이후에는 TDM 및 AI에 활용된 공급사의 콘텐츠를 모두 삭제할 것을 요구하는 조건이며, 모든 콘텐츠를 삭제했음을 밝히는 파기인증서

를 요구하는 경우도 있다. 다만 '사후 삭제' 조항에서도 단서 조항을 통해 학술적 증빙 목적의 보관은 허용하는 경우도 드물게 존재한다. 이와 같은 여섯 가지 세부 조건에 포함되지 않는 경우 '기타'로 구분하였다. '기타'에 해당하는 대표적인 사례로는 계약서에 명시된 특정 콘텐츠 범위 혹은 분량 기준 내에서만 TDM 및 AI 활용을 허가하는 조건이 명시된 경우가 있었으며, TDM 및 AI를 위한 API와 서비스 제공에 비용을 청구하는 사례 등이 존재한다. 본 연구를 통해 도출한 TDM 입력, AI 활용, AI 학습의 조항 분류 기준과 대표적인 문구를 정리한 내용은 <표 2>와 같다.

#### 4.1.3 TDM 출력 공개의 허용 정도 분류 기준 수립

TDM 출력 조항은 라이선스 콘텐츠를 학습하여 개발된 TDM의 산출물에 대해 어느 정도의 공개를 허용하는지에 주목하여 분류를 진행하였다. TDM 출력의 분류 기준을 제시한 선행연구도 존재하지 않으므로, 귀납적 방식을 통해 기준을 수립하였다. TDM 출력 조항의 기준을 수립하는 과정에서도 하나의 분류로 인정받기 위해서는 출현 빈도가 전체 계약서의 9% 이상이 되어야 한다는 조건을 동일하게 적용하였다. 그 결과, TDM 출력 조항은 '공개', '제한적 공개', '비공개'로 분류되었다. TDM 입력과 관련된 내용은 명시되어 있으나 TDM 출력에 대한 사항은 포함하지 않은 경우 '비연급'으로 구분하였으며, 하나의 항목으로 구분하기에는 빈도가 적은 내용에 대해서는 '기타'로 분류하였다. <표 3>은 본 연구에서 TDM 출력과 관련된 조항을 분류하고 대표적인 문구를 정리한 것이다.

〈표 2〉 TDM 입력, AI 도구, AI 학습 조항 분류

구분	설명
허용	콘텐츠에 대한 TDM 혹은 AI 도구 및 AI 학습(이하 AI 활용)을 별도의 조건 없이 허가하는 경우
조건부 허용	서면 허용 콘텐츠에 대한 TDM 혹은 AI 활용을 위해서는 공급사의 별도 서면 허용이 요구되는 경우 예) <i>"Customer and Authorized Users are prohibited, without publisher's express written permission..."</i>
	목적 계약서에 명시된 특정 목적으로 사용하는 경우에만, 혹은 특정 목적으로 사용하지 않는 경우에만 콘텐츠에 대한 TDM 혹은 AI 활용이 가능한 경우 예) <i>"AI tools must be used solely for non-commercial academic or research purposes."</i>
	환경 계약서에 명시된 특정 사용 환경을 충족하는 경우에만 콘텐츠에 대한 TDM 혹은 AI 활용이 가능한 경우 예) <i>"...download and store the Licensed Material Dataset on a secure server in a closed, hosted environment for use by the End-User only."</i>
	방식 계약서에 명시된 특정 방식으로만 TDM 혹은 AI 활용을 위한 수집이 가능한 경우 예) <i>"...through an approved API service such as CrossRef's TDM service or a publisher API."</i>
	사전 보고 콘텐츠에 대한 TDM 혹은 AI 활용 이전 공급사에 별도의 보고(통지, 허가, 등록 등)가 요구되는 경우 예) <i>"...the Authorized User registers each such Text And Data Mining activity by providing the details of such activity via an email to..."</i>
	사후 삭제 콘텐츠에 대한 TDM 혹은 AI 활용 이후 활용된 콘텐츠의 삭제 처리가 요구되는 경우 예) <i>"Copies of Licensed Material made under this clause shall be deleted promptly after the computational analysis has been completed."</i>
기타	그 외 콘텐츠에 대한 TDM 혹은 AI 활용과 관련하여 언급되는 조건이 있는 경우
불허	콘텐츠에 대한 TDM 혹은 AI 활용을 허가하지 않는 경우
비언급	계약서 내 TDM 혹은 AI 관련 조항이 존재하지 않는 경우

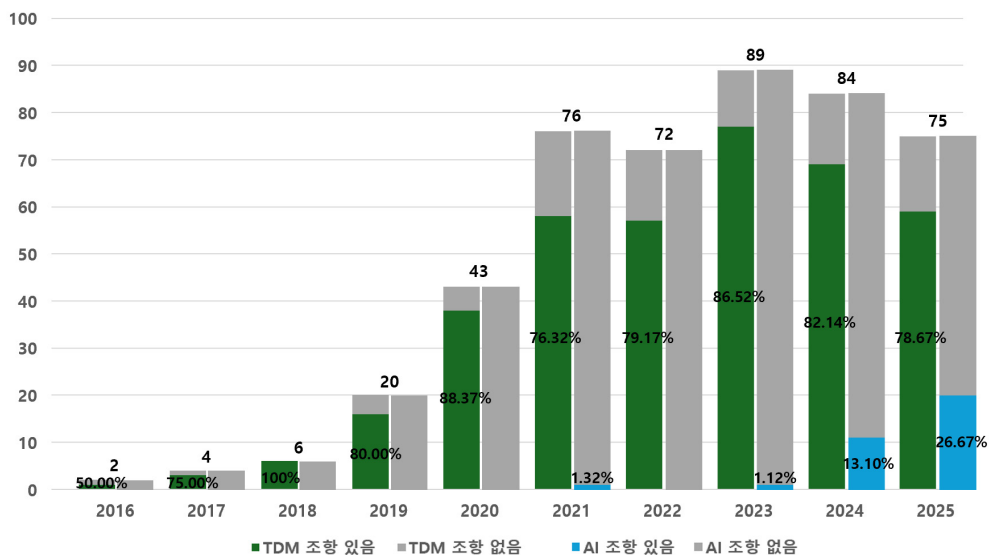
〈표 3〉 TDM 출력 조항의 허용 수준 분류

구분	설명
공개	콘텐츠를 기반으로 개발된 TDM의 생성물을 특정한 조건 없이 자유롭게 공개, 공유, 배포할 수 있는 경우 예) <i>"Members and Authorized Users may make the results available for use by others..."</i>
제한적 공개	콘텐츠를 기반으로 개발된 TDM의 생성물을 특정한 조건을 충족한 경우에만 공개, 공유, 배포할 수 있는 경우 예) <i>"TDM Output shall be licensed under a Creative Commons CC-BY NC ND licence, include a link to the DOI of the applicable article wherever practicable and include a maximum of 200 characters from the original article."</i> 예) <i>"The Authorized Users ... may utilize and share the results of text and/or data mining in their scholarly work and make the results available for use by others, so long as the purpose is not to create a product for use by third parties that would substitute for the Licensed Materials."</i>
비공개	콘텐츠를 기반으로 개발된 TDM의 생성물을 공개할 수 없다고 명시된 경우 예) <i>"Authorized Institution shall not share the results from the text and data mining with any third party, including, without limitation, Licensee or Participating Institutions."</i>
비언급	계약서 내 콘텐츠를 기반으로 개발된 TDM의 공개 혹은 활용에 대한 언급이 존재하지 않는 경우
기타	그 외 콘텐츠를 기반으로 개발된 TDM의 공개 혹은 활용과 관련하여 언급되는 조건이 있는 경우

‘공개’로 구분되는 조항은 생성물의 공개가 가능하다고 명시하고 특정한 제약 조건을 제시하지 않은 경우에 해당한다. 반면에 TDM 생성물의 공개를 불허하고 TDM 생성물을 공개할 수 있는 특수한 사례나 조건을 단서 조항으로 제시하지 않았다면 ‘비공개’로 구분하였다. ‘제한적 공개’는 생성물의 공개가 가능하다고 밝혀져 제약 조건을 둔 경우나, 생성물의 공개가 불가능하다고 명시하고 공개할 수 있는 단서 조항을 제시한 경우에 해당한다. TDM 입력과 마찬가지로 TDM 출력에 대해서도 제한적 공개의 세부적인 내용을 확인하였으나, 대부분이 생성물의 사용 목적에 대한 제약에 해당하고, 그 외 기준이 될 만큼의 빈도를 보이는 항목은 없었다. 따라서 ‘제한적 공개’에 대한 세부적인 구분을 하는 것이 유의미하지 않다고 판단하였다.

#### 4.2 분석 대상 계약서 및 TDM/AI 조항 현황

개별 조항의 TDM 및 AI 양상을 파악하기에 앞서, 전체 분석 대상 계약서와 TDM 및 AI 조항이 존재하는 계약서를 시계열적으로 분석하였다(〈그림 2〉 참조). ESAC Registry를 통해 공개되어 실질적으로 접근 가능한 OA 전환 계약서가 분석의 대상이므로, OA 전환계약의 계약서 공개 추이에 따라 연도별 분석 대상 계약서의 수가 상이하며, 2016년부터 2018년까지는 매년 10건 미만의 계약서만이 분석의 대상이 되고 있다는 점에서 한계가 있다. 그러나 EU의 DSM이 2019년 6월부터 발효되었고(European Parliament, 2018), 미국에서는 2021년 10월부터 디지털 밀레니엄 저작권법(Digital Millennium Copyright Act, 이하 DMCA)을 통해 TDM 예외 조항을 명시하고 있다(Sag, 2020). 이와 같은 국제적인 추세와 더불어, 2020년 이후부



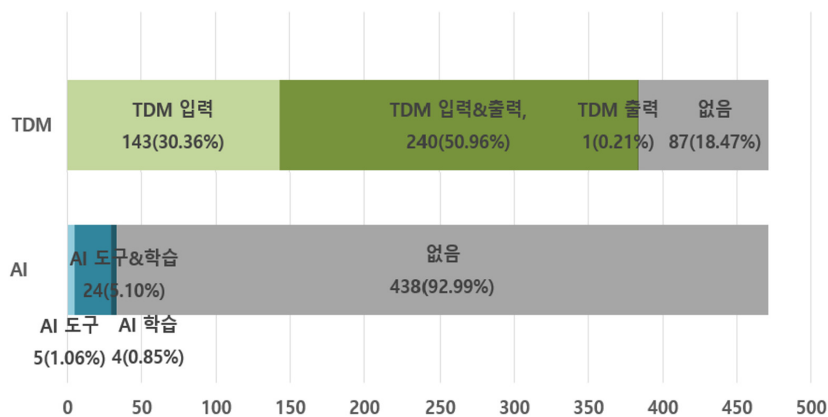
〈그림 2〉 TDM 및 AI 조항 존재 여부 시계열 그래프

터 분석 대상 계약서가 일정 수준 이상 존재한다는 점을 종합적으로 고려하여 분석을 진행하였다.

TDM과 관련한 법률 및 제도가 정착되기 이전인 2019년까지 TDM 조항이 존재하는 경우가 상당수 존재했으며, 2020년도 이후부터는 전체 계약서 중 80% 내외에서 TDM 조항을 포함하고 있다. 반면 AI가 일반적으로 사용되기 시작한 것은 2022년(Nobre & Hackett, 2025)부터이며, 이러한 배경으로 인해 전체 분석 대상 계약서 471건 중 AI와 관련한 규정이 포함된 계약서는 33건(7.01%)으로 매우 소수이다. 그러나 2021년도 1건, 2023년도 1건의 AI 조항 포함 계약이 체결된 데 반해 2024년도에는 11건, 2025년도에는 20건의 계약서에 AI 조항이 포함되면서 큰 폭으로 증가하는 추세를 보이고 있다. 따라서 분석 대상이 소수일지라도 AI 관련 조항이 확산되는 초기 단계의 특성을 조사하고 향후 변동을 추정하는 측면에서 분석을 진행하였다.

〈그림 3〉은 전체 분석 대상 계약서 471편에

서 TDM과 AI 관련 조항이 명시된 비율을 조사한 결과이다. 전체의 81.53%에 해당하는 384건의 계약서에 TDM 관련 내용이 명시되어 있으며, 이를 통해 TDM 관련 조항이 라이선스 계약의 일반적인 조항임을 알 수 있었다. TDM 입력은 383건(81.32%), TDM 출력은 241건(51.17%)의 계약서에 명시되어 있으며, TDM의 입력과 출력을 모두 다루는 사례가 240건(50.96%)으로 가장 많았다. TDM 입력과 관련된 내용 없이 TDM 출력만을 단독으로 다루는 경우는 1건뿐이었다. 분석 대상 계약서의 대부분이 TDM 관련 사항을 다루는 한편, AI와 관련된 조항이 포함된 계약서는 33건(7.01%)으로 매우 소수였다. 이들 중 29건(6.16%)의 계약서에서 AI 도구에 관한 내용을 포함하였고, AI 학습에 관한 내용을 다룬 계약서는 28건(5.95%)으로 확인되었다. TDM과 마찬가지로 AI 조항에서도 AI 도구 활용과 AI 학습 활용 두 가지 요소 모두를 포함하는 계약서가 24건으로 가장 많았다.



〈그림 3〉 TDM 및 AI 조항 존재 현황 그래프

### 4.3 TDM 관련 조항 분석

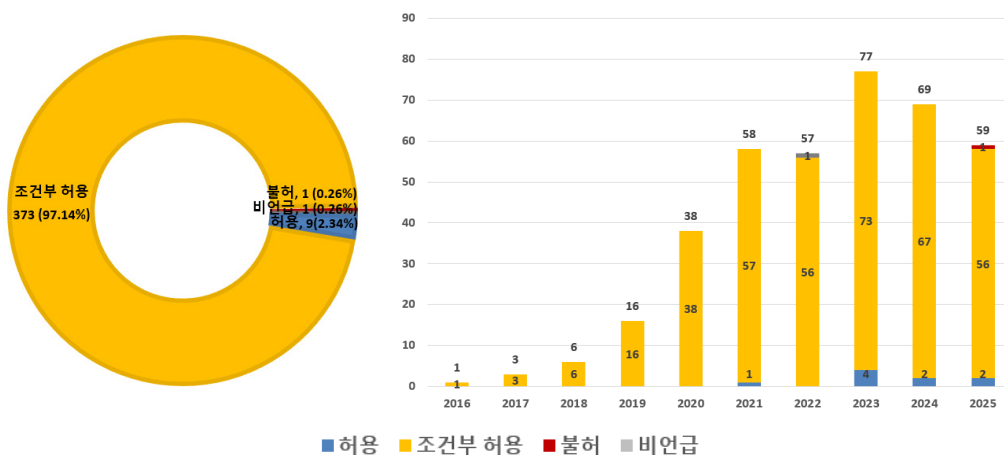
#### 4.3.1 TDM 조항의 특징

TDM 조항은 하나의 제목 조항으로 제시되거나 별첨 문서로 존재할 만큼 세부적인 내용이 기술된 경우가 많았다. TDM 입력 조항과 관련하여 TDM 모델 개발을 위해 라이선스된 콘텐츠를 활용하는 것에 대한 구체적이고 자세한 조건을 제시한다. TDM 입력 조항이 제시된 계약서 중 일부는 TDM 모델을 통해 산출된 생성물에 대한 권리와 공개 가능 범위를 다루는 TDM 출력 조항을 포함하기도 한다. TDM은 연구와 산업계에서 오래전부터 활용한 기술이고, 영미권과 유럽에서는 관련 규정 및 제도를 만들었으며 그 외 국가에서도 각 국가별 저작권법 내에서 TDM을 어떻게 규제 혹은 허가할 것인지에 대한 논의가 활발하다(이상용, 2021). 이러한 현황은 라이선스 계약서에도 반영되어 DSM과 DMCA를 비롯해서 각국의 저작권법을 준수한다는 내용이 포함되기도 한다.

또한 384개 계약서의 36.98%인 142개 계약서에서 생성물의 소유자를 명시하고 있으며, 공개 허용 수준과 무관하게 생성물의 소유권은 모두 허가된 이용자 혹은 허가된 이용자의 소속 기관에 있다고 밝혔다. 이는 TDM 출력의 소유권과 관련해서는 국제적인 합의가 형성된 것이라고 할 수 있다. TDM 관련 조항이 매우 구체적이고 상세하기 때문에 소수의 계약서에서는 구체적인 사항은 이메일로 문의하거나 공급사의 웹사이트를 참고하라고 명시되어 있기도 하다. 이 경우 공급사의 웹사이트 내용을 지속적으로 추적하기는 어려우므로, 웹사이트 내 TDM 관련 사항이 합의 없이 변동되지 않도록 단서조항이 추가될 필요가 있을 것이다.

#### 4.3.2 TDM 입력 관련 조항 분석

〈그림 4〉는 TDM을 위해 라이선스 콘텐츠를 입력하는 것이 계약서에서 어느 정도 허용되고 있는지를 정리한 것이다. TDM을 언급하는 384건의 계약서 중 아무런 조건 없이 활용을

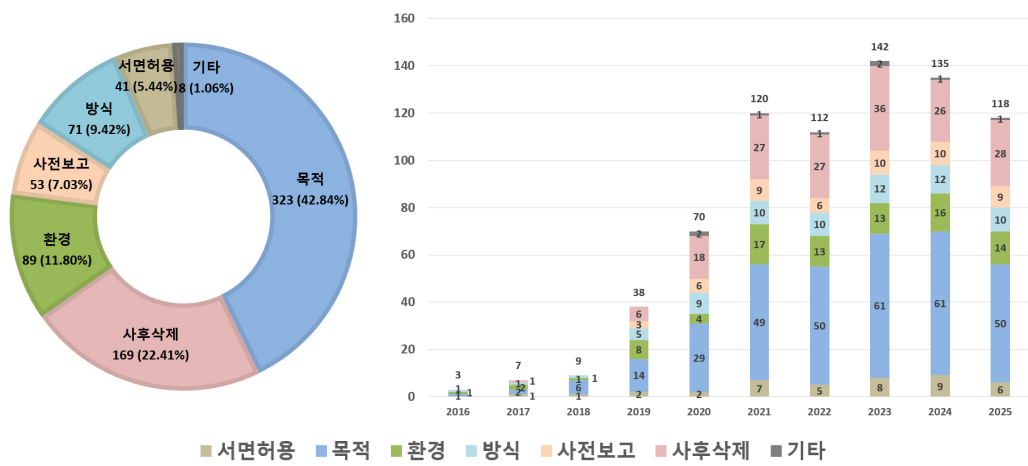


〈그림 4〉 TDM 입력 허용 여부의 분포 및 시계열적 분석

허용한 경우는 9건(2.34%)으로, 처음 ‘허용’이 나타난 2021년부터 2025년까지 매우 적은 비율만을 차지했다. TDM 입력을 전혀 허용하지 않는 계약도 1건(0.26%) 존재했으며, 2025년 헝가리에서 체결한 계약이다. 헝가리도 EU에 해당하여 DSM 지침의 영향을 받기 때문에 포괄적으로는 학술적 목적의 TDM을 허용하고 있으나, TDM에 대한 구체적인 내용을 기술하는 조항을 통해 전면적인 활용 금지를 밝혔다. ‘조건부 허용’은 전체의 97.14%(373건)를 차지하므로, TDM 입력과 관련해서 사실상 대부분의 계약이 ‘조건부 허용’을 택하고 있다는 점을 확인할 수 있었다.

‘조건부 허용’의 경우 하나의 계약서에서 한 가지 조건만 요구하는 경우도 있었으나, 본 연구에서 분류한 6가지 항목을 모두 요구하는 사례도 발견되었다. 이와 같이 하나의 계약서에 여러 조건을 요구하는 경우 중복 분류를 진행하였으며, ‘조건부 허용’에 대한 세부 조건의 분포를 확인한 결과는 <그림 5>와 같다.

‘조건부 허용’ 중 가장 높은 비율을 차지한 것은 ‘목적’ 조건(323건, 42.84%)이며, ‘사후 삭제’ 조건과(169건, 22.41%) ‘환경’ 조건(89건, 11.80%)이 뒤를 이었다. ‘조건부 허용’에 해당하는 373건 중 50건을 제외한 323건이 ‘목적’ 조건을 포함하고 있으며, ‘목적’ 조건은 2016년부터 2025년까지 항상 다수를 차지하고 있었다. 이러한 점을 통해 TDM 입력과 관련해서 공급사에서는 학술 및 연구 목적으로 제한하거나 상업적인 목적을 금지하는 등의 조건을 가장 일반적이고 기본적인 항목으로 간주한다는 점을 확인할 수 있었다. ‘사후 삭제’ 조항은 2019년부터 나타나기 시작해서 2023년까지 꾸준히 증가하는 추세를 보였다. 라이선스 종료 이후 혹은 TDM이 사용된 프로젝트 종료 이후 TDM을 위해 사용된 콘텐츠를 전부 삭제할 것을 요구하는 조항이 전체 ‘조건부 허용’의 45.31%를 차지한다는 것은 공급사에서 라이선스 콘텐츠의 유출을 우려하고 있기 때문으로 추정된다. ‘사전 보고’ 조건과 ‘방식’ 조건은 2019년도에



<그림 5> TDM 입력 조건부 허용 세부 조건 분포 및 시계열적 분석

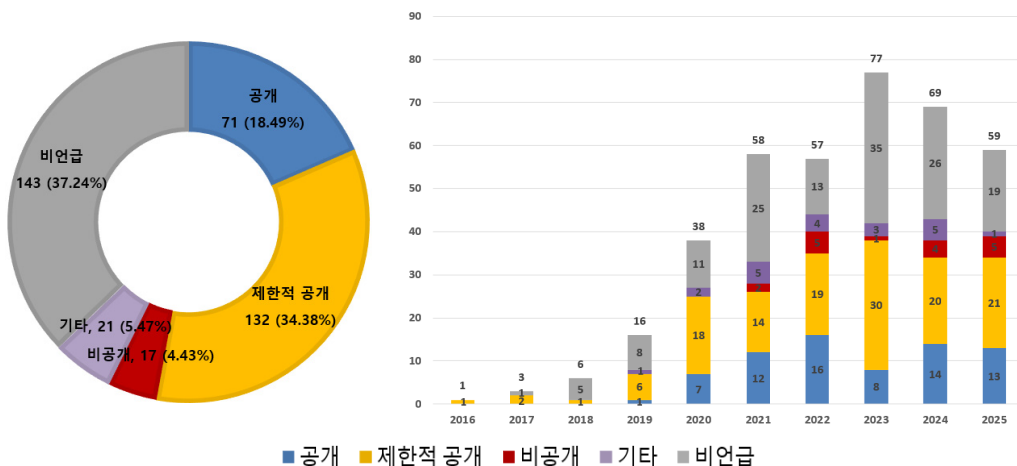
처음 나타났으며, 2021년도부터는 매년 10건 내외의 수치를 유지하고 있다. 높은 비중을 차지하는 '목적'과 '사후 삭제' 조항을 제외한 4개 조항의 빈도와 비율은 유사하게 나타나고 있으며, 이는 연도별 그래프에서도 마찬가지이다. 일정 수준의 규제와 법률이 정립된 TDM에 대해서는 공급사와 참가 기관이 다양한 단서 조항을 통해 각자의 권한을 확대하고자 하는 과정에서 다양한 세부 조건이 제시되고 있다고 할 수 있다.

#### 4.3.3 TDM 출력의 공개 허용 정도 분류

TDM을 통해 만들어진 생성물의 공개 허용 정도를 분류한 결과는 <그림 6>과 같다. TDM 입력 조항은 존재하나 TDM 출력 관련 내용은 존재하지 않는 '비연급'이 143건(37.24%)으로 가장 높게 나타났다. TDM 출력 조항이 없는 TDM 입력 조항 계약서는 1건뿐이었던 것을 고려하면 TDM 출력 조항은 아직 TDM 입력 조항만큼 일반적으로 포함되는 내용이 아니라

는 점을 알 수 있다.

'제한적 공개'가 132건(34.38%), '공개'가 71건(18.49%)으로, '공개'에 해당하는 비율이 전체의 과반수(52.87%)를 차지했다. TDM 입력과 마찬가지로 TDM 출력도 '제한적 공개'가 가장 높은 비율로 나타났으나, TDM 입력과 달리 TDM 출력은 대부분이 공개 및 활용 목적을 제한하는 방식이었다. 구체적으로, 연구 및 교육 목적의 사용은 허가하되 상업적 목적의 활용은 불허한다는 조항이 명시되어 있었다. 추가로, 공개할 수 있는 분량을 제한하거나 공개 시 저작권 혹은 출처를 밝히는 방법을 구체적으로 제시한 경우도 있었으나, 그 수가 많지는 않았다. TDM 출력물 공개 정도에서 '비공개'가 나타난 것은 2021년부터이며, 소수이지만 2023년부터 그 빈도와 비율이 증가하는 경향을 드러내고 있다.



<그림 6> TDM 출력 공개 허용 여부의 분포 및 시계열적 분석

#### 4.4 AI 도구 및 AI 학습 활용 관련 조항 분석

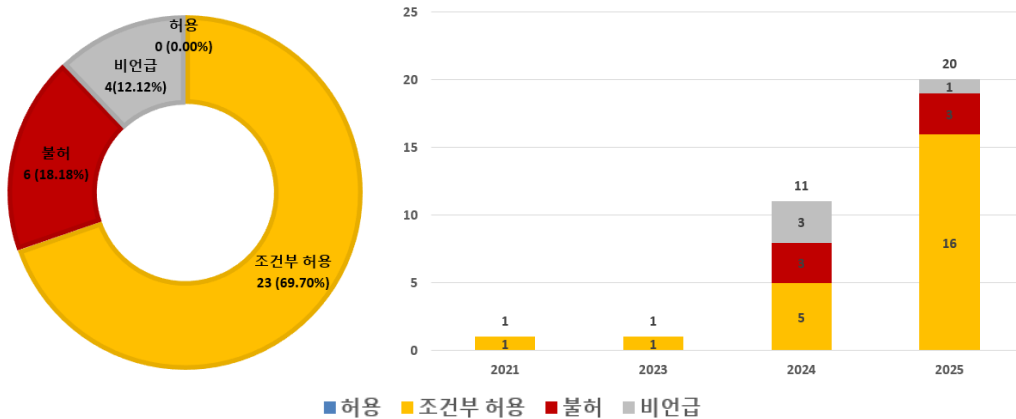
##### 4.4.1 AI 도구 및 AI 학습 활용 관련 조항의 특징

TDM 관련 조항은 별도의 제목 조항이나 별첨 문서로 제시되는 경우가 다수 존재하는 반면, AI 조항은 TDM과 함께 명시된 경우를 제외하면 대부분 세부 조항 혹은 단서 조항으로 기술되어 있다. 일반적으로 AI 관련 사항은 이용 제한사항(Restrictions on Use, Prohibited Uses 등)이나 이용 조건(Rights of Use, Permitted Uses 등)을 명시하는 제목 조항에 포함된다. AI와 관련된 내용을 언급하는 과정에서 LLM이 함께 기술된 경우도 존재하며, 향후 AI 기술의 변화에 대응하여 기관과 공급사가 적극적으로 협력하겠다는 내용이 명시되기도 한다. 이는 TDM 규정에 비해 AI 규정은 아직 구체적인 사항을 제시할 수 있는 단계에 이르지 못했으며, 현재 관련 법률과 제도의 수립을 대비하며 전반적인 사항을 조율하는 단계에 있음을 의미한다.

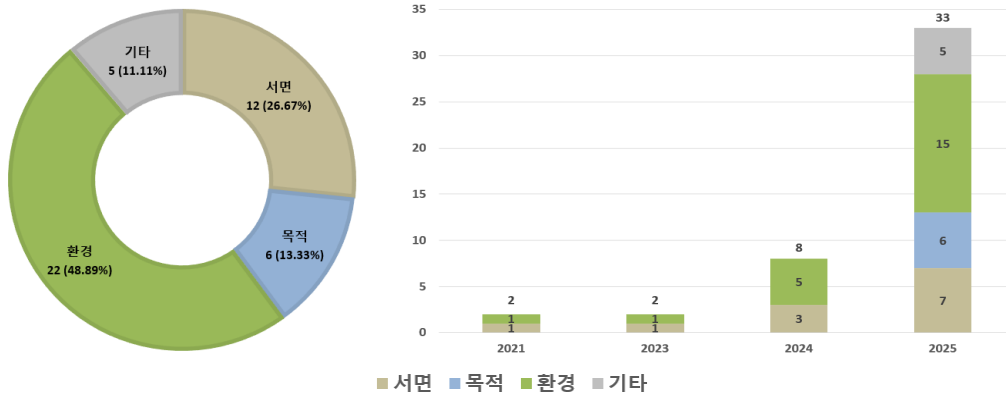
##### 4.4.2 AI 도구 활용 조항 분석

〈그림 7〉은 AI 도구 활용 관련 조항의 허용 정도 분포와 시계열 변화를 나타낸 것이다. AI 관련 조항 자체가 소수이므로, AI 조항이 있는 계약서 중에서 AI 도구 활용 조항이 존재하지 않는 사례만 '비언급'으로 표기하였다. AI 조항이 제시된 33개 계약서 중 AI 도구 활용을 조건부로 허용하는 사례는 23건(69.70%), 허용하지 않는 사례는 6건(18.18%)이며, '조건부 허용' 23건 중 16건이 2025년 계약서에 명시된 것으로 확인되었다. 이를 통해 AI 도구 활용 조항이 포함된 계약서가 증가하는 가운데, 향후 TDM과 마찬가지로 AI 도구 활용 조항에서도 '조건부 허용' 방식을 주로 채택하게 될 것이라는 점을 추정할 수 있다. 조건 제한 없이 콘텐츠의 AI 도구 활용을 허용하는 사례는 한 건도 존재하지 않았다.

허용 정도가 '조건부 허용'으로 구분된 23개 조항에서 세부적으로 어떤 조건을 요구하는지에 대한 분석을 진행하고, 세부 조건별 비율과 연도별 변화 양상을 조사하였다(〈그림 8〉 참조).



〈그림 7〉 AI 도구 활용 허용 여부의 분포 및 시계열적 분석



〈그림 8〉 AI 도구 조건부 허용 세부 조건 분포 및 시계열적 분석

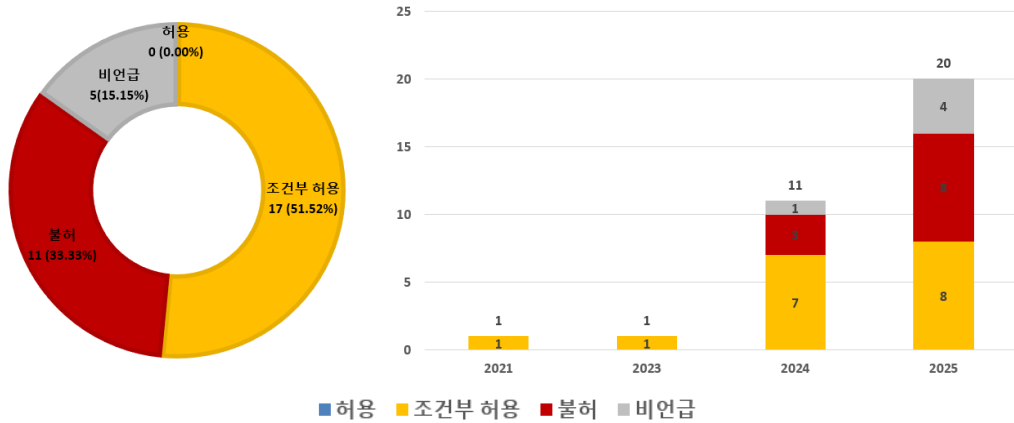
하나의 계약서에서 복수의 조건을 요구하는 경우 중복으로 분류하였으며, 그 결과 23건의 계약서에서 45개의 제한 조건을 확인하였다. AI 도구와 관련한 ‘조건부 허용’의 세부 조건은 TDM의 세부 조건과는 매우 다른 양상을 보였다. AI 도구를 사용하는 컴퓨팅 인프라 환경을 조건으로 제시하는 경우가 22건(48.89%)으로 가장 많았으며, 두 번째로 많은 ‘서면 허용’(12건, 26.67%)과 큰 차이를 보였다. ‘조건부 허용’에 해당하는 23개 계약서 중 1개만을 제외한 22개 계약서에서 ‘환경’ 조건을 채택할 만큼 AI 도구 사용과 관련해서는 컴퓨팅 환경을 중요시하고 있음을 확인할 수 있었다. TDM 입력 조항과 달리 ‘방식’이나 ‘사전 보고’, ‘사후 삭제’를 요구하는 경우는 없었으나, 이는 AI와 관련된 구체적인 사항이 결정되지 않은 영향으로 추정되며, 향후 TDM처럼 관련 규정이 구체화될 경우 세부적인 조건이 추가될 것으로 예상된다.

시계열 분석 결과, 2024년에 이르러서야 다양한 세부 조건이 제시되는 점이 확인되었다. 특히 2025년 ‘조건부 허용’에 해당하는 계약서가 16건인데 비해 세부 조건은 33건으로 한 계

약서당 평균적으로 2개 이상의 세부 조건을 요구한다는 점을 확인할 수 있었다. ‘목적’ 조건의 경우 2024년까지 서면 조건과 큰 차이를 보이지 않았으나 2025년 계약에서 급격하게 증가하는 추세를 보였으므로, 향후 AI 도구 활용과 관련해서는 도구를 활용하는 목적에 대한 중요성이 강조되고 관련 제약이 증가할 것으로 추정된다.

#### 4.4.3 AI 학습 활용 조항 분석

AI 도구 활용과 마찬가지로 AI 학습 활용 또한 관련 조항을 포함한 계약서의 개수와 허용 정도, 허용 정도의 시계열적 변화를 조사하였다(〈그림 9〉 참조). AI 도구 활용에서는 ‘조건부 허용’이 지배적인 경향이었던 반면에, AI 학습 활용에서는 ‘조건부 허용’(17건, 51.52%) 뿐 아니라 ‘불허’(11건, 33.33%)도 높게 나타났다. AI 도구 활용 조항에서 ‘조건부 허용’을 택했으나 AI 학습 활용은 ‘불허’로 결정한 경우가 7건(30.43%), AI 학습 활용에 대해서는 언급을 하지 않은 경우가 3건(13.04%)이었다. 특히 2025년에 AI 학습을 허용하지 않는 조항이



<그림 9> AI 학습 허용 여부의 분포 및 시계열적 분석

명시되는 경우가 급격히 증가하며 '조건부 허용'과 동일한 수치를 보였다. 17건의 '조건부 허용'에서 24건의 세부 조건이 제시되었다. AI 도구와 관련해서는 '환경' 조건이 절반을 차지했던 것과 달리, AI 학습 활용의 세부 조건에서는 사전에 서면을 통한 허가 및 합의를 요구하는 경우가 가장 많았으며(10건, 41.67%), '환경'(8건, 33.33%)과 '목적'(6건, 25.00%)이 뒤를 이었다. 분석을 통해 공급사에서는 AI 도구 활용에 비해 AI 학습 활용과 관련하여 더욱 방어적인 경향을 보이고 있음을 확인할 수 있었다. 다만 AI 학습 활용의 경우 현 시점에 관련 내용을 논의한 계약서가 소수이므로, 향후 AI 학습 활용을 다루는 계약서가 다수 축적된 시점에 다시금 분석을 진행할 필요가 있다.

### 5. 결론 및 시사점

본 연구는 국제적으로 문제가 제기되고 있으나 국내에서는 관련 논의가 미비한 학술정보

라이선스 계약서 내 TDM 및 AI 제한 조항에 대한 실증 연구이다. 학술정보 라이선스 계약서에서 TDM 및 AI 조항이 어떠한 양상으로 나타나는지를 분석함으로써 장기적으로 계약에서 TDM 및 AI와 관련한 이용자의 권익을 보호하는 데에 기여하고자 하였다. 이를 위하여 연구에서는 ESAC Registry를 통해 분석 대상 계약서 471건을 확보하고 조항을 분류별로 유형화 하였으며, 이를 토대로 양상을 분석하였다. 분석 결과는 다음과 같다.

학술정보 라이선스 계약서에 포함된 TDM 및 AI 관련 조항은 TDM 입력과 TDM 출력, AI 도구 활용과 AI 학습 활용으로 구분된다. TDM 입력, AI 도구 활용, AI 학습 활용 관련 내용은 콘텐츠에 대한 활용 측면인 반면 TDM 출력 내용은 생성물에 대한 공개 관련 규정으로 그 성격이 상이하므로, 각기 다른 방식으로 분석을 진행하였다.

TDM 조항이 포함된 계약서는 2016년도부터 꾸준히 존재하여 현재 전체 분석 대상 계약서의 81.53%에 해당한다. 반면 AI 조항이 최초

로 포함된 것은 2021년부터이며, 전체 대상의 7.01%만 해당할 정도로 소수였다. 다만 최근 2년간 급격한 증가세를 보이고 있으며, 향후 AI 기술이 발전할수록 AI 관련 조항은 더욱 증가할 것이다. 따라서 AI 관련 조항의 양상을 초기 단계부터 조사 및 분석한다는 측면에서 의의가 있다.

TDM 입력, AI 도구 활용, AI 학습 활용과 관련해서는 '허용', '조건부 허용', '불허', '비언급'의 4개 대분류를 제시하였으며, '조건부 허용'에 대해서는 7개 세부 조건 분류('서면 허용', '목적', '환경', '방식', '사전 보고', '사후 삭제', '기타')를 수립하였다. TDM 입력은 '조건부 허용'이 전체의 97.14%로 절대 다수를 차지했으며, '목적(42.84%)'과 '사후 삭제(22.41%)' 세부 조건의 비율이 높았다. AI 도구 활용 조항은 TDM 입력에 비해 '불허'의 비율이 높아지면서(18.18%) '조건부 허용'의 비율이 다소 낮아졌으나(69.70%) 여전히 다수를 차지하고 있었다. AI 학습 활용 조항도 '조건부 허용(51.52%)'이 가장 높게 나타나며 '불허(33.33%)'의 비중도 적지 않았다. TDM 출력 관련 조항은 '공개', '제한적 공개', '비공개', '비언급', '기타'의 5개 분류로 구분되었으며, '제한적 공개(34.38%)'와 '공개(18.49%)'의 비율이 높게 나타났다. 분석을 통해 TDM 조항은 어느 정도 규정이 정립된 반면, AI 조항은 아직 논의의 단계에 있음을 확인할 수 있었다. 또한 공급사는 TDM에 비해 AI와 관련해서 방어적인 태도를 보이고

있으며, AI 중에서도 AI 도구 사용보다는 AI 학습 활용에서 더욱 보수적인 경향을 확인할 수 있었다.

본 연구는 라이선스 계약서에서 명시되는 TDM 및 AI 조항의 유형 분류 기준을 제시하였으며, 실제 유형 분류를 통해 현황을 조사했다는 측면에서 의의가 있다. 본 연구에서 개발한 TDM 및 AI 조항에 대한 구분 기준은 향후 조항의 변화 양상을 일관성 있게 관찰하는 기준이 될 수 있을 것으로 기대된다. 또한 본 연구는 국내에서도 학술정보 라이선스 계약서 내 TDM 및 AI 조항에 대한 논의가 이루어지는 시발점이 된다는 측면에서도 의의가 있다. 다만 본 연구는 계약서 입수 문제로 인해 OA 전환계약서만을 분석의 대상으로 하고 있으며, 따라서 '공개된 OA 전환계약서'의 특성을 배제할 수 없다는 측면에서 한계가 있다. 또한 AI 조항을 포함한 계약서가 아직 소수이며 관련된 연구가 미비하므로 현황을 파악하고 분석하는 것에 중점을 두었다. 본 연구에서는 TDM 및 AI 제한 조건에만 국한하여 현황을 조사하였으나, 향후 관련 조항의 특성을 종합적으로 분석하여 라이선스 협상 전략 수립과 체크리스트 개발을 위한 연구를 진행할 예정이다. 또한 다양한 종류의 계약서가 공개되고 AI와 관련한 논의가 더욱 확대된 시점에 다시금 TDM 및 AI 조항의 현황을 파악하는 후속 연구를 진행하고자 한다.

## 참 고 문 헌

- 김경숙 (2025). 생성형 AI와 저작권 문제: 저작권 침해와 저작물성의 사례와 분석. *계간 저작권*, 38(2), 5-54.
- 김용주 (2020). 텍스트 및 데이터마이닝을 위한 저작권법 개정 방향. *법학연구*, 61(2), 283-314.  
<https://doi.org/10.35275/pnulaw.2020.61.2.011>
- 류시원 (2024). 인공지능 시대 저작권 정책 형성절차에 관한 제언. *법제*, 704, 109-146.  
<https://doi.org/10.23028/moleg.2024.704..004>
- 이상용 (2021). 데이터의 비계약적 이용: 데이터 마이닝을 위한 저작권 제한을 중심으로. *강원법학*, 65, 1-54.
- 차상욱 (2021). 저작권법상 인공지능 학습용 데이터셋의 보호와 쟁점: 텍스트·데이터마이닝(TDM) 면책규정을 중심으로. *경영법률*, 32(1), 1-55.
- Association of Research Libraries (2024). The enforceability of license restrictions on ingestion for artificial intelligence.  
[https://www.arl.org/wp-content/uploads/2024/04/Contract-override\\_AI.pdf](https://www.arl.org/wp-content/uploads/2024/04/Contract-override_AI.pdf)
- CNI (Coalition for Networked Information) (2023, December 12). Closing plenary: open access, open scholarship, & machine learning: a panel & community conversation. CNI fall 2023 membership meeting.  
<https://www.cni.org/events/membership-meetings/past-meetings/fall-2023/plenary-sessions-f23>
- European Parliament (2018). The exception for text and data mining (TDM) in the proposed directive on copyright in the digital single market - legal aspects (Study No. PE 604,941). Directorate-General for Internal Policies.  
[https://www.europarl.europa.eu/RegData/etudes/STUD/2018/604941/IPOL\\_STU\(2018\)604941\\_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2018/604941/IPOL_STU(2018)604941_EN.pdf)
- Fung, B. C. M., Wang, K., & Yu, P. S. (2003). Hierarchical document clustering using frequent item sets. *Proceedings of the 2003 SIAM International Conference on Data Mining*, 59-70.  
<https://doi.org/10.1137/1.9781611972733.6>
- ICOLC (International Coalition of Library Consortia) (2024, March 22). ICOLC statement on AI in licensing. ICOLC. <https://icolc.net/statements/icolc-statement-ai-licensing>
- Kim, K. (2024). Korean copyright issues in text data mining for generative AI. *Journal of AI Law and Regulation*, 1(1), 64-76. <https://doi.org/10.21552/aire/2024/1/8>

- Kumar, S., Vrabec, H. U., Xie, Z., Nikpoor, S., Zhao, D., Shinzato, L., & Eisenberg, R. AI Regulations in the context of natural language processing research (2023). SSRN. <http://dx.doi.org/10.2139/ssrn.4526253>
- Lazarova, A. (2024, December 28). Libraries, licenses, limitations: an empirical insight into the contractual conditions regulating text and data mining for research. Digital Republic. <https://digrep.bg/en/copyright/libraries-licenses-limitations/>
- Nobre, T. & Hackett, T. (2025, May 12). Unfair licensing practices: the library experience. COMMUNIA. <https://communia-association.org/publication/unfair-licensing-practices-the-library-experience/>
- Sag, M. (2020). Text and data mining under U.S. copyright law: landscape, flaws, and recommendations. *Journal of the Copyright Society of the U.S.A.*, 67(1), 1-34.
- Samberg, R. G. (2023, October 24). UC Berkeley Library to Copyright Office: protect fair uses in AI training for research and education. Berkeley Library Update. <https://update.lib.berkeley.edu/2023/10/24/uc-berkeley-library-to-copyright-office-protect-fair-uses-in-ai-training-for-research-and-education/>
- Samberg, R. G., Vollmer, T., & Teremi, S. (2024a, March 12). Fair use rights to conduct text and data mining and use artificial intelligence tools are essential for UC research and teaching. Office of Scholarly Communication, University of California. <https://osc.universityofcalifornia.edu/2024/03/fair-use-tdm-ai-restrictive-agreements/>
- Samberg, R. G., Vollmer, T., & Teremi, S. (2024b). Licensing research contest via agreement that authorize uses of artificial intelligence. Authors Alliance. <https://www.authorsalliance.org/2024/01/10/licensing-research-content-via-agreements-that-authorize-uses-of-artificial-intelligence/>
- Sandiumenge, I. (2023). Copyright implications of the use of generative AI. SSRN. <https://doi.org/10.2139/ssrn.4531912>
- Taplin, B. (2024, June 19). Guidance on resisting restrictive AI clauses in licences. National Centre for AI. Jisc. <https://nationalcentreforai.jiscinvolve.org/wp/2024/06/19/guidance-on-resisting-restrictive-ai-clauses-in-licences/>
- Temple, B. & Young, A. (2004). Qualitative research and translation dilemmas. *Qualitative Research*, 4(2), 161-178. <https://doi.org/10.1177/1468794104044430>
- Vrakas, G. (2024). A literature review of “lawful” text and data mining. Open Research Europe,

4(153). <https://doi.org/10.12688/openreseurope.18013.2>

Xalabarder, R. (2025). Regulating AI: differences between the U.S. and the EU. *The Columbia Journal of Law & The Arts*, 48(3), 325-341. <https://doi.org/10.52214/jla.v48i3.13869>

• 국문 참고자료의 영어 표기

(English translation / romanization of references originally written in Korean)

Cha, Sangyook (2021). Legal issues on protection of trainable datasets for artificial intelligence under copyright law: focusing on the copyright exception for text and data mining(TDM). *Journal of Business Administration & Law*, 32(1), 1-55.

Kim, Kyungsuk (2025). Generative AI and copyright issues: cases and analysis of copyright infringement and copyrightability. *Copyright Quarterly*, 38(2), 5-54.

Kim, Yongjoo (2020). On revision of the copyright law for text and data mining. *Law Studies*, 61(2), 283-314. <https://doi.org/10.35275/pnulaw.2020.61.2.011>

Lee, Sangyong (2021). Non-contractual use of data: focusing on limitations to copyright for text and data mining. *Kangwon Law Review*, 65, 1-54.

Ryu, Shiwon (2024). Recommendations on copyright policy making process in the age of artificial intelligence. *Legislation*, 704, 109-146. <https://doi.org/10.23028/moleg.2024.704..004>

