▣ 논 단

AI 환경에서의 주식회사 이사의 감시의무와 내부통제시스템의 실질화

– 개정 상법 제382조의3·인공지능기본법·Caremark–Marchand 법리의 통합적 분석 –

육 태 우

강원대학교 법학전문대학원 교수 / 법학박사

요 약 문

AI가 기업 경영의 핵심 의사결정 주체로 부상함에 따라 이사의 감시의무는 질적 전환을 요구받고 있다. 2025년 개정 상법 제382조의3은 이사의 의무 대상을 '회사'에서 '회사 및 주주'로 확장하여 AI 거버넌스 실패가 주주 공평대우 의무 위반으로 직결될 수 있는 법적 경로를 열었다. 2026년부터 시행된 인공지능기본법은 고영향 AI에 대한 사전 확인의무(제33조)·Human-in-the-loop 원칙(제34조제1항제4호)·영향평가(제35조)를 통해 이사의 선관주의의무 내용을 실정법 차원에서 구체화하였다. 특히 동법상 고영향 AI 조항은 Bingle 판결이 드러낸 '실정법 부재'의 공백을 메워, AI 리스크를 비즈니스 위험에서 컴플라이언스 리스크로 전환될 수 있는 규범적 근거를 제공한다.

이 연구는 Delaware Caremark·Marchand 법리를 AI 거버넌스에 기능적으로 재구성한다. 핵심 업무에 AI를 도입하는 기업에서 AI는 Mission Critical Risk에 해당할 수 있으므로 이사회는 AI 알고리즘의 투명성·공정성·보안을 실질적으로 모니터링할 의무를 부담하며, AI 경고신호에 대한 고의적 무지는 의식적 방치(conscious disregard)로 평가된다. 이사가 AI 결론을 탐문 없이 추인하는 이른바 '알고리즘 위험에 대한 의식적 방치'는 성실의무상 불성실(bad faith)에 해당하여 경영판단원칙의 보호를 배제한다. 비교법적으로 독일 주식법 제91조의 이분법과 1985년부터 2025년까지 '내부통제시스템의 실질화'로 수렴해 온 한국 대법원 판례 계보와의 합치를 논증하고, 입법론으로는 ① 준법지원인의 알고리즘 적법성 점검 기능 명문화, ② 알고리즘 영향평가(AIA)의 도입·공시 의무화, ③ AI 거버넌스 모범규준(Comply-or-Explain 방식)의 제정을 제언한다. Delaware SB21이 MFW 판례법의 소수파주주 보호 기능을 약화시킨 경위는, 한국에서 AI 거버넌스 절차적 안전항을 설계할 때 최소 기준을 타협해서는 안 된다는 비교법적 교훈을 제공한다.

주제어 : 이사의 감시의무, AI 내부통제시스템, AI 거버넌스, 개정 상법 제382조의3, Caremark 법리, 알고리즘 위험에 대한 의식적 방치, 고영향 AI, 인공지능기본법, 기업지배구조

〈目 次〉

Ⅰ. 서 론: AI 경영 시대와 지배구조의 패러다임 전환

현대 기업 경영에서 인공지능(AI)은 단순한 업무 보조 도구를 넘어 전략적 의사결정의 핵심 주체로 부상하였다. 기업들은 신용평가, 자산운용, 공급망 최적화, 인사결정 등 다양한 경영 영역에서 AI 알고리즘의 판단에 의존하고 있으며, 그 비중은 갈수록 확대되는 추세이다.[1] AI는 인간이 처리할 수 없는 방대한 데이터를 실시간으로 분석하여 높은 수익률과 운영 효율성을 제공하지만, 동시에 법적·윤리적 측면에서 심각한 도전을 야기한다.

이 중 가장 중대한 법적 문제는 알고리즘의 '불투명성(Black-box Problem)'이다. AI 시스템, 특히 딥러닝 기반의 모델은 그 판단 과정을 인간이 직관적으로 이해하거나 검증하기 어렵다.[2]

---

1) AI 의사결정 의존도에 관한 실증적 분석으로는 Erik Brynjolfsson & Andrew McAfee, The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies 89–112 (2014). 기업경영에서의 AI 활용 사례를 개관한 국내 문헌으로는 송옥렬, "인공지능과 이사회", 「기업법연구」 제35권 제2호, 93~96면 (2021) [이하 "송옥렬(2021)"] 참조. 이사회의 AI 도입·활용이 기업 경영에 미치는 영향을 이사회 중심으로 종합적으로 분석하면서, 이사의 선관주의의무·충실의무는 AI 도입 이후에도 변경되지 않는다는 원칙 아래 AI 거버넌스 프로세스의 적절한 구현이 이사회의 책무임을 논한 문헌으로는 홍복기, "인공지능이 회사제도에 미치는 영향 – 주식회사의 이사회를 중심으로 –", 「상사법연구」 제43권 제2호, 1~39면 (2024) [이하 "홍복기"] 참조.

이러한 기술적 속성은 이사회의 감독 기능을 근본적으로 위협한다. 이사가 AI의 결정 논리를 이해하지 못한 채 그 결과만을 수용한다면, 이는 선관주의의무(duty of care)와 충실의무(duty of loyalty)의 핵심 내용인 '정보에 입각한 판단(informed judgment)'을 포기하는 것과 다름없다.[3] 더욱이, 인간은 AI 결과물에 대해 '자동화 편향(automation bias)'을 나타내는 경향이 있어, 알고리즘의 결론이 직관적으로 의심스럽더라도 이를 최종 답으로 수용하는 경향이 있다. 이사도 예외가 아니며, AI 결과를 비판적으로 검토하지 않고 추인하는 것 자체가 선관주의의무 위반의 징표가 될 수 있다는 점은 각별히 유념하여야 한다.[4]

이러한 문제 상황은 2025년 상법 개정을 통해 법적 긴장감을 더욱 높이게 되었다. 개정 상법 제382조의3은 이사의 의무 대상을 종래의 '회사'에서 '회사 및 주주'로 확장하였으며, 주주에 대한 공평대우 의무를 명문화하였다.[5] 이로써 이사의 AI 거버넌스 실패는 단순히 회사에 대한 배임의 문제를 넘어, 주주에 대한 직접적인 의무 위반으로 구성될 수 있는 법적 경로가 열렸다.

이하에서는 이러한 문제의식 아래 다음의 순서로 논의를 전개한다. 먼저 개정 상법 제382조의3의 입법 취지와 그 AI 거버넌스 함의를 분석한다(Ⅱ). 이어서 미국 Delaware 법원의 Caremark 법리가 AI 내부통제 시스템 구축 의무로 어떻게 재구성될 수 있는지를 검토하고,

---

2) 딥러닝 모델의 불투명성 문제 일반에 관하여는 Finale Doshi-Velez & Been Kim, Towards a Rigorous Science of Interpretable Machine Learning (Feb. 28, 2017) (unpublished manuscript), https://www.semanticscholar.org/reader/5c39e37022661f81f79e481240ed9b175dec6513 (last visited June 5, 2026); 법적 맥락에서의 분석으로는 Frank Pasquale, The Black Box Society: The Secret Algorithms That Control Money and Information 1–18 (2015). 이를 회사법 관점에서 분석한 국내 문헌으로는 송옥렬(2021), 전게논문, 101면 ("현재의 딥러닝 기술에서는 인간이 적은 수의 코드만 설계하고 제한된 수의 규칙만 준 다음, 인공지능이 스스로 학습하여 인공신경망을 형성해 간다. 그 학습과정이 어떻게 이루어졌는지 인간은 알 수 없다"); 임철현, "AI 시대의 기업거버넌스 – AI의 기업법적 활용을 중심으로", 「서울법학」 제32권 제4호, 329~370면 (2025) [이하 "임철현(2025)"]; 최문희, "인공지능(Artificial Intelligence)과 회사법의 접점 – 인공지능의 활용에 수반되는 회사법적 논점 –", 「상사법연구」 제39권 제1호, 1~53면 (2020) [이하 "최문희"], 9면 및 32~33면 (딥러닝에서 "학습을 통해 만들어지는 심층신경망의 층과 연결망은 AI가 스스로 학습한 결과이지 인간이 모두 입력한 것이 아니"어서 최종 출력의 근거를 알 수 없는 블랙박스 문제를 회사법적 관점에서 최초로 분석) 참조. AI 시스템의 '인식론적 불투명성(epistemic opacity)' – 내부 작동이 사용자조차 해명하기 어려운 구조 – 이 이사의 fiduciary 판단의 전제를 침식하며, 이사가 AI 출력을 탐문 없이 추인하는 '인식론적 포기(epistemic abdication)' 위험에 관하여는 Maria Lillà Montagnani & Maria Lucia Passador, Fiduciary Duties and Business Judgment Rule 2.0 in the AI Act Age (June 9, 2025), https://ssrn.com/abstract=5714102 [이하 "Montagnani & Passador"], at 14–15 참조. 자동화 편향(automation bias) – 인간이 알고리즘의 결론을 직관적으로 의심하면서도 이를 최종 답으로 수용하는 경향 – AI 결과를 맹목적으로 추인하는 것이 이사의 선관주의의무 위반이 될 수 있다는 점에 관하여는 Alfred R. Cowger, Jr., Corporate Fiduciary Duty in the Age of Algorithms, 14 Case W. Reserve J.L. Tech. & Internet 136 (2022–2023) [이하 "Cowger"], at 140–142 참조.

3) 이사의 선관주의의무상 '정보에 입각한 판단' 요건에 관하여는 Smith v. Van Gorkom, 488 A.2d 858, 872–74 (Del. 1985) 참조. 동 판결은 이사가 충분한 정보 없이 합병에 찬성한 사안에서 선관주의의무 위반을 인정하였다.

4) Cowger, supra note 2, at 140, 163–165 참조.

5) 개정 상법 제382조의3의 입법 경위와 주요 내용에 관하여는 천경훈·정준혁, "이사의 충실의무와 주주이익 보호", 「상사법연구」 제43권 제3호, 1~65면 (2024); 이상훈, "이사의 주주에 대한 충실의무 — 비교법적 검토 —", 「비교사법」 제29권 제3호, 131~160면 (2022) 참조.

우리 대법원이 1985년 이래 발전시켜 온 내부통제시스템 감시의무 판례의 계보를 이와 접목한다(Ⅲ). 나아가 최근 미국 학계의 논의를 비교법적으로 검토한 뒤(Ⅳ), 위 논의를 종합하여 AI 환경에서 이사의 감시의무와 내부통제시스템 실질화를 위한 구체적 방향을 제시하는 것으로 결론을 맺는다(Ⅴ).

한편, 이 연구가 주된 분석 대상으로 삼는 AI는 학계의 분류를 따라 '보조형 인공지능(assisted AI)'과 '대체형 인공지능(autonomous AI)'의 스펙트럼 위에 놓인 것으로 이해한다. 보조형 AI는 이사의 판단을 보조하는 역할에 머무르고 최종 의사결정은 인간 이사가 담당하는 형태이며, 대체형 AI는 인간의 개입 없이 스스로 의사결정을 내리거나 인간이 형식적인 의결권만 보유하는 형태이다. 양 유형에 걸쳐 공통적으로 이사의 감시의무 문제가 발생하나, 그 내용과 강도는 달라진다. 특히 대체형 AI가 등장하면 종래 회사법이 전제하는 경영진의 이해상충 문제는 알고리즘의 오류·실패(algorithmic failure)라는 새로운 문제로 대체되고, 법적 강제의 중점 역시 이사의 사후적 책임 추궁에서 사전적 통제 체계의 구축으로 이전된다.[6] 이러한 변화는 이미 현실 기업 경영에서 실증되고 있다. AI를 의사결정 보조 수단으로 채택한 기업의 비중이 급격히 확대되고 있으며, 대체형 AI가 실질적 운영 주체로 기능하는 이른바 'AI 네이티브 기업'의 등장은 이사의 감시의무가 단순한 이론적 과제가 아니라 현행법의 즉각적 해석 과제임을 뒷받침한다.[7]

이 점에서 국내 회사법학에서 AI와 이사회의 관계를 국내에서 선도적으로 체계화한 최문희 교수의 선행연구는 본 논문의 중요한 이론적 기반을 제공한다. 해당 연구는 이사회에서 AI가 활용되는 양상을 ① 보조형(Assisted AI, 의사결정은 인간이 담당), ② 증강형(Augmented AI, 인간과 기계가 의사결정 분담), ③ 자율형(Autonomous AI, 기계가 의사결정권 자체를 보유)의 세 단계로 구분하고, 각 단계에서 이사의 의무와 경영판단원칙의 적용이 달라짐을 분석한다.[8]

---

6) 현행법상 AI 이사는 허용되지 않으며 이를 입법적으로 인정하기 위해서도 의무 구체화와 책임재산 부재라는 난관이 있음을 지적한 문헌으로는 최문희, 전게논문, 16~23면 참조. 보조형·대체형·완전 자율형의 단계적 구분과 각 단계별 책임 귀속 공백 문제에 관하여는 John Armour & Horst Eidenmueller, Self-Driving Corporations, ECGI Law Working Paper No. 457/2019 (August 2019), pp. 16, 26, 31–34 [이하 "Armour & Eidenmueller"] 참조. 동 문헌은 완전 자율형 단계에서 책임을 물을 인간 이사가 부재할 경우를 대비하여 최저자본제도·법인격부인론·책임보험 의무화라는 세 가지 대안을 제시하며, 알고리즘 목표함수의 경직성이 '알고리즘 실패(algorithmic failure)'를 야기할 수 있다는 점을 지적한다(p. 29).

7) 1인 창업자 비중 통계에 관하여는 Carta, State of Private Markets: Q4 2025 (2026), 스파크랩 김호민 공동대표 인터뷰, 동아일보 2026. 5. 8.자, "AI팀원 이끄는 1인 창업자, 유니콘 기업 만드는 시대 왔다" (last visited June 5, 2026). 메드비(Medvi) 사례에 관하여는 펀딩포유 매거진, "직원 2명으로 연매출 2조원, '1인 유니콘 기업' 시대의 서막", 2026. 5. 7.자 참조. 메드비는 첫 완전 회계연도(2025년)에 매출 약 4억 100만 달러(약 6,000억 원)를 기록하였고, 2026년 목표 매출은 약 18억 달러(약 2조 7,000억 원)이다. New York Times, Apr. 2, 2026; AI타임스 2026. 4. 7.자 참조.(https://www.funding4u.co.kr/community/insight_show/trend/203 (last visited June 5, 2026)). 스파크랩 공동대표의 발언("팀의 구성이 꼭 사람일 필요는 없게 된 것")은 동아일보, 전게 기사에서 인용.

8) AI 활용과 이사의 의무·경영판단원칙의 관계를 보조형·증강형·자율형 유형별로 체계화한 국내 선구적 문헌으로는 최문희, 전게논문, 1~53면 참조. 이른바 '이중층형' 의무 구조에 관한 비교법적 분석으로는 손창완, "상법

특히 AI 이사의 허용 여부에 관하여 현행법상 자연인 이사만을 허용하는 체계 아래에서 AI에게 이사의 지위를 부여하는 것은 허용되지 않으며, 입법론으로도 의무 구체화의 어려움과 책임재산 부재라는 근본적 난관이 있음을 지적한다. 나아가 이사는 AI의 작동에 관한 기초적 지식을 보유할 의무(숙지의무)를 부담하며, 이를 모른다는 사실 자체가 의무위반을 구성할 수 있다는 점을 명시한다. 본 논문의 연구는 이러한 선행연구의 성과를 계승하면서 2025년 개정 상법 제382조의3 및 Delaware Caremark-Marchand 법리와의 기능적 결합이라는 새로운 분석 축을 추가하여 논의를 심화한다.

Ⅱ. AI 감독의무의 법적 근거: 개정 상법 제382조의3의 함의

1. 개정의 배경과 입법 취지

2025년 상법 개정 이전까지 이사의 충실의무는 전통적으로 '회사'에 대한 것으로 한정되어 있었다. 판례와 다수 학설은 회사의 이익이 곧 주주 전체의 이익을 대변한다는 구조적 논리 아래, 주주에 대한 직접적인 의무를 이사에게 귀속시키는 것에 소극적이었다. 그러나 지배주주와 소수파주주 간의 이해충돌, 경영진의 사익 추구, 불공정한 합병·분할 과정에서 발생하는 소수파주주 피해 등이 누적되면서, 입법적 개입의 필요성이 강하게 대두되었다.[9)]

개정 상법 제382조의3 제1항은 이사가 "회사 및 주주를 위하여" 그 직무를 충실히 수행하여야 한다고 규정하고, 제2항은 이사가 주주를 공평하게 대우하여야 한다는 이른바 주주 공평대우 의무를 명시하였다. 이는 이사의 의무 구조를 종래의 '단층형'(회사에 대한 의무)에서 '이중층형'(회사 및 주주에 대한 의무)으로 전환시킨 중대한 변화이다.

개정의 입법취지를 분석하면 두 가지 핵심 목표가 드러난다. 첫째(목표 ①)는 회사 자체의 손해를 발생시킨다고는 보기 어려우나 주주 전체에게 귀속되는 몫을 감소시키는 사안－합병비율 불공정, 불리한 공개매수 대응 등－에서도 주주 전체의 이익을 보호하는 것이다. 둘째(목표 ②)는 주주들 사이에서, 특히 일반주주로부터 지배주주로 부당한 부(富)의 이전이 일어나는

개정의 현황과 평가 — 제21대 국회의 상법 개정안을 중심으로 —", 「상사법연구」 제43권 제1호 (2024) [이하 "손창완"], 271~280면. Delaware 회사법상 이사의 의무 구조와의 비교에 관하여는 Lyman Johnson, Unsettledness in Delaware Corporate Law: Business Judgment Rule, Corporate Purpose, 38 Del. J. Corp. L. 405, 412–18 (2013) 참조.

9) 개정 전 이사의 충실의무를 회사에 대한 의무로 한정하여 온 판례·다수설의 상황과 그에 대한 비판적 논의에 관하여는 천경훈·정준혁, 전게논문, 15~22면; 이상훈, 전게논문 참조. 개정의 배경이 된 지배주주와 소수파주주 간 이해충돌 문제에 관하여는 천경훈, "2025년 개정상법상 이사 충실의무 조항의 해석론", 「상사법연구」 제44권 제2호, 299~302면 (2025) [이하 "천경훈(2025)"] 참조.

것을 방지하는 것이다. 이에 대응하여 제2항 전단의 '총주주 이익 보호의무'는 목표 ①을, 후단의 '전체 주주 공평대우 의무'는 목표 ②를 각각 규율한다.[10]

이러한 입법적 전환은 AI 거버넌스의 맥락에서 특별한 의의를 지닌다. AI가 기업의 핵심 의사결정을 담당하는 현실에서, AI 시스템의 설계·운영·감독 방식이 특정 주주 집단에게 불리하게 작용할 가능성은 더 이상 이론적 우려에 그치지 않는다.[11] 개정 상법은 이러한 AI 관련 위험을 이사의 법적 책무 영역으로 명확히 포섭하는 법적 기반을 마련하였다.

2. 이중 기준의 도입: 회사와 주주의 이익 조화

개정 상법하에서 이사는 AI 시스템이 (i) 회사의 가치를 극대화하는지와 (ii) 총주주의 이익을 보호하는지라는 이중 기준을 동시에 충족하는지 감시해야 한다. 두 기준은 대부분의 경우 수렴하지만, AI를 활용하는 경영 환경에서는 일치하지 않을 수 있다.

예컨대, AI가 단기적 주가 상승에 최적화된 자산 매각 전략을 제안하더라도, 그 과정에서 소수파주주가 적정한 정보를 제공받지 못하거나 지배주주에 비해 불리한 조건에 처하게 된다면, 이사는 이러한 AI의 제안을 그대로 수용해서는 안 된다. AI의 예측 결과가 수익성 면에서 우월하더라도, 그 절차가 특정 주주에게 불리하다면 이사는 이를 통제해야 할 법적 책무를 진다는 점에서 개정법의 규범적 함의가 드러난다.

나아가 이사는 AI 시스템이 주주 간 이익 충돌을 어떻게 처리하는지 그 알고리즘적 논리를 이해하고, 필요한 경우 개입할 수 있는 내부 절차를 갖추어야 한다. 이러한 의무는 단순한 결과 점검을 넘어, AI의 의사결정 과정 자체에 대한 실질적 감독을 요구한다는 점에서 종래의 감독의무와 질적으로 다르다.

3. 주주 공평대우 의무와 알고리즘 편향

개정법 제2항의 주주 공평대우 의무는 AI 환경에서 특히 강한 규범적 함의를 지닌다. 알고리즘 편향(Algorithmic Bias)은 AI 시스템이 학습 데이터의 편향성이나 설계 의도에 따라 특정 집단에 불리한 결과를 산출하는 현상을 말한다. 금융 AI의 경우 이러한 편향이 지배주주 또는

---

10) 천경훈(2025), 전게논문, 299~302면. 나아가 영국 회사법 제172조 제1항의 "for the benefit of its members as a whole" 문언과 기능적으로 대응함에 관하여는 동 논문, 302면 참조.

11) AI 시스템이 주주 간 이익 충돌에 미치는 영향에 관한 분석으로는 임철현(2025), 전게논문, 345~352면; 같은 저자, "주주에 대한 충실의무 조항의 해석 – 독일법상 계약의 제3자보호효를 실마리로 하여 –", 「비교사법」 제33권 제1호, 145~171면 (2026) [이하 "임철현(2026)"], 151~153면 (AI 알고리즘이 매개하는 자본거래·조직재편을 개인법적 주주이익 침해의 전형적 사례로 제시). 알고리즘 편향의 법적 의미에 관하여는 Sandra G. Mayson, Bias In, Bias Out, 128 Yale L.J. 2218, 2225–34 (2019) 참조.

내부자에게 유리한 방향으로 작용할 수 있으며, 이는 소수파주주에 대한 불공평한 대우로 귀결된다. 이러한 결과가 발생하는 것은, 신용평가·투자추천·리스크 산정 등 금융 AI의 산출값이 결국 내부자가 설정한 목적함수와 가중치에 따라 결정되기 때문이다. 지배주주나 경영진은 알고리즘의 설계, 학습 데이터의 선정, 매개변수의 조정 과정에 사실상 영향력을 행사할 수 있는 반면, 소수파주주는 그 과정에 접근할 수단도 이를 검증할 정보도 갖지 못한다. 그 결과 알고리즘의 외견상 중립성 뒤에서 특정 주주집단에 유리한 편향이 구조적으로 재생산될 수 있다.

이사회는 AI 시스템이 주주 집단 간 이해관계에 중립적으로 작동하는지를 정기적으로 검증하고, 편향이 발견된 경우 즉시 시정 조치를 취할 의무를 부담한다. 이를 위해 이사회 차원에서 알고리즘 감사(Algorithmic Audit) 체계를 도입하고 그 결과를 주주에게 공시하는 제도적 장치가 필요하다.[12] 이때 알고리즘 감사는 AI 개발사나 경영진이 제출하는 자기보고형 성능·공정성 지표를 그대로 신뢰하여서는 안 되며, 해당 업종의 특성과 데이터 환경상 편향·오류가 개입되기 쉬운 영역에 대해서는 독립적 감사기구가 더욱 엄격한 검증을 수행하여야 한다. 다만 알고리즘 감사의 구체적 제도 설계 — 감사 주체의 독립성과 전문성 확보 방안, 비용 부담 구조, 비상장회사에의 적용 가능성 등 — 에 관하여는 후속 연구에서 상세히 검토한다.[13]

이와 관련하여 2025년 3월 Delaware 주지사가 서명한 DGCL 제144조 개정은 AI 거버넌스 설계에 중요한 비교법적 시사를 제공한다. 동 개정에서 신설된 안전항(safe harbor)은 이사의 이해관계 보유 사실 자체를 이유로 한 청구만을 차단하므로, AI 특별위원회의 실질적 독립성 결여나 공시 부실을 이유로 한 충실의무 위반 청구는 안전항 밖에 위치할 수 있다. 이는 이사회가 AI 시스템 도입 시 형식적 절차 준수에 더하여 실질적 독립성과 충실한 공시를 병행해야 함을 확인하는 것이다.[14]

---

12) 알고리즘 감사(Algorithmic Audit) 제도에 관하여는 EU AI Act, Regulation (EU) 2024/1689 of the European Parliament and of the Council, arts. 9–11 (2024) 참조. 인공지능 투명성을 AI 작동 프로세스 단계에 따라 ① 데이터 투명성(학습 데이터의 품질·출처·편향 공개), ② 알고리즘 투명성(모델 내부 작동의 이해가능성·해석가능성), ③ 표시 의무(AI 작동 여부 및 생성물 고지)의 세 유형으로 체계화하고, EU AI법이 각 유형별로 부과하는 의무의 내용과 한계를 비판적으로 분석한 문헌으로는 김민주·김현경, "인공지능 투명성 규제에 대한 비판적 고찰 — EU AI법 투명성 의무 조항 분석을 중심으로 —", 「선진상사법률연구」 통권 제110호, 59~108면 (2025) [이하 "김민주·김현경"], 70~76면 참조. 특히 동 논문은 EU AI법 제53조제1항(d)에 따른 범용 인공지능 모델의 학습 데이터 요약 공개 의무와 제10조의 고위험 AI 데이터세트 관리 요건이 영업비밀 침해 위험 및 기술적 이행 불능 문제를 내포한다고 비판하며, 나아가 알고리즘 투명성 측면에서 AI법이 요구하는 기술문서화가 실질적 이해관계자 참여 향상으로 이어지지 않을 가능성과 표시 의무 측면에서 워터마킹 등 기술적 한계를 상세히 분석한다(동 논문, 87~99면).

13) STX조선해양 분식회계 사건(대법원 2022. 7. 28. 선고 2019다202146·202153 판결, 병합)은 감사인(회계법인)의 주의의무와 관련하여, "경영자의 진술이나 피감회사가 제출한 자료 등을 그대로 신뢰하여서는 안 되고, 업종의 특성·피감회사가 속한 경영상황 등에 비추어 부정이나 오류가 개입되기 쉬운 사항이 있으면 더욱 엄격하게 감사하여야 한다"고 판시하였다. 이 직업적 회의주의 기준은 AI 시스템의 독립적 알고리즘 감사에서 개발사·경영진의 자기보고에 대한 검증 강도를 정하는 기준으로 유추 적용될 수 있다.

14) Michael J. Barrie & Andrew D. Kinsey, Delaware Amends DGCL Section 144 to Add Safe Harbors for Interested and Controlling Stockholder Transactions, Benesch Law (Apr. 9, 2025), available at

이 맥락에서 합병 거래를 중심으로 개정 상법 제382조의3의 실무적 함의를 검토한 학설은 주목할 만한 분석을 제시한다. 주식회사 합병은 개정 상법상 이사의 주주이익 보호의무가 1차적으로 적용되는 거래 유형으로서, 합병 과정에서 대상회사 자체의 순자산 증가에도 불구하고 불공정한 합병 비율로 소수파주주가 손해를 입는 이른바 '파이 배분 불공정' 구조는 전통적 충실의무 법리만으로는 포착되지 않는다. 이러한 구조적 문제는 AI 거버넌스에서도 동일하게 재현된다. 즉 AI 알고리즘이 회사 전체의 수익성을 높이면서도 특정 주주집단에 불리한 방향으로 작동하는 경우, 그 불이익은 회사의 손해로 환원되지 아니하므로 '회사에 대한 충실의무' 법리만으로는 포착되지 않는다. 아울러 합병에서의 이사 의무 준수가 실체적 정당성(경영상 필요성·합병 비율 적정성)과 절차적 정당성(이사회 결의·특별위원회·주주총회 결의)의 이중 기준에 따라 판단된다는 논거가 제시되는데, 이 이중 기준은 AI 거버넌스에서도 그대로 적용된다. 특별위원회 승인 및 소수파주주 다수결을 한국 상법에 직접 이식하기에는 현행법상 구조적 한계(상법 제393조의2 제1항 제1호의 위임 제한 등)가 있다는 지적도 있다.[15] AI 특별위원회 모델 역시 동일한 구조적 제약에 직면하며, 현행 상법 범위 내에서 이사회가 AI 특별위원회의 검토 결과를 후속적으로 수용하는 방식이 잠정적 해법이 될 수 있지만, 궁극적으로는 이에 대한 제도적 보완이 필요해 보인다.[16]

학계에서는 이러한 의무의 법적 성격과 범위에 대하여 이견이 있다. 먼저 개정법이 이사와 주주 사이의 직접적인 신인관계(fiduciary relationship)를 창설하는 창설적 입법이라는 견해가 있는 반면, 개정 전 상법하에서도 이사는 주주 전체의 이익을 보호하고 주주간 공평을 도모할 의무를 원래부터 부담하였는바 개정법은 이를 명문화한 확인적 입법에 불과하다는 견해도 유력하다.[17] 확인적 입법설은 개정 조문이 법적 의미에서 당사자들의 권리의무를 근본적으로 변

---

https://www.beneschlaw.com/insight/delaware-amends-dgcl-section-144-to-add-safe-harbors-for-interested-and-controlling-stockholder-transactions/ (last visited June 5, 2026) [hereinafter Barrie & Kinsey]. 안전항은 이사의 이해관계 보유 사실 자체를 이유로 한 청구만을 차단하므로 AI 특별위원회 설계의 비교법적 참고로 활용할 수 있다. SB21의 단일 보호 면책규정 도입이 AI 내부통제·감시의무 맥락에서 갖는 의미에 관하여는 김지현, "델라웨어주 회사법의 최근 동향과 한국 상법에 대한 시사점 – 지배주주 거래 규제와 절차적 안전장치를 중심으로 –", 「기업법연구」 제39권 제4호(2025)[이하 "김지현"], 336~341면 참조.

15) 특별위원회 승인 및 소수파주주 다수결의 한국 상법 이식에 따르는 구조적 한계(상법 제393조의2 제1항 제1호의 위임 제한, 이사회내 위원회의 보수 한도, 자문사 선임 권한의 제약 등)에 관하여는 김지평, "이사의 주주이익 보호의무 관점에서의 합병 계약의 쟁점", 「기업법연구」 제39권 제4호(한국기업법학회, 2025. 12.)[이하 "김지평"], 268~275면 참조.

16) 김지평, 전게논문, 236~238면, 243면, 268~275면. 주식회사 합병이 개정 상법상 주주이익 보호의무의 1차 적용 대상임에 대하여는 동 논문, 236~237면; '파이 배분 불공정' 구조(합병 과정에서 회사에는 순자산 증가가 발생하더라도 소수파주주가 불공정한 합병 비율로 손해를 입는 문제)에 대하여는 동 논문, 238면; 이사 의무 준수의 실체적 정당성(Fair Price)·절차적 정당성(Fair Dealing) 이중 기준에 대하여는 동 논문, 243면; 특별위원회 및 소수파주주 다수결의 한국 상법 구조적 이식 한계(상법 제393조의2 제1항 제1호의 위임 제한, 이사회내 위원회 보수 한도(제388조), 자문사 선임 권한(현행법상 감사위원회에만 인정, 제415조의2 제5항), 소수파주주 다수결 실현 방안의 법적 논란)에 대하여는 동 논문, 268~275면; 연성 가이드라인 또는 입법을 통한 제도적 보완의 필요성에 대하여는 동 논문, 275면.

경하는 의미는 크지 않을 수 있더라도, 회사 및 시장의 참여자들에게 '주주를 중시하라'는 강력한 신호효과를 발신한다는 점을 강조한다. 어느 견해에 따르더라도, 개정 상법이 이사와 주주 사이에 위임계약을 직접 성립시키지 않음은 분명하다. 그러나 AI 거버넌스 실패로 인한 주주 피해에 대한 이사의 책임은 개정법을 통해 보다 명확한 법적 근거를 갖게 되었으며, 이사는 개정법 제382조의3만이 아니라 제399조 또는 제401조에 기하여 주주에 대한 책임을 부담할 수 있게 된 것이다.[18]

이와 관련하여, 2025년 개정 상법 제382조의3의 법적 성격에 관하여, 동 조항을 독일 민법학의 '계약의 제3자보호효(Vertrag mit Schutzwirkung für Dritte)' 이론을 회사법 영역에 입법화한 것으로 파악하면서, 주주이익을 '단체법적 주주이익'(회사이익을 거쳐 배당 등 단체법 절차로 귀속되는 이익)과 '개인법적 주주이익'(자본거래·조직재편 과정에서 주주에게 직접·개별적으로 귀속되는 이익)으로 구별하고, 개정 조항은 후자를 새로운 보호대상으로 명시한 것이라고 해석하는 견해가 있다.[19] 이러한 해석론은 AI 거버넌스 맥락에서 특히 중요한 함의를 지닌다. AI 알고리즘이 매개하는 물적분할 결정·알고리즘 기반 공개매수에서의 정보 비대칭·AI 주도의 주주 간 차별적 처우는, 회사에 대한 손해를 수반하지 않더라도 특정 주주집단의 개인법적 주주이익을 침해하는 충실의무 위반으로 구성될 수 있기 때문이다. 이에 따르면, 이사의 AI 감시의무가 형해화될 경우, 이사는 상법 제399조(유추적용)와 제401조의 경합적 적용으로 주주에 대한 직접 책임을 부담할 수 있게 된다.[20]

이러한 개정의 비교법적 위치를 이해하기 위해서는, 유럽 주요국의 법 현황을 폭넓게 조망할 필요가 있다. 최근 발표된 비교법 연구에 따르면, 조사 대상 28개 유럽 국가 모두 이사의 의무는 원칙적으로 '회사에 대한 것'으로 보고 있으며, 영국·프랑스·아일랜드·사이프러스 4개국만이 판례를 통해 이사와 주주 사이에 특별한 사실관계(special factual relationship)가 있는 매우 예외적인 경우에만 주주에 대한 이사의 직접적 충실의무를 인정한다.[21] 독일 주식법은 이사와 주주 사이에 특별한 결합관계(Sonderverbindung)가 존재하지 않는다는 이유로 이사의 대내적

---

17) 이사와 주주 간 신인관계 성립 여부에 관한 학설 대립에 관하여는 천경훈·정준혁, 전게논문, 15~22면. 확인적 입법설 대 창설적 입법설의 대립에 관하여는 천경훈(2025), 전게논문, 311~312면 참조. 또한 비교법적으로는 Jonathan R. Macey & Geoffrey P. Miller, Corporate Stakeholders: A Contractual Perspective, 43 U. Toronto L.J. 401 (1993) 참조.

18) 임철현(2025), 전게논문, 164~166면(주주 충실의무 위반에 대해 상법 제399조·제401조의 경합적 적용을 인정하는 책임 체계론). 동 논문의 핵심 논거는, 회사와 이사 간 위임계약에서 주주는 제3자이지만 ① 주주의 계약과의 밀접성, ② 회사의 주주 보호에 관한 특별한 이익, ③ 이사의 인식 가능성이라는 제3자보호효의 세 요건이 충족되므로 주주가 이사에게 직접 채무불이행책임을 물을 수 있다는 것이다.

19) 임철현(2025), 전게논문, 145~148면, 151~156면 참조.

20) 임철현(2025), 전게논문, 151~156면(단체법적 주주이익과 개인법적 주주이익의 구별론 및 개정 상법 제382조의3의 해석 체계).

21) 정대익, "주주에 대한 이사의 충실의무 부존재 하의 독일법상 주주 보호 수단 – 유럽 국가 및 영연방 국가와의 비교를 겸하여 –", 「상사법연구」 제44권 제1호(2025), 35~85면 [이하 "정대익"], 58면.

책임(Binnenhaftung) 원칙을 채택하며, 주주 보호는 사원권관계(Mitgliedschaftsverhältnis)에 기초한 회사의 채무불이행책임 및 사원권의 절대권성(absolutes Recht)에 근거한 불법행위책임이라는 간접적 경로로 달성된다.[22] 유럽 주요국은 대체로 이사의 의무를 회사에 대한 것으로 보고, 주주에 대한 직접적 의무는 예외적으로만 인정한다.[23] 이와 달리 2025년 개정 상법 제382조의3은 이사가 주주에게 지는 직접적 충실의무를 – 개별 사안이 아니라 일반조항의 형식으로 – 명문화하였다. 이는 유럽의 일반적 규율과 뚜렷이 구별되는 적극적 선택이다. 다만 이 조항이 AI 거버넌스 영역에서 실효적으로 기능하려면 세 가지 이론적 과제를 먼저 정리해야 한다. 즉 ① 주주의 '직접 손해'를 어떻게 정의할 것인지, ② 남소를 막기 위해 어떤 요건을 둘 것인지, ③ 이사 책임의 법적 성격을 어떻게 볼 것인지이다. 이 과제들에 해석론적 답을 마련하는 것이 바로 이 연구의 중심 과제이며, Caremark 감시의무 법리와의 기능적 결합과 AI Red-team 등 절차적 장치의 설계가 그 핵심을 이룬다.[24]

Ⅲ. Caremark 법리와 AI 내부통제 시스템의 실질화

1. Caremark 법리의 연혁과 AI 시대의 재해석

1996년 Delaware 형평법원의 **In re Caremark International Inc. Derivative Litigation** 판결은 이사의 감시의무에 관한 현대적 기준을 정립한 선도적 판례이다. 이 판결에서는 이사가 기업의 법규 준수를 위한 합리적인 정보 및 보고 시스템을 구축하지 않거나, 시스템의 존재에도 불구하고 이를 의식적으로 무시하는 경우 선관주의의무 위반이 성립한다고 판시하였다.[25]

---

22) 정대익, 전게논문, 44~48면 참조. 상세는 후술 각주 24 참조.

23) 유럽 주요국 및 영연방 국가의 규율 상황에 관한 비교법적 검토로는 정대익, 전게논문, 58면 이하 참조.

24) 독일법의 이사 대내적 책임 원칙(Binnenhaftung)을 뒷받침하는 11개 논거의 상세한 분석은 정대익, 전게논문, 44~48면 참조. 사원권의 절대권성(absolutes Recht, 독일 민법 제823조 제1항)에 근거한 불법행위책임 및 사원권관계(Mitgliedschaftsverhältnis)를 통한 회사의 채무불이행책임이라는 간접 보호 구조에 관하여는 동 논문, 50면, 55~58면 참조. 나아가 동 논문은 이사의 주주에 대한 충실의무를 일반조항으로 입법화하는 경우 ① 신인관계적 요소 존재 여부, ② 적용 범위 확정, ③ 직접 손해의 개념 정의, ④ 남소 방지 수단, ⑤ 충실의무 위반의 책임 성격 등을 사전에 검토해야 한다는 입법론적 경고를 담고 있어(동 논문, 71~75면), 2025년 개정 상법 제382조의3의 AI 거버넌스 적용을 위한 해석론적 기초를 제공한다.

25) In re Caremark Int'l Inc. Derivative Litig., 698 A.2d 959, 970 (Del. Ch. 1996). 동 판결에 관한 상세한 분석으로는 Stephen M. Bainbridge, Caremark and Enterprise Risk Management, 34 J. Corp. L. 967, 969–74 (2009). 이후 동 법리는 Stone v. Ritter, 911 A.2d 362, 370 (Del. 2006)에 의하여 성실의무(duty of good faith) 맥락에서 재확인되었다. EU AI Act 시대에 Caremark 법리가 'AI 상당한 주의의무(duty of AI due care)'로 재구성되어야 한다는 논거 – 이사는 알고리즘 시스템의 설계 전제·인식론적 한계·운영 결과를 적극 탐문해야 하며, 절차적 형식만으로는 fiduciary 판단을 다한 것이 아니라는 점 – 에 관하여는 Montagnani & Passador, supra note 2, at 13–21, 33–35 참조. AI에 전적으로 위임하는 경우 BJR 전제 요건 미충족(unintelligent/unadvised

이후 **Stone v. Ritter**(2006) 판결은 Caremark 의무를 성실의무(duty of good faith)의 맥락으로 재위치시키면서, 이사의 의식적 방치(conscious disregard)가 인정되는 경우 경영판단원칙(Business Judgment Rule)의 보호를 받을 수 없다고 선언하였다. 주목할 점은 Caremark 법리가 등장 후 20여 년간 실질적으로 기능하지 못하다가 2019년 이후 비로소 연속된 판결들에서 실효성을 발휘하기 시작하였다는 것이다. 학설은 그 원인으로 델라웨어 회사법 제220조(주주열람권) 해석의 변화를 지목한다. 즉, 비공식 이메일까지 열람 범위로 인정되면서 이사의 주관적 불성실(bad faith)에 대한 입증이 용이해진 것이 Caremark 기준 실효화의 관건이었다. AI 환경에서 이 입증 구조는 더욱 복잡해지는데, 알고리즘 결정 로그·XAI 보고서·이사회 의사록이 이사의 불성실 또는 성실(good faith)을 증명하는 핵심 증거로 기능하기 때문이다.[26][27]

AI 시대에 Caremark 법리는 새로운 도전에 직면한다. 비교법적으로 주목할 점은, Marchand 판결이 제시한 'mission-critical' 기준의 입법적 선례가 이미 1998년 독일 콘트라법(KonTraG)에 의해 신설된 독일 주식법 제91조 제2항에 존재하였다는 사실이다. 동 조항은 이사회에게 "회사의 존속을 위협할 만한 상황전개"를 조기에 인식하기 위한 조기경보시스템(Frühwarnsystem) 구축을 의무화하였는데, 이는 Delaware 판례법보다 21년 앞서 'mission-critical' 개념을 성문화한 것으로 평가된다.[28] 전통적 내부통제 시스템은 인간 구성원의 행동을 규율하고 모니터링하는 것을 전제로 설계되어 있으나, AI는 인간의 의도와 독립적으로 스스로의 학습과 추론을 통해 결과를 산출한다. 따라서 AI를 포함한 기업 운영의 내부통제 시스템은 (i) AI의 알고리즘적 의사결정 과정을 모니터링하고, (ii) AI가 산출하는 경고신호(Red Flags)를 포착하며, (iii) 필요

---

judgment) 및 전체적 공정성(entire fairness)의 입증 불가에 관하여는 Cowger, supra note 2, at 165–167 참조.

26) Stone, 911 A.2d at 370 ("We hold that Caremark articulates the necessary conditions predicate for director oversight liability."). '의식적 방치(conscious disregard)' 기준에 관하여는 In re Walt Disney Co. Derivative Litig., 906 A.2d 27, 67 (Del. 2006) 참조.

27) Caremark 법리가 등장 후 20여 년간 실질적으로 기능하지 못하였다가 2019년 Marchand 판결 이후 비로소 실효성을 발휘하게 된 원인에 관하여, 송옥렬 교수는 델라웨어 회사법 제220조(주주의 회계장부·회사기록 열람권) 해석의 변화를 핵심 요인으로 지목한다. 법원이 비공식 인터넷 교신(이사들 사이의 사적 이메일·메시지)까지 열람 범위로 인정하면서 이사의 주관적 요소(scienter)에 대한 원고 측 입증이 용이해졌고, 이것이 감시의무 위반 인정의 현실적 관문을 낮추었다는 것이다. 송옥렬, "이사의 감시의무와 내부통제시스템 구축의무", 「기업법연구」 제36권 제1호, 한국기업법학회, 9~43면 (2022) [이하 "송옥렬(2022)"], 32면 참조. 전통적 법리인 "의심할 만한 사유(red flag)"와 "신뢰의 원칙(principle of trust)"이 미국·영국·일본·한국에서 현재도 기본 법리로 기능하고 있음에 관하여는 동 논문, 14~23면; 신뢰의 원칙이 AI 환경에서 알고리즘 불투명성으로 인해 이사의 "고의적 무지"를 은폐하는 방패로 남용될 위험이 있으므로 Mission Critical AI 영역에서는 그 적용이 제한되어야 한다는 논거는 동 논문의 mission critical 판단기준론과 본고 Ⅲ.2의 논의를 결합한 것이다.

28) 김정호, "내부통제시스템과 이사의 법적 책임 – 특히 독일주식법의 상황을 소개하며 –", 「상사법연구」 제41권 제3호, 한국상사법학회, 155~214면 (2022) [이하 "김정호(2022)"], 159~160면 참조. 나아가 독일 주식법 제91조 제3항은 2021년 FISG법(Wirecard 사태 계기)으로 신설되어 상장사에 대해 비필수영역(非必須領域)에서도 유효성(Wirksamkeit)과 상당성(Angemessenheit) 원칙에 따른 내부통제시스템 구축을 의무화하되, 여기서는 비용·효용 형량이 허용된다. 동 논문, 163~165면.

한 경우 인간이 개입할 수 있는 메커니즘을 갖추는 방향으로 재설계되어야 한다.[29] 특히 AI 결과에 전적으로 위임하는 경우, '정보에 입각한 판단'이라는 경영판단원칙의 전제 요건이 충족되지 않아 BJR의 보호를 받을 수 없으며(unintelligent or unadvised judgment), 블랙박스 특성상 'wholly fair' 요건도 입증이 불가능해진다는 점은 미국의 유력한 학설이 명확히 논증한 바 있다.[30] 또한 AI 도구 제조사를 '정보수탁자(Information Fiduciary)'로 일반화하여 기업 데이터를 부적절하게 사용하는 경우 별도의 신인의무를 부담시키는 또 다른 책임 구조도 AI 거버넌스 실행력을 높이는 수단으로 검토되어야 한다.[31]

2. Mission Critical Risk로서의 AI 관리

2019년 Delaware 대법원의 **Marchand v. Barnhill** 판결은 Caremark 법리를 'Mission Critical Risk' 이론으로 발전시켰다. 법원은 기업의 핵심 사업 위험에 관하여는 이사회 수준에서 직접적이고 실질적인 모니터링 시스템이 존재해야 한다고 판시하였다. 단순히 경영진이 보고하는 정보를 수동적으로 수령하는 것만으로는 부족하며, 이사회가 해당 위험을 능동적으로 추적할 수 있는 구조가 필요하다.[32]

이 법리를 AI 거버넌스에 적용하면, AI를 핵심 의사결정 도구로 활용하는 기업에서 AI 관련 위험은 Mission Critical Risk에 해당할 수 있다. 이사회는 AI 알고리즘의 설계 원칙, 학습 데이터의 품질, 모델의 오류율, 보안 취약점 및 편향성 등을 실시간으로 모니터링할 수 있는 합리적인 보고 체계를 구축할 의무를 부담한다.[33]

---

29) AI 내부통제 시스템 재설계의 필요성에 관하여는 임철현(2025), 전게논문, 348~352면; Anna Gressel, Avi Gesser & Lily Coad, AI Oversight Is Becoming a Board Issue, Harv. L. Sch. F. on Corp. Governance (Apr. 26, 2022), https://corpgov.law.harvard.edu/2022/04/26/ai-oversight-is-becoming-a-board-issue/.

30) Cowger, supra note 2, at 165–167 참조.

31) Information Fiduciary(정보 수탁자) 개념은 데이터를 수집·활용하는 자가 데이터 제공자에 대해 신인의무를 부담한다는 이론으로, Balkin & Zittrain이 소셜미디어 맥락에서 처음 제안하였다. Cowger는 이 개념을 기업 AI 도구 개발사·제조사에 확장하여, AI 도구가 기업 데이터를 부적절하게 처리하거나 알고리즘 결함으로 손해를 야기한 경우 제조사가 별도의 신인의무를 부담해야 한다고 주장한다. 또한 완전 자율형 AI 법인체(Algorithmic Entities, AEs)의 위험성과 이를 금지해야 할 필요성에 관하여는 Cowger, supra note 2, at 185–192, 198–202 참조.

32) Marchand v. Barnhill, 212 A.3d 805, 822 (Del. 2019) ("When a company operates in an environment where externally imposed regulations define the most important risks to the company's business, the board's oversight function must be particularly attentive to regulatory compliance."). 동 판결의 AI 거버넌스 함의에 관한 분석으로는 Bainbridge, supra note 25, at 971–76 참조.

33) 핵심 임무 개념이 Caremark 기준의 판단에서 어떻게 기능하는지에 관한 체계적 분석으로, 류지민, "케어마크 청구에서의 미션 크리티컬 – 이사의 감시의무 영역의 범위와 한계 –", 「상사법연구」 제41권 제4호, 한국상사법학회, 351~421면 (2023) [이하 "류지민"] 참조. 류지민은 Marchand 판결(식품안전), Clovis 결정(의약품 안전성), Chou 판결(계열회사 의약품 안전성), Armstrong 판결(송유관 건전성), Boeing 결정(항공기 안전) 등 케어마크 청구가 '성공'한 사례들에서 법원이 해당 리스크를 핵심 임무로 인정한 핵심 요인을 분석한다. 특히 Boeing

구체적으로, 이사회 수준의 AI 거버넌스 시스템은 다음의 4가지 요소를 포함해야 한다. 첫째, AI 리스크 관리 전담 위원회(AI Risk Committee)의 설치와 정기적 보고 체계. 둘째, AI 모델의 성능·편향·보안을 정기적으로 검증하는 알고리즘 감사 절차. 셋째, 예외적 상황 발생 시 이사회에 즉각 보고하는 에스컬레이션(Escalation) 프로토콜. 넷째, AI 관련 중요 의사결정에 대한 이사회 승인 절차.[34] 이러한 요소들이 유기적으로 결합될 때, Caremark가 요구하는 '합리적인 감독 시스템'의 기준을 충족할 수 있다.

한편, 이러한 맥락에서, 규제 준수 및 감독 기능에 AI 기술을 접목한 레그테크(RegTech)와 섭테크(SupTech)는 이사의 감시·감독의무 이행을 위한 내부통제시스템의 실질적 구현 수단으로서 특히 주목된다.[35] AI를 활용한 내부통제 시스템의 구체적 모습은 분야별로 다양하게 나타난다. 첫째, 컴플라이언스 모니터링 영역에서는 전자메일·사내 메시지의 자동 스크리닝, 전화 통화 내용 분석, 품질검사 자동화 등 AI가 인간 감사인의 역할을 대체·보완하는 방향으로 발전하고 있다. 둘째, 계약업무 지원 영역에서는 AI를 활용한 계약서 초안 작성·수정·검토, 계약 조항의 체크리스트 자동화, 전자계약 관리 등이 실무에 도입되고 있다. 셋째, 법률상담 지원 영역에서는 법무부문이 리걸리서치 서비스나 AI 챗봇을 활용하여 사내외 법률 문의에 초동 대응하는 사례가 증가하고 있다. 넷째, 인사(HRtech) 영역에서는 채용 심사, 업무 자기분류, 근로자 유형(고용·자영업·업무위탁 등) 판별에 AI가 활용되고 있으며, 이 경우 유럽 AI 규정이 고위험 AI 시스템으로 분류하여 엄격한 규제를 부과하는 점에서 이사회의 특별한 감독이 요구된다.[36]

결정에서 법원이 (i) 항공기 안전 전담 이사회내 위원회의 부재, (ii) 안전 문제에 대한 정기적 심의 절차의 결여, (iii) 경영진의 선별적 보고 구조라는 세 가지 구조적 실패를 Caremark 제1요건 충족의 근거로 삼은 점(동 논문, 382~386면)은, 본고가 AI 거버넌스 시스템의 필수 요소로 제안하는 AI 리스크 위원회 – 알고리즘 감사 – 에스컬레이션 프로토콜 체계와 구조적으로 대응한다.

34) AI 거버넌스 시스템의 구성요소에 관하여는 National Institute of Standards and Technology, Artificial Intelligence Risk Management Framework (AI RMF 1.0) 9–18 (2023) [hereinafter NIST AI RMF]. 데이터 거버넌스의 수립·유지가 AI 시대 이사의 선관주의의무의 핵심 내용이 됨을 논한 문헌으로는 송옥렬(2021), 전게논문, 100~101면 ("적절한 데이터 거버넌스의 수립 및 유지는 인공지능 시대에 이사의 중요한 선관주의의무의 내용이 된다. 데이터 거버넌스는 CEO를 포함한 경영진에게 일차적 의무가 있지만, 그 적정성의 감독 역시 이사회의 감시기능의 범위에 포함된다").

35) 레그테크(RegTech)는 규제 준수 의무 이행을 촉진하기 위해 고안된 혁신적 기술을 지칭하고, 섭테크(SupTech)는 감독당국의 감독 역량을 향상시키는 기술을 의미한다. 이사의 감시·감독의무 이행을 위한 내부통제시스템에 레그테크·섭테크를 도입하면 대규모 회사에서 비상근 사외이사도 상시적으로 감시·감독의무를 이행할 수 있게 된다는 점, 그리고 이사회 개최 없이도 위험요소를 이사에게 보고하여 의도적 보고 누락 관행을 타파할 수 있다는 점에 관하여는 김민정, "이사의 감시·감독의무 이행을 위한 레그테크(RegTech)·섭테크(SupTech) 도입에 관한 연구", 「경영법률」 제33집 제3호, 한국경영법률학회, 1~38면 (2023) [이하 "김민정"], 28~29면 참조. 레그테크에 적용되는 기술로는 빅데이터 분석, 리스크 및 규정 준수 모니터링, 모델링·시각화 기술, 기계학습 및 인지기술 등이 거론된다. 동 논문, 28면.

36) 松尾剛行, 「AIとガバナンス――企業統治の高度化・効率化にAIを役立てるという観点からの検討」, 商事法務 No. 2297 (2022), 26–31頁 [이하 「松尾(2022)」]. 계약업무 지원·법률상담 AI에 관하여는 同論文 28頁; HRtech 분야 AI에 관하여는 同論文 27頁; 유럽 AI 규정의 고위험 AI 분류에 관하여는 同論文 27頁 참조.

다만 이러한 AI 활용이 항상 안전한 것은 아니다. 특히 주목해야 할 것은 AI의 '오검지'(誤檢知) 즉, '잘못된 검지'(false positive/negative) 문제이다. AI가 잘못된 결과를 산출하였음에도 이를 적시에 발견·정정하지 못하면, 기업은 해당 오류에 기초한 위법 행위를 지속하거나 오히려 심화시키는 위험에 노출된다. 실제로 국제적 플랫폼 사업자가 이용규약 위반 여부의 판정에 AI를 활용한 사안에서, AI가 잘못된 판정을 하였음에도 사업자가 인간의 실질적 개입 없이 그 판정을 반복하다가, 이용자의 제소에 이르러서야 비로소 이를 철회한 사례가 있다. 이는 AI가 리스크 자체를 제거하는 것이 아니라, 부적절한 AI 운용이 새로운 거버넌스 리스크—즉 'AI 자체가 이사의 감독 해태를 야기하는 위험'—를 창출할 수 있음을 보여준다. 이사회는 AI를 내부통제 수단으로 도입하는 것과 동시에, AI 시스템 자체에 대한 검증·감독 체계를 별도로 구축해야 하는 이중적 의무를 부담한다.[37]

한편 2026년 1월 22일부터 시행된 인공지능기본법은 Caremark 법리가 요구하는 '합리적인 감독 시스템'의 내용을 한국법 차원에서 구체화하는 중요한 외부 규범 기준으로 기능한다. 특히 동법 제2조제4호는 '고영향 인공지능'을 정의하면서 '채용, 대출 심사 등 개인의 권리·의무 관계에 중대한 영향을 미치는 판단 또는 평가'(사목)를 그 적용 영역으로 명시하였다. 이는 채용 알고리즘·신용평가 모델을 운용하는 기업의 이사회에 직접적인 규범 기준을 부여한다. 동법 제33조부터 제35조는 고영향 AI에 대한 사전 확인(제33조), 위험관리방안·설명방안·이용자 보호방안 수립 및 '고영향 인공지능에 대한 사람의 관리·감독' 의무(제34조제1항제4호)를 법정화하였다. 특히 제34조제1항제4호의 '사람의 관리·감독' 조항은 Human-in-the-loop 원칙을 실정법 수준으로 격상한 것으로, 이사회가 AI 시스템에 의사결정권을 전적으로 위임하는 행위를 명시적으로 금지하는 근거로 기능한다. 다만 제35조의 고영향 AI 영향평가는 '노력' 의무로 규정되어 있어 강행 의무는 아니나, 시행령 제28조는 영향평가에 포함되어야 할 사항 7가지를 구체화함으로써 이를 이행한 사업자와 이를 해태한 사업자 간 이사 책임의 분화를 유도하는 구조를 형성한다. 회사의 핵심 업무에 사용되는 AI 시스템이 고영향 AI에 해당하는지를 이사회

37) 松尾(2022), 前揭論文, 29頁. 同論文은 국제적 위법행위에 AI가 활용된 사안으로서 AI가 정확한 결과를 산출하지 못하였음에도 피해자 및 법원에 그 결과가 제시된 사례(注22)를 소개하며, AI 자체가 리스크 관리를 불능화하는 새로운 거버넌스 리스크를 지적한다. 위 사례의 전거는 松尾剛行, 「プラットフォーム事業者によるアカウント凍結等に對する私法上の救濟について」, 情報法制研究 第10号 (2021), 66–78頁이다. 同論文이 소개하는 사안은, 국제적 동영상 공유 플랫폼 사업자가 스팸 판정 AI의 오판정에 기초하여 Vtuber인 이용자에게 스팸 행위를 하였다는 통고를 하였고, 이용자의 대리인이 재판 외 이의절차를 통하여 여러 차례 재검토를 요구하였음에도 사업자가 인간의 실질적 개입 없이 AI의 판정을 반복하였으므로, 결국 이용자가 도쿄지방재판소에 제소하기에 이른 사건이다(同論文 69–72頁). 제소 직후 사업자가 스팸 인정을 철회함으로써 사건이 종결되었다는 점에서, 同論文은 AI 판정에 대한 인간의 개입을 확보하는 절차의 중요성을 지적한다(同論文 73頁, 76頁). 이러한 'AI 리스크의 이중성'—AI가 내부통제 수단인 동시에 감독 해태를 야기할 수 있는 리스크 원천—은 이사회가 AI 도입과 AI 감독을 분리하여 이중적으로 의무를 부담함을 시사한다. 松尾 역시 AI의 판정에 대하여 이용자가 이의를 제기한 단계에서는 인간이 실질적으로 개입하는 이른바 '켄타우로스 모델'에 따른 인간과 AI의 협업이 개시되어야 한다고 지적하는바, 이는 아래 Ⅲ. 4.에서 제시하는 절차적 안전항의 요건, 특히 경고신호가 보고된 경우 지체 없이 외부 검증을 의뢰하여야 한다는 요청과 직접 연결된다.

가 확인하고서도 법정 사업자의 책무이행 여부를 감독하지 않는 경우, 이는 Marchand 법리상 'Mission Critical Risk'에 관하여 실질적 모니터링을 해태한 것과 기능적으로 동일하게 평가될 수 있다. 나아가 동법 제31조(투명성 확보 의무)·제32조(안전성 확보 의무)는 이사의 선관주의의무 위반 여부를 판단하는 외부 준거 기준으로서, 법원이 이사의 AI 감독 소홀을 심사할 때 참조할 수 있는 객관적 행위 기준이 된다는 점에서 중요하다.[38]

이러한 분석은 Delaware 법원의 최근 동향과 대비하면 그 의의가 더욱 선명해진다. Delaware 형평법원은 **2022년 Construction Industry Laborers Pension Fund v. Bingle** 판결(이하 'Bingle 판결')에서, 사이버보안이 온라인 소프트웨어 회사(SolarWinds)에게 핵심 임무(Mission Critical)에 해당하는 리스크임을 인정하면서도, 사이버보안을 규율하는 실정법상 의무가 충분히 정립되어 있지 않다는 이유로 해당 리스크를 '컴플라이언스 리스크가 아닌 비즈니스 리스크'로 분류하고 케어마크 청구를 기각하였다.[39] 이 판결은 핵심 임무에 해당하는 위험이라 하더라도 이사의 감시의무 위반이 성립하려면 그 위험을 구체적으로 규율하는 실정법상 의무의 존재가 선행되어야 함을 시사한다. AI 리스크가 Bingle 판결에서 사이버보안 리스크가 처했던 것과 동일한 논리적 위치에 놓일 수 있다는 점에서, 한국법의 대응은 주목된다. 인공지능기본법 제33조부터 제35조의 고영향 AI 조항은 채용·금융·의료 등 핵심 영역에서 AI 리스크를 단순한 비즈니스 위험의 영역에서 실정법상 의무가 수반되는 컴플라이언스 리스크의 영역으로 전환시킬 수 있는 제도적 기반을 마련함으로써, Bingle 판결이 드러낸 간극을 실정법 차원에서 메우는 기능을 수행한다. 즉 이 조항들이 적용되는 고영향 AI 시스템을 운용하는 회사의 이사회가 해당 법정 책무의 이행 여부를 감독하지 않는 경우, 그 해태는 Delaware 법원이라면 비즈니스 리스크 감시 소홀에 불과하다고 판단할 수 있는 사안을 한국 법원이 컴

38) 인공지능기본법 제33조~제35조의 고영향 AI 체계가 이사의 감독의무와 결합하는 구조에 관하여는 성승제·정규, "인공지능과 회사법의 법정책에 대한 검토 — 인공지능과 법인 이사 검토 등을 중심으로 —", 「법과 정책연구」 제25집 제4호, 한국법정책학회, 227~270면 (2025) [이하 "성승제·정규"], 244~245면. 비교법적으로, 일본이 2025년 6월 제정·시행한 인공지능기본법은 강행 규제보다 기본법 방식을 채택하면서 이노베이션 촉진을 우선 목표로 설정하였고, 이 점에서 EU AI법(위험 기반 사전 규제)과 대비되는 절충적 모델로 평가된다(김유정, "일본 AI법(人工知能關連技術の研究開發及び活用の推進に關する法律) 제정 과정에 관한 연구", 「日本學報」 第146輯, 313~315면 (2026)).

39) Construction Industry Laborers Pension Fund on behalf of SolarWinds Corporation v. Mike Bingle, et al., 2022 WL 4102492 (Del. Ch. Sept. 6, 2022). 이 판결에서 법원은 온라인 소프트웨어 회사에게 사이버보안이 "핵심 임무"에 해당하는 리스크임을 인정하면서도, 2018년 미국 증권거래위원회(SEC) 해석지침이나 뉴욕증권거래소(NYSE) 사이버보안 안내문은 실정법(positive law)이 아니라는 이유로, 이사들의 감시 소홀을 컴플라이언스 리스크 관련 케어마크 기준 위반이 아닌 비즈니스 리스크 평가의 실패로 분류하였다. 이로써 Bingle 판결은 핵심 임무 개념이 케어마크 기준을 자동으로 활성화하지 않으며, 그 전제로 이사의 행위를 규율하는 실정법상 의무의 존재가 필요함을 확인하였다. 류지민, 전게논문, 388~397면 참조. Bingle 판결과 같은 해 선고된 Hamrock 판결(City of Detroit Police & Fire Retirement System, Inc. v. Hamrock, 2022 WL 2387653 (Del. Ch. Jun. 30, 2022))은 반대로, 회사가 핵심 컴플라이언스 리스크를 담당하는 실질적 이사회내 위원회를 갖추고 운영하였음이 문서로 확인된 경우에는 핵심 임무에 해당하는 리스크(가스공급관 안전)에 관한 케어마크 청구도 기각될 수 있음을 보여주어, 절차적 증거의 중요성을 재확인한다.

플라이언스 리스크에 관한 감시의무 위반으로 인정할 수 있는 규범적 근거를 제공한다. 이러한 구조는 이미 금융규제 영역에서 선례를 찾을 수 있다. 「금융회사의 지배구조에 관한 법률」 제24조는 내부통제기준의 '마련의무'를 규정하고 있으나, 이를 기준의 제정 의무로만 좁게 해석하면 기준이 실질적으로 준수되지 아니하여 발생한 사고에 대하여 임원의 관리책임을 묻기 어렵다는 문제가 지적되어 왔다. 후술하는 바와 같이(Ⅴ) 우리은행 DLF 불완전판매 사건의 행정소송은 '마련의무'와 '준수의무'의 구별이 갖는 실천적 함의를 잘 보여준다. AI 내부통제시스템에서도 동일한 문제가 반복될 수 있으므로, 인공지능기본법상 책무를 형식적으로 이행하는 것만으로는 이사의 감시의무가 충족되지 아니한다.

이 법리의 실천적 함의는 최근 미국에서 제기된 알고리즘 담합(Algorithmic Collusion) 사례에서 극명하게 드러난다. AI 알고리즘이 기업의 핵심 가격 결정 기제로 자리잡으면서, 명시적 합의 없이도 경쟁사 간 가격이 동조화되는 새로운 형태의 경쟁법 위반이 현실화되고 있다. 테네시 중부 연방지방법원의 **In re RealPage, Inc.** 사건은 그 대표적 사례이다. 피고 RealPage사는 임대용 부동산 관리 소프트웨어인 'YieldStar'를 다수의 임대인에게 제공하면서, 각 임대인의 실제 임대료·공실률 등 비공개 경쟁 데이터를 중앙 서버에 집적(Pooling)하고 이를 학습한 알고리즘이 산출한 '최적 권장 가격'을 각 임대인에게 제시하는 허브 앤 스포크(Hub-and-Spoke) 구조를 구축하였다.[40] 미국 법무부(DOJ)는 임대인들이 가격 결정 권한을 공통의 알고리즘(Hub)에 위임하고 경쟁사들도 동일한 알고리즘을 사용한다는 점을 인식하면서 행동하였다면, 이는 셔먼법(Sherman Act) 제1조가 금지하는 수평적 합의를 대체하는 것이라고 논증하였다.[41]

이 사건은 상법적 맥락에서도 중대한 시사점을 제공한다. 알고리즘 담합은 일회성 일탈이 아닌 구조적 위법 행위이므로, 이사의 내부통제시스템 구축 의무는 단순한 이상 징후 포착을 넘어 알고리즘 자체의 설계 단계에서부터 경쟁법 준수 여부를 검증하는 사전적 통제로 확장된다. 특히 가격 결정이 기업의 '핵심 임무(Mission Critical)'에 해당하는 이상, 이사회는 AI 가격 알고리즘이 경쟁사 데이터를 공유하거나 가격 결정권을 제3의 플랫폼에 위임하는 구조를 취하지 않도록 능동적으로 감시하여야 한다.[42]

---

40) In re RealPage, Inc., Rental Software Antitrust Litig., Case No. 3:23-md-03071 (M.D. Tenn.). 사건의 개요 및 YieldStar의 허브 앤 스포크 구조에 관하여는 강선희, "알고리즘 가격결정과 상법상 이사의 감시의무 – 미국 RealPage 사건과 한국 대법원 내부통제시스템 법리를 중심으로", 「법학논문집」 제49집 제3호, 중앙대학교 법학연구원, 110~111면 (2025) [이하 "강선희"] 참조. 알고리즘 담합의 유형론 및 허브 앤 스포크 구조의 경쟁법적 분석에 관하여는 OECD, Algorithms and Collusion: Competition Policy in the Digital Age (2017), p. 22; Ariel Ezrachi & Maurice E. Stucke, Virtual Competition: The Promise and Perils of the Algorithm-Driven Economy, Harvard University Press (2016), pp. 45–50 참조.

41) U.S. Department of Justice, Statement of Interest in In re RealPage, Inc., Case No. 3:23-md-03071 (M.D. Tenn. Nov. 15, 2023); 강선희, 전게논문, 111~112면 참조.

42) 강선희, 전게논문, 120~121면.

3. 데이터 거버넌스와 이사의 감시의무

Caremark 법리가 요구하는 합리적 모니터링 시스템의 구체적 내용은 AI 시대에 있어 '데이터 거버넌스(data governance)'의 수립·유지 의무로 구체화된다. 회사의 AI 시스템이 산출하는 판단의 질은 학습에 사용된 데이터의 적정성에 결정적으로 의존하기 때문이다.[43]

데이터 거버넌스의 수립·유지에 관한 일차적 의무는 CEO를 포함한 경영진에게 귀속되나, 그 적정성에 대한 감독은 이사회의 감독의무 범위에 포함된다.[44] 나아가 데이터 거버넌스는 관련 기술 및 데이터에 관한 전문적 지식을 요하므로, 이를 갖춘 사외이사로 구성된 이사회 하부위원회를 통해 실질적 감독 체계를 갖추는 것이 바람직하다. 이 맥락에서 사외이사는 AI 내부통제시스템을 직접 구축할 수는 없으나, 이사회 또는 대표이사에게 AI Red-team·기술심의기구의 설치를 '촉구'하는 최소한의 의무를 부담한다고 해석함이 타당하다. 사외이사가 이러한 촉구 조치를 취하였음에도 경영진이 응하지 않은 경우에는 사외이사의 감시의무 위반을 부정함이 상당하다.[45]

특히 AI 내부통제 시스템의 재설계와 관련하여 주목해야 할 점은, 전통적 회사법이 경영진의 이해상충이나 성실성 유무에 집중하는 것과 달리, AI 환경에서는 알고리즘의 오류 또는 실패가 주된 문제로 부상한다는 것이다.[46] 이는 이사의 감시의무가 사람(경영진)에 대한 감시에서 시스템(AI 알고리즘·데이터)에 대한 감시로 그 성격이 질적으로 전환됨을 의미한다. 데이터 편향성 문제는 이제 단순한 기술적 사안이 아니라, 이사의 신인의무 위반 여부와 직결되는 법적 사안으로 다루어져야 한다.

4. 경고신호(Red Flags)와 고의적 무지의 법리

Caremark·Stone v. Ritter 법리의 핵심은 경고신호에 대한 이사의 반응 의무이다. AI 시스템이 비정상적인 작동 패턴을 보이거나, 예측 불가능한 방식으로 주주 이익을 침해하거나, 사이버보안 취약성 경보를 발생시키는 경우, 이는 이사가 반드시 주목해야 할 경고신호에 해당한다. 이사가 기술적 난해함(technical complexity)을 핑계로 이러한 신호를 방치하는 것은 **Stone**

---

43) NIST AI RMF, supra note 34; 송옥렬(2021), 전게논문, 100~101면.

44) 송옥렬(2021), 전게논문, 101면; Marchand, 212 A.3d at 824.

45) 사외이사가 내부통제시스템 구축의무에서 자동적으로 면책되지 않는다는 점에 관하여, 송옥렬(2022)은 사외이사가 회사의 내부조직을 통해 내부통제시스템을 직접 구축할 수는 없으나 대표이사 또는 이사회에 필요한 시스템 구축을 "촉구"하는 형태의 의무는 부담한다고 정리한다. 사외이사가 이를 촉구하였음에도 대표이사 등이 아무 조치도 취하지 않았다면 사외이사는 감시의무 위반으로 인한 손해배상책임을 지지 않는다는 논리이다. 송옥렬(2022), 전게논문, 35~36면 참조. AI Red-team 및 독립적 기술심의기구의 설치를 "촉구"하는 것이 AI 환경에서 사외이사 감시의무의 최소 이행 단위가 된다는 본고의 논증은 이 법리 구조와 정합적이다.

46) 임철현(2025), 전게논문, 348~352면; 송옥렬(2021), 전게논문, 103면.

v. Ritter가 경계한 '의식적 방치(conscious disregard)'로 평가될 수 있다.[47]

나아가 이사가 핵심적 의사결정을 AI에 전적으로 위임하는 것은 경영의 핵심에 놓인 의무를 위임할 수 없다는 위임금지의 법리(Grimes v. Donald 법리)에도 반한다. AI는 기계학습을 통해 스스로 변화하므로, 이사가 처음 선택한 알고리즘과 실제로 가동 중인 알고리즘이 동일하다고 볼 수 없어 위임의 통제 요건이 충족되지 않는다. 따라서 이사가 AI의 결론에 탐문 없이 복종하는 것－이른바 '알고리즘 위험에 대한 의식적 방치'－은 성실의무상 '불성실(bad faith)'에 해당하며, 경영판단원칙에 의한 보호와 면책 규정 모두의 적용을 배제한다는 점을 명심하여야 한다.[48].

특히 주목해야 할 것은 '고의적 무지(Willful Ignorance)' 또는 '눈 감기(Ostrich Problem)'의 문제이다. AI의 기술적 복잡성을 이유로 이사가 스스로 AI 관련 정보에 접근하지 않거나, 이를 이해하려는 노력을 기울이지 않는 경우, 법원은 이를 '의식적 방치(Conscious Disregard)'와 동일하게 평가할 가능성이 높다.[49] 법적 의무는 이사에게 AI 전문가 수준의 기술 지식을 요구하지 않으나, 합리적인 이사라면 AI 시스템의 기본 작동 원리와 주요 위험 요소를 이해하려는 성실한 노력을 기울여야 한다는 것이 현대 기업지배구조 법학의 일반적 입장이다. 이사의 감독책임은 단순한 실수나 오판에 대한 책임이 아니라, 위험을 인식하고도 감독과정을 작동시키지 않은 경우에 한하여 인정되는 특별한 책임이다. 따라서 AI 시대 내부통제는 단순한 규정준수 체계로는 충분하지 않고, 오히려 질문하고 검증하고 학습하는 동적 감독체계(dynamic oversight system)로 발전할 필요가 있다

---

47) Stone v. Ritter, 911 A.2d at 370; In re Caremark, 698 A.2d at 971. 경고신호 법리의 발전에 관하여는 Leo E. Strine, Jr. et al., Loyalty's Core Demand: The Defining Role of Good Faith in Corporation Law, 98 Geo. L.J. 629, 652–58 (2010) 참조. AI 거버넌스 실패를 ① '절차적 방치(procedural neglect)', ② '인식론적 포기(epistemic abdication)', ③ '구조적 공모(structural complicity)'의 세 유형으로 분류하고, 부작위(omission)가 이 세 유형의 연결고리이자 책임 귀속의 핵심 경로임을 논증하는 문헌으로는 Montagnani & Passador, supra note 2, at 24–27 참조. 이 분류체계는 Delaware 법원의 경고신호 및 고의적 무지 법리의 기능적 연장선에 있다. AI에 핵심 의사결정을 위임하는 것이 Grimes v. Donald, 673 A.2d 1207, 1214 (Del. 1996)의 위임금지 법리에 반하며, 기계학습으로 인해 위임된 알고리즘과 가동 중인 알고리즘이 상이해지면 통제 요건도 충족되지 않는다는 점에 관하여는 Cowger, supra note 2, at 181–185 참조.

48) Grimes, 673 A.2d at 1214; Cowger, supra note 2, at 181–185 참조. 김정호(2022)는 good faith를 주의의무·충실의무와 구별되는 제3의 독립적 신인의무인 '성실의무'로 체계화할 것을 제안하며(김정호(2022), 전게논문, 181면 이하), 영미법상 good faith를 한국 민법의 '선의(무지)'와 혼동하지 않도록 '성실'로 번역할 것을 권고한다(동 논문, 194면). 본고는 이 제안을 수용하여 bad faith를 '불성실'로, good faith를 '성실'로 번역한다.

49) '고의적 무지(Willful Ignorance)' 법리에 관하여는 United States v. Jewell, 532 F.2d 697, 700 (9th Cir. 1976); 기업법 맥락에서의 적용에 관하여는 Leo E. Strine, Jr., The Dangers of Denial: The Need for a Clear-Eyed Understanding of the Power and Accountability Structure Established by the Delaware General Corporation Law, 50 Wake Forest L. Rev. 761, 775–82 (2015) 참조. 이사가 기존의 법리를 유지하면서도 AI에 대해 이해 노력을 기울여야 하는 논거로는 송옥렬(2021), 전게논문, 99~100면 (Zetzsche 교수의 세 가지 논거, 즉 ① AI 수준의 불완전성, ② 경쟁전략상 이유, ③ 위험 인수의 소극성 방지를 소개하면서 이사의 재량범위를 분석); 임철현(2025), 전게논문, 358~365면; 최문희, 전게논문, 26~27면 참조.

한국 법의 맥락에서, 개정 상법 제382조의3과 제382조 제2항을 결합하여 보면, 이사가 AI 시스템에서 발생하는 경고신호를 합리적으로 인식하였거나 인식할 수 있었음에도 이를 무시한 경우, 회사 및 주주에 대한 선관주의의무 및 충실의무 위반으로 손해배상책임이 인정될 수 있다. 이는 Delaware 법리의 Caremark 책임 구조와 기능적으로 대응하는 것으로서, 양 법체계의 수렴적 발전 방향을 보여준다.

한국 대법원 판례의 흐름 역시 이러한 방향으로 수렴하고 있다. 이미 1985년에 대법원은 업무집행을 담당하지 아니하는 평이사도 다른 이사의 업무집행이 위법하다고 의심할 만한 사유가 있음에도 이를 방치한 경우 감시의무 위반의 책임을 진다고 판시하여, 한국 이사 감시의무 법리의 출발점을 이루었다(대법원 1985. 6. 25. 선고 84다카1954 판결).[50] 이러한 토대 위에서, 2008년 대우그룹 분식회계 사건(대법원 2008. 9. 11. 선고 2006다68636 판결)에서 대법원은 이사가 "합리적인 정보 및 보고시스템과 내부통제시스템을 구축하고 그것이 제대로 작동하도록 배려할 의무"를 부담한다고 최초로 명시하였다. 그 후 대법원은 이 내부통제시스템 구축의무를 가격담합 사건에 적용하여, 대표이사가 담합을 직접 지시하거나 알지 못하였더라도 담합 위험이 높은 업종에서 이를 방지할 내부통제시스템을 실질적으로 구축하지 아니한 채 장기간에 걸친 조직적 담합을 방치한 경우 감시의무 위반이 성립한다고 보아, 이를 부정한 원심을 파기환송하였다(유니온스틸 철강제품 가격담합 사건, 대법원 2021. 11. 11. 선고 2017다222368 판결).[51] 나아가 2022년 대우건설 사건(대법원 2022. 5. 12. 선고 2021다279347 판결)은 이 감시의무를 대표이사·업무집행이사를 넘어 사외이사를 포함한 모든 이사에게로 확장하여, 내부통제시스템이 구축되어 있지 않거나 제대로 운영되지 않는다고 의심할 만한 사유가 있음에도 이를 외면·방치한 경우 사외이사도 감시의무 위반의 책임을 진다고 판시하였다.[52] 같은 해 선고된 STX조선해양 분식회계 사건(대법원 2022. 7. 28. 선고 2019다202146·202153 판결)은 한 걸음 더 나아가, "특정 제도나 직위가 회사에 도입된 것만으로 합리적인 내부통제시스템의 구축·운영을 긍정할 수는 없고, 그 제도나 직위의 내용과 실질적 운영 여부 등을 살펴 판단하여야 한다"

50) 대법원 1985. 6. 25. 선고 84다카1954 판결. 동 판결은 업무집행을 담당하지 아니하는 평이사(平理事)도 다른 이사의 업무집행이 위법하다고 의심할 만한 사유가 있음에도 감시의무를 위반하여 이를 방치한 때에는 손해배상책임을 진다고 보아, 후속 내부통제시스템 법리(2006다68636 판결 이하)의 효시를 이룬다.

51) 대법원 2021. 11. 11. 선고 2017다222368 판결(유니온스틸 철강제품 가격담합 주주대표소송). 동 판결은 이사의 내부통제시스템 구축의무를 가격담합(공정거래법 위반) 사안에 적용하여 대표이사의 감시의무 위반을 인정한 최초의 대법원 판결로서, 위 2006다68636 판결의 법리를 담합 영역으로 확장하였다. 대법원은 대표이사가 담합행위를 직접 지시하거나 알지 못하였더라도, 철강산업의 특성상 담합 위험이 높았음에도 이를 방지할 내부통제시스템을 갖추지 아니한 채 장기간·다수 품목에 걸친 조직적 담합을 방치하였고, 피고가 내세운 내부회계관리제도·윤리규범 등이 실질적인 위법행위 통제장치로 기능하지 못하였다는 이유로, 감시의무 위반을 부정한 원심을 파기환송하였다(공정거래위원회 부과 과징금 합계 약 320억 원). 사실관계 및 법리 분석과 알고리즘 담합에의 함의에 관하여는 강선희, 전게논문, 117~119면 참조.

52) 대법원 2008. 9. 11. 선고 2006다68636 판결(대우그룹 분식회계 이사 감시의무 사건) 및 대법원 2022. 5. 12. 선고 2021다279347 판결(대우건설 사건 – 대표이사·업무집행이사를 넘어 사외이사를 포함한 모든 이사의 감시의무 위반을 인정한 최초의 대법원 판결).

고 판시하여, 내부통제시스템의 형식적 구비가 아니라 그 실질적 운영을 책임 판단의 기준으로 삼았다. 이는 본고가 강조하는 '내부통제시스템의 실질화' 요청을 정면으로 뒷받침하며, Marchand 판결이 요구한 실질적 모니터링 시스템 기준과 기능적으로 수렴한다.[53] 이러한 흐름은 2025년에도 재확인되었다. 대법원은 가격담합 관련 주주대표소송에서 위 감시의무·내부통제시스템 법리를 다시 확인하는 한편, 이사의 법령위반으로 회사가 위법한 이익을 얻었더라도 그 이익을 손해배상액에서 공제할 수 없다고 판시하였다(대법원 2025. 6. 26. 선고 2024다316391 판결).[54]

이 판결의 법리는 AI 알고리즘 담합 사건에 그대로 적용될 수 있다. 특히 위 유니온스틸 판결(2017다222368)이 정립한 담합 내부통제시스템 법리는, 명시적 합의 없이 알고리즘을 매개로 경쟁사 간 가격이 동조화되는 알고리즘 담합 사안에서도 동일하게 작동한다. AI의 블랙박스적 특성으로 인해 이사가 알고리즘의 구체적 작동 메커니즘을 사후적으로 '인지'하는 것은 기술적으로 불가능에 가깝다. 그러나 대법원이 대우건설 사건에서 '조직의 복잡성'을 면책 사유가 아닌 시스템 구축의 필요성 근거로 삼았듯이, AI의 기술적 복잡성 역시 이사 면책의 방패가 아닌 더욱 고도화된 알고리즘 준법감시 시스템(Algorithmic Compliance System)을 구축해야 할 이유로 해석되어야 한다. 또한 이 의무는 결과 발생을 100% 차단하라는 결과채무(Obligation de résultat)가 아니라, 리스크를 최소화할 합리적 조치를 다하라는 수단채무(Obligation de moyens)의 성격을 지닌다는 점도 분명히 해 두어야 한다.[55]

다만 이 의무가 수단채무의 성격을 지닌다는 점은 경영판단원칙과의 관계에서도 중요한 함의를 가진다. Delaware 판례법은 Caremark 기준이 적용되는 감시의무 사안에서도 경영판단원

---

53) 전게 STX조선해양 분식회계 사건 판결(대법원 2022. 7. 28. 선고 2019다202146·202153). 동 판결은 자본시장법 제162조에 따른 대표이사의 손해배상책임에서 '상당한 주의' 항변과 관련하여, "특정 제도나 직위가 회사에 도입된 것만으로 합리적인 내부통제시스템의 구축·운영을 긍정할 수는 없고, 그 제도나 직위의 내용과 실질적 운영 여부 등을 살펴 판단하여야 한다"고 판시하였다. 이는 내부통제시스템의 형식적 구비가 아니라 그 실질적 운영을 책임 판단의 기준으로 삼은 것으로서, 본고가 강조하는 내부통제시스템의 실질화 요청 및 Marchand 판결의 실질적 모니터링 요구와 기능적으로 수렴한다.

54) 대법원 2025. 6. 26. 선고 2024다316391 판결. 동 판결은 가격담합 관련 주주대표소송에서 위 감시의무·내부통제시스템 법리(2017다222368, 2021다279347, 2019다202146 판결 등)를 재확인하는 한편, 이사의 법령위반으로 회사가 위법한 이익을 얻었더라도 그 이익을 손해배상액에서 공제할 수 없다고 판시하였다. 위법한 이익의 유보를 허용하는 결과가 되어 손해배상제도의 본질적 기능에 반하기 때문이다.

55) 내부통제시스템 구축 의무가 수단채무(Obligation de moyens)의 성격을 지닌다는 점에 관하여는 강선희, 전게논문, 122면 (리스크 최소화를 위한 합리적 조치의무를 수단채무로 파악하고, 조기 발견 시 감면받았을 과징금과 실제 부과된 과징금의 차액을 손해로 보는 차액설을 제안) 참조. 나아가 비교법적 시각에서, 유럽연합의 2024년 제조물책임 지침(Directive 2024/2853)이 소프트웨어·AI를 '제품'에 명시 편입하고 AI 자율 학습 이후 결함에 대해서도 책임을 귀속시키는 구조를 채택한 것과, 프랑스에서 의사가 AI를 활용한 경우 알고리즘 결과물은 의사의 지적 판단을 거쳐야 '의료행위'로 완성된다는 점에서 의사의 AI 오류에 대한 면책이 허용되지 않으며 이는 여전히 수단채무 영역에서 전문가 주의의무로 처리되어야 한다는 프랑스 학설의 입장에 관하여는 이상헌, "인공지능으로 인한 침해에 대한 민사책임 – 유럽연합과 프랑스에서의 논의를 중심으로 –", 「저스티스」 통권 제213호, 한국법학원, 154~183면 (2026) [이하 "이상헌"], 170~172면 참조.

칙이 여전히 작동할 수 있는 이중적 구조를 정립해왔다. 구체적으로, 내부통제시스템이 전혀 구축되지 않았거나 의도적으로 운영을 방기한 경우에는 이사의 성실의무상 불성실(bad faith)이 인정되어 경영판단원칙의 보호 범위에서 벗어나게 된다. 그러나 반대로 합리적인 내부통제시스템이 구축되어 실질적으로 운영되고 있는 상황에서, 이사회가 예견하지 못한 사업상 위험으로 인해 회사에 손해가 발생하더라도, 그 손해는 단순히 경영상 위험 변수의 실현으로 평가될 뿐 경영판단원칙에 의해 사법심사가 자제될 수 있다. Delaware 형평법 법원은 Citigroup 판결(2009)에서 이사의 감시의무 위반 주장이 실질적으로는 사업상 위험 인식 실패나 결과적 책임 추궁에 불과한 경우 – 즉, 이사가 충분한 정보에 기초하여 성실하게 경영판단을 내렸고 실효적 내부통제시스템이 운영되고 있었던 경우 – 에는 경영판단원칙이 사법적 개입을 차단한다고 명확히 선언하였다.[56] 나아가 ProAssurance 판결(2023)은 사업상 위험에 대한 평가와 전략적 결정은 이사회의 '핵심적 기능(the quintessential board function)'으로서 경영판단원칙의 보호 대상이며, Caremark 기준에 따른 감시의무의 범위는 이러한 경영판단 영역에까지 확장되지 않는다는 점을 재확인하였다.[57] 또한 Clem 판결(2024)은 정부 조사 개시나 대규모 합의금 지급 등의 사정만으로 Caremark 기준에 따른 감시의무 위반이 곧바로 성립하지는 않으며, 그러한 사정만을 근거로 한 소송 제기는 오히려 경영판단원칙이 보호하고자 하는 경영상 자율성을 침해하고 회사 자원을 소모하는 이중의 폐해를 초래할 수 있음을 경고하였다.[58]

AI 환경에서 이러한 법리 구조는 다음과 같이 적용된다. 이사회가 AI 내부통제시스템 – AI 리스크 위원회, 알고리즘 감사 체계, XAI 보고서 기반 이사회 심의 – 을 실질적으로 구축하고 운영해왔다면, AI 알고리즘이 예상치 못한 방식으로 오작동하여 회사에 손해가 발생하더라도, 이는 경영판단원칙에 의해 보호되는 영역에 놓일 수 있다. 반면 그러한 시스템이 형식적으로만 존재하거나 이사회가 AI 경고신호를 의식적으로 방치한 경우에는 Caremark 기준상 불성실(bad faith)이 인정되어 경영판단원칙의 보호 밖에 놓이게 된다. 이 이중 구조는 곧, AI 내부통제시스템의 구축과 실질적 운영이 단순히 이사의 의무 이행 증거에 그치는 것이 아니라, 장래 이사 책임 추궁 시 경영판단원칙을 원용할 수 있는 전제 조건으로 기능한다는 점을 시사한다.[59]

---

56) *In re Citigroup Inc. S'holder Derivative Litig.*, 964 A.2d 106, 124–130 (Del. Ch. 2009)(감시의무 위반 주장이 실질적으로는 사업상 위험에 관한 경영판단의 사후적 평가에 불과한 경우 경영판단원칙이 사법심사를 차단한다고 선언; 사업상 위험 평가와 수익-위험 간 경영상 재량 판단은 이사의 재량 영역임을 확인).

57) *In re ProAssurance Corp. S'holder Derivative Litig.*, 2023 WL 6426294, at *13–15 (Del. Ch. Oct. 2, 2023)(사업상 위험 평가는 이사회의 '핵심적 기능(the quintessential board function)'으로서 경영판단원칙의 보호 대상이며, Caremark 기준에 따른 감시의무 범위는 이러한 경영판단 영역에 확장되지 않는다고 재확인; 합법적 전략적 판단이 사후에 손실을 낳았어도 성실의무 위반이 아님).

58) *Clem v. Skinner*, 2024 WL 668523, at *11 (Del. Ch. Feb. 19, 2024).

59) 문정해, "내부통제시스템에 관한 이사의 감시의무와 경영판단원칙 – 미국 델라웨어주 판례법상 Caremark 기준과 경영판단원칙의 관계를 중심으로 –", 「상사판례연구」 제38권 제2호, 한국상사판례학회, 107~158면(2025) [이하 "문정해"], 152~153면.

이 이중 구조의 실천적 함의는 이사의 책임 면책 기준과 직결된다. 예를 들면, 레그테크·섭테크 도구를 내부통제시스템에 도입하여 최선의 기술을 적용하고 관련 운영 매뉴얼을 갖추어 이를 준수하고 있었다면, 다른 이사 또는 임직원의 불법행위로 회사에 손해가 발생하더라도 이사의 감시·감독의무 위반으로 인한 책임은 면책될 가능성이 높다. 반면 레그테크·섭테크 도구를 형식적으로 도입하였더라도 위험요소 발견을 위한 모니터링이 실질적으로 이루어지지 않고 알고리즘에 대한 주기적 검증과 수정 절차가 결여되어 있었다면, 이사의 감시의무를 제대로 이행하였다고 보기 어렵다.[60] 이 점에서 AI 기반 내부통제시스템의 구축은 단순한 기술 도입에 그치지 않고, 그 실질적 운영과 주기적 검증까지 포함하는 계속적 의무로 파악하여야 한다.

이와 관련하여 비교법적으로 주목할 것은, 독일 주식법이 내부통제시스템의 구축 정도를 '필수핵심영역'과 '비필수영역'으로 이분하고 의무위반의 주관적 구성요건도 이원화하는 구조를 취하고 있다는 점이다. 독일 주식법 제91조 제2항(필수핵심영역)에서는 비용·효용 형량이 허용되지 않으며 경과실을 포함한 고의·과실의 전 범위에 걸쳐 책임이 인정되는 반면, 제91조 제3항(비필수영역)에서는 상당성(Angemessenheit) 원칙에 따른 비용·효용 형량이 허용된다. 나아가 한국의 대우건설 판결이 사내이사와 사외이사의 책임발생요건을 별도로 설시한 것은 독일 법리의 방향과도 일치한다. 구체적으로, 대우건설 판결은 업무집행을 담당하는 사내이사에 대해서는 내부통제시스템 미구축과 부실운영 모두에 걸쳐 고의·과실의 전 범위에서 책임을 인정한 반면, 업무집행을 담당하지 않는 사외이사에 대해서는 ① 내부통제시스템이 전혀 구축되지 않은 경우에는 그 구축을 '촉구'하지 않은 부작위(과실 포함), ② 시스템이 구축되었더라도 제대로 운영되지 않는다고 의심할 만한 사유가 있는데도 이를 '외면하고 방치'한 고의적 부작위로 책임 발생요건을 이원화하였다. 이는 독일 법리에서 사내이사 성격의 경영이사회(Vorstand)에게는 경과실을 포함한 전면적 책임을 부과하되, 사외이사 성격의 감사회 또는 감독이사회(Aufsichtsrat)에 대해서는 상대적으로 좁은 요건 하에 예외적으로만 책임을 인정하는 이원적 구조와 기능적으로 수렴한다.[61]

---

60) 김민정, 전게논문, 29~30면.

61) 김정호(2022), 전게논문, 191~196면. 동 문헌은 Delaware 판례법이 이사회 구성원의 절대다수가 사외이사인 미국적 특수 상황(monitoring board, director primacy)을 반영하여 불성실(bad faith)이라는 고도의 주관적 구성요건을 요구하는 반면, 독일과 일본은 고의·과실의 전 범위에 걸쳐 책임을 인정하고 대신 책임제한의 탄력적 법리를 가동한다고 비교법적으로 분석하면서, 한국은 사외이사 비율이 약 60%로 미국과 일본의 중간에 있으므로 사내이사는 고의·과실 전범위로, 사외이사는 고의 또는 별도의 객관적 요건으로 이원화하는 것이 타당하다고 논한다. 또한 내부통제시스템기준이 경영판단기준과 마찬가지로 원칙적으로 이사의 면책을 돕는 사법심사기준임을 확인하는 문헌으로도 동 논문, 191면 참조. 나아가 미국 델라웨어주 회사법상 신인의무의 심사기준 체계(경영판단원칙은 사실상 책임면제 기능, 완전한 공정성 기준은 피고 입증책임의 엄격 심사)와 한국에서의 경영판단원칙(미국과 달리 이사의 임무해태 여부를 실질적으로 판단하는 기준으로 기능하여 경영판단원칙이 적용된 사안에서도 책임이 인정되는 경우가 있음)의 비교에 관하여는 박준선, "미국 회사법상 경영판단의 원칙과 완전한 공정성 기준의 적용 법리 – 미국 델라웨어주를 중심으로 –", 「사법」 제75호(2026)

이와 관련하여 AI 활용과 이사의 선관주의의무 문제를 다음과 같이 세 가지 유형으로 정밀하게 분석하는 견해가 있다.[62]

① AI를 활용하지 않는 경우: 비용·여건을 고려한 불활용이 곧 의무위반은 아니나, AI가 보편화된 환경에서 만연한 불활용은 경영판단원칙의 보호를 받기 어렵다. 이때 인간 이사에게 AI와 동등한 정보처리 능력을 요구할 수는 없으므로 "인간 수준에서" 입수 가능한 정보를 수집·검토한 것으로 충분하다.

② AI의 조언에 따라 판단한 경우: 학습 데이터의 질과 블랙박스 문제로 인해 AI 판단의 투명성이 의무 준수의 핵심 판단 기준이 된다. 복수의 AI 또는 전문 감정인 의견을 병행 활용한 경우 경영판단원칙의 보호를 받기 수월하다.

③ AI의 조언과 다른 판단을 한 경우: 경제적 합리성 이외의 여러 사정을 고려한 이사의 독자적 판단이 AI 조언과 배치된다 하여 곧 의무위반은 아니나, 장차 AI가 일반화되면 합리적 이탈 근거를 제시해야 경영판단원칙의 보호를 받을 수 있다. 이러한 유형론은 본 논문이 제안하는 Caremark–XAI 연계 감독 모델의 이론적 토대를 제공한다.

이에 더하여, AI 시대 이사의 합리적 행동 기준으로는 ① AI 도구 평가 능력 보유, ② 장단점 인지, ③ 자동화 편향 회피, ④ 알고리즘 편향 식별, ⑤ 알고리즘 자기감사 기능의 이해, ⑥ 최종 결정권 인간 보유, ⑦ 기업 차원 AI 리스크 정책 수립이라는 7가지 요건이 제시되고 있으며, 이는 기업 내부통제시스템의 AI 거버넌스 매뉴얼에 반영되어야 할 구체적 실천 규범으로 기능한다.[63]

이러한 유형론은 이사의 책임 분화라는 측면에서도 중요한 실천적 함의를 지닌다. 이사가 AI를 활용하여 의사를 결정한 경우 그 결론에 따른 경우와 따르지 아니한 경우 각각에 대해 이사에게 유불리가 어떻게 달라지는지가 경영판단원칙의 틀 안에서 분석된다.[64] 이사가 AI의 제안을 따른 경우에는 이를 경영판단원칙의 요건 충족을 뒷받침하는 사정으로 활용할 수 있는 반면, 이를 따르지 아니한 경우에는 이사가 그 이탈의 정당한 사유를 제3의 전문가 의견이나

[이하 "박준선"], 123~125면, 140~142면 참조.

62) 최문희, 전게논문, 28~34면.

63) Cowger는 AI 시대 합리적 이사(reasonable corporate fiduciary)의 7가지 행동 기준으로 ① AI 도구 평가 능력 보유, ② AI의 장단점 인지, ③ 자동화 편향 회피, ④ 설계·데이터 편향 식별, ⑤ 알고리즘 자기감사 기능의 이해, ⑥ 최종 결정권을 인간이 보유, ⑦ 기업 차원의 AI 리스크 정책 수립을 제시하며, 이에 대응하는 기업의 내부 AI 활용 정책 7가지(리스크 분석, 성과지표, 수정 메커니즘, 버전관리, 책임기준, 인증, 인간검토절차)도 함께 제안한다. Cowger, supra note 2, at 203–206 참조.

64) 남궁주현, "인공지능(AI)과 회사법의 연결에 관한 소고 – 이사의 책임 부담 가능성을 중심으로", 「성균관법학」 제36권 제4호, 성균관대학교 법학연구원, 41~69면 (2024) [이하 "남궁주현"], 55~59면. 이사의 AI 활용 여부 및 결론 채택 여부에 따른 책임 분화 분석에 관하여는 Mosco, Gian Domenico, AI and Board Within Italian Corporate Law: Preliminary Notes, European Company Law 17(3) (2020), at 95–96; Mertens, Floris, The Use of Artificial Intelligence in Corporate Decision-Making at Board Level, Financial Law Institute Working Paper Series 2023-01 (2023), at 20–23도 함께 참조.

별도의 인공지능 검증을 통해 주장·증명해야 할 부담을 진다. 더 나아가 AI가 보편화된 환경에서 이를 활용하지 않은 이사는 선관주의의무 판단에서 불리한 위치에 놓일 수 있다. 이러한 분석은 AI 활용이 이사의 의무 이행을 기계적으로 담보하는 것이 아니라 책임 공방에서 하나의 '절차적 준거'로 기능함을 보여 준다는 점에서 주목할 만하다.

한편 이러한 논의는 비전문가인 이사에게 어느 정도의 숙지와 탐문이 요구되는지의 문제로 이어진다. 알고리즘의 기술적 내용을 스스로 검증할 수 없다는 사정만으로 이사의 책임이 당연히 면제되는 것은 아니지만, 반대로 모든 이사에게 알고리즘 자체에 대한 전문적 이해를 요구하는 것도 현실적이지 않다. 감시의무의 이행 여부는 이사가 알고리즘을 스스로 이해하였는지가 아니라, 이해할 수 있는 상태를 조직적으로 확보하였는지를 기준으로 판단하여야 한다. 이러한 관점에서 ① AI 리스크의 식별·평가 결과를 정기적으로 이사회에 보고하는 체계를 갖추고, ② 독립적 알고리즘 감사 또는 외부 전문가의 검증 보고서를 형식적으로 접수하는 데 그치지 아니하고 이사회가 실질적으로 심의하며, ③ 설명가능 인공지능(XAI) 기법에 기초한 설명을 요구하고 그 설명의 한계까지 의사록에 기록하며, ④ 이상 징후나 경고신호가 보고된 경우 지체 없이 외부 검증을 의뢰하는 등의 후속 조치를 취하였다면, 이사는 알고리즘 위험에 대한 의식적 방치의 책임을 면할 수 있다고 보아야 한다. 이는 산출 결과의 정확성을 담보하는 실체적 안전항이 아니라, 감독 절차의 충실성을 담보하는 절차적 안전항(procedural safe harbor)으로 기능한다.[65]

Ⅳ. 최근 미국 학계의 논의와 시사점

최근 미국 회사법학에서는 AI와 이사의 의무에 관한 일련의 중요한 연구들이 발표되어 이사 감시의무론의 지평을 크게 확장하고 있다. 이하에서는 그 중 이 연구와 직접 연관되는 논의들을 선별하여 검토한다.

첫째, AI가 기업 경영 전반을 대체할 경우의 회사법적 함의를 분석하면서, 기업 이사회가 '융합 이사회(fused board)'로 진화하고 이사·임원의 이분법적 구조가 소멸한 '융합 경영(fused management)' 단계에서는 현행 신인의무 체계가 더 이상 작동할 수 없다는 견해가 있다.[66] 이

65) 절차적 안전항의 구체적 요건과 그 사법심사 기준에 관한 상세한 검토는 후속 연구에서 다룬다.

66) Martin Petrin, Corporate Management in the Age of AI, 2019 Colum. Bus. L. Rev. 965, 970 – 971, 996 – 1002, 1007 – 1008, 1018 – 1022 (2019) [이하 "Petrin"]. '융합 이사회(fused board)' 및 '융합 경영(fused management)' 개념에 관하여는 Id. at 1002 – 1008; AI 경영 단계에서의 책임 귀속 세 유형(AE 책임(Artificial Entity Liability: AI를 독립된 법적 주체로 인정하여 AI 자체에 법적 책임을 귀속시키는 방안. 현행법에서는 인정되지 않으나 미래 입법론으로 제시된다), 책임 체계 폐지, 제조물책임 유추)에 관하여는 Id. at 1013 – 1018; AI의 복수 목표 동시 최적화 가능성에 관하여는 Id. at 1021 – 1022 참조.

연구는 나아가 AI 경영 단계에서의 책임 귀속을 (i) AI 법인체 자체를 피고로 하는 방안, (ii) 경영자 책임 체계 폐지, (iii) AI 소프트웨어 개발·제공사를 주된 책임 주체로 하는 제조물책임 유추 적용의 세 가지로 유형화하고, 제조물책임 체계로의 전환이 가장 유력하다고 예측하였다. 또한 AI는 인간 경영자와 달리 복수 목표(주주 이익 + 이해관계자 이익)를 동시에 최적화할 수 있어 주주 이익 극대화 원칙을 근본적으로 변화시킬 가능성을 열어놓는다는 점도 주목할 만하다. 이러한 전망은 '이사' 중심의 AI 감시의무론이 장기적으로 AI 소프트웨어 개발사에 대한 제조물책임론과 병존하는 이중 책임 구조로 발전해야 함을 시사한다.

둘째, AI 외부 규제의 네 가지 원칙(AI 무지 면책 불가·역사적 차별 방지·책임 소급 추급·기업 정치 자금 규제)을 제시하면서, AI 사용으로 피해가 발생한 경우 기업이 AI의 작동 방식을 이해하지 못했다는 사실은 책임의 면제 근거가 아니라 오히려 더 강화된 감독 의무의 근거라는 견해가 있다.[67] 또한 기업 내부 지배구조 개선 방안으로 이사회와 최고 경영진이 AI를 단순 감독 대상이 아닌 자신들의 의사결정을 개선하는 도구로 직접 활용해야 하며, '이 회사는 AI를 통해 어떻게 돈을 버는가?'라는 수익 모델에 대한 근본적 이해가 이사회의 핵심 역량이 되어야 한다고 강조하였다. 특히 현행 감사위원회가 재무 준법 기능과 AI를 포함한 각종 비재무적 리스크를 함께 담당하는 과부하 구조를 지적하고, 기업의 핵심 수익 창출 방식에서 비롯되는 법적·윤리적 리스크를 전담하는 '산업 특화 위원회(industry-specific committee)'의 신설을 제안하였다.[68] 이는 본고가 제안하는 AI 리스크 위원회(AI Risk Committee)와 방향을 같이 하는 것이다.

셋째, AI의 블랙박스 특성이 현행 신인의무 체계를 근본적으로 무력화한다는 견해가 있다. 이 견해에 따르면, 이사·임원이 AI의 블랙박스 작동 원리를 이해하기 사실상 불가능한 상황에서는 주의의무의 '정보에 입각한 판단' 요건이 구조적으로 충족될 수 없으며, 알고리즘 편향은 특정 주주·이해관계자에게 이익 또는 불이익을 초래하는 새로운 형태의 충실의무 위반을 야기한다는 것이다.[69] 특히 AI 기업에게 인간과 동등한 헌법적 정치적 언론 자유를 허용하는 것은 민주주의에 대한 인간의 주권을 알고리즘 법인체에 넘기는 것과 다름없다는 급진적 문제 제기도 포함되어 있어, 기업 법인격론과 AI의 교차점에서 발생하는 제도적 긴장을 예리하게 드러낸다.[70] 아울러 동 논문이 제시한 미래 기업법의 8가지 원칙 중 '투명성·책임성(transparency

---

67) Leo E. Strine, Jr., Using Experience Smartly to Ensure a Better Future: How the Hard-Earned Lessons of History Should Shape the External and Internal Governance of Corporate Use of Artificial Intelligence, 50 J. Corp. L. 1265, 1281–1292, 1295–1311 (June 2025) [이하 "Strine"]. AI 외부 규제 원칙에 관하여는 Id. at 1281–1292.

68) 이사회 내부 지배구조 5가지 제안(수익 모델 이해·산업 특화 위원회·이사회 자기 평가·AI 직접 활용·직원 목소리 제도화)에 관하여는 Id. at 1295–1311 참조.

69) Michael R. Siebecker, Reconceiving Corporate Rights and Regulation in the AI Era, 89 Mo. L. Rev. 941, 956–968, 972–978, 988–995 (Summer 2024) [이하 "Siebecker"]. 블랙박스 특성으로 인한 주의의무의 구조적 무력화에 관하여는 Id. at 956–960. 알고리즘 편향에 의한 충실의무 위반 가능성에 관하여는 Id. at 962–963.

and accountability)' 원칙은 AI 공시 의무화론의 이론적 토대를 제공한다.[71]

넷째, AI 거버넌스가 씨름하는 문제들이 기업이론(corporate theory)이 수천 년에 걸쳐 고민해 온 대리비용·기업목적 논쟁의 구조와 본질적으로 동일하다고 논증하면서, 'AI 거버넌스는 기업이론의 숙제를 베낄 수 있다'는 명제를 제시하는 견해도 있다.[72] 특히 이 논문은 AI 정렬 문제를 '이용자-AI 정렬(end-user alignment)'과 '사회-AI 정렬(societal alignment)'으로 이분하고, 전자는 기업이론의 대리비용 이론으로, 후자는 기업목적 논쟁(Berle vs. Dodd 논쟁)의 AI 버전으로 각각 분석하였다.[73] 나아가 주주(이용자)·이사회 수준 AI·임원 수준 AI의 3층 구조로 구성되는 '다층적 AI 지배구조(multitiered AI governance)'를 제안하였는데, 이는 본고가 제안하는 AI 리스크 위원회·AI Red-team·이사회의 삼중 감독 구조와 기능적으로 대응한다.[74] '누가 감독자를 감독하는가(who monitors the monitors)'의 문제도 기업이론과 동일한 방식 – 감독 AI의 행동 범위 제한 또는 복수의 이사회급 AI 병렬 배치 – 으로 해결될 수 있다는 시사도 주목할 만하다.[75]

이 네 편의 연구가 공통으로 수렴하는 결론은, AI 시대의 이사 감시의무가 단순한 사후적 책임 추궁 체계에서 사전적 알고리즘 설계 단계부터의 실질적 개입·통제 체계로 그 중심을 이동해야 한다는 것이다. 이사회 자기 평가와 세대 교체를 통한 AI 역량 이사 풀 확충, 인간-기계 간 대리비용의 사전적 통제 기준 설계, AI 의사결정의 투명성과 설명가능성의 공시 체계 반영이 이 연구가 제안하는 Caremark-XAI 연계 감독 모델의 장기적 발전 방향이다.

다만 이러한 논의를 우리 법에 수용할 때에는 책임의 경합 문제를 정리할 필요가 있다. AI 소프트웨어 개발사가 부담하는 책임과 이사가 부담하는 감시의무 위반 책임은 그 발생 근거와 보호법익을 달리한다. 전자는 결함 있는 제품 또는 서비스의 공급이라는 사실에, 후자는 회사 내부의 감독체계 구축·운영의 해태에 각각 기초한다. 따라서 양자는 어느 하나가 다른 하나를 흡수하는 관계가 아니라, 동일한 손해에 대하여 각자 전부의 책임을 지는 부진정연대채무의 관계로 구성하는 것이 타당하다. 이렇게 볼 때 주주나 제3자는 개발사와 이사 중 어느 쪽에 대하여도 손해 전부의 배상을 청구할 수 있고, 내부적으로는 각자의 기여도에 따라 구상관계를 정리하게 된다. 다만 개발사의 책임을 제조물책임으로 구성할 수 있는지, 그 경우 소프트웨어의 제조물성을 어떻게 볼 것인지 등은 회사법의 범위를 넘는 문제로서 별도의 연구 과제로 남긴다.

---

70) AI 기업의 정치적 언론 자유 제한 필요성에 관하여는 Id. at 978 – 988.
71) 미래 기업법 8가지 원칙(투명성·책임성 원칙 포함)에 관하여는 Id. at 988 – 995 참조.
72) Paul Weitzel, Governing Systems Through Corporate Theory, 92 Tenn. L. Rev. 169, 183 – 196, 201 – 215, 218 – 227 (Fall 2024) [이하 "Weitzel"].
73) Id. at 183 – 196.
74) Id. at 201 – 215.
75) Id. at 218 – 227.

Ⅴ. 결 론

이 연구는 2025년 개정 상법 제382조의3과 2026년 시행 인공지능기본법이 결합하여 이사의 AI 감시의무에 어떤 새로운 규범 구조를 만들어내는지를 Delaware Caremark·Marchand 법리 및 독일 주식법 제91조와의 비교법적 분석을 통해 검토하였다. 핵심 결론을 정리하면 다음과 같다.

첫째, 개정 상법은 이사의 의무 구조를 '단층형'(회사)에서 '이중층형'(회사+주주)으로 전환함으로써, AI 거버넌스 실패가 주주 공평대우 의무 위반으로 직결될 수 있는 법적 경로를 열었다. 알고리즘 편향이 특정 주주 집단에 불리하게 작용하는 경우, 이사는 개정법상 직접적인 법적 책임을 부담할 수 있다.

둘째, Caremark·Marchand 법리를 기능적으로 재구성하면, 핵심 업무에 AI를 도입하는 기업에서 AI는 Mission Critical Risk에 해당할 수 있으므로 이사회는 AI 알고리즘의 투명성·공정성·보안을 모니터링할 실질적 내부통제 시스템을 구축해야 할 의무를 부담한다. AI 경고신호에 대한 고의적 무지는 의식적 방치(conscious disregard)와 동일하게 평가되어 경영판단원칙의 보호 밖에 놓일 수 있다.

셋째, 인공지능기본법상 고영향 AI 조항(제33조~제35조)은 Bingle 판결이 드러낸 '실정법 부재'의 공백을 메워, 이사의 AI 감독 해태를 단순한 비즈니스 리스크 판단 실패에서 컴플라이언스 리스크 위반으로 전환시킬 수 있는 규범적 근거를 제공한다. 특히 제34조제1항제4호의 Human-in-the-loop 의무와 시행령 제27조의 5년 문서 보관 요건은 이사의 의무 이행 여부를 사후적으로 판단하는 기준이 된다.[76]

넷째, 비교법적으로 독일 주식법 제91조의 필수핵심영역·비필수영역 이분법과 사내이사·사외이사 주관적 구성요건 이원화는 한국 판례의 흐름과도 수렴한다. 우리 대법원은 1985년 평이사 감시의무 판결(84다카1954)에서 출발하여 2008년 내부통제시스템 구축의무의 최초 명시(2006다68636), 가격담합 영역으로의 확장(유니온스틸, 2017다222368), 사외이사를 포함한 모든 이사로의 확대(대우건설, 2021다279347), 형식적 제도 구비가 아니라 실질적 운영을 책임 판단의 기준으로 삼은 STX조선해양 판결(2019다202146·202153), 그리고 2025년의 재확인(2024다316391)에 이르기까지 '내부통제시스템의 실질화'를 향해 일관되게 발전하여 왔으며, 이는

76) 인공지능기본법 시행령 제27조 제2항은 고영향 AI 사업자에게 '고영향 인공지능에 대한 사람의 관리·감독' 관련 조치사항, 위험관리방안 등 제34조 제1항 각호의 책무 이행에 관한 기록을 5년간 보존하도록 의무화한다. 이 기록은 사후적 이사 책임 추궁 시 감시의무 이행 여부를 입증하거나 반박하는 핵심 증거로 기능한다. 성승제·정규, 전게논문, 244~245면; 각주 38 참조.

Caremark·Marchand 법리 및 독일 주식법 제91조와 기능적으로 합치한다. 이러한 국내 판례 계보는 우리 법원이 향후 AI 감시의무 사안에서 직접 원용할 수 있는 규범적 기반을 제공한다. 입법론으로는 ① 준법지원인의 알고리즘 적법성 점검 기능 명문화, ② 알고리즘 영향 평가(Algorithmic Impact Assessment; AIA)의 도입·공시 의무화, ③ AI 거버넌스 모범규준(Comply-or-Explain 방식)의 제정을 제언한다. 모범규준은 이사회가 AI 시스템 도입·운영에 관하여 일정한 거버넌스 요건을 준수하거나, 준수하지 못할 경우 그 이유를 공시하도록 하는 연성규범(soft law) 형태의 규율 체계를 의미한다.[77]

명문화의 필요성(①)은 실무에서도 확인된다. 금융 내부통제 분야에서는 지배구조법 제24조의 '내부통제기준 마련의무'가 실무상 단순한 기준 제정 의무로 축소 해석되어, 임직원의 기준 준수 위반으로 중대한 금융사고가 발생하더라도 관리의무를 제대로 이행하지 못한 임원에 대한 직접 제재 근거로 기능하지 못한다는 비판이 제기되어 왔다. 실제로 우리은행 DLF 불완전판매 사건의 행정소송에서 서울고등법원은 "내부통제기준 마련 위반과 내부통제기준 준수의무 위반은 분명히 구별되는 바", 임직원들의 기준 준수 위반으로 중대한 금융사고가 발생하더라도 그 관리의무를 제대로 하지 못한 임원들에 대한 직접적인 제재규정이 없다는 이유로 제재처분을 취소하였다.[78] 동 조항을 '내부통제시스템 구축 및 운영 의무'로 개정하여야 한다는 입법론은 알고리즘 준법감시 기능의 명문화를 통해 AI 내부통제시스템 구축 의무를 실정화하려는 본고의 제언과 동일한 방향에 있다. 종래 기준 마련 의무에 그쳤던 규율 체계를 시스템 구축 및 운영 의무로 격상하는 것은, 이사의 AI 감시의무가 형식적 내부통제기준의 제정에 그치는 것이 아니라 해당 시스템이 실질적으로 작동하도록 배려할 의무를 포함한다는 본고의 핵심 논지와 정합한다.

이 연구는 AI 시대 이사 감시의무의 규범적 기반을 정립하는 데 초점을 맞추었으나, 여기서 다루지 못한 주제들—AI 의사결정에 대한 절차적 정당화 모델, 주요 기업의 자율 거버넌스 사례의 비교법적 수용 방안 및 AI 거버넌스 모범규준의 구체적 설계 방안 등—은 앞으로의 연구에서 순차적으로 다루어져야 할 과제로 남는다. ▧

(논문접수 : 2026. 06. 17. / 심사개시 : 2026. 07. 10. / 게재확정 : 2026. 07. 21.)

77) 준법지원인 제도의 현황과 내부통제기준 마련의무의 한계에 관한 상세한 실무적 분석으로는 임현정, "금융회사 내부통제제도의 발전과 과제 – 지배구조법 제정 이후 10년간의 변화와 개선방안 –", 「상사판례연구」 제38권 제4호, 2025, 413~416면 참조. 동 논문은 현행 지배구조법 제24조의 '내부통제기준 마련의무'를 '내부통제시스템 구축 및 운영 의무'로 개정하여야 한다고 주장하는바, 이는 본고가 제안하는 준법지원인의 알고리즘 적법성 점검 기능 명문화와 동일한 입법 방향에 있다.

78) 서울고등법원 2022. 7. 22. 선고 2021누60238 판결; 대법원 2022. 12. 15. 선고 2022두54047 판결(심리불속행 기각); 임현정, 전게논문, 415~416면 참조.

참 고 문 헌

[국내 문헌]

강선희, “알고리즘 가격결정과 상법상 이사의 감시의무 – 미국 RealPage 사건과 한국 대법원 내부통제시스템 법리를 중심으로”, 「법학논문집」 제49집 제3호, 중앙대학교 법학연구원 (2025).

김민정, “이사의 감시·감독의무 이행을 위한 레그테크(RegTech)·섭테크(SupTech) 도입에 관한 연구”, 「경영법률」 제33집 제3호, 한국경영법률학회 (2023).

김민주·김현경, “인공지능 투명성 규제에 대한 비판적 고찰 – EU AI법 투명성 의무 조항 분석을 중심으로 –”, 「선진상사법률연구」 통권 제110호 (2025).

김유정, “일본 AI법(人工知能關連技術の研究開發及び活用の推進に關する法律) 제정 과정에 관한 연구”, 「日本學報」 第146輯 (2026).

김정호, “내부통제시스템과 이사의 법적 책임 – 특히 독일주식법의 상황을 소개하며 –”, 「상사법연구」 제41권 제3호, 한국상사법학회 (2022).

김지평, “이사의 주주 이익 보호의무 관점에서의 합병 계약의 쟁점”, 「기업법연구」 제39권 제4호, 한국기업법학회 (2025. 12.)[이하 “김지평”].

김지현, “델라웨어주 회사법의 최근 동향과 한국 상법에 대한 시사점 – 지배주주 거래 규제와 절차적 안전장치를 중심으로 –”, 「기업법연구」 제39권 제4호(2025).

남궁주현, “인공지능(AI)과 회사법의 연결에 관한 소고 – 이사의 책임 부담 가능성을 중심으로”, 「성균관법학」 제36권 제4호, 성균관대학교 법학연구원 (2024).

류지민, “케어마크 청구에서의 미션 크리티컬 – 이사의 감시의무 영역의 범위와 한계 –”, 「상사법연구」 제41권 제4호, 한국상사법학회 (2023).

문정해, “내부통제시스템에 관한 이사의 감시의무와 경영판단원칙 – 미국 델라웨어주 판례법상 Caremark 기준과 경영판단원칙의 관계를 중심으로 –”, 「상사판례연구」 제38권 제2호, 한국상사판례학회 (2025).

박준선, “미국 회사법상 경영판단의 원칙과 완전한 공정성 기준의 적용 법리 – 미국 델라웨어주를 중심으로 –”, 「사법」 제75호(2026).

성승제·정규, “인공지능과 회사법의 법정책에 대한 검토 — 인공지능과 법인 이사 검토 등을 중심으로 —”, 「법과 정책연구」 제25집 제4호, 한국법정책학회 (2025).

손창완, "상법 개정의 현황과 평가 — 제21대 국회의 상법 개정안을 중심으로 —", 「상사법연구」 제43권 제1호 (2024).

송옥렬, "인공지능과 이사회", 「기업법연구」 제35권 제2호(통권 제85호) (2021).

송옥렬, “이사의 감시의무와 내부통제시스템 구축의무”, 「기업법연구」 제36권 제1호(통권 제88호) (2022).

이상헌, “인공지능으로 인한 침해에 대한 민사책임 – 유럽연합과 프랑스에서의 논의를 중심으로 –”, 「저스티스」 통권 제213호, 한국법학원 (2026).

이상훈, "이사의 주주에 대한 충실의무 — 비교법적 검토 —", 「비교사법」 제29권 제3호 (2022).

임철현, "AI 시대의 기업거버넌스 — AI의 기업법적 활용을 중심으로 —", 「서울법학」 제32권 제4호 (2025).

임철현, “주주에 대한 충실의무 조항의 해석 – 독일법상 계약의 제3자보호효를 실마리로 하여 –”, 「비교사법」 제33권 제1호 (2026).

임현정, “금융회사 내부통제제도의 발전과 과제 – 지배구조법 제정 이후 10년간의 변화와 개선방안 –”, 「상사판례연구」 제38권 제4호 (2025).

정대익, “주주에 대한 이사의 충실의무 부존재 하의 독일법상 주주 보호 수단 – 유럽 국가 및 영연방 국가와의 비교를 겸하여 –”, 「상사법연구」 제44권 제1호(2025).

천경훈, “2025년 개정상법상 이사 충실의무 조항의 해석론”, 「상사법연구」 제44권 제2호 (2025).

천경훈·정준혁, “이사의 충실의무와 주주이익 보호”, 「상사법연구」 제43권 제3호 (2024).

최문희, "인공지능(Artificial Intelligence)과 회사법의 접점 — 인공지능의 활용에 수반되는 회사법적 논점 —", 「상사법연구」 제39권 제1호 (2020).

홍복기, “인공지능이 회사제도에 미치는 영향 – 주식회사의 이사회를 중심으로 –”, 「상사법연구」 제43권 제2호, 한국상사법학회 (2024).

[외국 문헌]

1. Books

Brynjolfsson, Erik & Andrew McAfee, The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies (2014).

Ezrachi, Ariel & Stucke, Maurice E., Virtual Competition: The Promise and Perils of the Algorithm-Driven Economy, Harvard University Press (2016).

Pasquale, Frank, The Black Box Society: The Secret Algorithms That Control Money and Information (2015).

2. Articles

Armour, John & Eidenmueller, Horst, Self-Driving Corporations, ECGI Law Working Paper No. 457/2019 (2019), available at https://ssrn.com/abstract=3447037.

Bainbridge, Stephen M., Caremark and Enterprise Risk Management, 34 J. Corp. L. 967 (2009).

Barrie, Michael J. & Andrew D. Kinsey, Delaware Amends DGCL Section 144 to Add Safe Harbors for Interested and Controlling Stockholder Transactions (Benesch Law, Apr. 9, 2025).

Cowger, Alfred R., Jr., Corporate Fiduciary Duty in the Age of Algorithms, 14 Case W. Reserve J.L. Tech. & Internet 136 (2022–2023).

Doshi-Velez, Finale & Been Kim, Towards a Rigorous Science of Interpretable Machine Learning (2017), https://www.semanticscholar.org/reader/5c39e37022661f81f79e481240ed9b175dec6513.

Gressel, Anna, Avi Gesser & Lily Coad, AI Oversight Is Becoming a Board Issue, Harvard Law School Forum on Corporate Governance (Apr. 26, 2022), https://corpgov.law.harvard.edu/2022/04/26/ai-oversight-is-becoming-a-board-issue/.

Johnson, Lyman, Unsettledness in Delaware Corporate Law: Business Judgment Rule, Corporate Purpose, 38 Del. J. Corp. L. 405 (2013).

Macey, Jonathan R. & Geoffrey P. Miller, Corporate Stakeholders: A Contractual Perspective, 43 U. Toronto L.J. 401 (1993).

Mayson, Sandra G., Bias In, Bias Out, 128 Yale L.J. 2218 (2019).

Mertens, Floris, The Use of Artificial Intelligence in Corporate Decision-Making at Board Level: A Preliminary Legal Analysis, Financial Law Institute Working Paper Series 2023-01 (2023).

Montagnani, Maria Lillà & Passador, Maria Lucia, Fiduciary Duties and Business Judgment Rule 2.0 in the AI Act Age (June 9, 2025), available at https://ssrn.com/abstract=5714102.

Mosco, Gian Domenico, AI and Board Within Italian Corporate Law: Preliminary Notes, European Company Law, Vol. 17, No. 3 (2020).

Petrin, Martin, Corporate Management in the Age of AI, 2019 Colum. Bus. L. Rev. 965 (2019).

Siebecker, Michael R., Reconceiving Corporate Rights and Regulation in the AI Era, 89

Mo. L. Rev. 941 (Summer 2024).

Strine, Leo E., Jr. et al., Loyalty's Core Demand: The Defining Role of Good Faith in Corporation Law, 98 Geo. L.J. 629 (2010).

Strine, Leo E., Jr., The Dangers of Denial: The Need for a Clear-Eyed Understanding of the Power and Accountability Structure Established by the Delaware General Corporation Law, 50 Wake Forest L. Rev. 761 (2015).

Strine, Leo E., Jr., Using Experience Smartly to Ensure a Better Future: How the Hard-Earned Lessons of History Should Shape the External and Internal Governance of Corporate Use of Artificial Intelligence, 50 J. Corp. L. 1265 (June 2025).

Weitzel, Paul, Governing Systems Through Corporate Theory, 92 Tenn. L. Rev. 169 (Fall 2024).

松尾剛行, 「AIとガバナンス－－企業統治の高度化·効率化にAIを役立てるという觀点からの檢討」, 商事法務 No. 2297 (2022).

松尾剛行, 「プラットフォーム事業者によるアカウント凍結等に對する私法上の救濟について」, 情報法制研究 第10号 (2021).

3. Regulatory Documents & Reports

European Parliament & Council, Regulation (EU) 2024/1689 (EU AI Act) (2024).

National Institute of Standards and Technology, Artificial Intelligence Risk Management Framework (AI RMF 1.0) (2023).

OECD, Algorithms and Collusion: Competition Policy in the Digital Age, OECD Publishing (2017).

U.S. Department of Justice, Statement of Interest in In re RealPage, Inc., Case No. 3:23-md-03071 (M.D. Tenn. Nov. 15, 2023).

Abstract

Directors' Oversight Duties in the Age of AI: Substantive Internal Control

- An Integrated Analysis of the Amended Korean Commercial Act Article 382-3, the AI Basic Act, and the Caremark-Marchand Doctrine -

Yook, Taewoo

As artificial intelligence increasingly assumes a central role in corporate decision-making, directors' oversight duties have undergone a qualitative transformation. The 2025 amendment to Article 382-3 of the Korean Commercial Act extended directors' fiduciary obligations from the corporation alone to "the corporation and its shareholders," opening a direct legal pathway under which AI governance failures may constitute a breach of the duty of fair treatment of shareholders. The AI Basic Act (in force January 2026) gives concrete statutory content to directors' duty of care through the obligation to confirm high-impact AI status (Article 33), the Human-in-the-loop mandate (Article 34(1)(iv)), and the AI impact assessment (Article 35). These provisions fill the regulatory gap exposed by the Delaware Bingle decision, providing a normative basis for potentially converting AI risks from business risk to compliance risk for Korean corporate law purposes.

This article functionally reconstructs the Delaware Caremark–Marchand framework as a duty of substantive AI internal control. Because AI may constitute a Mission Critical Risk when deployed in core business functions, boards must establish real monitoring systems covering the transparency, fairness, and security of AI algorithms; willful ignorance of AI red flags constitutes conscious disregard. A director who approves AI outputs without independent scrutiny engages in "algorithmic conscious disregard," amounting to bad faith and excluding Business Judgment Rule protection. The article further demonstrates convergence between the German Stock Corporation Act § 91 framework and the Korean Supreme Court's line of oversight decisions, which from 1985 to 2025 has steadily

converged toward the substantive operation of internal control systems, and proposes three legislative reforms: (i) codification of algorithmic legality review within the compliance officer regime; (ii) mandatory Algorithmic Impact Assessment with public disclosure; and (iii) an AI Governance Code of Best Practice on a comply-or-explain basis. Delaware's SB21, which weakened MFW minority-shareholder protections, offers a comparative benchmark against which Korean procedural safe harbor design must not compromise minimum standards; detailed analysis is reserved for subsequent research.

Key Words : Directors' Oversight Duty, AI Internal Control System, AI Governance, Amended Korean Commercial Act Article 382-3, Caremark Doctrine, Algorithmic Conscious Disregard, High-Impact AI, AI Basic Act (Korea), Corporate Governance