

RESEARCH ARTICLE

Protecting AI Model Weights as Trade Secrets: Statutory Requirements and Infringement Standards under Korea's Unfair Competition Prevention Act

Na-Rae Kim

Associate, DLG Law Corporation, Republic of Korea

Corresponding Author: Na-Rae Kim (nrkim1217@gmail.com)

ABSTRACT

The core assets of the AI industry are shifting toward model weights, which are the final products of model training. Because model weights are difficult to protect under copyright or patent law, their protection in practice relies on trade secret protection under Korea's Unfair Competition Prevention Act. Building on the technical characteristics of model weights, this study examines each statutory requirement for trade secret protection. Model weights constitute technical information; in closed-source deployment, secrecy can be maintained despite the theoretical possibility of extraction; and large-scale model licensing transactions demonstrate their independent economic value. Regarding secrecy management, the most contested requirement in determining trade secret status, this study argues that technical, contractual, and organizational measures must be comprehensively established and that their adequacy should be assessed using objective recognizability as the essential criterion. Drawing on recent U.S. disputes involving model extraction and knowledge distillation, the study further argues that violations of terms of service constitute improper means only insofar as the violated provisions function as secrecy-management measures. It also proposes a four-factor framework for assessing extraction: the legitimacy of access, the mode of conduct, the purpose and result, and awareness of protective measures. Finally, the study argues that transparency obligations and trade secret protection can coexist through controlled disclosures designed according to the recipient, scope, and conditions of disclosure.

KEYWORDS

artificial intelligence, model weights, trade secret, Unfair Competition Prevention Act, knowledge distillation, reverse engineering, secrecy management

Open Access

Received: July 31, 2026**Revised:** August 03, 2026**Accepted:** September 04, 2026**Published:** September 30, 2026

Funding: The author received manuscript fees for this article from Korea Institute of Intellectual Property.

Conflict of interest: No potential conflict of interest relevant to this article was reported.

© 2026 Korea Institute of Intellectual Property



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.

원저

AI 모델 가중치의 영업비밀 보호에 관한 연구: 부정경쟁방지법상 보호 요건과 침해 판단 기준을 중심으로

김나래

변호사, 법무법인 DLG

교신저자: 김나래 (nrkim1217@gmail.com)

차례

1. 서론
2. AI 모델 가중치의 본질
 - 2.1. 가중치의 의미
 - 2.2. 배포 형태에 따른 가중치 분류
 - 2.3. 가중치 추출 및 역공학 기법
 - 2.4. 선행연구의 검토
3. 부정경쟁방지법상 가중치의 영업비밀성
 - 3.1. 영업비밀로서의 보호 필요성
 - 3.2. 비공지성
 - 3.3. 경제적 유용성
 - 3.4. 비밀관리성
 - 3.5. 관련 판례의 시사점
 - 3.6. 소결
4. 해외 분쟁 사례와 규제 논의
 - 4.1. 미국
 - 4.2. EU
5. 한국법에의 시사점과 개정 방향
 - 5.1. 약관 위반과 부정한 수단의 관계
 - 5.2. 적법한 역공학과 부정한 수단의 구별 기준
 - 5.3. 투명성 규제와 영업비밀 보호의 조화
 - 5.4. 실무적 대응
6. 결론

국문초록

인공지능 산업의 핵심 자산은 학습의 결과물인 모델 가중치로 이동하고 있으나, 가중치는 저작권·특허에 의한 보호가 어려워 실무상 부정경쟁방지법상 영업비밀에 의존하게 된다. 본 연구는 가중치의 기술적 본질에서 출발하여 영업비밀의 각 요건을 검토하였다. 가중치는 기술상 정보에 해당하고, 폐쇄형 배포에서는 추출 가능성에도 불구하고 비공지성이 유지되며, 대규모 모델 라이선스 거래가 경제적 유용성을 입증한다. 영업비밀성 판단 시 가장 문제되는 비밀관리성의 경우 그 본질적 요건인 객관적 인식 가능성을 기준으로 기술적·계약적·조직적 조치를 종합적으로 갖추는 것이 필요하다. 나아가 미국의 최신 분쟁례 분석을 토대로, 약관 위반은 그것이 비밀 관리 조치로 기능하는 범위에서만 부정한 수단 판단에 산입되고, 추출 행위는 접근 자격의 적법성, 행위의 태양, 목적과 결과, 관리 조치에 대한 인식이라는 네 가지 징표로 판단하는 틀을 제시하였다. 투명성 규제와의 관계에서는 공개의 상대방·범위·조건을 설계하는 통제된 공개가 비밀관리성과 양립할 수 있음을 논증하였다.

주제어

인공지능, 모델 가중치, 영업비밀, 부정경쟁방지법, 지식 종류, 역공학, 비밀관리성

1. 서론

인공지능 산업에서 기업의 핵심 자산이 소스코드에서 학습의 결과물인 모델 가중치(weights)로 옮겨가고 있다. 막대한 연산 비용과 정제된 데이터를 투입한 결과가 가중치 파일 하나로 모이고, 시장은 이미 이를 독립된 거래 대상으로 다룬다. 2024년 Microsoft는 회사 지분을 인수하지 않은 채 Inflection의 모델 라이선스 대가로 약 6억 2,000만 달러를 지급하였고,¹⁾ Google 역시 Character.AI와 약 27억 달러 규모의 지분 인수 없는 기술 라이선스 계약을 체결한 것으로 보도되었다.²⁾ 그런데 가중치는 사람이 작성한 표현물이 아니라 학습 과정에서 자동으로 생성되는 수치의 집합이어서 저작권법의 창작성 요건이나 특허법의 발명 성립성·명세서 기재요건을 충족하기 어렵고, 공개를 전제로 하는 특허제도는 비공개를 원하는 기업의 이해와 맞지 않는다. 결국 실무에서 가중치의 보호는 부정경쟁방지 및 영업비밀보호에 관한 법률(이하 ‘부정경쟁방지법’)상 영업비밀에 의존하게 된다.³⁾

문제는 영업비밀 법리가 가중치라는 자산을 전제로 만들어지지 않았다는 데 있다. 가중치는 파일을 통째로 입수해도 내용을 해석하기 어려운 반면, 파일에 접근하지 않고 API 질의만으로 기능을 복제할 수도 있다. 실제로 미국에서는 모델 추출과 지식 증류(knowledge distillation)가 영업비밀 침해인지 적법한 역공학인지를 다투는 분쟁이 시작되었고⁴⁾⁵⁾ 인공지능의 안전성 검증을 위한 투명성 요구가 확대되면서 영업비밀 보호와의 간극도 커지고 있다.⁶⁾ 2026년 1월 시행된 인공지능 발전과 신뢰 기반 조성 등에 관한 기본법(이하 ‘인공지능기본법’)으로 인하여 이 문제가 현실화되고 있다.⁷⁾

본 연구는 한국 부정경쟁방지법의 해석론을 중심으로, 미국·EU의 관련 법제와 판례를 분석한다. 이하 제2장에서 가중치의 기술적 본질과 배포·추출 방식을 확인하고, 제3장에서 부정경쟁방지법상 영업비밀의 요건을 가중치에 대입하여 검토하며, 제4장에서 미국의 분쟁 사례와 EU의 투명성 규제 논의를 분석한 뒤, 제5장에서 우리 법의 해석 기준과 실무·입법적 개선방안을 제시하고 제6장에서 결론을 도출한다.

2. AI 모델 가중치의 본질

2.1. 가중치의 의미

인공지능망은 수많은 연산 단위(노드)를 층층이 연결한 구조다. 각 노드는 입력값에 일정한 수치를 곱하고 더해 다음 층으로 넘긴다. 이때 곱해지는 수치들이 가중치(weights)이며, 편향(bias) 등과 함께 매개변수(parameters)를 이룬다. 모델 하나의 매개변수는 수억 개에서 수천

1) Shirin Ghaffary & Rachel Metz, “Microsoft to Pay Inflection AI \$650 Million After Scooping Up Most of Staff”, Bloomberg, <<https://finance.yahoo.com/news/microsoft-pay-inflection-ai-650-210933932.html>>, 작성일: 2024. 3. 22.

2) Kenrick Cai, “Google hires top talent from startup Character.AI, signs licensing deal”, Reuters, <<https://www.reuters.com/technology/artificial-intelligence/google-hires-characterai-cofounders-licenses-its-models-information-reports-2024-08-02/>>, 작성일: 2024. 8. 2.

3) Camilla A. Hrdy, “Keeping ChatGPT a Trade Secret While Selling It Too”, *Berkeley Technology Law Journal*, Vol.40 No.1(2025), pp. 93-94.

4) OpenEvidence, Inc. v. Pathway Medical, Inc., No. 1:25-cv-10471 (D. Mass. filed Feb. 26, 2025).

5) Camilla A. Hrdy, “Trade Secrecy Meets Generative AI”, *Chicago-Kent Law Review*, Vol.100 No.1(2025), p. 318.

6) Perry Keller & Tanya Aplin, “Reconciling Trade Secrets and AI Public Transparency”, *The Cambridge Law Journal*, Vol.85 No.1(2026), pp. 153-154.

7) 인공지능 발전과 신뢰 기반 조성 등에 관한 기본법(법률 제20676호, 2025. 1. 21. 제정, 2026. 1. 22. 시행).

억 개에 이르는 숫자의 집합, 즉 거대한 행렬이다.⁸⁾

주목할 점은 이 숫자들을 사람이 정하지 않는다는 것이다. 개발자가 설계하는 것은 신경망의 구조(아키텍처)와 학습 방법이다. 가중치 값 자체는 학습 과정에서 만들어진다. 모델이 내놓은 답과 정답의 차이를 계산하고 그 차이가 줄어드는 방향으로 수치를 고쳐나가는 과정을 수없이 반복하면, 데이터에 담긴 패턴이 가중치 값으로 새겨진다. 즉 가중치는 사람이 작성한 표현물이 아니라, 사람이 설계한 절차에 따라 데이터로부터 자동으로 도출된 결과물이다. 사람이 직접 작성하지 않았다는 이 특성이 저작권법상 창작성 인정을 어렵게 만드는 출발점이고, 별도의 정책적 보호 필요성을 야기한다.

가중치는 인공지능 개발에 대한 투자가 응축되는 최종 산출물이다. 대규모 모델 하나를 학습시키는 데에는 막대한 연산 비용, 정제된 학습데이터, 그리고 수많은 시행착오가 들어가고, 그 전부가 마지막에는 가중치 파일 하나로 모인다. 같은 아키텍처를 쓰더라도 가중치가 다르면 전혀 다른 성능의 모델이 되므로 기업 간 경쟁력의 차이는 곧 가중치의 차이로 나타날 수밖에 없다.

시장은 이미 가중치를 독립된 거래 대상으로 취급하고 있다. 서론에서 본 것처럼 Microsoft는 2024년 회사 지분을 인수하지 않은 채 Inflection의 모델 라이선스 대가로 약 6억 2,000만 달러를 지급하였고,⁹⁾ Google도 지분 인수 없이 Character.AI의 기술 라이선스에 약 27억 달러 규모로 보도된 대가를 지급하였다.¹⁰⁾ 법인 전체가 아니라 학습된 모델에 값이 매겨진 거래들은, 가중치가 회사와 분리하여 평가·이전될 수 있는 독자적 자산임을 보여준다.

2.2. 배포 형태에 따른 가중치 분류

가중치의 법적 보호를 논하려면 그것이 시장에 어떤 형태로 나오는지에 대한 구분이 선행되어야 한다. 배포 형태에 따라 비밀로 유지되는 대상과 범위가 달라지기 때문이다.

첫째, 폐쇄형(closed-source)이다. 가중치 파일을 개발사의 서버에만 두고, 이용자에게는 API나 웹 서비스를 통한 입력·출력만 허용하는 방식이다. ChatGPT가 대표적이다. 이용자는 모델과 대화할 수 있을 뿐 가중치 파일 자체에는 접근할 수 없고, 보통 이용약관으로 역공학, 역설계 등을 금지하고 있다.¹¹⁾ 이 형태에서 가중치는 서버 안에 은닉되어 있으므로, 쟁점은 ‘서비스를 공개적으로 제공하면서도 가중치의 비공개성이 유지되는가’이다.

둘째, 공개 가중치형(open-weights)이다. Meta의 Llama처럼 가중치 파일 자체를 내려받을 수 있게 공개하는 방식이다.¹²⁾ 주의할 점은 이것이 소스코드 전부를 공개하는 전통적 오픈소스와 다르다는 것이다. 공개되는 것은 학습이 끝난 가중치이고, 그 가중치를 만들어낸 학습데이터의 구성이나 정제 방법, 학습 과정의 세부는 공개되지 않는 경우가 많다. 즉 공개 가중치형에서도 비밀로 남는 정보가 존재한다.

셋째, 두 형태의 중간에 실무적으로 중요한 유형이 있다. 공개된 가중치를 기초로 자사 데이터로 추가 학습하여 만든 가중치다. 국내외 다수 기업이 Llama 등 공개 모델 위에 자사 특화 모델을 구축하는 현실을 고려하면, 공개된 모델을 토대로 만들어진 가중치의 독자적 비밀성이 인정될지, 인정된다면 공개된 모델과의 구분은 어떻게 할지가 역시 쟁점이 된다.

8) Camilla A. Hrdy, 앞의 논문(주 3), p. 107.

9) Krystal Hu & Harshita Mary Varghese, 앞의 글(주 1).

10) Kenrick Cai, 앞의 글(주 2).

11) Camilla A. Hrdy, 앞의 논문(주 3), pp. 79-80.

12) Khari Johnson, “Meta’s Open Source Llama Upsets the AI Horse Race”, WIRED, <<https://www.wired.com/story/metass-open-source-llama-upsets-the-ai-horse-race/>>, 작성일: 2023. 7. 26.

2.3. 가중치 추출 및 역공학 기법

가중치에는 서로 반대 방향의 두 특성이 공존한다. 먼저 파일을 통째로 입수하더라도 그 내용을 해석하기 어려운데, 이는 수십억 개의 숫자 중 어느 것이 어떤 지식이나 기능을 담당하는지 사람이 읽어낼 방법이 없기 때문이다. 흔히 블랙박스라 불리는 이 성질 때문에 가중치는 설계도나 소스코드처럼 ‘읽고 이해하는’ 방식의 이용이 불가능하다. 그런데도 가중치 파일에 접근하지 않은 채 모델의 가치를 가져가는 것은 가능하며, 그 방법이 아래의 추출 기법들이다.

첫째, 모델 추출 공격(model extraction attack)이다. 공개된 API에 질의를 대량으로 보내고 입력-출력 세트 데이터를 수집하면, 그 데이터로 별도의 모델을 학습시켜 원본과 비슷하게 작동하는 모델을 만들 수 있다.¹³⁾

둘째, 지식 증류(knowledge distillation)다. 큰 모델의 출력을 교사 삼아 작은 모델을 학습시키는 기법으로, 원래는 모델 경량화를 위해 개발된 정상적 기술이다.¹⁴⁾ 그런데 이것이 타인의 모델을 상대로 행해지는 경우 이 행위가 약관 위반에 그치는지 영업비밀 침해에 이르는지가 논쟁되고 있다.¹⁵⁾

셋째, 프롬프트 주입(prompt injection)을 통한 내부 정보 추출이다. 정교하게 설계한 입력으로 모델의 제한을 우회하여, 시스템 프롬프트 등 서비스 내부의 비공개 정보를 출력하게 만드는 기법이다.

종합하면 가중치는 열어도 보이지 않고, 열지 않아도 베낄 수 있는 자산이다. 전통적 영업비밀 법리는 비밀 정보를 입수하여 내용을 파악하는 침해를 전제로 설계되었다. 그러나 가중치의 세계에서는 파일을 입수한 자가 아무것도 알아내지 못할 수 있고, 파일에 손대지 않은 자가 가치를 통째로 가져갈 수 있다. 영업비밀의 요건을 어떻게 심사할 것인지, 그리고 위 추출 기법들을 침해로 볼 것인지가 문제되는 이유다.

2.4. 선행연구의 검토

인공지능과 부정경쟁방지법에 관한 국내 연구는 세 갈래로 정리할 수 있다. 첫째, 데이터 보호 연구다. 2021년 제2조 제1호 카목 신설을 전후하여 일본의 한정제공데이터 제도와 비교하며 데이터 부정사용 규정의 해석과 개정 방향을 다룬 연구가 대표적이다.¹⁶⁾¹⁷⁾ 둘째, 알고리즘 공개 요구와 영업비밀 보호의 모순을 다룬 연구다.¹⁸⁾ 셋째, 생성형 인공지능을 쓰는 과정에서 회사의 기존 영업비밀이 새어 나가 비밀관리성을 잃을 위험을 다룬 연구다.¹⁹⁾ 각각 데이터의 보호, 알고리즘의 공개, 기존 비밀의 유출 방지를 다루고 있으나, 인공지능 산업의 핵심 자산인 가중치 그 자체가 영업비밀로 보호될 수 있는지, 그 요건과 침해 판단이 가중치의 기술적 특성과 맞는지를 정면으로 다룬 국내 연구는 확인되지 않는다.

13) Florian Tramèr et al., “Stealing Machine Learning Models via Prediction APIs”, 25th USENIX Conference on Security Symposium, 2016, pp. 601-618.

14) Houman Asefi, “The OpenAI vs DeepSeek Knowledge Distillation Dispute: Technical and Legal Implications”, Medium, <<https://houman-asefi.medium.com/the-openai-vs-deepseek-knowledge-distillation-dispute-technical-and-legal-implications-1e69d646b928>>, 작성일: 2025. 1. 30.

15) Camilla A. Hrdy, 앞의 논문(주 3), p. 86.

16) 김시열, “일본의 한정제공데이터 보호 규정 재논의와 데이터에 관한 우리 부정경쟁방지법의 개정 방안”, 『지식재산연구』, 제18권 제1호(2023), 125-131면.

17) 차상욱, “부정경쟁방지법상 데이터의 보호방안 - 일본 부정경쟁방지법에 대한 비교법적 고찰과 시사점을 중심으로”, 『IP & Data 法』, 제1권 제1호(2021), 3면.

18) 김윤명, “알고리즘 공개와 영업비밀 보호 간의 긴장관계”, 『사법』, 통권 제66호(2023), 867-868면.

19) 이경민, “생성형 AI 활용과 영업비밀 보호”, 『법이론실무연구』, 제12권 제4호(2024), 206면.

미국에서는 Levine이 생성형 인공지능이 영업비밀의 생성·유출·소멸에 미치는 영향을 분석한 이래,²⁰⁾ Hrdy가 모델을 상용 서비스로 판매하면서 동시에 영업비밀로 유지하는 구조의 쟁점과,²¹⁾ 모델 추출·증류에 대한 영업비밀법의 대응을 검토하였고,²²⁾ Keller와 Aplin은 투명성 확보를 위해 영업비밀 보호에 일괄적 공개 예외를 두는 방안의 한계를 지적하며 대안적 구조를 제시하였다.²³⁾ 본 연구는 이러한 논의에 기반하여, 한국의 부정경쟁방지법을 기준으로 가중치의 영업비밀성을 요건별로 검토하고 침해 판단을 위한 기준을 제시한다는 점에서 선행연구와 구별된다.

3. 부정경쟁방지법상 가중치의 영업비밀성

3.1. 영업비밀로서의 보호 필요성

가중치를 보호할 수 있는 법적 수단을 차례로 짚어보면, 영업비밀이 사실상 현실적인 선택지로 남는다.

저작권법은 인간의 사상 또는 감정을 표현한 창작물을 보호한다. 그런데 제2장에서 본 것처럼 가중치는 사람이 작성한 표현물이 아니라 학습 절차가 데이터로부터 자동으로 도출한 수치의 집합이다. 창작성은 물론 인간 저작자의 존재 자체를 인정하기 어렵다.

특허법은 발명의 성립성과 명세서 기재요건에 부딪힌다. 가중치가 왜 그러한 성능을 내는지 설명하기 어려운 블랙박스성 때문에, 통상의 기술자가 재현할 수 있을 정도로 기재하는 것부터가 난제다. 더 근본적인 것으로, 특허는 공개의 대가로 독점을 주는 제도라는 점이다. 가중치를 공개하는 순간 경쟁자는 그것을 그대로 내려받아 쓸 수 있으므로, 공개를 전제로 한 보호는 가중치 보유자가 원하는 보호가 아니다.

부정경쟁방지법 안에서도 영업비밀 아닌 조항들은 가중치와 맞지 않는다. 제2조 제1호 카목의 데이터 부정사용 규정은 “업으로서 특정인 또는 특정 다수에게 제공되는 것으로서, 전자적 방법으로 상당량 축적·관리되고 있으며, 비밀로서 관리되고 있지 아니한” 정보를 대상으로 한다.²⁴⁾ 가중치는 타인에게 제공하기 위해 축적한 데이터가 아니라 스스로 사용하는 자산이고, 무엇보다 비밀로 관리되는 정보는 카목의 정의에서 명시적으로 제외된다. 즉 카목은 비밀로 관리되지 않는 제공용 데이터를 위한 조항이고, 비밀로 지키려는 가중치는 처음부터 그 보호 범위에서 제외된다. 파목의 성과물 도용 규정은 상당한 투자나 노력으로 만들어진 성과의 무단 사용을 규제하므로 문언상 가중치를 포섭할 수 있으나, 보충적 일반조항으로서 그 인정 범위가 좁기 때문에 가중치의 주된 보호 수단이 되기는 어렵다.

결국 가중치 보호의 중심은 제2조 제2호의 영업비밀이다. 법은 영업비밀을 ‘공공연히 알려져 있지 아니하고 독립된 경제적 가치를 가지는 것으로서, 비밀로 관리된 생산방법, 판매방법, 그 밖에 영업활동에 유용한 기술상 또는 경영상의 정보’로 정의한다.²⁵⁾ 판례에 따르면 ‘공공연히 알려져 있지 아니하다’는 것은 그 정보가 간행물 등의 매체에 실리는 등 불특정 다수인에게 알려져 있지 않기 때문에 보유자를 통하지 아니하고는 그 정보를 통상 입수할 수 없는 것을 말하고,²⁶⁾ ‘독립된 경제적 가치를 가진다’는 것은 그 정보의 보유자가 그 정보의 사용을 통해 경쟁자

20) David S. Levine, “Generative Artificial Intelligence and Trade Secrecy”, *Journal of Free Speech Law*, Vol.3(2023), pp. 559-588.

21) Camilla A. Hrdy, 앞의 논문(주 3), pp. 78-87.

22) Camilla A. Hrdy, 앞의 논문(주 5), pp. 347-351.

23) Perry Keller & Tanya Aplin, 앞의 논문(주 6), pp. 179-184.

24) 부정경쟁방지 및 영업비밀보호에 관한 법률 제2조 제1호 카목.

25) 같은 법 제2조 제2호.

에 대하여 경쟁상의 이익을 얻을 수 있거나 그 정보의 취득이나 개발을 위해 상당한 비용이나 노력이 필요하다는 것을 말하며,²⁷⁾ 비밀관리성이란 그 정보가 비밀이라고 인식될 수 있는 표시를 하거나 고지를 하고, 접근할 수 있는 대상자나 접근 방법을 제한하거나 접근한 자에게 비밀 준수 의무를 부과하는 등 객관적으로 그 정보가 비밀로 유지·관리되고 있다는 사실이 인식 가능한 상태인 것을 말한다.²⁸⁾

가중치가 기술상 정보에 해당한다는 점에는 다툼의 여지가 크지 않다. 법문은 정보의 형식을 제한하지 않으며, 기계만 읽을 수 있는 목적코드도 영업비밀의 대상이 되어 온 이상 수치 행렬이라는 형식이 장애가 되지 않는다. 따라서 이하에서는 실질적으로 다투어질 세 요건인 비공지성, 경제적 유용성, 비밀관리성의 해당 가능성을 차례로 검토한다.

3.2. 비공지성

제2.2절의 배포 형태별 분류 중 폐쇄형에서는 비공지성이 유지된다고 보아야 한다. 이용자에게 공개되는 것은 모델의 입력과 출력, 즉 서비스이지 가중치 파일이 아니다. 이는 소프트웨어를 판매하더라도 목적코드만 배포되는 한 소스코드의 비공지성이 유지된다고 보아 온 법리와 같은 구조다.²⁹⁾ 다만 제2.3절에서 본 추출 기법에 의해 API 질의만으로 기능이 근사하게나마 복제될 수 있다면, 가중치는 ‘보유자를 통하지 아니하고도 통상 입수할 수 있는’ 정보가 되어버린 것은 아닌지 문제된다.

그러나 추출로 얻어지는 것은 원본 가중치 값이 아니라 그와 기능이 유사한 별개의 가중치이므로, 원본 정보 자체가 알려진 상태가 되는 것이 아니다. 또한 통상 입수할 수 없는 것인지는 입수에 드는 시간·비용·노력을 고려하여 판단하는데, 상용 규모 모델의 기능을 의미 있게 복제하려면 막대한 질의량과 연산 자원, 상당한 기술이 필요하다. 역설계가 이론적으로 가능하다는 사실만으로 비공지성이 부정되지 않는다는 종래의 법리가 여기에도 적용된다고 봄이 타당하다.³⁰⁾³¹⁾ 물론 추출 비용이 낮아질수록 비공지성 인정 가능성 역시 낮아지므로, 그 판단은 해당 시점의 기술 수준에 따른 개별 판단에 의할 수밖에 없다.³²⁾

반면 공개 가중치형에서는 공개된 가중치 자체의 비공지성은 소멸한다. 그러나 제2장에서 본 것처럼 공개되는 것은 결과물이고, 그 가중치를 만들어낸 학습데이터 구성·정제 방법·학습 세부는 여전히 비밀로 남을 수 있다. 비공지성 판단의 대상을 가중치 파일 그 자체와 그것을 만든 방법으로 나누어 보아야 하는 이유다.

미세조정형은 공개 모델을 토대로 추가 학습한 가중치 안에 공개 부분과 비밀 부분이 섞여 있다. 그러나 공지된 요소들의 조합이라도 조합 자체가 알려져 있지 않고 독자적 가치를 지니면 영업비밀이 될 수 있다는 법리에 비추면, 미세조정으로 달라진 부분이 보유자를 통하지 않고 입수될 수 없는 한 비공지성을 인정할 수 있다.³³⁾

실무적으로 가중치 파일이 제한 없는 다운로드 형태로 외부에 배포되는 순간 비공지성 입증은 급격히 어려워진다. 반대로 API를 통해서만 제공하거나, 고객·협력사에게 파일을 제공하

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

27) 대법원 2008. 2. 15. 선고 2005도6223 판결.

28) 대법원 2008. 9. 11. 선고 2008도5383 판결.

29) 대법원 2008. 7. 24. 선고 2007도11409 판결; 서울중앙지방법원 2015. 1. 14. 선고 2012고단2014, 2013고단7936 판결.

30) 대법원 1999. 3. 12. 선고 98도4704 판결.

31) 양인수·박선하, “영업비밀의 비공지성에 관한 연구”, 『지식재산연구』, 제20권 제1호(2025), 28-29면.

32) 대법원 2022. 6. 30. 선고 2018도4794 판결.

33) 대법원 2024. 4. 12. 선고 2022도16851 판결.

라도 접근 권한을 제한하고 비밀유지의무를 부과하는 구조라면 비공지성이 유지될 가능성이 높다. 이때 계약과 접근통제 등이 함께 갖추어져야 하는데, 판례도 계약상 비밀유지의무가 부과된 상태에서 소프트웨어가 공급된 사정을 비공지성 판단 시 고려하고 있다.³⁴⁾ 배포 구조와 계약 설계가 요건 판단에 직결된다는 점에서, 비공지성은 사후에 다투는 요건이 아니라 사전에만 들어 두는 요건이라 할 수 있다.

3.3. 경제적 유용성

가중치는 그 개발에 막대한 연산 비용과 데이터 투자가 들고, 성능 좋은 가중치를 가진 자가 시장에서 앞선다. 지분 인수 없이 모델 라이선스에만 수익~수십억 달러가 지급된 거래들은 이 가치가 관념이 아니라 시장 가격으로 존재함을 보여준다.

가중치의 경우 경제적 유용성의 주장 구조가 비교적 명확하다. 학습에 투입된 비용과 시간, 정확도·추론 비용·안전성에서의 성능 우위가 그대로 경쟁력으로 연결되기 때문이다. 특히 경쟁자가 가중치를 확보하면 재학습·미세조정·실험을 생략할 수 있다는 점이 중요한데, 판례 역시 경쟁자가 해당 정보를 취득하면 독자 개발에 비해 시행착오를 줄여 시간적·경제적 비용을 절약할 수 있다는 사정을 경제적 유용성의 판단 요소로 고려한다.³⁵⁾ 실무적으로도 영업비밀성의 세 요건 중 경제적 유용성은 상당한 개발 비용의 투입이나 경쟁상 이익이 소멸되면 큰 다툼 없이 인정되는 경우가 많다.

다만 미국의 신경망 연산 가속 알고리즘의 영업비밀 침해가 문제된 사건(Neural Magic 사건)에서, 법원은 원고가 아직 제품 출시 전이고 매출도 없다는 점 등을 들어 경제적 가치의 소명이 부족하다고 보았다.³⁶⁾ 상용화 전 단계라면 경제적 유용성이 자동으로 인정되지 않을 수 있다는 것이다. 따라서 개발 비용 내역, 경쟁사와의 성능 비교, 라이선스 협상 기록 등 가치를 객관화할 자료를 평소에 축적해 둘 필요가 있다.

3.4. 비밀관리성

실무적으로 가장 빈번하게 문제되는 비밀관리성 요건은 두 차례 완화된 바 있다. ‘상당한 노력’에서 2015년 ‘합리적인 노력’으로, 2019년 개정에서는 노력 문구를 삭제하고 ‘비밀로 관리된’으로 바뀌었다.³⁷⁾ 중소기업 등이 관리 수준의 미비로 보호에서 탈락하는 것을 줄이려는 취지다.

그런데 판례가 세운 판단 요소를 가중치에 형식적으로 대입하면 곳곳에서 어긋난다. 수백 기가바이트의 바이너리 파일에 대외비 표시를 한다는 것이 무엇을 의미하는지부터 상정하기 어렵다. 파일명이나 저장소에 표지를 붙일 수는 있으나, 정작 가중치는 학습 클러스터와 추론 서버 사이를 데이터 형태로 오가며 복제·분산된다. 접근 제한의 측면에서도, 가중치는 금고에 넣어두는 정보가 아니라 서비스 운영을 위해 상시 가동되는 정보다. 개발·운영 인력의 접근이 계속 필요하고 클라우드 인프라 위에서 관리되며 API를 통해 불특정 다수의 외부 이용자와 상시 연결되어 있다. 즉 접근 방법의 제한은 물리적 차단이 아니라 계정 권한 관리, 암호화, 접근 기록, 호출량 제한 같은 기술적 통제의 형태를 띌 수밖에 없다.

그렇다면 판례가 요구하는 것의 본질은 특정한 형식이 아니라 객관적 인식 가능성, 즉 이 정

34) 서울중앙지방법원 2016. 1. 15. 선고 2014가합563704 판결.

35) 부산지방법원 2016. 9. 29. 선고 2015노3718 판결.

36) Neural Magic, Inc. v. Facebook, Inc., No. 1:20-cv-10444 (D. Mass. May 29, 2020).

37) 서울동부지방법원 2023. 11. 10. 선고 2022노1404 판결 참조.

보가 비밀로 관리되고 있음을 접하는 자가 알 수 있는 상태로 보아야 할 것이다. 가중치 환경에서 그 상태는 기술적·계약적·조직적 조치가 결합될 때 달성된다. 기술적 조치로는 저장 시 암호화, 접근 권한의 차등 부여와 기록, API 인증과 호출량 제한을 들 수 있다. 계약적 조치로는 임직원·협력사와의 비밀유지약정에 더하여, 불특정 이용자와의 관계에서 이용약관의 역공학·추출·증류 금지 조항이 해당 서비스의 기반이 되는 정보가 비밀로 관리되고 있음을 고지하는 기능을 수행한다. 조직적 조치로는 가중치를 비밀 자산으로 지정·관리하는 내부 규정과 담당 체계를 갖추는 것이 여기에 해당한다.

구체적으로는 가중치 파일 자체를 단위로 하여 (i) 접근 권한의 제한, (ii) 반출 승인 절차, (iii) 비밀 표시 내지 등급 지정, (iv) 비밀유지서약과 퇴직 시 반납·폐기 의무, (v) 다운로드·열람 기록의 보존이 갖추어지면 비밀관리성의 입증은 한층 수월해진다. 판례 중에도 접근 권한 관리, 저장장치 접근 시 인증 절차, 반출 승인 절차, 비밀유지약정 체결 등을 비밀 관리의 정황으로 인정한 예가 있다.³⁸⁾

여기서 불특정 다수에게 서비스를 공개하면서 약관만으로 비밀준수의무를 부과하는 것이 판례가 말하는 비밀준수의무의 부과에 해당하는지도 문제된다. 개별적 교섭 없이 약관만으로 성립하는 의무라는 점에서 전통적인 비밀유지약정과 결이 다르다. 그러나 목적코드만 배포되는 소프트웨어에서 라이선스 계약상 역공학 금지 조항이 비밀 관리 조치의 하나로 평가되어 온 것에 비추어 볼 때, 기술적 접근 통제와 결합된 약관상 금지 조항은 비밀관리성을 구성하는 유효한 조치로 보아야 할 것이다. 다만 약관 위반이 곧바로 영업비밀 침해가 되는지는 별개의 문제로, 아래 제4장과 제5장에서 다룬다.

마지막으로, 인공지능기본법 등에 따라 학습데이터의 개요나 모델에 관한 일정 정보를 공개하게 될 가능성이 있으나, 공개 대상은 법이 정한 범위의 정보이지 가중치 자체가 아니다. 따라서 법령상 요구되는 최소한을 공개하면서 그 밖의 부분에 대한 관리 조치를 유지하는 한, 그러한 의무의 이행이 비밀관리성의 포기로 평가될 수는 없다. 오히려 무엇을 공개하고 무엇을 비밀로 남기는지를 구분하여 관리하는 체계 자체가 비밀관리성을 입증하는 증거가 된다.

3.5. 관련 판례의 시사점

국내 판례는 소스코드·실행파일·설정파일·데이터구조 같은 무형의 소프트웨어 구성요소를 영업비밀로 인정해왔다. 판례는 프로그램 소스파일·실행파일·설계 문서·설정파일 등으로 구성된 파일들이 비공지성·경제적 유용성·비밀관리성을 갖추면 영업비밀성을 인정하고, 특히 타인의 영업비밀을 참조하여 시행착오를 줄이거나 필요한 실험을 생략하는 등 시간·비용을 절약하는 행위도 영업비밀의 사용에 해당할 수 있다고 판시하기도 한다.³⁹⁾ 이러한 판시 취지를 고려하였을 때, 타인의 가중치를 그대로 탑재하는 경우는 물론, 지식 증류와 같이 이를 참조하여 자신의 학습 과정을 단축하는 행위까지 사용으로 평가될 수 있을 것이다.

실무적으로는 소송에서의 영업비밀 특징이 문제될 수 있다. 영업비밀 침해를 주장하는 자는 침해 대상인 비밀을 다른 정보와 구별할 수 있을 정도로 특정해야 하는데,⁴⁰⁾ 가중치는 어느 부분이 비밀인가를 내용적으로 나누어 설명하기 쉽지 않다. 다만 가중치는 특정 시점의 학습 완료본 파일로 존재할 수 있으므로, 파일 단위·버전 단위로 해시값 등을 통해 물리적으로 특정하는 방법이 가능할 것이다.

38) 서울고등법원 2009. 6. 19. 선고 2008노2801 판결.

39) 서울고등법원 2017. 6. 1. 선고 2014나4592 판결.

40) 정대순, “영업비밀 침해소송에서 입증책임 완화를 위한 추정규정의 도입검토: 최근의 중국과 일본 사례를 중심으로”, 『지식재산연구』, 제19권 제4호(2024), 98면.

3.6. 소결

가중치는 기술상 정보에 해당하고, 폐쇄형에서는 추출 가능성에도 불구하고 비공지성이 유지되며, 시장에서의 실제 거래사례를 통해 경제적 유용성 역시 인정될 수 있다. 비밀관리성의 경우, 문서 시대의 판단 요소를 형식적으로 대입하면 가중치는 관리되지 않는 정보처럼 보이지만 객관적 인식 가능성을 기준으로 기술적·계약적·조직적 조치를 종합 평가하면 충족될 가능성이 높다. 문제는 이러한 관리 체계를 무력화하거나 우회하는 추출 행위를 침해로 볼 수 있는 것이며, 장을 바꾸어 미국과 EU의 논의를 살핀다.

4. 해외 분쟁 사례와 규제 논의

4.1. 미국

4.1.1. 미국 연방 영업비밀보호법

미국 연방 영업비밀보호법(Defend Trade Secrets Act, 이하 ‘DTSA’)은 ‘부정한 수단(improper means)’에 절취, 뇌물, 허위표시, 비밀유지의무의 위반 또는 그 유인, 전자적 수단에 의한 스파이 행위를 포함시키면서, 역공학(reverse engineering)과 독자 개발, 그 밖의 적법한 취득 수단은 부정한 수단에 해당하지 않는다고 명시한다.⁴¹⁾ 정당하게 입수한 제품을 분해·분석하여 그 안의 정보를 알아내는 것은 영업비밀 침해가 아니라 허용된 경쟁이라는 것이다.

문제는 이 원칙이 만들어질 때 상정된 역공학과, 제2.3절에서 본 추출 기법 사이의 거리다. 전통적 역공학은 시장에서 구매한 물건을 직접 분해·분석하는 행위였다. 반면 모델 추출과 증류는 보유자의 서버에 계속 접속하여, 보유자가 정한 이용조건 아래에서만 열려 있는 통로(API)로 대량의 질의를 보내는 행위다. 취득의 장소가 침해자의 작업대가 아니라 보유자의 시스템이라는 점, 그리고 그 통로에 이용약관이라는 계약이 걸려 있다는 점에서, 적법하게 입수한 물건의 분석이라는 역공학의 전제가 그대로 들어맞지는 않는다.

4.1.2. Compulife 판결

이 거리를 처음 좁힌 것이 Compulife Software, Inc. v. Newman 판결이다. 피고는 원고의 보험료 산출 시스템이 웹사이트에서 공개적으로 제공하는 건적 정보를 자동화된 봇을 통해 수백만 건 긁어모아 경쟁 서비스를 만들었다. 이에 대하여 법원은 각각의 개별 건적이 공개된 정보라 하더라도, 인간이 통상적으로 열람하는 수준을 넘어 자동화된 수단으로 이를 대량 추출하는 행위는 부정한 수단에 의한 취득이 될 수 있다고 판단하였다.⁴²⁾ 특히 법원은 이 사안을 공중에서 몰래 항공 촬영으로 공장 설계를 알아낸 행위를 부정한 수단으로 본 과거의 판례와도 유사한 취지로 판시하였다.⁴³⁾ 요컨대 API가 공개되어 있다는 사정만으로 추출이 적법해지는 것이 아니라, 질의의 자동화와 규모, 서비스의 통상적 이용 목적에 부합하는지 여부 등이 부정한 수단 판단 시 고려 대상이 될 수 있다.

4.1.3. OpenEvidence 사건

Compulife의 법리가 생성형 AI에서 실제로 정면으로 다투어진 것이 OpenEvidence, Inc.

41) 18 U.S.C. § 1839(6).

42) Compulife Software, Inc. v. Newman, 959 F.3d 1288, 1314 (11th Cir. 2020).

43) E.I. duPont deNemours & Co. v. Christopher, 431 F.2d 1012, 1014 (5th Cir. 1970).

v. Pathway Medical, Inc. 사건이다.⁴⁴⁾ 의료용 생성형 AI를 운영하는 원고는, 경쟁사인 피고가 타인의 의사 면허번호를 도용해 의료인 전용 서비스에 접근한 뒤 수십 차례의 프롬프트 주입 공격으로 서비스의 핵심인 시스템 프롬프트를 추출했다며 DTSA 위반 등을 주장하였다. 이에 대해 피고는 공개된 인터페이스에 질문을 입력해 답을 받은 것은 적법한 역공학일 뿐이며 비공개 정보에 접근한 바 없다고 주장하였다. 당초 Pathway Medical을 상대로 제기된 소송은 Doximity가 Pathway를 인수함에 따라 취하되었고, 같은 주장은 Doximity를 상대로 한 소송에서 계속 다투어지고 있으나 아직 법원의 판단은 나오지 않았다.

이 사건은 질의를 통한 추출이 역공학으로서 적법하다는 항변이 실제 소송에서 방어방법으로 등장하였다는 점에서, 추출 행위를 침해로 구성하려면 이 항변을 넘어설 논리가 필요하다는 것이 확인되었다고 볼 수 있다. 또한, 시스템 프롬프트라는, 가중치는 아니지만 같은 방식으로 은닉되고 추출되는 정보가 영업비밀로 소구되었다는 점에서, 법원이 이 사건에서 세우는 기준은 가중치 추출 사건에도 그대로 적용될 가능성이 높을 것으로 보인다.

4.1.4. 약관 위반과 부정한 수단

미국 판례는 이용약관 위반이 결합되면 취득이 부정한 수단으로 평가될 가능성이 높아진다는 태도를 보여 왔고, 실무에서도 역공학 금지 약관과 영업비밀 주장을 결합하는 방식이 폐쇄형 소프트웨어 분야에서부터 활용되어 왔다.⁴⁵⁾ OpenAI가 DeepSeek에 대해 제기한 증류 의혹은 이 문제가 생성형 AI 영역에서 정면으로 제기된 사례다. OpenAI의 이용약관은 출력물을 이용한 경쟁 모델 개발을 금지하는데, 이를 위반하여 행해진 증류가 단순한 계약 위반에 그치는지, 아니면 비밀유지의무 위반에 준하는 부정한 취득 및 사용이 되는지가 문제되는 것이다.⁴⁶⁾ 약관 위반만으로 역공학의 적법성 원칙을 우회할 수 있게 하면 사실상 계약에 의해 영업비밀법의 적용 범위가 확장되는 결과가 된다는 의견도 있으나⁴⁷⁾, 이 쟁점에 관한 미국법의 태도는 아직 확립되지 않았다.

4.2. EU

미국의 논의가 사인 간 탈취를 둘러싼 것이라면, EU의 논의는 국가가 요구하는 공개와 영업비밀 간 간극에 관한 것이다. EU는 2016년 영업비밀지침으로 회원국의 영업비밀 보호를 정비하는 한편,⁴⁸⁾ AI법(AI Act)을 통해 범용 AI 모델 제공자에게 기술문서 작성·제공 등 투명성 관련 의무를 부과하였다.⁴⁹⁾ 모델의 안전성·적법성을 검증하려는 규제요청과 모델의 핵심 정보를 비밀로 유지하려는 기업의 이익이 하나의 제도 안에서 충돌하는 구조다.

이 충돌의 해법으로 흔히 제안되는 것이 공익 영역에 대한 영업비밀 보호의 일괄 배제이다. 그러나 Keller와 Aplin은 이러한 접근이 무엇이 배제 대상인지에 관한 불확실성, 필요 범위를 넘는 과잉 공개, 그리고 실효성 부족을 낳는다고 비판한다.⁵⁰⁾ 이들이 제시하는 대안은 정보를 일반에 공개하는 대신, 비밀유지의무를 지는 인가된 제3자(감사인, 연구자, 규제기관)가 검증에 필요한 범위에서 접근하도록 하는 구조다. 공개의 상대방과 범위를 통제함으로써 검증과 비

44) OpenEvidence, Inc. v. Pathway Medical, Inc., 앞의 판례(주 4).

45) Camilla A. Hrdy, 앞의 논문(주 3), pp. 80-81.

46) Camilla A. Hrdy, 앞의 논문(주 5), p. 349.

47) Camilla A. Hrdy, 앞의 논문(주 3), p. 134.

48) Directive (EU) 2016/943 (영업비밀지침).

49) Regulation (EU) 2024/1689 (AI법).

50) Perry Keller & Tanya Aplin, 앞의 논문(주 6), pp. 179-184.

밀 유지를 양립시키려는 것이다.

이 논의는 제3.4절에서 본 비밀관리성 문제와 연관된다. 법령상 의무 이행으로서의 정보 제공이 통제된 상대방에 대한 통제된 범위의 공개로 설계되는 한, 그것은 비밀 관리의 포기가 아니라 오히려 관리 체계의 일부가 된다. 투명성과 영업비밀은 양자택일의 관계가 아니라 공개의 상대방과 공개 범위, 조건을 어떻게 설계할 것인지의 문제가 되는 것이다.

5. 한국법에의 시사점과 개정 방향

5.1. 약관 위반과 부정한 수단의 관계

부정경쟁방지법 제2조 제3호 가목은 “절취, 기망, 협박, 그 밖의 부정한 수단으로 영업비밀을 취득하는 행위”를 침해행위로 규정한다. 판례는 부정한 수단을 절취·기망·협박 등 형법상 범죄를 구성하는 행위뿐만 아니라, 비밀유지의무의 위반 또는 그 위반의 유인 등 건전한 거래질서의 유지 내지 공정한 경쟁의 이념에 비추어 이에 준하는 선량한 풍속 기타 사회질서에 반하는 일체의 행위나 수단으로 넓게 인정하여 왔다.⁵¹⁾

이 정의는 미국법과의 중요한 차이를 드러낸다. DTSA는 역공학의 적법성을 명문으로 규정하고 있으므로 약관 위반을 매개로 이를 배제하려는 시도는 연방법과의 충돌 문제를 발생시키지만, 우리 법에는 역공학에 관한 명문 규정이 없고 이 문제는 부정한 수단이라는 일반조항의 해석으로 처리된다. 따라서 우리 법에서는 약관 위반이 부정한 수단 판단 시 고려 대상이 되는 데 구조적 장애가 없다. 검토가 필요한 것은 그 반대 방향, 즉 약관 위반이 있으면 곧 부정한 수단에 해당한다고 볼 수 있는가이다.

판례가 예시하는 비밀유지의무 위반은 정보의 비밀성을 전제로 그 유지 의무를 부담하는 자의 의무 위반을 뜻하나, 불특정 다수와의 계약으로 성립하는 이용약관상 의무가 전부 이 수준의 의무에 해당하는 것은 아니다. 결제 조건이나 이용 시간 제한을 위반한 자를 영업비밀 침해자로 평가할 수는 없기 때문이다. 즉, 그 판단 기준은 위반된 조항의 성격에서 찾아야 한다. 역공학·추출·증류 금지 조항과 같이 서비스의 기반이 되는 정보의 비밀 유지를 직접 목적으로 하는 조항의 위반은 비밀유지의무 위반에 준하는 사정으로 평가할 수 있는 반면, 비밀 보호와 무관한 조항의 위반은 계약상 책임의 문제에 그친다. 결국 약관 위반은 해당 조항이 비밀 관리 조치로 기능하는 범위에서 부정한 수단 판단에 포함된다. 이러한 해석에 따르면 비밀관리성 요건과 부정한 수단 요건이 동일한 조항을 중복하여 평가하는 것이 아니라, 관리 조치의 존재가 이를 위반한 취득행위의 위법성 판단의 근거가 되는 관계에 있게 된다.

5.2. 적법한 역공학과 부정한 수단의 구별 기준

약관이 없거나 그 위반이 불분명한 경우에도 추출 행위 자체를 부정한 수단으로 평가할 수 있는지가 문제된다. 제4장에서 본 Compulife의 법리와 우리 판례의 일반조항 해석을 종합하면, 접근 자격의 적법성, 행위의 태양, 행위의 목적과 결과, 보유자의 관리 조치에 대한 인식을 종합하여 판단 기준을 제시할 수 있다.

접근 자격의 적법성 측면에서, 타인의 자격증명을 도용하거나 신분을 속여 접근 제한을 회피하였다면 그 취득은 기망에 의한 취득으로 평가될 수 있다. OpenEvidence 사건에서 의사 면허 번호의 도용이 사건의 성격을 좌우한 것이 그 예다.

51) 대법원 2011. 7. 14. 선고 2009다12528 판결.

행위의 태양 측면에서는, 인간의 통상적 이용 범위를 넘는 자동화된 대량 질의, 탐지 회피를 위한 계정 분산이나 신원 은폐가 통상적인 서비스 이용이 아니라 체계적인 추출 작업임을 보여주는 사정이 된다.

행위의 목적과 결과 측면에서는, 서비스가 예정한 효용을 얻기 위한 이용과 서비스의 기반이 되는 정보를 획득하여 경쟁 모델을 개발하기 위한 이용이 구별된다. 수집된 질의응답 데이터가 경쟁 모델의 학습에 투입되었다면, 타인의 영업비밀을 참조하여 시행착오를 줄이고 시간과 비용을 절약하는 행위도 영업비밀의 사용에 해당한다고 본 제3.5절의 법리에 포섭된다.⁵²⁾

마지막으로 보유자의 관리 조치에 대한 인식 측면에서, 호출량 제한을 우회하고 금지 조항의 존재를 알면서 추출을 실행하였다면, 객관적으로 인식 가능한 비밀 관리 체계를 의도적으로 침해한 것으로서 그 위반이 인정되기 쉽다.

반대로 경쟁사 서비스에 통상적 규모의 질의를 보내 성능을 비교·평가하는 행위, 공개된 출력물을 학술적으로 분석하는 행위, 자신이 적법하게 내려받은 공개 가중치를 분석하는 행위는 위 징표들에 해당하지 않는 한 허용되는 경쟁 정보의 수집으로 평가될 수 있다. 이 구별 기준의 목적은 추출 행위 일반을 금지하는 데 있는 것이 아니라, 허용되는 경쟁 행위와 부정한 취득 행위 사이의 판단 기준을 예측 가능하게 제시하는 데 있다.

5.3. 투명성 규제와 영업비밀 보호의 조화

인공지능기본법을 비롯한 투명성 규제의 이행은 영업비밀 보호와 충돌하는 것처럼 보이지만, 제3.4절과 제4.2절의 분석을 종합하면 이는 공개의 상대방·범위·조건을 어떻게 설계하는지에 따라 좌우된다.

먼저 법령상 의무에 따른 정보 제공은 비밀관리성의 포기로 평가되지 않으므로, 의무의 범위를 넘지 않는 제공, 특히 비밀유지의무를 부담하는 규제기관이나 인가된 검증 주체에 대한 제공은 통제된 공개로서 비밀 관리 체계와 양립 가능하다. 또한, 규제 이행 과정에서 취득한 정보를 제3자가 유출·사용하는 행위는 전형적인 영업비밀 침해로 규율될 수 있다. 따라서 기업으로서 는 정보 제공 시 해당 정보가 영업비밀임을 지정하고 취급 조건을 명시해 둘 필요가 있다.

입법론 차원에서, 투명성 관련 법령에 제출 정보의 비밀 취급 절차와 담당자의 비밀유지의무를 명문화하는 것이 EU 논의가 시사하는 정비 방향이다. 검증에 필요한 접근은 확대하되 공개의 범위는 제한하는 것이 투명성 확보와 영업비밀 보호를 양립시키는 방법이 된다는 것이다.

5.4. 실무적 대응

실무적으로 가중치 보호는 분쟁 발생 후의 소송 수행 방법이 아니라 사전적 설계의 문제다. 그 설계는 배포 구조의 결정, 관리 조치의 구축과 문서화, 증거의 축적이라는 세 단계로 정리할 수 있다.

배포 구조의 결정 단계에서는, 제3.2절에서 본 것처럼 비공지성이 배포 구조의 설계에 의하여 사전에 확보되는 요건이라는 점이 출발점이 된다. 폐쇄형·공개형·미세조정형 중 어느 형태를 택할지, 파일을 제공한다면 누구에게 어떤 의무를 부과하여 제공할지가 이후 보호 수준의 상한을 결정한다.

관리 조치의 구축과 문서화 단계에서는, 기술적 조치(암호화, 권한 차등, 접근 기록, 호출량

52) 서울고등법원 2017. 6. 1. 선고 2014나4592 판결.

제한), 계약적 조치(임직원·협력사 비밀유지약정, 약관상 역공학·추출·증류 금지 조항), 조직적 조치(비밀 자산 지정, 반출 승인 절차)를 갖추되, 분쟁 시 입증 자료로 활용할 수 있도록 각 조치의 이행을 기록으로 남겨야 한다. 특히 약관의 금지 조항은 제5.1절의 해석론상 부정한 수단 판단의 근거가 되므로, 금지 대상 행위(자동화 질의, 출력물의 경쟁 모델 학습 이용 등)를 구체적으로 특정하여 규정할 필요가 있다.

증거 축적의 단계에서는, Neural Magic 사건의 시사점을 고려하여 개발 비용 내역, 성능 지표, 라이선스 협상 기록으로 경제적 유용성을 객관화해 두고, 비정상적인 대량 질의 패턴을 탐지·보존하는 로그 체계를 갖추어 추출 행위의 태양을 입증할 수 있도록 대비한다. 미세조정형의 경우 공개 원본과의 차분을 산출·보관하여 비밀 부분의 특징에 대비하는 것도 이 단계에 포함된다.

6. 결론

본 연구는 인공지능 산업의 핵심 자산으로 부상한 모델 가중치가 부정경쟁방지법상 영업비밀로 보호될 수 있는지를 그 기술적 본질에서 출발하여 검토하였다. 가중치는 파일을 입수하더라도 내용을 해석할 수 없는 한편, 파일에 접근하지 않고도 질의만으로 기능이 복제될 수 있는 자산이다. 이 두 특성이 전통적 영업비밀 법리와 정합하지 않는 지점마다, 본 연구는 영업비밀성 요건의 판단 및 해석 기준을 제시하고자 하였다. 즉, 비공지성은 배포 구조의 설계 단계에서 확보되는 요건이라는 점, 경제적 유용성은 대규모 라이선스 거래로 실증되거나 상용화 전 단계에서는 이를 객관화할 자료의 축적이 필요하다는 점, 비밀관리성은 객관적 인식 가능성을 기준으로 기술적·계약적·조직적 조치를 종합 평가하여 판단하여야 한다는 것이다. 침해 판단에 관하여는, 약관 위반은 비밀 관리 조치로 기능하는 조항의 위반인 경우에 한하여 부정한 수단 판단에 산입된다는 기준을 제시하였고, 접근 자격의 적법성, 행위의 태양, 행위의 목적과 결과, 관리 조치에 대한 인식이라는 조건이 적법한 역공학과 부정한 추출 행위를 구별하는 판단 기준이 될 수 있음을 논증하였다.

다만 국내에는 가중치의 영업비밀성을 직접 판단한 판례가 아직 없으므로 본 연구가 제시한 기준은 해석론적 제안의 성격을 가지며, 추출에 소요되는 비용과 기술 수준이 변화하면 비공지성 판단의 전제도 달라질 수 있다. 가중치 침해에 대한 손해배상액 산정과 형사적 규율의 범위는 본 연구가 다루지 못한 과제로, 후속 연구를 통한 검토가 필요하다.

참고문헌

학술지(국내 및 동양)

- 김시열, “일본의 한정제공데이터 보호 규정 재논의와 데이터에 관한 우리 부정경쟁방지법의 개정 방안”, 「지식재산연구」, 제18권 제1호(2023).
- 김윤명, “알고리즘 공개와 영업비밀 보호 간의 긴장관계”, 「사법」, 통권 제66호(2023).
- 양인수·박선하, “영업비밀의 비공시성에 관한 연구”, 「지식재산연구」, 제20권 제1호(2025).
- 이경민, “생성형 AI 활용과 영업비밀 보호”, 「법이론실무연구」, 제12권 제4호(2024).
- 정대순, “영업비밀 침해소송에서 입증책임 완화를 위한 추정규정의 도입검토: 최근의 중국과 일본 사례를 중심으로”, 「지식재산연구」, 제19권 제4호(2024).
- 차상욱, “부정경쟁방지법상 데이터의 보호방안 - 일본 부정경쟁방지법에 대한 비교법적 고찰과 시사점을 중심으로”, 「IP & Data 法」, 제1권 제1호(2021).

학술지(서양)

- Camilla A. Hrdy, “Keeping ChatGPT a Trade Secret While Selling It Too”, *Berkeley Technology Law Journal*, Vol.40 No.1(2025).
- Camilla A. Hrdy, “Trade Secrecy Meets Generative AI”, *Chicago-Kent Law Review*, Vol.100 No.1(2025).
- Perry Keller & Tanya Aplin, “Reconciling Trade Secrets and AI Public Transparency”, *The Cambridge Law Journal*, Vol.85 No.1(2026).
- David S. Levine, “Generative Artificial Intelligence and Trade Secrecy”, *Journal of Free Speech Law*, Vol.3(2023).

판례

- 대법원 1999. 3. 12. 선고 98도4704 판결.
- 대법원 2004. 9. 23. 선고 2002다60610 판결.
- 대법원 2008. 2. 15. 선고 2005도6223 판결.
- 대법원 2008. 7. 24. 선고 2007도11409 판결.
- 대법원 2008. 9. 11. 선고 2008도5383 판결.
- 대법원 2011. 7. 14. 선고 2009다12528 판결.
- 대법원 2022. 6. 30. 선고 2018도4794 판결.
- 대법원 2024. 4. 12. 선고 2022도16851 판결.
- 부산지방법원 2016. 9. 29. 선고 2015노3718 판결.
- 서울고등법원 2009. 6. 19. 선고 2008노2801 판결.
- 서울고등법원 2017. 6. 1. 선고 2014나4592 판결.
- 서울동부지방법원 2023. 11. 10. 선고 2022노1404 판결.
- 서울중앙지방법원 2015. 1. 14. 선고 2012고단2014, 2013고단7936 판결.
- 서울중앙지방법원 2016. 1. 15. 선고 2014가합563704 판결.
- Compulife Software, Inc. v. Newman, 959 F.3d 1288 (11th Cir. 2020).
- E.I. duPont deNemours & Co. v. Christopher, 431 F.2d 1012 (5th Cir. 1970).
- Neural Magic, Inc. v. Facebook, Inc., No. 1:20-cv-10444 (D. Mass. May 29, 2020).
- OpenEvidence, Inc. v. Pathway Medical, Inc., No. 1:25-cv-10471 (D. Mass. filed Feb. 26, 2025).

인터넷 자료

- Houman Asefi, “The OpenAI vs DeepSeek Knowledge Distillation Dispute: Technical and Legal Implications”, Medium, <<https://houman-asefi.medium.com/the-openai-vs-deepseek-knowledge-distillation-dispute-technical-and-legal-implications-1e69d646b928>>, 작성일: 2025. 1. 30.
- Kenrick Cai, “Google hires top talent from startup Character.AI, signs licensing deal”, Reuters, <<https://www.reuters.com/technology/artificial-intelligence/google-hires-characterai-cofounders-licenses-its-models-information-reports-2024-08-02/>>, 작성일: 2024. 8. 2.
- Khari Johnson, “Meta’s Open Source Llama Upsets the AI Horse Race”, WIRED, <<https://www.wired.com/story/metas-open-source-llama-upsets-the-ai-horse-race/>>, 작성일: 2023. 7. 26.
- Shirin Ghaffary & Rachel Metz, “Microsoft to Pay Inflection AI \$650 Million After Scooping Up Most of Staff”, Bloomberg, <<https://finance.yahoo.com/news/microsoft-pay-inflection-ai-650-210933932.html>>, 작성일: 2024. 3. 22.

기타 자료

- Florian Tramèr et al., “Stealing Machine Learning Models via Prediction APIs”, 25th USENIX Conference on Security Symposium, 2016.