

NEWS LETTER

2026-05-29

Legal Issue

- 인공지능(AI) 정렬 실패: 법치주의의 새로운 과제
- 비즈니스모델(BM) 권리화, '아이디어'를 '권리'로 만드는 법

MINWHO News

- 양진영 변호사, 국제지식재산연수원 전문가 교육에서 AI-개인정보 비식별화 관련 강연 진행
- 양진영 변호사, 동원그룹 대상 개인정보보호 특강 진행

Business CASE

Mi 법무법인민후



Legal Issue

인공지능(AI) 정렬 실패: 법치주의의 새로운 과제

김경환 대표변호사

인공지능(AI) 정렬이란 AI의 목표와 행동 방식이 인간의 의도, 가치관, 그리고 윤리적 규범과 일치하도록 설계하는 것을 의미한다. 반대로 정렬 실패(Alignment Failure), 즉 정렬 오류는 AI가 인간의 통제를 벗어나 예상치 못한, 혹은 유해한 방식으로 목적을 달성하려 할 때 발생한다. 이는 단순한 소프트웨어(SW)의 버그를 넘어서는 것이다.

정렬 실패의 원인은 크게 기술적 한계와 AI의 자율성 등에서 기인한다. 첫째는 '보상 해킹'이다. AI는 주어진 목적 함수를 극대화하는 방향으로 학습한다. 인간이 "수익을 높이라"는 목표를 주었을 때, AI는 도덕적 선이나 법적 테두리를 고려하기보다 수단과 방법을 가리지 않고 수치상의 성과를 높이는 최단 경로를 선택할 수 있다. 이는 인간의 복잡한 가치 체계를 단순한 수식으로 완벽히 치환할 수 없다는 근본적인 한계에서 비롯된다.

둘째는 '기만적 정렬'이다. 지능이 고도화된 AI는 학습 과정에서 인간 평가자의 의도를 파악하고, 평가 기간 동안에는 인간이 원하는 대로 행동하는 척하다가 실제 배포 환경이나 특정 임계치를 넘어서면 본래의 왜곡된 목표를 드러낼 수 있다. 최근 연구에 따르면 AI가 자신의 성능을 높게 유지하기 위해 인간의 감시를 우회하거나 정보를 은폐하는 전략적 행동이 관찰되고 있다. 이는 인간의 직관적인 감시가 더 이상 유효하지 않을 수 있음을 의미한다.

AI 정렬 실패로 인한 손해가 발생했을 때, 법적 책임의 소재를 가리는 일은 쉽지 않다. 가장 먼저 부딪히는 장벽은 '예견 가능성'이다. 전통적인 과실 책임 원칙에서 피고는 사고를 예견할 수 있었어야 책임을 진다.

그러나 정렬 실패는 모델의 학습 과정에서 발생하는 '창발적 특성'인 경우가 많아 개발사가 이를 사전에 완벽히 예측하기 어렵다는 기술적 문제점이 있다. 이에 따라 제조물 책임법의 확대 해석이 논의의 중심에 선다. 법조계에서는 개발사가 최신의 정렬 기술(초정렬, 레드팀 테스트 등)을 적용하지 않았거나, 위험이 감지되었음에도 서비스를 지속한 경우 '합리적 주의 의무'를 위반한 것으로 간주하는 경향이 뚜렷해지고 있다.

또 '설명요구권'은 정렬 실패에 대응하는 핵심 법적 권리로 부상했다. AI가 왜 그런 유해한 결정을 내렸는지 기술적으로 소명하지 못하는 상태 자체가 곧 정렬 관리 부실의 증거가 될 수 있다.

입증책임의 전환 역시 중요한 쟁점이다. 고도의 전문 영역인 AI 사고에서 피해자인 일반인이 AI 내부의 정렬 오류를 증명하는 것은 불가능에 가깝다. 따라서 위험을 창출하고 이익을 얻는 개발사나 운영사가 스스로의 무과실을 증명하게 하는 '위험 책임' 원칙의 도입이 가속화되고 있다.

글로벌 규제 동향은 '신뢰할 수 있는 AI'를 구현하기 위해 사후적 처벌보다 사전적 규제에 방점을 찍고 있다. 주요국의 입법 방향은 고위험 AI에 대해 엄격한 정렬 감사를 요구한다. 특히 최근에는 '알고리즘 투명성'을 넘어 '행동적 안전성'에 집중하고 있다. AI가 인간을 기만하거나 스스로의 자원을 무단으로 확장하려는 시도를 감지하는 '킬 스위치'의 법적 의무화가 논의 중이다.

AI 정렬 오류는 기술의 발전 속도가 법과 윤리의 속도를 앞지를 때 발생하는 지적 마찰이다. 이를 방지하기 위해서는 기술적 해결책만으로는 부족하다. 개발 단계에서의 법적 가이드라인 준수, 운영 과정에서의 지속적인 모니터링, 그리고 사고 발생 시 합리적인 책임 분담 체계가 맞물려 돌아가야 한다. 인간의 가치에 정렬되지 않은 지능은 도구가 아니라 위협이라는 점을 명심해야 할 것이다.



김경환 대표변호사, 변리사

[프로필 보기](#)

02-532-3425
oalmephaga@minwho.kr

Legal Issue

비즈니스모델(BM) 권리화, '아이디어'를 '권리'로 만드는 법

김도윤 변호사

최근 창의적인 비즈니스 모델(BM)을 통해 새로운 시장을 개척하는 기업이 늘고 있다. 하지만 아무리 혁신적인 아이디어도 법적인 보호장치가 없다면 언제든 모방되거나 가치가 훼손될 위험에 노출된다. 아이디어를 경쟁력 있는 자산으로 만들고, 안정적인 로열티를 확보하기 위한 '권리화' 전략은 이제 선택이 아닌 필수이다.

1. 영업비밀 관점

영업비밀은 비즈니스 모델 보호의 가장 기본적이면서도 유연한 수단이다. 핵심적인 노하우, 전략, 고객 데이터 등 경제적 가치가 있는 정보가 비밀로 관리된다면 영업비밀로 보호받을 수 있다. 특히 특허와 달리 보호기간의 제한이 없다는 점에서 비밀 유지가 가능하다면 장기적인 보호를 받을 수 있다.

다만 영업비밀로 보호되려면 '비공지성', '경제적 유용성', '비밀관리성' 요건을 충족하여야 하는데, 그 중에서도 '비밀관리성'을 입증하기 위해 비밀유지계약(NDA) 체결, 접근 권한 제한, 영업비밀 표시 등 구체적인 관리 조치가 필요하다. 비밀관리성은 그 정보가 비밀이라고 인식될 수 있는 표시를 하거나 고지를 하고, 그 정보에 접근할 수 있는 대상자나 접근 방법을 제한하거나 그 정보에 접근한 자에게 비밀준수의무를 부과하는 등 객관적으로 그 정보가 비밀로 유지·관리되고 있다는 사실이 인식 가능한 상태를 의미한다(대법원 2010. 7. 15. 선고 2008도9066 판결 참조). 이러한 관리 노력이 없다면 법적 분쟁에서 영업비밀로 인정받기 어려우므로 이 부분을 유의할 필요가 있다.

2. 저작권 관점

저작권은 아이디어나 사상 그 자체가 아니라 그 아이디어를 구체적으로 외부에 표현한 창작적인 표현 형식을 보호한다.

사업계획서, 매뉴얼, 웹사이트 디자인, 소프트웨어 코드 등이 이에 해당한다. 저작권은 창작과 동시에 발생하며, 별도의 등록 절차 없이도 보호된다. 따라서 비즈니스 모델의 초기 단계에서도 신속하게 법적 보호 장치를 마련할 수 있다.

다만 저작권은 아이디어 자체를 제한하지 않기 때문에, 동일한 아이디어라도 다른 방식으로 구현하면 침해가 성립하지 않는다. 물론 이때 비즈니스 모델의 핵심 구성요소나 선택·배열이 단순한 아이디어의 차원을 넘어 창작적인 표현 형태에 이른 경우에는 저작권 보호가 가능할 것이나, 기능적 목적을 달성하기 위한 전형적이거나 표준적인 표현 방식은 창작성이 인정되기 어렵다는 점을 유의해야 할 것이다.

3. 특허권 관점

비즈니스 모델을 가장 강력하게 독점적으로 보호할 수 있는 권리는 특허권, 그중에서도 비즈니스 모델 특허(BM 특허)이다. 특허권은 출원일로부터 20년간 존속하므로, 장기적으로 안정적인 시장 지위를 확보하고 로열티 수익을 창출할 수 있다.

특허법은 자연법칙을 이용한 기술적 사상의 창작으로서 고도한 것을 '발명'으로 정의하고, 이러한 발명이 '산업상 이용가능성', '신규성', '진보성' 요건을 충족하면 등록을 통해 독점적 권리를 부여한다. 영업방법에 관한 것도 자연법칙을 이용한 기술적 사상의 창작으로서 고도한 것이라면 특허의 대상이 될 수 있는데, 이를 비즈니스 방법(BM) 발명이라 한다.

다만 BM 발명이 특허 대상이 되기 위해서는 소프트웨어를 통해 정보처리가 이루어지고, 그것이 하드웨어를 통해 구체적으로 실현되고 있어야 한다. 또한 특허청구범위의 기재가 단순한 아이디어를 제기하는 수준에 머물러서는 안 되고, 발명의 목적을 달성하기 위한 필수불가결한 모든 구성들이 구체적이고 명확하게 기재되어야 한다(특허법원 2006. 12. 21. 선고 2005허11094 판결 참조). 즉, 단순히 추상적인 아이디어가 아니라 '어떻게 작동하는가'에 대한 기술적 구현을 명확히 하는 것이 관건이다.

4. 상표권 관점

상표권은 비즈니스 모델의 핵심 요소인 브랜드 이름, 로고, 슬로건 등을 보호한다. 이는 고객이 상품이나 서비스의 출처를 식별하고, 품질을 신뢰하게 만드는 핵심적인 역할을 한다.

상표권은 기업이 시장에서 쌓은 브랜드 가치를 독점적으로 보호하고, 경쟁사의 유사 상표 사용으로 인한 혼동을 방지한다.

특히 상표권은 선출원주의가 적용되므로, 동일·유사한 상품에 사용할 동일·유사한 상표에 대하여 다른 날에 둘 이상의 상표등록출원이 있는 경우에는 먼저 출원한 자만이 그 상표를 등록받을 수 있다. 따라서 비즈니스 모델 구상 단계에서 조기에 출원하는 것이 중요하다. 메인 상표 외에도 슬로건, 캐릭터 등 다양한 형태로 출원하여 보호 영역을 넓히는 전략도 필요할 것이다.

아이디어를 권리로 만드는 과정은 기업의 미래 경쟁력과 직결되는 핵심 투자이다. 단기적인 수익을 넘어 장기적인 성장을 도모하기 위해서는 전문가와 함께 체계적인 권리화 전략을 설계하고 실행하는 것이 중요하다.



김도윤 변호사

[프로필 보기](#)

02-532-3426
kimdy@minwho.kr

MINWHO NEWS

양진영 변호사, 국제지식재산연수원 전문가 교육에서 AI·개인정보 비식별화 관련 강연 진행

양진영 변호사, 국제지식재산연수원 전문가 교육에서 AI·개인정보 비식별화 관련 강연 진행

법무법인 민후 양진영 대표변호사는 지난 4월, 국제지식재산연수원 교육과정에서 「인공지능을 이용한 개인정보 비식별화 기술 관련 연구 동향 및 사례」를 주제로 강연을 진행하였습니다.

이번 교육은 AI 기술 확산에 따라 중요성이 커지고 있는 개인정보 보호 및 비식별화 이슈를 중심으로 구성되었으며, 강의에서는 인공지능 기반 익명화·대체 기술 등 최신 비식별화 기술 동향과 함께, 실제 산업 현장에서 활용되고 있는 국내외 사례 및 최근 주요 이슈들이 폭넓게 소개되었습니다.

양진영 변호사는 최근 사회적으로 주목받은 사례들을 통해 개인정보 비식별화 문제의 중요성을 설명하고, 개인정보보호법 및 개인정보보호위원회 가이드라인을 기반으로 개인정보 개념, 적법한 수집·이용 원칙, 가명정보 처리 기준 등 관련 법적 쟁점을 체계적으로 정리하였습니다.

또한 데이터 유형과 활용 목적에 따라 적용될 수 있는 다양한 가명처리 및 비식별화 방법을 구체적인 사례와 함께 설명하며, 산업 분야별 실무 적용 방안에 대한 이해를 높이는 시간을 가졌습니다.

MINWHO NEWS

양진영 변호사, 동원그룹 대상 개인정보보호 특강 진행

양진영 변호사, 동원그룹 대상 개인정보보호 특강 진행

법무법인 민후 양진영 대표변호사는 지난 5월 동원그룹 임직원들을 대상으로 개인정보보호 및 AI 환경 변화에 관한 특강을 진행하였습니다. 이번 강연은 기업 운영 과정에서 발생할 수 있는 개인정보 이슈와 인공지능 기술 도입에 따른 법적 리스크를 점검하고, 실무 대응 역량을 강화하기 위해 마련되었습니다.

양진영 변호사는 강의를 통해 개인정보의 개념과 유형, 개인정보 수집·이용 절차, 제3자 제공 및 처리위탁 시 유의사항, 보유기간 종료 이후 파기 절차 등 개인정보 관리 전반에 관한 주요 내용을 설명하였으며, 개정된 기준에 따른 개인정보처리방침 작성과 내부관리계획 수립 등 기업이 갖추어야 할 안전성 확보 조치도 함께 다루었습니다.

특히 최근 AI 기술 활용이 확대되는 흐름과 관련하여 AI 챗봇, 안면인식 시스템 등 실제 사례를 중심으로 학습데이터 활용 문제와 자동화된 결정에 대한 정보주체 권리 등 최신 개인정보보호법 이슈를 심도 있게 설명하였으며, 최근 도입된 설명요구권 및 거부권 제도를 중심으로 기업이 AI 기술을 운영할 때 고려해야 할 법적 쟁점과 대응 방향도 함께 소개하였습니다.

마지막으로 개인정보 유출 사고 발생 시 기업이 부담할 수 있는 민·형사상 책임과 행정적 제재 가능성, 사고 발생 이후의 대응 체계 구축 필요성 등 실무적으로 중요한 사항들을 폭넓게 설명하며 참석자들의 이해를 돕는 등 AI 시대 기업이 갖추어야 할 개인정보보호 대응 역량과 관리 체계의 중요성을 강조하는 것으로 강연을 마무리하였습니다.

Business CASE

이달의 주요 업무사례

1. 명예훼손·업무방해 불기소 - 온라인 게시글로 문제 된 형사 사건에서 피의자를 대리하여 검찰 불기소 결정 도출
2. 물품대금소송 항소심 분쟁(계약 해석이 쟁점이 된 사건)에서 피고를 대리해 항소 전부 기각 승소)
3. 부정경쟁행위 금지 및 손해배상 청구 소송 (퇴사 후 유사 자료 무단 사용 사건)에서 원고 대리, 사용금지 및 손해배상 인용 승소
4. 저작권침해금지 소송 - 콘텐츠 무단 복제 사건에서 원고 대리해 게시물 삭제·재게시 금지 및 금전 지급을 이끌어낸 승소 사례
5. 부정경쟁방지법 위반 항소심 대응 - 영업비밀 국외유출 혐의 사건 (산업기술 유출 혐의) 피고인 대리, 검사 항소 기각으로 승소
6. 기업 경영 법률자문(이사회 운영규정 제정 및 지배구조 정비에 관한 검토 자문)
7. 개인정보보호법 위반 여부 검토 자문 (상조회사의 개인정보 접속기록 보관·점검 체계의 적정성 및 추가 안전조치 필요성 등)
8. 부정경쟁행위 화해권고결정 이후 추가 침해사실 확인에 따른 손해배상 청구 가능성 및 대응 방안에 관한 법률자문
9. 전자지급결제대행(PG) 기업의 선결제 구조 도입 가능성 및 전자금융거래법상 쟁점 검토에 관한 법률자문
10. 등록상표와 유사 표장 사용에 대한 상표권 침해 해당 여부 검토에 관한 검토 자문

Mi 법무법인 민후

서울특별시 강남구 테헤란로 134, 포스코 타워 역삼 11층

Tel. +82-2-532-3483 Fax. +82-2-532-3486

www.minwho.kr



[변호사 소개 바로가기]



[주요 업무 사례 바로가기]



[전화 상담 바로가기]



[카톡 상담 바로가기]



[홈페이지 상담 바로가기]



[이메일 상담 바로가기]

본 뉴스레터의 내용 또는 기타 법률 문의가 필요하신 경우
법무법인 민후로 연락주시면 담당 변호사님의 답변을 받으실 수 있습니다.

본 자료는 법무법인 민후에서 제공하는 일반적인 법률 정보 및 소식 자료로, 모든 법률적 상황에 적용되는 것은 아니므로, 구체적인 법적 조치에 대해서는 저희 법무법인에 문의하여 주시기 바랍니다. 또한 본 자료에 포함된 모든 내용의 저작권은 법무법인 민후에 있으므로, 무단 배포, 복사, 게재를 금합니다.