

AI를 전제하지 않은 계약, AI 시대에 어떻게 읽히는가

생성형 인공지능(Generative AI)이 콘텐츠 생태계를 재편하면서 AI 학습을 염두에 두지 않고 체결된 기존 계약과 약관이 해석 갈등의 한가운데 놓였습니다. 권리자는 자신의 저작물이 무단 학습되었다고 주장하고 사업자는 포괄적 권원 조항을 들어 면책을 시도합니다. 이 구도는 이미 여러 법정에서 되풀이되고 있습니다. 이번 TL에서는 학습 이용 사실의 입증, 침해 대상의 특정, 손해액 산정, 기존 약관 포괄 조항의 적용이라는 네 갈래 쟁점을 검토하고 권리자와 AI 사업자가 각각 점검할 사항을 체크리스트로 정리합니다.

1. 학습에 이용된 사실의 입증 문제

저작권 침해 소송에서 권리자가 가장 먼저 부딪히는 장벽은 자신의 저작물이 실제로 AI 모델의 학습 데이터로 쓰였다는 사실의 입증입니다. 학습 데이터의 구성과 선정은 사업자 내부에서 이루어지므로 권리자가 직접 접근할 수 없고 산출물에 출처가 명시되는 경우도 드뭅니다. 사업자가 보유한 자료에 의존하지 않고서는 주장을 구체화하기 어렵습니다. 저작물이 어떤 경로로 수집되었는지 알 수 없는 경우뿐만 아니라, 계약에 따라 권리자가 스스로 콘텐츠를 제공한 경우에도 실제로 학습에 투입된 범위를 외부에서 확인할 길이 없습니다. 결국 주장을 증거로 뒷받침할 수 있기 위해서는 증거개시나 자료제출 절차를 통해 사업자 내부 자료에 대한 접근이 확보되어야 합니다. 최근 해외 사건에서 그 접근이 좌절된 경우와 재판 절차로 확보된 경우가 어떻게 갈렸는지 확인할 수 있습니다.

영국 Getty Images v. Stability AI 사건에서 피고는 자사 모델의 학습에 원고 웹사이트의 이미지가 사용되었다는 점을 인정했고 다툼은 그 이미지의 수에 관한 것이었습니다. 그러나 원고의 청구는 학습이 영국 내에서 이루어졌음을 입증해야 하는 지점에서 좌절되었습니다. 법원은 원고가 해당 정보에 접근할 수 없는 이상 증거개시 전에는 추론에 기초한 주장 외의 것을 기대할 수 없다고 보아 피고의 약식판결 신청을 기각했습니다¹. 그럼에도 불구하고

Contact

구태연 변호사

T. 02-3477-8695
E. tekoo@law-lin.com

안서연 변호사

T. 02-3477-8695
E. syahn@law-lin.com

이상강 변호사

T. 02-3477-8695
E. [sang kang.lee@law-lin.com](mailto:sangkang.lee@law-lin.com)

¹ Getty Images (US) Inc & Ors v Stability AI Ltd [2023] EWHC 3090 (Ch) (2023. 12. 1.)

원고는 끝내 증거를 확보하지 못해 최종 변론에 앞서 학습·산출물·데이터베이스권에 관한 청구를 모두 포기했습니다². 정보 접근의 비대칭을 법원이 인정해 추론에 의한 주장을 허용했으나, 이를 뒷받침하지 못해 청구가 유지되지 않은 사례입니다.

조달 경로가 다른 유형도 있습니다. 사업자가 제휴 계약으로 콘텐츠를 확보한 경우인데, 여기서는 확보의 적법성보다 그 중 학습에 이용된 범위가 쟁점이 됩니다. 권리자가 콘텐츠를 직접 제공했더라도 그 사용 범위를 스스로 확인하기는 어렵고, 침해 사실의 입증책임을 지는 권리자로서는 정보 비대칭을 어떤 절차로 메울지 가 관건입니다.

국내에서 같은 입증을 하려면 다음 세 가지 경로를 검토해볼 수 있습니다. 첫째, 민사소송법 제344조에 따른 문서제출명령입니다. 학습 데이터의 목록과 구성에 관한 자료라면 실제 신청 근거는 제2항의 일반적 제출의무입니다. 대법원은 제2항 각 호의 제출거부사유에 해당하지 않으면 소지인이 원칙적으로 제출의무를 진다고 판시했습니다³. 그래서 해당 자료가 제2항제1호가 가리키는 제1항제3호 다목의 문서, 즉 제315조제1항제2호의 기술 또는 직업의 비밀이 적힌 문서에 해당하는지, 아니면 제2항제2호의 자기이용문서에 해당하는지가 실무상 쟁점이 됩니다. 데이터의 선별과 전처리 기준은 모델 성능과 직결되므로 사업자가 비밀이라고 주장할 여지가 큼니다. 다만 같은 사건에서 대법원은 영업상 비밀을 담고 있다는 이유로 문서제출신청을 기각한 원심결정을 파기했고, 이는 비밀에 해당한다는 사정만으로 제출의무가 면제되지는 않는다는 것을 시사합니다.

둘째, 저작권법 제129조의2에 따른 정보제공명령입니다. 정당한 이유 없이 명령에 따르지 않으면 법원은 정보에 관한 당사자의 주장을 진실한 것으로 인정할 수 있습니다. 그러나 영업비밀 보호가 거부사유로 명시되어 있고 제공 대상도 침해 행위나 불법복제물의 생산·유통에 관련된 자를 특정할 수 있는 정보와 불법복제물의 생산·유통 경로에 관한 정보로 한정됩니다. 학습 데이터의 목록보다는 데이터 조달 경로가 문제되는 사안에서 사용할 수 있는 수단입니다.

셋째, 감정입니다. AI 모델의 출력 분석, 학습 데이터 추론, 프롬프트 엔지니어링을 이용한 재현 실험 같은 기술적 감정이 입증을 보완합니다. 다만 감정인이 모델 내부에 직접 접근하지 못하는 경우가 많아, 감정의 실효성은 사업자의 협조에 의해 크게 좌우됩니다.

² Getty Images (US) Inc & Ors v Stability AI Ltd [2025] EWHC 2863 (Ch) (2025. 11. 4.)

³ 대법원 2008. 4. 14.자 2007마725 결정

2. 침해 대상 특정 문제

대량의 저작물이 문제되는 AI 학습 사건에서는 침해 대상을 어느 수준까지 특정해야 하는지가 핵심 쟁점입니다. 개별 특정을 엄격히 요구하면 권리자의 권리 행사가 사실상 막히고 포괄적 특정을 허용하면 피고가 개별 저작물 단위로 다룰 수 없어 방어권이 좁아집니다.

특정을 요구하는 논거는 명확합니다. 저작권 침해를 청구원인으로 삼는 이상 어떤 저작물의 어떤 권리가 침해되었는지는 개별 저작물 단위로 판단할 수밖에 없고 대상이 확정되어야 피고가 방어권을 행사할 수 있습니다. 학습에 투입된 자료에는 저작물성이 인정되지 않는 것이 섞여 있을 수 있고 인용된 부분이 창작적 표현인지도 개별적으로 따져야 합니다. 판결의 집행 가능성을 확보하려면 침해금지 대상이 특정되어야 한다는 점도 같은 맥락입니다.

특정의 완화를 요구하는 쪽에도 근거는 있습니다. 학습은 개별 저작물 단위로 이루어지지 않습니다. 전처리를 거친 데이터의 집합이 대상이므로, 개별 저작물 단위의 접근은 실제 이용 형태와 어긋나는 지적이 가능합니다. 무엇보다 어느 저작물이 학습에 투입되었는지에 관한 정보가 사업자에게 물려 있어, 권리자에게 개별 특정을 그대로 요구하면 권리 행사 자체가 막힐 수 있습니다.

실무에서는 양측에 대한 요구를 조정하는 방식으로 이 문제를 다루게 될 것입니다. 권리자에게는 청구 대상의 특정을 요구하되, 사업자에게는 보유한 자료의 범위에서 학습에 이용한 저작물을 밝히도록 해 입증 부담을 나누는 접근입니다. 다만 대량 저작물 사건에서 개별 특정을 어디까지 완화할 수 있는지, 저작물 유형별·창작적 표현 단위별 특정으로 충분한지는 판례가 쌓이면서 정립될 문제입니다.

3. 손해액 산정 문제

국내에는 생성형 AI 학습과 관련한 손해액 산정 방법을 판단한 선례가 없습니다. 학습 행위 자체의 대가를 어떻게 평가할지, 대량의 저작물이 문제되는 사건에서 저작물 단위 산정이 실제로 작동하는지가 앞으로 다투어질 쟁점입니다.

저작권법의 산정방식은 각기 다른 제약에 부딪힙니다. 제125조제1항은 침해로 얻은 이익의 액을 손해액으로 추정하지만 개별 저작물이 모델의 수익에 기여한 정도를 떼어내기 어렵습니다. 제2항은 그 권리의 행사로 일반적으로 받을 수 있는 금액에 상응하는 액을 청구할 수 있도록 합니다. 판례는 구 법의 통상 받을 수 있는 금액을 침해자가 이용허락을 받았더라면 대가로 지급하였을 객관적으로 상당한 금액이라고 해석하는 한편, 침해행위와 유사한 형태의 이용계약이 있으면 그 이용료를, 그런 계약을 맺거나 이용료를 받은 사례가 전혀 없으면 업계에서 일반화되어 있는 이용료를 기준으로 삼고 그것으로도 산정하기 어려우면 제126조에 따라 상당한 손해액을 인정할 수 있다고 보았습니다⁴. 그러나 AI 학습을 기존 계약과 유사한 형태의 이용으로 볼 수 있는지가 분명하지 않고 학습 라이선스 시장도 형성 중이어서 앞의 두 기준을 적용하기 어려운 경우가 많습니다. 제125조의2의 법정손해배상은 침해행위가 일어나기 전에 저작물이 등록되어 있을 것을 요구하므로, 사전 등록이 일반적이지 않은 유형의 저작물에는 적용하기 어렵습니다.

2026년 2월 10일 개정으로 신설되어 같은 해 8월 11일 시행된 제125조 제4항과 제5항도 함께 보아야 합니다. 법원은 침해행위가 고의적인 것으로 인정되면 제1항부터 제3항까지 및 제126조에 따라 손해로 인정된 금액의 5배를 넘지 않는 범위에서 배상액을 정할 수 있고 그 판단에서 고의의 정도, 권리자가 입은 피해 규모, 침해자가 얻은 경제적 이익, 침해행위의 기간과 횟수 등을 고려합니다. 개정 규정은 시행 이후 침해행위가 발생한 경우부터 적용됩니다⁵. 권리자가 학습 거부 의사를 밝혔음에도 불구하고 학습이 이어진 사안이라면 고의성 다툼이 배상액의 규모를 좌우하게 되므로, 사업자에게는 학습 제외 요청의 접수와 반영 이력을 남기는 것이 곧 손해액 산정 단계에서의 방어수단을 확보해 두는 길입니다.

특히 제2항과 관련해서는 기존 콘텐츠 이용계약의 대가를 AI 학습에 대한 대가의 준거로 삼을 수 있는지가 쟁점입니다. 대법원은 뉴스통신사가 언론기관과 맺은 전재계약의 전재료를 성격이 다른 상대방의 침해행위에 대해서까지 통상 받을 수 있는 금액으로 볼 수는 없다고 판단한 바 있습니다⁶. 이용의 목적·범위·기간이 다른 별개의 거래로 보아야 한다는 반론이 여기서 힘을 얻고 이 판단은 계약 해석 문제와도 맞물립니다. 학습 단계와 산출물 이용 단계의 손해를 나누어 산정한다면 전자는 데이터 이용의 대가를, 후자는 시장 대체 효과를 기초로 하는 이원적 구조를 상정할 수 있으나, 중복 산정을 막을 기준이 함께 필요합니다.

⁴ 대법원 2001. 11. 30. 선고 99 다 69631 판결; 대법원 2009. 5. 28. 선고 2007 다 354 판결; 대법원 2010. 3. 11. 선고 2009 다 80637 판결. 모두 구 저작권법 제 93 조제 2 항의 “통상 받을 수 있는 금액”에 관한 판시로, 현행 제125 조제 2 항은 “일반적으로 받을 수 있는 금액”으로 문언이 바뀌었습니다(2021. 5. 18. 개정).

⁵ 저작권법(법률 제21336호, 2026. 2. 10. 일부개정, 2026. 8. 11. 시행) 제125조 제4항제5항 및 부칙 제2조

⁶ 대법원 2009. 5. 28. 선고 2007다354 판결

해외 사례에서도 명확한 산정 기준을 찾기는 어렵습니다. Bartz v. Anthropic 사건에서 15억 달러 규모로 이루어진 합의는 학습 행위에 대한 대가가 아니라 불법 복제물의 취득과 보관에 관한 청구를 대상으로 한 것입니다⁷. 담당 법원이 적법하게 취득한 복제물로 한 학습은 공정이용에 해당한다고 앞서 판단했기 때문인데, 사건이 합의로 끝나 그 판단은 선례로 남지 않았습니다. GEMA v. OpenAI 사건에서 독일 뮌헨지방법원은 금지청구를 인용하고 침해 이용의 범위에 관한 정보제공을 명하면서 손해배상 의무의 존재를 확인했으나, 구체적인 금액은 정하지 않았습니다⁸. 확보된 자료를 토대로 금액을 다투는 절차가 남아 있는 구조입니다. 두 사례는 손해의 규모를 확정하기 전에 침해행위의 범위에 관한 자료를 확보하는 절차가 앞서야 함을 보여줍니다.

4. 기존 약관 포괄 조항의 적용 문제

콘텐츠 제휴 계약과 이용약관에는 사업자가 “서비스 개선 및 신규 서비스 개발을 위한 연구”에 정보를 이용할 수 있다는 취지의 포괄 조항이 들어 있는 예가 있습니다. 이런 조항이 생성형 AI 학습을 예정한 것인지는 그 자체로 분명하지 않고 이용 목적·범위·기간이 특정되지 않은 경우도 많습니다.

계약 해석의 관점에서 핵심 쟁점은 이 조항을 생성형 AI 학습까지 포함하는 것으로 읽을 수 있는지입니다. 목적이 “서비스 개선 및 신규 서비스 개발을 위한 연구”로만 적혀 있다면, 저작물을 AI 모델의 학습 데이터로 활용하는 행위와 그 산출물의 상업적 이용까지 여기에 넣어 확장 해석할 수 있는지가 문제됩니다. 문언만으로 판단하기 어려운 경우가 많으므로, 계약 체결 당시의 교섭 경과나 사업자가 이후 약관을 개정해 AI 학습에 관한 별도의 동의 절차를 신설했는지가 해석의 근거가 됩니다.

이 문제는 계약 해석을 넘어 공정이용 법리 다툼으로도 이어집니다. 문화체육관광부가 2026년 2월 26일 발간한 「생성형 인공지능의 저작물 학습에 대한 저작권법상 “공정이용” 안내서」는 저작물 수집 과정에서 이용약관을 어기거나 허락되지 않은 방식으로 접근하면 저작권 침해가 될 수 있고 접근 자체는 가능하나 AI 학습이 허용되지 않은 저작물을 학습한 경우에는 허락된 이용 범위를 넘어선 것으로 보아 공정이용 제1요소 판단에 불리하게 작용할 수 있다고 설명합니다⁹. 대법원이 공정이용 여부를

⁷ Bartz v. Anthropic PBC, No. 3:24-cv-05417-AMO (N.D. Cal.). 적법하게 취득한 복제물에 의한 학습이 공정이용에 해당한다는 판단은 2025. 6. 23.자 order이고, 15억 달러 합의에 대한 최종 승인 및 종국판결은 2026. 7. 20.자입니다.

⁸ LG München I, 42 O 14139/24 (2025. 11. 11. 판결). OpenAI가 뮌헨고등법원(OLG München)에 항소해 확정되지 않았습니다.

⁹ 문화체육관광부, 「생성형 인공지능의 저작물 학습에 대한 저작권법상 “공정이용” 안내서」(2026. 2. 26)

판단할 때 저작권법 제35조의5 제2항 각 호를 예시로 보고 이용의 경위나 방법처럼 열거되지 않은 사정도 고려할 수 있다고 판시한 것과 같은 방향입니다¹⁰. 다만 안내서는 스스로 유권해석이 아니라고 밝히고 있어 최종 판단은 법원에 달려 있습니다.

권리자로서는 기존 계약의 포괄 조항이 AI 학습을 포함하는 것으로 해석될 수 있는지를 문언과 교섭 경과를 중심으로 점검해야 합니다. 사업자로서는 확보한 권원이 학습 단계와 그 결과물의 상업적 이용에까지 미치는지 확인하고 제3자로부터 데이터를 구매하거나 제공받는 경우에는 AI 학습용이라는 범위가 명확한지와 데이터셋에 대한 권리 보유 여부를 함께 확인해야 합니다.

5. 콘텐츠 권리자와 AI 사업자 양측의 체크리스트

앞의 네 쟁점을 전제로 권리자와 AI 사업자가 각각 점검할 사항을 정리하면 다음과 같습니다.

구분	권리자	AI 사업자
기존 계약	포괄 조항이 AI 학습까지 포함하는 것으로 읽힐 수 있는지 문언·교섭 경과 중심으로 점검	확보한 권원이 학습과 그 결과물의 상업적 이용에까지 미치는지 적용범위 확인
신규 계약	AI 학습 목적을 별도로 분리하고 대가와 이용 범위를 명시	이용 목적·범위·기간을 특정하여 사후 분쟁 소지 축소
취득 경로	저작물의 유통 경로와 접근 제한 조치를 정비하고 그 이력을 관리	데이터의 취득 경로별로 적법성을 확인하고 근거 자료를 보존
입증 자료	보유 저작물의 목록·창작 시점·권리 귀속을 개별 단위로 조회할 수 있도록 관리하고 법정손해배상을 염두에 둔 사전 등록 여부 점검	학습에 사용한 저작물의 출처와 범위를 기록·보존하고 영업비밀 보호와 자료 제출을 병행할 수 있는 관리 체계 마련
대가 산정	이용계약의 대가 산정 근거를 계약서에 남기고 조건을 일관되게 운용	이용허락을 받은 범위와 대가 지급 내역을 저작물 단위로 정리하고 학습 제외 요청의 접수·반영 이력을 보존해 고의 침해 주장에 대비
제3자 관계	제휴처의 재이용·재제공 권한 범위를 계약에 명확히 규정	데이터 공급사로부터 적법성 보증과 면책을 확보하고 AI 학습용이라는 범위가 명확한지 확인

¹⁰ 대법원 2024. 7. 11. 선고 2021다272001 판결. 구 저작권법 제35조의3(현행 제35조의5)에 관한 판시입니다.

6. 결론: 계약이 읽히는 방식의 변화

입증·특정·산정·해석의 네 쟁점은 서로 연결되어 있습니다. 학습 사실의 입증이 침해 대상의 특징으로 이어지고 특정된 대상이 손해액 산정의 단위가 되며 기존 약관이 보장하는 권원의 범위는 그 산정의 준거이자 공정이용 판단의 요소로 다시 작용합니다. 이 연쇄 위에서 권리자와 사업자가 명시적인 합의와 투명한 기록으로 권리의무를 정립하는 것이, AI를 전제하지 않은 계약을 AI 시대에도 예측 가능하게 작동시키는 가장 확실한 경로입니다.

APT 그룹의 관점 — ‘선사용 후분배’ 구조의 설계

본 그룹은 이 문제의 해법이 개별 침해 판단을 정교화하는 데 있지 않고 거래 구조를 설계하는 데 있다고 봅니다. 앞의 네 쟁점이 반복적으로 부딪히는 지점은 하나입니다. 현행 구제수단은 개별 저작물 단위로 설계되어 있는 반면 학습은 데이터의 집합을 기반으로 이루어진다는 것입니다. 대량의 저작물이 문제되는 현실에서 학습 단계의 이용을 일일이 사전 허락에 걸어 두는 방식이 작동하기 어려운 이유도 여기에 있습니다.

이용을 허용하되 산출된 수익을 사전에 합의된 규칙에 따라 권리자에게 배분하는 ‘선사용 후분배’ 구조는 입증·특정·산정이라는 세 병목을 소송의 문제에서 계약과 정산 설계의 문제로 옮겨 놓습니다. 어느 저작물이 학습에 투입되었는지를 사후에 다투는 대신 이용 내역을 기록으로 남기고 저작물 단위의 손해액을 사후에 추정하는 대신 배분 요율을 미리 정하는 것이 필요합니다.

여기서의 ‘분배’는 침해가 확인된 뒤에 지급하는 사후 보상과 다릅니다. 배분 규칙이 이용에 앞서 합의되어야 하고 그 전제로 학습데이터 계보의 기록과 배분 요율의 투명한 공개가 확보되어야 합니다.

법무법인(유한) 린의 APT 그룹은 개별 계약의 사후 검토에 그치지 않고 이러한 배분 체계를 설계하는 단계에서부터 자문을 제공해드리고 있습니다.

법무법인(유) 린 APT(AI, Platform, Technology) 그룹 미디어플랫폼팀은 국내 기업들의 규제 이슈에 전략적인 원스톱(One-Stop) 토탈 솔루션(Total Solution)을 제공하고 있습니다.

상기 내용에 대해 문의사항이 있으시면 언제든지

법무법인(유) 린 APT(AI, Platform, Technology) 그룹 미디어플랫폼팀(Tel. 02-3477-8695)에 문의해 주시기 바랍니다.