경북대학교 법학연구원『법학논고』
제95집 (2026.10) 1~31면.
Kyungpook Natl. Univ. Law Journal
Vol.95 (Oct 2026) pp.1~31.
http://doi.org/10.17248/knulaw..95.202610.1

자율주행자동차 사이버보안에 관한 헌법적 검토*

- 인공지능기본법상 고영향 인공지능과의 관계를 중심으로 -

문 의 빈**

〈국문초록〉

오늘날 자동차는 수십 개의 전자제어장치와 소프트웨어, 통신기술이 결합된 '바퀴 달린 컴퓨터'가 되었고, 운전이라는 인간 고유의 영역은 인공지능에 의해 대체되고 있다. 이러한 패러다임 전환은 사이버 위험의 성질 자체를 변화시킨다. 자율주행자동차에 대한 사이버 공격은 정보유출이라는 전통적 피해에 머무르지 않고, 제동·조향 체계의 무력화나 센서 인식의 기만을 통하여 탑승자와 제3자의 생명·신체에 대한 직접적 침해로 확장되기 때문이다. 적대적 공격, 센서 스푸핑, 학습데이터 오염과 같이 인공지능의 인식·판단 과정 자체를 겨냥하는 위협은 차량이라는 제품의 관점에서 설계된 기존 자동차 법제만으로는 충분히 포착되지 않는다. 이 논문은 자율주행자동차의 사이버보안이 기술적·산업적 측면을 넘어 헌법적 규율의 대상이라는 문제의식에서 「자동차관리법」상 사이버보안 관리체계와 인공지능기본법상 고영향 인공지능 규율을 함께 검토하였다.

먼저 조건부 완전자율주행 이상의 자율주행시스템은 교통수단의 주요한 작동 및 운영에 활용되어 생명·신체의 안전에 중대한 영향을 미칠 우려가 있으므로 고영향 인공지능에 해당한다는 점을 확인하였다. 다음으로, 고영향 인공지능에 해당하는지 여부가 인공지능사업자의 판단과 가이드라인에 의존할 수 밖에 없다는 점은 명확성원칙에 비추어 문제가 있고, 인공지능기본법상 생명·신체에 대한 위험과 그 밖의 기본권 위험에 대하여 동일한 책무가 부과되어 위험의 중대성·심각성에 비례하는 규율이 이루어지지 않으며, 인공지능사업자 책무의 보호 대상이 '이용자'에 한정되어 보행자 등 '영향받는 자'가 보호에서 누락된다는 헌법적 쟁점에 대하여 검토하였다.

나아가 「자동차관리법」상 사이버보안 관리체계가 생명·신체 침해 위험으로부터 국민을 보호하기 위한 국가의 기본권보호의무를 구체화함으로써, 생명권과 개인정보자기결정권을 실효적으로 보장하는 헌법적 안전장치로 기능한다는 점에 주목하였다.

주제어 : 자율주행자동차, 사이버보안 관리체계, 인공지능기본법, 고영향 인공지능, 기본권보호의무

• 투고일 : 2026.10.07. / 심사일 : 2026.10.27. / 게재확정일 : 2026.10.27.

---

* 이 논문은 2025년 상반기 한국사이버안보법정책학회 학술세미나의 발표문 "자율주행자동차의 사이버안보 강화 방안"을 수정·보완한 것임.

** 국민대학교 법과대학 조교수.

Ⅰ. 서론

오늘날 자동차는 수십 개의 전자제어장치와 다양한 소프트웨어 그리고 통신 기술이 결합된 '바퀴 달린 컴퓨터'[1]가 되었다. 자율주행 기술의 진전과 더불어, 운전이라는 인간 고유의 영역이 인공지능에 의해 대체되는 패러다임 전환이 본격화되고 있는 것이다. 운전자의 개입 없이 차량이 스스로 인지·판단·제어하는 자율주행은 교통사고의 획기적 감소와 이동권의 확장이라는 사회적 편익을 예고하는 한편, 그 작동을 가능하게 만드는 인공지능과 통신체계가 사이버 공격의 표적이 되면서 기존의 자동차에서는 상정할 수 없었던 새로운 위험이 발생하고 있다.

사이버 공격이 자율주행자동차의 제동·조향 체계를 무력화하거나 센서 인식을 기만하여 오작동을 야기하는 순간, 그 결과는 정보유출이라는 전통적인 사이버 위험을 벗어나 탑승자와 제3자의 생명·신체에 대한 직접적인 침해로 확장된다. 이는 사이버보안이 정보보호의 차원을 넘어 생명·신체의 안전과 직결되고 있음을 보여준다. 자율주행자동차 시대의 사이버보안은 더 이상 정보유출 피해에 대응하는 기술적·산업적 과제에 머무르는지 않고, 생명과 신체의 안전이라는 기본권을 보호할 국가의 기본권보호의무가 작동하는 영역이 되었다.

이 전환의 핵심에는 인공지능이 있다. 자율주행자동차는 하나의 알고리즘이 아니라 위치추정, 인지, 계획, 제어, 인간-차량 상호작용에 걸친 여러 인공지능 시스템이 상호 연관되어 자동차의 형태로 구현된 것이다.[2] 그리고 이들 인공지능 구성요소는 전통적 소프트웨어와 다른 규칙을 따른다. 기계학습은 대량의 데이터에 대한 통계적 분석에 기초한 암묵적 규칙에 의존하기 때문에, 그 복잡성을 공격자가 악용할 새로운 기회가 열린다.[3] 그러므로 자율주행자동차의 사이버보안을 헌법적으로 검토하려면 자동차 법제뿐 아니라 인공지능 법제를 함께 보아야 한다.

국내에서도 자동차 사이버보안 관리체계(Cyber Security Management System) 인증을 의무화하는 내용으로 「자동차관리법」이 개정되었고,[4] 「인공지

---

1) 임성호, "자동차는 '바퀴 달린 컴퓨터' … 디스플레이, 핵심부품으로 부상", 연합뉴스(2026. 6. 7.) <https://www.yna.co.kr/view/AKR20260605066600003> (최종확인 2026. 10. 28.)

2) David Fernández Llorca and Emilia Gómez, "Trustworthy Autonomous Vehicles", Publications Office of the European Union, 2021, p.2 & 5.

3) Georgia Dede et al., "Cybersecurity Challenges in the Uptake of Artificial Intelligence in Autonomous Driving", Publications Office of the European Union, 2021, p.6.

4) 「자동차관리법」(법률 제20298호, 2025. 8. 14. 시행) 제30조의9부터 제30조의12까지 및 제34

능 발전과 신뢰 기반 조성 등에 관한 기본법」(이하 '인공지능기본법')에서는 「교통안전법」에 따른 교통수단의 '주요한 작동 및 운영'에 활용되는 인공지능을 고영향 인공지능의 한 영역으로 규정하여 자율주행자동차를 고영향 인공지능으로 규율할 수 있는 근거를 마련하였다.[5] 그러나 자율주행자동차를 규율하는 이러한 법률체계가 어떠한 헌법적 근거에서 정당화되며, 기본권 보호의 관점에서 적절한지에 대한 헌법학적 검토는 부족한 실정이다. 특히 고영향 인공지능이라는 새로운 규율 범주가 자율주행자동차에 적용될 때 어떤 헌법적 쟁점이 생기는지는 거의 논의되지 않았다.

이에 본 연구는 먼저, 자율주행자동차에서 인공지능이 어떤 역할을 하는지, 그리고 그 인공지능 때문에 사이버보안 위협의 양상이 어떻게 달라지는지를 살핀다(Ⅱ). 이어서 자율주행자동차가 인공지능기본법상 고영향 인공지능에 해당하는지를 검토하고, 그 판단구조와 법적 효과가 안고 있는 헌법적 쟁점을 분석한다(Ⅲ). 다음으로 「자동차관리법」상 사이버보안 관리체계의 도입이 갖는 헌법적 의미와 자율주행자동차로 인하여 발생할 수 있는 기본권 관련 쟁점을 분석하고, 이를 통해 자율주행자동차의 사이버보안 관리체계가 단순한 기술적 안전장치가 아니라 인공지능 시대에 국민의 생명·신체와 개인정보를 보호하는 헌법적 안전장치로 기능한다는 점을 밝힌다(Ⅳ).

Ⅱ. 자율주행자동차에 발생할 수 있는 사이버 보안 위협

1. 자율주행자동차의 기술구조와 위협 환경

(1) 자율주행 시스템의 구조와 인공지능의 역할

자율주행자동차에 대한 사이버 위협을 규명하기 위해서는 그 기술적 작동구조에 대한 이해가 선행되어야 한다. 자율주행 시스템은 ① 인지(센서 데이터 처리·물체 탐지·차선 인식), ② 예측(주변 객체의 미래 움직임 예측), ③ 계획(주행 경로 결정), ④ 제어(조향·가속·제동 실행)라는 4단계로 구분된다.[6]

조의5.

5) 「인공지능 발전과 신뢰 기반 조성 등에 관한 기본법」(법률 제20676호, 2026. 1. 22. 시행) 제2조 제4호아목.

6) Kathrin Grosse and Alexandre Alahi, "A Qualitative AI Security Risk Assessment of Autonomous Vehicles", *Transportation Research Part C: Emerging Technologies* Vol.

인공지능 기술은 이 모든 과정에 핵심적으로 개입한다. 인지 단계에서는 차량에 탑재된 다중 센서가 주변 환경과 객체를 감지하고, 예측과 계획 단계에서는 인공지능이 수집된 정보를 바탕으로 충돌 회피·경로 생성 등의 의사결정을 수행하며, 제어 단계에서는 그 판단 결과에 따라 실제 차량의 운행이 이루어진다. 이는 위협이 다양한 차원에서 발생할 수 있음을 보여주며, 사이버보안이 자율주행 시스템의 특정 단계가 아니라 시스템 전체의 관점에서 검토되어야 함을 시사한다. 이러한 점에서 자율주행자동차는 카메라·라이다(LiDAR)·레이더·GPS 등 다양한 센서로 주행환경을 인지하고, 이를 처리하여 판단한 뒤, 제동·조향 등 물리적 제어를 수행하는 사이버물리시스템이라고 할 수 있다.[7]

또한, 자율주행자동차의 인공지능 소프트웨어는 여러 센서에서 얻은 다양한 데이터와 통계·수학 등으로부터 도출된 인공지능 방법론의 결합이다.[8] 현행법을 살펴보면, 「자율주행자동차 상용화 촉진 및 지원에 관한 법률」(이하 '자율주행자동차법')에서는 '자율주행시스템'을 운전자 또는 승객의 조작 없이 주변상황과 도로 정보 등을 스스로 인지하고 판단하여 자동차를 운행할 수 있게 하는 자동화 장비, 소프트웨어 및 이와 관련한 모든 장치로 규정하고 있고,[9] 인공지능기본법에서는 '인공지능시스템'을 다양한 수준의 자율성과 적응성을 가지고 주어진 목표를 위하여 실제 및 가상환경에 영향을 미치는 예측, 추천, 결정 등의 결과물을 추론하는 인공지능 기반 시스템으로 규정하고 있다.[10] '스

169, Elsevier, 2024, pp.3-4. 자율주행시스템을 1) 디지털 지도에 대한 차량의 위치와 방향을 정하는 위치추정(localisation), 2) 주변의 동적 요소를 탐지·해석하고 다른 교통참여자의 미래 행동과 궤적을 예측하는 인지(perception and prediction), 3) 위치·교통규칙·신호·예측 궤적에 기초하여 주행 궤적과 속도를 정하는 계획(planning), 4) 가속·제동과 조향을 수행하는 제어(control), 5) 차량과 사람 사이의 인간-차량 상호작용(human-vehicle interaction), 6) 고장 관리·데이터 관리·사고기록 등을 담당하는 시스템 관리(system management)의 6단계로 구분하기도 한다. 이에 관하여는 David Fernández Llorca et al., "Testing Autonomous Vehicles and AI: Perspectives and Challenges from Cybersecurity, Transparency, Robustness and Fairness", *European Transport Research Review* Vol.17, Springer, 2025, pp.3-5 참조.

7) Mashrur Chowdhury, Mhafuzul Islam and Zadid Khan, "Security of Connected and Automated Vehicles", *The Bridge*, Vol.49, No.3, 2019, pp.48-51. 자율주행에 관여하는 인공지능을 위치추정·인지·계획·제어·인간-차량 상호작용·시스템 관리 등 여섯 개의 '운영단계'로 구분하고, 각 계층마다 인공지능이 담당하는 역할과 그에 따른 위험을 분석하는 연구로 David Fernández Llorca et al., *op. cit.*, vol. 17, 2025.

8) Georgia Dede et al., *op. cit.*, p.6.

9) 「자율주행자동차 상용화 촉진 및 지원에 관한 법률」 제2조제1항제2호. '자율주행자동차'는 「자동차관리법」 제2조 제1호의3에 따라 운전자 또는 승객의 조작 없이 자동차 스스로 운행이 가능한 자동차를 말한다(같은 항 제1호).

10) 인공지능기본법 제2조제2호.

스로 인지하고 판단하여' 운행하는 자율주행시스템은 실제 환경에 영향을 미치는 결정을 자율적으로 추론하는 시스템이므로, 자율주행자동차법에 따른 자율주행시스템은 인공지능기본법상 인공지능시스템에 해당할 수 있다.

(2) 자율주행 단계와 인간·인공지능의 역할 배분

인공지능이 운전에서 차지하는 비중은 자율주행의 단계에 따라 다르다. 국제적으로 통용되는 미국 자동차공학회(SAE)의 분류에서는 운전 자동화를 자동화가 전혀 없는 레벨 0부터 운전자가 필요 없는 완전 자동화의 레벨 5까지 여섯 단계로 나눈다.[11] 우리 자율주행자동차법은 SAE의 6단계가 아니라 작동한계상황 등 필요한 경우 운전자의 개입을 요구하는 '부분 자율주행자동차'와 자율주행시스템만으로 운행할 수 있어 운전자가 없거나 운전자 또는 승객의 개입이 필요하지 아니한 '완전 자율주행자동차'의 두 종류로 구분한다.[12] 이 구분은 운전이라는 행위를 사람과 인공지능 사이에 어떻게 배분하는지를 정한다는 점에서 의미가 있다. 보조 단계에서 인공지능은 운전자가 당연히 인지하고 있어야 할 사항을 보조하는 데 그치지만, 자율 단계에서는 인지·판단·제어의 전 과정을 인공지능이 수행하고 사람은 그 결과를 받아들이는 위치에 놓인다. 자동화의 레벨이 오를수록 운전자의 역할은 단계적으로 줄어들게 되고, 운행에 대한 책임은 제작사에게 귀속된다.[13]

(3) 다양한 센서의 융합과 학습데이터 의존성

자율주행자동차에서 이루어지는 '인지' 기능은 하나의 센서가 아니라 다양한 센서의 융합에 기반하며, 학습데이터의 품질 및 무결성에 의존한다.[14] 즉, 자율주행자동차의 인지 기능은 단일 센서가 아니라 카메라,[15] 라이다(LiDAR:

11) Georgia Dede et al., *op. cit.*, p.9; Bill Canis, "Issues in Autonomous Vehicle Testing and Deployment (CRS Report R45985)", *Congressional Research Service*, 2021, pp.2-3. 레벨 1-2를 운전자가 책임지는 '보조(assisted)' 운전, 레벨 3을 예비 운전자가 책임지는 '자동화(automated)' 운전, 레벨 4-5를 탑승자만 있거나 무인인 '자율(autonomous)' 운전으로 부르기도 한다.

12) 자율주행자동차법 제2조제2항.

13) 송문호, "자율주행자동차법제 개선방안", 「법학논총」 제32권 제2호, 조선대학교 법학연구원, 2025, 433면.

14) GPS, 카메라, LiDAR 등 개별 센서가 직면한 재밍·스푸핑·환각·실명 공격에 대한 취약성을 표로 일목요연하게 정리하고, 이에 대한 방어 전략으로 다중 GPS 수신기 활용, 중복 센서를 통한 카메라 데이터 교차 검증, 그리고 적대적 데이터(Adversarial Data)로 훈련된 기계학습 모델 적용 등의 구체적인 대응책을 제안하는 연구로, Mashrur Chowdhury, Mhafuzul Islam and Zadid Khan, *op. cit.*, pp.48-49.

Light Detection and Ranging),[16] 레이더(Radar: Radio Detection and Ranging),[17] 초음파센서 등 다양한 센서를 융합함으로써 구현된다. 이들 센서가 상호 보완적으로 결합되어 주행환경에 대한 종합적 인식이 이루어진다.[18] 나아가 고정밀 3차원 지도는 주행환경 정보를 차선 단위로 제공하여, 실시간 센서 정보와 결합함으로써 보다 정확한 주행 판단을 가능하게 한다. 그러나 GPS와 라이다에는 재밍과 스푸핑이, 카메라에는 착시·실명 공격이 가능하다. 그 대응책으로는 다중 GPS 수신기, 중복 센서를 통한 카메라 데이터의 교차 검증, 적대적 데이터로 훈련된 기계학습 모델의 적용 등이 제안된다.[19] 자율주행자동차에서 인공지능의 판단은 센서가 수집한 입력 데이터의 정확성에 전적으로 의존하므로, 센서 데이터의 위·변조나 학습데이터의 오염은 곧 인공지능의 오판으로 이어진다. 지도정보나 위치정보가 갱신되지 않거나 전송 과정에서 오류가 생기는 것만으로도 사고가 날 수 있다는 점에서,[20] 데이터 의존성은 그 자체로 공격에 노출되는 지점이 된다. 따라서 자율주행자동차에서 발생하는 위협은 전통적인 정보통신시스템의 위협과 달리 인공지능의 인식·판단 과정 자체를 겨냥하는 새로운 유형의 공격이라는 특수성을 갖는다.[21]

(4) 통신을 통한 연결의 확대

오늘날의 자동차는 자율주행기능을 갖추고 있을 뿐 아니라, 무선 정보통신기술을 활용하여 정보를 송신·수신하는 이른바 '커넥티드자동차'[22]이기도 하다. 과거에는 자동차를 탈취하려면 물리적인 접근이 필수적이었으나, 커넥티드자동차의 경우 차량에 탑재된 무선랜(Wi-Fi)·이동통신(LTE) 모듈, 근거리무선통신 기반 디지털 키, 무선 소프트웨어 업데이트, 차량-사물 통신 등 연결 통로가 대폭

15) 카메라는 교통표지판, 차선, 신호등, 보행자 등 시각적으로 구별 가능한 객체를 식별하는 기능을 수행한다. 예를 들어, 신호등의 색상 인식이나 앞 차량의 브레이크등 점등 여부를 감지한다.

16) 라이다는 빛을 발사해 반사 시간을 측정함으로써 차량 주변 3D 지도를 생성하는 기능을 수행한다. 장애물 거리 측정, 공간 인식에 강점을 가지며, 자율주행자동차의 핵심 센서로 사용된다.

17) 레이더는 전파를 이용해 이동 물체의 속도와 거리를 감지하는 기능을 수행한다. 악천후나 먼지, 안개 등에서 카메라나 라이다보다 안정적이라는 특징이 있다.

18) Mashrur Chowdhury, Mhafuzul Islam and Zadid Khan, *op. cit.* p.48.

19) *Ibid.,* pp.48-49.

20) Martin Ebers, "Civil Liability for Autonomous Vehicles in Germany", SSRN Scholarly Paper, 2022, pp.3-4, <https://doi.org/10.2139/ssrn.4027594>.

21) Dede et al., *op. cit.* pp. 6-7; Kathrin Grosse and Alexandre Alahi, *op. cit.*, p.4.

22) 「자동차관리법」 제2조제1호의7.

늘어났기 때문에, 공격자에게는 이러한 통로 하나하나가 원격 침투 경로가 된다. 자율주행자동차는 센서와 자동차 제어 시스템, 차량 간(Vehicle-to-Vehicle) 및 차량-인프라 간(Vehicle-to-Infrastructure) 통신이라는 세 가지 구성요소에 기반하기 때문에 오래된 지도 정보나 부정확한 교통 상황과 같은 잘못된 정보를 수신하면 사고가 발생할 수 있다.[23] 또한, 무선 업데이트도 새로운 공격 통로로 악용될 수 있다.[24] 자동차가 인포테인먼트 시스템을 통해 인터넷에 연결되고, 내비게이션을 위해 GPS를 수신하며 인프라와 다른 차량 등 외부요소와 통신하게 되면서 무선 소프트웨어 업데이트가 일반화되었고, 그만큼 악성코드를 설치하기도 쉬워졌기 때문이다.[25]

첨단 교통수단이 운행과 교통관리의 효율성 향상을 위하여 외부와의 상호연결성을 높이면 높일수록 그에 비례하여 사이버 공격을 받을 위험성도 증대되며, 해커 등이 첨단차량의 사이버 보안 취약성을 악용할 경우 야기될 수 있는 위험은 탑승자의 불편이나 사소한 주행 장애에 그치지 아니하고 생명·신체에 대한 중대한 침해로 확장될 수 있다.[26] 이는 통신을 통한 연결의 확대가 효율성과 위험을 동시에 증폭시키는 양면적 구조임을 보여주며, 사이버보안이 자율주행자동차를 설계할 때 고려해야 할 필수적이고 본질적 요건이라는 근거가 된다.

2. 위협 유형별 분석

(1) 차량 내·외부 통신망 공격

차량 내부 통신망에 대한 공격은 가장 고전적이면서도 치명적인 위협이다. 차량 내부의 CAN(Controller Area Network) 버스는 방화벽 등 기초적 보안기능을 갖추고 있으나 암호화되지 않은 메시지에 대한 접근에 취약하고, 데이터 폭주를 통한 서비스거부(denial of service, DoS) 공격에 노출될 수 있다.[27] CAN 버스의 해킹을 통해 제동·조향과 같은 핵심 차량 기능에 대한 제어권을 탈취할 수 있고, 인포테인먼트 시스템이 CAN 버스로 진입하는 백도어

---

23) Martin Ebers, *op. cit.*, pp.3-4.
24) Mashrur Chowdhury, Mhafuzul Islam and Zadid Khan, *op. cit.* p.48.
25) Nynke E Vellinga, "Connected and Vulnerable: Cybersecurity in Vehicles", *International Review of Law, Computers & Technology*, Vol.36 No.2, 2022, pp.162-163.
26) 류병운, "사이버 공격으로부터 첨단 교통 시스템 보호 – 관련 입법례를 중심으로", 「홍익법학」, 제24권 제3호, 홍익대학교 법학연구소 2023, 1-3면.
27) Mashrur Chowdhury et al., *op. cit.*, p.49.

(backdoor)[28]로 기능할 수 있다.[29] 2015년 두 연구자가 지프 페로키(Jeep Cherokee)의 인포테인먼트 시스템을 통해 원격으로 제어계통에 접근하여 고속도로를 주행 중인 차량을 감속시킨 사례이 대표적이다.[30] 차량 외부 통신망 역시 다양한 공격에 노출된다. 블랙홀·그레이홀 공격[31], 서비스 거부(DoS) 공격, 재밍[32]처럼 안전에 필요한 정보가 제때 전달되지 못하게 하는 공격이 가능하고, 허위 메시지 주입과 시빌(Sybil)[33] 공격처럼 차량이 잘못된 정보를 신뢰하게 만드는 공격이나 도청처럼 차량 간에 오가는 정보를 몰래 수집하는 공격 방식도 존재한다.[34]

(2) 소프트웨어 무선 업데이트 경로에 대한 공격

무선 업데이트 경로는 보안 패치를 신속히 배포하는 핵심 도구이나 그 자체가 변조·재생 공격의 표적이 된다. 자율주행자동차는 그 속성상 소프트웨어 중심 차량(Software Defined Vehicle)이기 때문에 기능의 추가 및 수정이 무선 소프트웨어 업데이트로 이루어지고, 그 결과 무선 소프트웨어 업데이트 자체가 중대한 공격 경로가 되었다.[35] 예를 들어 무선 소프트웨어 업데이트 과정에서 전송되는 소프트웨어가 변조되거나, 악성 코드가 차량의 핵심 제어 시스템에 직접 주입될 수 있는 것이다. 이는 자율주행자동차의 가속·제동·조향과 같은 핵심 기능조차 소프트웨어가 제어하는 환경에서 업데이트 경로에 대한 공격으로 인해 곧바로 자율주행자동차의 안전이 붕괴될 수 있음을 의미한다.[36] 인공지능 모델의 업데이트는 새로운 취약점을 들여올 수 있으므로 그 보안성과 견고성을 체계적으로 검증해야 한다.[37]

---

28) 시스템 설계자가 만들어 놓은 보안 절차를 우회하는 비밀 통로를 말한다. (한경경제용어사전)
29) Anastasios Giannaros et al., “Autonomous Vehicles: Sophisticated Attacks, Safety Issues, Challenges, Open Topics, Blockchain, and Future Directions”, *Journal of Cybersecurity and Privacy* Vol.3 No.3, MDPI, 2023, p.509.
30) Nynke E Vellinga, *op. cit.*, p.165
31) 해킹당한 차량은 다른 CAV로 전송되는 모든 패킷(블랙홀) 또는 일부 패킷(그레이홀)의 전달을 중단하여, 다른 CAV들이 전방 충돌 경고와 같은 안전에 중요한 정보를 수신하지 못하도록 하는 공격을 의미한다.
32) 데이터를 훼손하거나 통신 채널을 무력화하기 위해 신호를 송출하는 공격을 의미한다.
33) 하나의 공격 차량이 여러 개의 가짜 신원을 만들어 마치 여러 대의 차량이 존재하는 것처럼 행세하여 다른 차량과 교통관리 시스템에 혼동을 주는 공격방식을 의미한다.
34) 이상의 다양한 공격방식에 관하여는, Chowdhury, Islam and Khan, *op. cit.*, p.49.
35) *Ibid.*, p.48.
36) 펌웨어 업데이트에 대한 적극적 개선조치가 필요하고, 그 업데이트 내용은 사전에 관할 행정청에 보고해야 한다는 견해로, 김상태, “자율주행자동차에 관한 법적 문제”, 「경제규제와 법」, 제9권 제2호, 서울대학교 공익산업법센터, 2016, p.184.

(3) 인공지능 특유의 위협

자율주행자동차의 사이버 위협이 전통적 정보통신 시스템의 위협과 본질적으로 구별되는 지점은 인공지능의 인지·판단 과정 자체를 겨냥하는 새로운 유형의 공격이 존재한다는 데 있다. 인공지능에 대한 공격은 시점과 목표에 따라 나뉜다. 훈련 시점의 공격으로는 학습데이터에 트리거를 심어 특정 입력에서 공격자가 원하는 출력을 내게 하는 백도어(backdoor)와 학습데이터를 오염시켜 성능을 떨어뜨리는 포이즈닝(poisoning)이 있다. 학습이 끝난 모델이 실제로 운용되는 추론 시점의 공격으로는 사람이 알아차리기 어려운 미세한 교란을 입력에 가해 모델의 오판을 유도하는 회피 공격(evasion attack), 연산량과 처리 시간을 급격히 늘리는 입력을 주입하여 시스템의 응답을 지연시키거나 무력화하는 스펀지 공격(sponge attack), 그리고 반복적인 질의를 통해 모델의 기능이나 매개변수를 복제하는 모델 탈취(model stealing)가 있다.[38] 이러한 공격은 시스템 내부에 침입하지 않고 환경을 조작하는 방식만으로도 인공지능의 판단을 교란할 수 있다는 점에서 특히 위험하다. 대표적인 사례는 다음과 같다.

먼저, 도로변 전광판에 보행자 이미지를 짧은 순간 투사하는 이른바 '팬텀 공격(phantom attack)'이나, 도로 표지판에 특수한 스티커를 부착하여 속도 제한 표지판을 오인하게 만드는 적대적 공격이 있다. 인공지능 객체 탐지 모델은 인간에게는 인지되지 않는 미세한 시각적 교란에도 오류를 일으킬 수 있기 때문에 정지 표지판을 제한 속도 표지판으로 오인하도록 하는 등 치명적 결과를 초래할 수 있다. 다음으로, 라이다(LiDAR) 센서에 가짜 신호를 주입함으로써 존재하지 않는 장애물을 만들어 차량을 급정거하게 만들거나, 반대로 실제 장애물을 인식에서 지워버리는 센서 스푸핑(sensor spoofing)도 심각한 문제를 발생시킨다. 마지막으로, 인공지능 모델의 학습 데이터를 손상하거나 악의적인 데이터를 주입하여 인공지능 모델의 내부 논리 자체를 근본적으로 변경하는 데이터 오염(data poisoning) 공격이 있다. 데이터 오염 공격은 시스템의 코드를 노리는 여타의 공격과 달리 학습 과정 자체를 표적으로 삼는다는 특징이 있다. 예를 들어, 학습 과정에 개입한 공격자가 정지 표지판 하단에 작은 스티커를 붙인 이미지를 학습데이터에 섞어 넣으면, 모델이 같은 스티커가 붙은 정지 표지판의 90% 이상을 제한 속도 표지판으로 오인하게 만드는 백도어 공격

37) Georgia Dede et al., *op. cit.*, p.43.
38) Kathrin Grosse and Alexandre Alahi, *op. cit.*, pp.4-6.

이 가능하다.[39] 이러한 공격은 정상적인 입력에 대한 정확도에는 거의 영향을 주지 않으므로 통상적인 성능 검증으로는 발견하기 어렵다.

3. 현행 「자동차관리법」상 사이버보안 관리체계의 내용

현행 「자동차관리법」은 '자동차 사이버공격 · 위협'을 해킹, 컴퓨터바이러스, 서비스 거부, 전자기파 등 전자적 수단으로 자동차의 부품 · 장치 · 정보통신기기 또는 이와 관련된 정보시스템을 침입 · 교란 · 마비 · 파괴하거나 자동차의 소프트웨어, 자동차제어 정보 등을 위조 · 변조 · 훼손 · 유출하는 행위 및 그와 관련된 위협으로 정의하고, '자동차 사이버보안 관리체계'를 그러한 공격 · 위협으로부터 자동차를 보호하기 위한 관리적 · 기술적 · 물리적 보호조치를 포함한 종합적 관리체계로 정의한다.[40]

「자동차관리법」 제30조의9는 "자동차 사이버보안 관리체계 인증 등"이라는 제명 아래 인증제도의 골격을 규정한다. 동조 제1항은 자동차의 종류 · 생산수량 · 기능 등을 고려하여 국토교통부령으로 정하는 자동차에 대하여 자동차자기인증을 하려는 자동차제작자등은 자동차 사이버보안 관리체계를 수립하여 국토교통부장관의 인증을 받도록 규정한다. 제3항은 국토교통부장관이 인증기준을 고시로 정할 수 있도록 하면서 그 인증기준에 관리적 · 기술적 · 물리적 보호조치가 포함되어야 한다고 규정하고, 제4항은 인증기준을 정할 때 미리 과학기술정보통신부장관과 협의하도록 한다.[41] 나아가 같은 법 제30조의10은 국토교통부장관이 자동차 사이버보안 관리체계의 안전성 · 신뢰성 확보를 위하여 필요하다고 인정하는 경우에는 자동차 사이버보안 관리체계 수립 및 운영에 관한 자료의 제출을 요구할 수 있도록 한다. 제30조의11은 인증을 받은 제작자 등이 인증기준에 적합하지 아니하게 된 경우 등에 대하여 국토교통부장관이 인증을 취소하거나 그 효력을 정지시킬 수 있도록 하여, 인증의 실효성을 담보하는 사후적 제재수단을 마련하고 있으며, 제30조의12는 사이버공격 · 위협과 관련된 사고가 발생하면 즉시 신고하도록 한다.[42]

또한, 「자동차관리법」은 제34조의5를 신설하여 소프트웨어 업데이트에 관한

---

39) Tianyu Gu, Brendan Dolan-Gavitt and Siddharth Garg, "BadNets: Identifying Vulnerabilities in the Machine Learning Model Supply Chain", arXiv preprint arXiv:1708.06733, 2017, Section 5.

40) 「자동차관리법」 제2조제4호의3 · 제4호의4.

41) 「자동차관리법」 제30조의9제1항 · 제3항 · 제4항.

42) 「자동차관리법」 제30조의10부터 제30조의12까지.

안전관리 의무를 법률에 명문화하였다. 같은 조 제1항은 제작자등이 소프트웨어 업데이트를 실시할 때 준수하여야 할 사항으로 다음 다섯 가지를 열거하고, 그 밖의 사항은 국토교통부령에 위임한다. 업데이트 후에도 자동차의 모든 장치와 기능이 정상적으로 작동하여야 하고(제1호), 업데이트된 자동차가 안전기준에 적합하여야 하며(제2호), 업데이트는 사이버공격·위협으로부터 보호되는 상태에서 안전하게 실시되어야 한다(제3호). 또한 제작자등은 업데이트에 관한 정보를 사용자에게 사전 및 사후에 제공하여야 하고(제4호), 업데이트의 내용과 이력을 기록·보관하면서 그 훼손과 위·변조를 방지하여야 한다(제5호). 업데이트는 취약점을 보완하는 수단이면서 동시에 새로운 공격 통로가 될 수 있다는 점에서 업데이트 경로의 위험을 직접 규율하고 있는 제3호는 특히 중요하다.

4. 인공지능 요소와 자동차 법제의 한계

「자동차관리법」의 사이버보안 관리체계는 차량이라는 제품의 관점에서 설계되었다. 그러나 자율주행자동차의 사이버 위험 가운데 가장 주목해야 할 부분은 인공지능의 학습과 추론 과정을 겨냥한 위협이다. 기존의 자동차 사이버보안 표준과 기능안전 표준은 이러한 위협에 충분히 대응하지 못한다. 자동차에 탑재되는 전기·전자 시스템의 오작동으로 인한 사고와 인명 피해를 방지하기 위해 제정된 자동차 기능안전 국제표준(ISO 26262)과 같은 영역별 안전 표준은 고위험 인공지능 시스템에 요구되는 사항의 일부를 다루지만, 안전·투명성·인간 감독에 관한 인공지능 특유의 요소는 다루지 않는다.[43] 「자동차관리법」이 완성된 자동차를 규율 대상으로 삼아 그 제작자의 관리체계를 인증한다면, 인공지능기본법은 차량에 탑재된 인공지능시스템을 규율 대상으로 삼아 인공지능사업자에게 위험관리 책무를 부과한다. 따라서 자율주행자동차의 사이버보안 관리체계를 인공지능과의 관계에서 검토하기 위해서는 「자동차관리법」과 함께 인공지능기본법을 살펴보아야 한다.

Ⅲ. 자율주행자동차의 고영향 인공지능 해당 여부와 헌법적 쟁점

43) Jessica Kelly et al., “Navigating the EU AI Act: A Methodological Approach to Compliance for Safety-critical Products”, 2024., p.1.

1. 인공지능기본법의 규율 체계 개관

(1) 인공지능기본법의 제정 경위와 입법 목적

인공지능이 산업 전반을 넘어 사회구조 자체를 재편하는 기반기술로 부상하면서, 그 잠재적 혜택과 함께 기술적 한계 및 오·남용에서 비롯되는 위험에 대한 규범적 대응의 필요성이 전 세계적으로 제기되어 왔다. 이러한 국제적 흐름 속에서 우리나라는 국가 차원의 인공지능 경쟁력 제고와 신뢰할 수 있는 인공지능 활용 기반 조성이라는 두 가지 정책목표를 동시에 달성하기 위한 법적 토대를 마련하고자 하였다. 그 결과 인공지능기본법이 2024년 12월 26일 국회 본회의를 통과하여, 2025년 1월 21일 법률 제20676호로 제정·공포되었으며, 1년의 경과기간을 거쳐 2026년 1월 22일부터 시행되었다. 이로써 대한민국은 2024년 인공지능법(AI Act)을 제정한 유럽연합(EU)에 이어 세계에서 두 번째로 인공지능 일반에 관한 포괄적 법제를 갖춘 국가가 되었다.[44]

인공지능기본법은 그 제명에서 드러나듯 '발전'과 '신뢰 기반 조성'이라는 두 축을 병렬적으로 표방한다. 제1조는 "인공지능의 건전한 발전과 신뢰 기반 조성에 필요한 기본적인 사항을 규정함으로써 국민의 권익과 존엄성을 보호하고 국민의 삶의 질 향상과 국가경쟁력을 강화하는 데 이바지"하는 것을 입법목적으로 명시하고 있다. 즉 인공지능기본법은 한편으로 인공지능에 관한 국가 차원의 거버넌스 체계를 정립하고 산업을 체계적으로 육성하면서, 다른 한편으로는 인공지능으로 인하여 발생할 수 있는 위험을 사전에 예방함으로써 국민의 기본권을 보호한다는 이중적 지향을 갖는다. 이러한 입법태도는 권리 보호를 위한 예방적 규제에 무게를 두는 EU 인공지능법과 달리, 산업 진흥과 신뢰 확보를 절충한 이른바 혼합형(hybrid) 규율모델에 해당하며 하나의 범주에 해당하면 단일 주무부처가 일관되게 규제하는 수평적 규제체계를 채택한 것으로 평가된다.[45] 다만, 이 법에 대해서는 그 자체로 완결된 법이 아니라 '토대와 얼개만 마련하여 내용의 보충이 필요한 법이며, 규제와 진흥을 모두 다루되 진흥에 주안점을 둔 법'이라는 비판적 평가도 존재한다.[46]

---

44) 이희옥, "인공지능의 위험에 관한 기본권 논의와 입법 대응", 「공법연구」, 제53권 제3호, 한국공법학회, 2025, 253면.

45) 김나루, "우리나라의 인공지능 기본법 제정에 따른 법적 과제", 「헌법학연구」, 제31권 제2호, 한국헌법하회, 2025, 264면.

46) 정찬모, "인공지능기본법 주요 내용의 검토와 향후 과제", 인하대학교 「법학연구」, 제28권 제1호, 인하대학교 법학연구소, 2025, 85면.

(2) 고영향 인공지능 개념의 구조와 비판

인공지능기본법 제2조제4호는 고영향 인공지능을 “사람의 생명, 신체의 안전 및 기본권에 중대한 영향을 미치거나 위험을 초래할 우려가 있는 인공지능시스템으로서 다음 각 목의 어느 하나의 영역에서 활용되는 것”으로 정의한다. 세부적으로는 에너지 공급, 먹는물 생산, 보건의료, 의료기기와 디지털의료기기, 핵물질과 원자력시설, 범죄 수사·체포를 위한 생체인식정보의 분석·활용, 채용·대출 심사 등 권리·의무 관계에 중대한 영향을 미치는 판단·평가, 교통수단·교통시설·교통체계의 주요한 작동 및 운영, 국가기관등의 의사결정, 초·중등 학생 평가에 관한 영역이 규정되어 있다.[47] 이 정의는 인공지능이 활용되는 영역의 특수성과 그 활용으로 인하여 초래될 수 있는 법익 침해의 중대성이라는 두 개의 표지를 포함하고 있다. 즉 고영향성은 인공지능 기술 자체의 성능이나 규모가 아니라, 그것이 어디에 그리고 어떻게 활용되는가라는 맥락적 요소에 의하여 결정된다. 이 점에서 고영향 인공지능 규율은 인공지능 모델의 학습 연산량[48]이라는 정량적 기준에 의존하는 안전성 확보 의무(제32조)의 규율 구조와 본질적으로 구별된다.

이러한 정의를 구체화하기 위하여 과학기술정보통신부는 「고영향 인공지능 판단 가이드라인」(이하 “가이드라인”이라 함)을 통하여 고영향 인공지능 해당 여부를 단계적으로 판단하는 2단계 판단 구조를 제시하였다. 이는 ① 인공지능 시스템이 특정 영역에서 활용되는지 여부를 먼저 판단하고(제1단계), ② 그 활용이 법익에 중대한 영향을 미치거나 위험을 초래할 우려가 있는지 여부를 다시 평가하는(제2단계) 구조로서, 두 단계를 모두 충족하는 경우에 비로소 고영향 인공지능으로 분류된다.[49] 고영향 해당 여부에 대한 판단은 결국 각 영역의 특성을 고려하여 생명 등에 대한 위험을 판단하는 기준이 수립되어 있어야 가능한데,[50] 그 기준과 적용 예시는 가이드라인으로 제공된다.[51]

인공지능기본법이 채택한 고영향 인공지능 개념에 대해서는 그 용어 선택을 둘러싸고 학계의 비판이 제기된다. EU 인공지능법은 고위험과 고영향을 구분

47) 인공지능기본법 제2조 제4호가목부터 차목까지.
48) 인공지능기본법 시행령 제24조제1항제1호에서는 학습에 사용된 누적 연산량이 10의 26승 부동소수점연산 이상일 것을 요구하고 있다.
49) 과학기술정보통신부, 「고영향 인공지능 판단 가이드라인」, 2026. 1., 20 - 21면.
50) 채은선, “우리나라 「인공지능기본법」의 주요 내용 및 관련 고려사항”, 「경제규제와 법」 제18권 제1호, 서울대학교 공익산업법센터, 2025, 105면.
51) 인공지능기본법 제33조제3항.

하여 고위험을 중심으로 규제 체계를 잡았으나, 우리나라의 경우 입법의 마지막 단계에서 고위험(high-risk)을 고영향(high-impact)으로 단순 치환한 것에 불과하여 그 타당성이 부족하다는 것이다.[52] EU 인공지능법이 인공지능을 "금지된 위험 · 고위험 · 제한적 위험 · 최소 위험"의 4단계로 분류하는 것과 달리 우리 법은 고영향과 생성형으로 이분하는 구조적 차이를 지적하면서, 고영향 인공지능 정의의 모호성이 법적 안정성과 사회적 수용성을 저해할 수 있다는 비판도 제기된다.[53]

(3) 고영향 인공지능 규율의 내용

인공지능기본법이 고영향 인공지능에 관하여 규율하는 내용은 고영향 인공지능 확인, 고영향 인공지능과 관련한 사업자의 책무 및 고영향 인공지능 영향평가이다. 먼저, 인공지능사업자는 인공지능 또는 이를 이용한 제품 · 서비스를 제공하는 경우 그 인공지능이 고영향 인공지능에 해당하는지를 사전에 검토하여야 하고, 필요한 경우 과학기술정보통신부장관에게 확인을 요청할 수 있다. 과학기술정보통신부장관은 요청이 있으면 고영향 인공지능에 해당하는지 여부를 확인해야 하고, 필요한 경우 전문위원회의 자문을 받을 수 있으며, 고영향 인공지능의 기준과 예시 등에 관한 가이드라인을 수립 · 보급할 수 있다.[54] 다음으로, 고영향 인공지능 또는 이를 이용한 제품 · 서비스를 제공하는 인공지능사업자는 위험관리방안의 수립 · 운영, 최종결과 · 주요 기준 · 학습용데이터 개요에 대한 설명 방안의 수립 · 시행, 이용자 보호 방안의 수립 · 운영, 고영향 인공지능에 대한 사람의 관리 · 감독, 조치 내용을 확인할 수 있는 문서의 작성 · 보관을 포함하는 조치를 이행하여야 한다. 과학기술정보통신부장관은 그 구체적 사항을 고시하고 준수를 권고할 수 있으며, 사업자가 다른 법령에 따라 이에 준하는 조치를 이행한 경우에는 이 조치를 이행한 것으로 본다.[55] 마지막으로, 고영향 인공지능을 이용한 제품 · 서비스를 제공하는 사업자는 사전에 사람의 기본권에 미치는 영향을 평가하기 위하여 노력하여야 한다.[56] 국가기관등이 영향평가를 실시한 제품 · 서비스를 우선적으로 고려하도록 함으로써

52) 인공지능기본법상 고영향 인공지능의 정의가 실제로는 EU 인공지능법의 고위험 인공지능과 다를 바 없다는 견해로, 정찬모, 앞의 글, 76-77&91-93면.
53) 신의수 · 한재경, "인공지능(AI) 활용 적정화를 위한 법제도 고찰 – 「AI 기본법」을 중심으로", 「공공사회연구」, 제15권 제1호, 한국공공사회학회, 2025, 213면.
54) 인공지능기본법 제33조.
55) 인공지능기본법 제34조 제1항부터 제3항까지.
56) 인공지능기본법 제35조.

영향평가를 받을 유인을 제공하고 있다.

2. 자율주행자동차의 고영향 인공지능 해당성 검토

(1) 자율주행 단계와 운전자 개입 정도의 규범적 의미

가이드라인은 자율주행자동차에 적용된 인공지능을 ① 자율주행시스템, ② 자율주행시스템 이외에 차에 적용된 인공지능, ③ 교통시설·교통체계에 관한 기술로 나누어 고영향 해당성을 판단한다. 자율주행시스템에 관하여 가이드라인은 ㉠ 대체로 도로 주행을 목적으로 하는 기술로서 ㉡ 인공지능이 스스로 판단하여 자동차를 운전하는 데 필요한 의사결정을 내리고 ㉢ 도로 주행 중 사고 및 그로 인한 인명 피해 가능성이 있는 경우, 생명·신체의 안전에 중대한 영향을 미칠 가능성이 높아 원칙적으로 고영향에 해당한다고 보았다.[57] 가이드라인에 따르면 지정 도로와 같은 특정 조건에서 운전자의 개입 없이 자동차의 운행을 전제하는 조건부 완전자율주행시스템 이상은 고영향에 해당한다. 조건부 완전자율주행시스템은 인공지능이 주행 제어를 담당하고 급박한 위기 상황에서도 운전자의 개입이 제한되기 때문이다.[58] 이러한 단계별 구분은 운전에서 인간의 상시적 감독이 줄어들수록 고영향 인공지능에 해당해야 한다는 규범적 발상으로 이해된다.

(2) 자율주행자동차의 고영향 인공지능의 해당 여부

자율주행자동차가 고영향 인공지능에 해당하는지 여부를 판단하기 위해서는 먼저 자율주행기술의 단계적 분류와 그에 따른 운전 제어권의 귀속을 이해할 필요가 있다. 앞서 살펴본 SAE의 6단계 분류에 따르면, 레벨0부터 레벨2까지는 시스템이 주행을 보조하는 데 그치고 주행의 제어와 주행 중 변수 감지 및 그에 대한 책임이 모두 운전자에게 귀속되는 반면, 레벨3부터는 특정 조건의 구간에서 주행 제어와 변수 감지를 시스템이 담당하되 위험 상황에서는 운전자의 개입이 요청된다. 레벨4에 이르면 제한적 운행설계 영역에서 운전자의 개입 없이 시스템이 돌발상황까지 대처하며, 레벨5에서는 모든 조건에서 운전자가 불필요하게 된다. 이러한 단계 구분에서 규범적으로 결정적인 기준은 '운전자의 개입이 필요한지 여부'이다. 앞에서 살펴본 고영향 인공지능 판단의 제2

57) 과학기술정보통신부, 앞의 가이드라인, 90면.
58) 과학기술정보통신부, 앞의 가이드라인, 90면.

단계, 특히 자율성 · 의사결정 능력 요소와 인간 개입에 의한 제외 기준은 이러한 자율주행단계와 밀접한 관련이 있다.

SAE의 분류에서 레벨 3과 레벨 4 이상을 가르는 핵심은 개입 요청에 대응할 운전자가 필요한지 여부이다. 레벨 3은 자율주행자동차가 중요한 주행을 수행하되 운전자는 언제라도 주도권을 되찾을 수 있는 단계이고, 레벨 4는 특정 조건에서 자율주행자동차가 모든 주행을 수행하되 시스템 요청 시 운전자가 개입할 수 있는 단계, 레벨 5는 모든 여정 및 주행 조건에 걸쳐 어떠한 형태의 인간 개입도 없이 모든 운전 기능을 수행하는 단계로 정의된다.[59] 이처럼 레벨 4 이상에서 사람의 감독 및 개입이 거의 존재하지 않게 됨에 따라 인공지능시스템이 주요한 작동 및 운영에 전면적으로 관여하게 되며, 이는 시스템 자체의 안전성 · 견고성 · 사이버보안에 대한 규범적 요구를 한층 강화한다.

자율주행시스템에 인공지능기본법 상의 고영향 인공지능에 적용하면, 자율주행시스템은 도로 주행을 목적으로 하는 교통수단의 작동 및 운영에 활용되므로 인공지능기본법 제2조제4호아목에 해당하고, 자율주행시스템은 주행 중 사고와 그로 인한 인명 피해의 가능성을 내포하므로 생명 · 신체의 안전에 중대한 영향을 미칠 우려가 있다. 이러한 점에서 자율주행자동차는 인공지능의 사용으로 인간의 생명이나 신체에 위해를 초래할 수 있는 인공지능의 대표적 예로 거론되고 있다.[60] 유럽연합에서도 자율주행자동차의 활용 맥락이 본질적으로 고위험이라는 점이 널리 받아들여지고 있다.[61] 실제 EU 인공지능법에서도 자율주행자동차는 고위험 인공지능으로 분류되고 있다.[62]

(3) “주요한 작동 및 운영”에 대한 관여의 판단

가이드라인은 자율주행시스템을 인지, 판단, 제어의 세 부분으로 분류하고, 인공지능이 판단 · 제어에 관여하면 교통수단의 주요한 작동 및 운영에 관한

---

59) Anastasios Giannaros et al., *op. cit.*, p.498.

60) 이희옥, 앞의 논문, 258-259면.

61) Fernández Llorca and Gómez, *op. cit.*, p.5. 다만 EU 인공지능법은 자율주행자동차를 형식승인 법제(부속서 I 제B절)와 연결하여 고위험 범주에 두면서도, 고위험 요건을 직접 적용하지 않고 형식승인 법제의 위임·이행법령을 제정할 때 그 요건을 ‘고려’하도록 하는 간접적 방식을 취한다.

62) Michael Borrelli, Sean Musch and Shabbi Khan, “Applying the EU AI Act to the Automotive Industry: Ensuring Explainability and Transparency”, SSRN Scholarly Paper, 2024, pp.5-6, <https://doi.org/10.2139/ssrn.4819226>; Mert Keser, Youssef Shoeb and Alois Knoll, “How Could Generative AI Support Compliance with the EU AI Act? A Review for Safe Automated Driving Perception”, arXiv preprint, 2024, p.1, <https://doi.org/10.1109/ICVES61986.2024.10928135>.

의사결정에 관여하는 것을 의미하므로 고영향에 해당할 가능성이 높다고 본다.[63] 반면 인공지능이 인지 단계에만 적용되고 실질적 판단은 규칙 기반 알고리즘에 따르는 경우, 인공지능의 산출물이 의사결정을 좌우하는 주된 근거가 아니라면 고영향에 해당하지 않을 수 있으며, 사람에 의한 안전조치로 인공지능에 의한 의사결정을 적시에 교정·번복하여 생명·신체의 안전에 대한 중대한 영향 가능성을 극히 낮출 수 있는 경우에는 고영향 인공지능이 아닌 것으로 분류된다.[64] 가이드라인은 별도로 마련한 기타 확인 기준에서도 ① 인공지능의 사용 목적, ② 주요한 작동·운영 의사결정에 미치는 영향, ③ 사고 시 잠재적 피해의 정도, ④ 별도로 규정된 안전기준의 존재 여부를 종합하여 고영향 인공지능 여부를 검토하도록 하고 있다.[65] 고영향 인공지능에 해당하는지 여부에 대한 판단은 자율주행자동차에 탑재된 모든 인공지능 기능에 일률적으로 적용되는 것이 아니라, 개별 기능의 역할과 자율성의 정도에 따라 달라질 수 있다.[66]

3. 고영향 인공지능 규율에 관한 헌법적 검토 사항

(1) 고영향 인공지능 판단 기준의 불명확성

첫 번째 쟁점은 고영향 인공지능에 해당하는지 여부를 어떠한 기준으로 판단하는지가 불명확하다는 점이다. 인공지능기본법 제33조제1항에서는 인공지능사업자가 고영향 인공지능에 해당하는지 여부를 '사전에 검토하여야' 하고, '필요한 경우' 과학기술정보통신부장관에게 확인을 요청할 수 있다고 규정하고 있다. 즉, 고영향 인공지능에 해당하는지 여부에 관한 확인 요청은 선택적 사항이므로, 원칙적으로 인공지능사업자가 자율적으로 고영향에 해당하는지를 검토하여 분류할 수 있는 것이다. 이 분류 결과에 따라 인공지능기본법 제31조제1항에 따른 고지의무, 제34조에 따른 안전성·신뢰성 확보 조치, 제35조에 따른 영향평가 노력 의무가 적용되는데, 제34조제1항 위반은 사실조사와 시정명령 및 그 불이행에 대한 과태료 부과 처분으로 이어질 수 있다.[67] 또한, 고영향

63) 과학기술정보통신부, 앞의 가이드라인, 91면.
64) 과학기술정보통신부, 앞의 가이드라인, 91면.
65) 과학기술정보통신부, 앞의 가이드라인, 93-94면.
66) 가이드라인에 따르면 운전자의 착석 여부 및 안전띠 착용 여부 등을 감지하는 운전자 상태감시시스템에 대하여 운전자가 당연히 인지하고 있어야 할 사항을 보조하는 데 그치므로 고영향 인공지능에 해당하지 않는다.
67) 인공지능기본법 제40조 및 제43조제1항제3호.

인공지능에 해당하는지 여부를 판단하는 영역은 법률에 특정되어 있으나 '생명·신체의 안전 및 기본권에 중대한 영향을 미치거나 위험을 초래할 우려'는 수범자 입장에서 스스로 판단하기에 명확한 기준이 되기 어렵다. 그런데 법령에 대강의 기준도 없이 가이드라인으로 고영향 인공지능 여부가 결정되고 그 결과, 시정명령과 과태료 처분 대상의 범위까지 가이드라인으로 정해지는 것이 적절한지는 검토가 필요하다.[68] 시정명령 및 과태료 부과의 대상이 되는 의무의 적용 범위가 법률이 아니라 가이드라인에서 비로소 구체화된다면, 수범자는 법률 문언만으로는 자신이 규율 대상인지 예측하기 어렵기 때문에 명확성원칙에 비추어 볼 때 바람직하지 않다. 인공지능 기술의 변화 속도에 비추어 활용 영역별 세부 기준을 법률에 모두 규정하는 것은 불가능하다는 점을 고려하여, 적어도 판단의 구조와 핵심 표지, 그리고 제외 사유의 유형을 법률 또는 시행령에 규정하고, 가이드라인은 그 적용 예시를 제공하는 데 그치도록 할 필요가 있다.[69]

(2) 생명·신체에 미치는 영향에 따른 차등적 규율의 필요성

두 번째 쟁점은 고영향 인공지능 책무의 내용이 그 인공지능이 사람의 생명·신체의 안전에 미치는 위험의 크기에 비례하지 않는다는 점이다. 인공지능기본법 제2조제4호는 생명·신체의 안전에 대한 위험과 그 밖의 기본권에 대한 위험을 하나의 표지로 묶고, 같은 법 제34조제1항은 모든 고영향 인공지능에 동일한 안전성·신뢰성 확보조치를 이행하도록 책무를 부여하고 있다. 이러한 구조에서는 생명·신체에 대한 위험이 큰 원자력시설이나 자율주행자동차에 활용되는 인공지능시스템과 채용, 대출 심사처럼 생명·신체에 직접적인 영향을 주지 않는 인공지능시스템에 동일한 수준의 책무가 적용된다. 따라서, 생명·신체에 대한 위험과 그 밖의 기본권에 대한 위험을 구분하고, 전자에 대해서는 엄격한 수준의 규제와 위험관리를 하되, 후자에 대해서는 구체적 사안에 따라 차등적으로 규율하는 것이 타당하다.[70]

(3) 보호 대상의 공백

---

68) 가이드라인은 처분 대상자의 기준을 정하는 것이므로 「행정절차법」 제20조제1항의 처분기준을 정하는 성격을 가지며, 사전 의견수렴과 공포 등 행정규칙의 절차를 준수할 필요가 있다는 지적으로, 채은선, 앞의 논문, 106면.

69) 다만, 자율주행자동차는 생명·신체에 대한 위험이 영역의 성격상 분명하기 때문에, 레벨 4 이상의 자율주행시스템이 고영향에 해당하는지 여부가 상대적으로 명확할 수 있다.

70) 같은 취지로, 이희옥, 앞의 논문, 277면.

세 번째 쟁점은 고영향 인공지능의 안전성·신뢰성을 확보하는 조치를 통해 보호하는 대상이 지나치게 협소하다는 점이다. 인공지능기본법 제34조제1항에서는 고영향 인공지능의 안전성·신뢰성을 확보하기 위한 조치의 하나로 이용자 보호 방안(제3호)을 규정하고 있으나, 인공지능기본법 상 '이용자'는 인공지능제품 또는 인공지능서비스를 제공받는 자에 불과하다. 인공지능기본법은 인공지능제품 또는 인공지능서비스에 의하여 자신의 생명, 신체의 안전 및 기본권에 중대한 영향을 받는 자를 '영향받는 자'로 별도로 규정하고 있으나,[71] '영향받는 자'는 인공지능기본법 제3조의 기본원칙에서 설명을 제공받을 수 있어야 한다는 선언적 규정의 주체로만 규정되어 있을 뿐, 고영향 인공지능에 대한 인공지능사업자의 투명성 확보 의무 및 안전성·신뢰성 책무에서 보호 대상으로 명시되어 있지 않다.[72]

예를 들어, 자율주행자동차의 이용자는 운전자나 탑승자이지만, 사이버 공격의 피해자 또는 영향을 받는 자는 보행자, 자전거 이용자, 다른 차량의 운전자 및 탑승자가 될 수 있다. 그런데 보행자, 자전거 이용자, 다른 차량의 운전자 및 탑승자는 인공지능기본법상 '이용자'가 아니라 '영향받는 자'이다. 그 결과 인공지능기본법 제34조제1항제3호의 이용자 보호 방안에서 자율주행자동차에 대한 사이버 공격으로 가장 큰 위험에 놓이는 사람들이 제외되는 불합리한 결과가 발생한다.

4. 고영향 인공지능의 법적 효과와 자율주행자동차 사이버보안의 관계

(1) 안전성 확보의무와 사이버보안 위험관리의 규범적 연계

고영향 인공지능사업자의 책무, 특히 인공지능기본법 제34조제1항제1호에 따른 위험관리방안의 수립·운영은 자율주행자동차의 사이버보안과 직접적으로 연결된다. 자율주행자동차에 대한 사이버 공격은 차량의 주행 제어를 교란하여 탑승자 및 제3자의 생명·신체에 직접적 위협을 초래할 수 있고, 주행데이터·위치정보의 탈취를 통하여 개인정보자기결정권을 침해할 수 있다는 점에서, 고영향 인공지능이 초래할 수 있는 위험의 본질적 일부를 구성한다. 따라서 자율주행자동차에 대한 사이버 위협의 식별·분석·평가·처리는 인공지능기본법 제34조가 요구하는 위험관리방안의 핵심 내용이 된다.

---

71) 인공지능기본법 제2조제9호.
72) 정찬모, 앞의 논문, 78-79면; 채은선, 앞의 논문, 100면.

이러한 관련성은 자율주행자동차에 대한 사이버보안 관리체계를 단순한 기술적 과제가 아니라 고영향 인공지능 규율과 관련된 법적 의무로 격상시킨다. 즉 사이버보안 관리체계의 구축은 자율주행자동차 제작자가 고영향 인공지능사업자로서 부담하는 안전성·신뢰성 확보 조치의 이행으로 평가될 수 있다. 이러한 점에 비추어 볼 때, 「자동차관리법」상 사이버보안 관리체계는 인공지능기본법상 안전성·신뢰성 확보에 관한 책무를 기술적으로 구현하는 역할을 한다.

(2) 사이버보안 관리체계 인증에 의한 위험관리 책무의 대체

자율주행자동차의 제작자등은 인공지능기본법상 고영향 인공지능사업자의 책무, 「자동차관리법」상 사이버보안 관리체계 인증 의무, 나아가 자율주행자동차법상의 성능인증 및 안전운행 조치 의무를 동시에 부담해야 하는 중복 규제를 받을 수 있다. 이러한 이중·삼중의 규제는 사업자에게 불합리한 부담을 줄 수 있으므로 신중한 접근이 필요하다. 인공지능기본법은 이러한 중복규제의 문제를 입법적으로 해소하기 위한 장치를 두고 있다. 인공지능기본법 제34조제3항은 인공지능사업자가 다른 법령에 따라 안전성·신뢰성 확보 의무에 준하는 조치를 이행한 경우 해당 책무를 이행한 것으로 인정하도록 규정하고 있고, 인공지능기본법 시행령 [별표 1]은 그 구체적 이행조치 인정기준과 절차를 제시하고 있다. 특히 인공지능기본법 시행령 [별표 1] 제1호나목 및 다목에서는 인공지능사업자가 자율주행자동차법상의 성능인증 및 변경인증을 받고 「자동차관리법」상의 사이버보안 관리체계 인증 및 변경인증을 받아 같은 법 제34조의5에 따른 책무를 모두 이행한 경우에는 인공지능기본법 제34조제1항제1호의 위험관리방안에 관한 사업자 책무를 이행한 것으로 본다고 규정하고 있다.

자율주행자동차에 관한 이러한 이행간주 규정은 인공지능기본법의 위험관리 책무와 「자동차관리법」의 사이버보안 관리체계가 동일한 역할을 한다고 입법자가 평가한 것으로 이해된다. 다시 말해, 「자동차관리법」상 사이버보안 관리체계 인증의 취득과 이행은 고영향 인공지능에 해당하는 자율주행자동차에 대한 인공지능기본법상 위험관리 책무를 대신할 수 있게 되는 것이다.[73]

---

73) 인공지능기본법이 일반법으로서의 위상에 더하여 인공지능 규범체계의 구심점 역할을 하는 '가교적 입법(Bridging Legislation)'으로 기능해야 한다는 주장으로, 박종준, "인공지능기본법과 개별 법령 간 체계정합성에 관한 공법적 고찰", 「외법논집」 제50권 제1호, 한국외국어대학교 법학연구소, 2026, 247면 이하.

Ⅳ. 「자동차관리법」상 사이버보안체계 도입의 헌법적 의의

1. 사이버보안 관리체계 도입의 헌법적 의미

(1) 국가의 기본권보호의무

국가의 기본권보호의무는 국가가 국민의 기본권을 실현하고 보호하여야 할 의무로서 사적 주체에 의한 기본권침해에 있어 국가의 기본권보호의무가 인정되기 위해서는 보호대상성, 위험원의 사인성, 위법성(위해의 발생 또는 발생우려)가 요구된다.[74] 자율주행자동차에 대한 사이버 공격은 해커·악의적 제3자라는 사적 주체에 의한 침해 위험이다. 차량 제어권의 탈취, 제동·조향 시스템의 무력화는 이용자는 물론, 제3자의 생명과 신체에 직접적 위해를 가할 수 있는바, 이는 전형적으로 국가의 기본권보호의무가 발동되어야 하는 삼각관계(국가-가해자-피해자)를 형성한다. 즉 사이버보안 관리체계의 도입은 국가가 사적 주체에 의한 생명·신체 침해 위험으로부터 국민을 보호하기 위한 기본권보호의무의 이행으로 평가될 수 있는 것이다. 따라서 「자동차관리법」이 사전인증 방식의 사이버보안 관리체계를 선택한 것은 보호의무 이행을 위한 입법형성권의 정당한 행사로 이해될 수 있으며, 오히려 입법자가 사전적·계속적 위험관리체계를 구축함으로써 과소보호금지원칙이 요구하는 최소한의 보호수준을 넘어서는 적극적 보호조치를 마련한 것으로 평가할 수 있다.

국가는 자율주행자동차와 관련한 인공지능의 위험으로부터 국민의 생명·신체를 보호할 헌법적 의무를 부담한다고 볼 수 있다.[75] 인공지능 기반 자율주행자동차에 대한 해킹은 해당 자동차의 탑승자 및 제3자의 안전에 큰 위험을 초래할 수 있기 때문에 위험을 개인의 책임에만 맡길 수 없고 국가의 법적 규율이 필요하다.[76][77] 자율주행 인공지능에 대한 사이버 침해는 사망·중상으로 직

---

74) 정재황, 「헌법학」, 박영사, 2022, 645, 652면.

75) 헌법적 기본권 보호의 관점에서 인공지능 기술이 기본권을 침해할 우려가 커지고 있음을 전제로, 유럽연합과 미국이 시민의 기본권을 보호하기 위한 규제 체계를 구축하고 있으며 우리나라도 기술 혁신과 기본권 보호의 균형을 유지하는 한국형 규제모델이 필요하다는 견해로 박혜란, "인공지능 기술의 기본권 침해 대응 방안 연구", 「성균관법학」, 제36권 제3호, 성균관대학교 법학연구원, 2024, 85면.

76) 신의수·한재경, 앞의 논문, 212면.

77) 인간의 생명·신체에 위해를 초래할 수 있는 인공지능의 대표적 예로 자율주행자동차를 들면서, 그 오작동으로 인한 사고 위험을 위험사회에서의 위험관리 대상으로 파악하고 국가가 이에 대한 사전적·예방적 규율을 통해 생명·신체를 보호할 책무를 부담해야 한다는 견해로, 이희옥, 앞의 논문, 258-259면.

결된다는 점을 고려할 때,[78][79] 자율주행자동차에 대한 사이버 위협으로부터 국민의 생명·신체를 보호할 국가의 헌법적 의무를 인정하는 것이 타당하다.

나아가, 헌법 제34조제6항은 "국가는 재해를 예방하고 그 위험으로부터 국민을 보호하기 위하여 노력하여야 한다"고 규정하여, 위험으로부터의 국민 보호를 국가의 헌법적 의무로 명시한다. 헌법 전문이 "우리들과 우리들의 자손의 안전과 자유와 행복을 영원히 확보할 것"을 선언하고 있는 점과 더불어, 이 조항은 안전 보장이 단순한 정책목표가 아니라 헌법 차원의 국가적 책무임을 분명히 한다. 사이버 공격으로 인한 자율주행자동차의 대규모 오작동은 개별적 교통사고의 차원을 넘어선다. 자율주행자동차에 대한 조직적 사이버 공격은 인공지능시스템의 취약점을 노린 구조적 위험이며, 그 규모와 파급효과를 고려할 때 사회재난[80]의 성격을 갖는다. 자율주행자동차에 대한 사이버보안 관리체계의 구축은 헌법 제34조제6항에서 규정하고 있는 재해예방의무의 구체적 이행으로 볼 수 있을 것이다.[81]

(2) 교통안전 및 공공 인프라에 대한 위험 해소

자율주행자동차에 대한 사이버 공격은 운전자뿐 아니라, 보행자, 그 밖의 차량 및 관련 인프라의 안전에 직접적인 영향을 미칠 수 있다. 도로교통은 다수의 참여자가 상호 신뢰를 전제로 작동하기 때문에, 차량 한 대의 오작동이라 할지라도 연쇄적으로 다른 이용자에게 위험이 전가될 수 있다. 따라서 자율주행자동차에 대한 사이버 공격은 개별적 기본권 침해를 넘어 일반적 교통안전이라는 공공의 법익에 대한 구조적 위험으로 확장된다. EU 인공지능법은 도로교통의 관리 및 운영에서 안전 구성요소로 사용되는 인공지능을 중요기반시설 영역의 고위험 인공지능으로 분류하고 있는바,[82] 이는 도로교통 인프라에 통합된 인공지능의 침해가 공공안전 전반을 위협할 수 있다는 규범적 인식을 반영한다.

---

78) Kathrin Grosse and Alexandre Alahi, *op. cit.*, p.23.

79) 자율주행자동차 시대에 운전자에게만 손해배상책임을 부담시키는 것은 형평의 원칙상 논란의 소지가 있어서 자동차제작자 등에 대해서도 보험가입의무를 부과해야 한다는 견해로, 김상태, "자율주행자동차에 관한 법적 문제", 「경제규제와 법」, 제9권 제2호, 서울대학교 공익산업법센터, 2016, 186면.

80) 「재난 및 안전관리 기본법」 제3조제1호에 따르면 재난은 자연재난과 사회재난으로 구분된다.

81) 「자동차관리법」상 사이버보안 관리체계 도입 이전의 자율주행자동차의 안전법제에 대하여 헌법 제34조제6항을 언급하고 있는 연구로, 윤성현, "자율주행자동차 시대 국민의 생명·신체의 안전보호를 위한 공법적 검토", 「헌법학연구」, 제22권 제3호, 한국헌법학회, 2016, 277면.

82) Regulation (EU) 2024/1689 (Artificial Intelligence Act), Annex III.

위험의 대규모성은 자율주행자동차 사이버보안 관리체계의 필요성에 대한 평가를 질적으로 변화시킨다. 개별 차량에 대한 산발적 공격이 개별 기본권 침해의 문제라면, 특정 차종이나 차량군 전체에 대한 동시다발적 공격은 다수의 생명·신체와 공공 교통질서를 동시에 위협하는 구조적 위험의 성격을 띠기 때문이다. 위험의 발생 확률은 개별적으로는 낮더라도 그 결과의 중대성과 광범위성을 고려하면, 사후적 손해배상이나 형사처벌만으로는 법익 보호가 불충분할 수 있다. 따라서 교통안전 및 공공 인프라에 대한 위험을 해소할 수 있다는 점 역시 자율주행자동차의 사이버보안 관리체계 도입의 헌법적 필요성을 뒷받침한다.

2. 개별 기본권 측면

(1) 생명권 보장의 실현

생명권은 우리 헌법에 명문의 규정은 없지만, 다른 모든 기본권이 행사되기 위한 전제로서 인정된다.[83] 헌법재판소는 생명권을 인간의 생존본능과 존재목적에 바탕을 둔 선험적·자연법적 권리로서 헌법에 규정된 모든 기본권의 전제가 되는 '기본권 중의 기본권'으로 파악하고 있다.[84] 국가는 국민의 생명권을 침해하지 않아야 한다는 소극적 의무는 물론, 국민의 생명권을 보호해야 하는 적극적 의무도 부담한다. 특히 오늘날 자주 발생하고 있는 테러행위로부터 국민의 생명을 보호해야 할 의무까지도 도출될 수 있다는 점[85]에서 자율주행자동차에 대한 사이버보안관리체계의 구축은 중요한 의미를 갖는다.

기존의 사이버 위협은 정보의 유출이나 훼손이라는 무형의 손해에 머물렀고, 그로 인한 생명 및 신체의 침해를 상정하기 어려웠다. 그렇기 때문에 사이버보안은 오랫동안 개인정보 보호와 재산적 손해를 중심으로 논의되었다. 그러나 자율주행자동차의 사이버보안에 대한 위협은 자동차 탑승자의 생명 및 신체에 직접적인 위해를 가할 수 있다. 앞서 살펴본 바와 같이 자율주행자동차는 인지·예측·계획·제어의 전 과정이 소프트웨어와 통신에 의하여 매개되는 사이버물리시스템이므로, 이에 대한 침해는 곧바로 물리적 위험으로 나타난다. 예를 들어, 공격자가 CAN 버스를 통하여 제동 명령을 발신하거나, 라이다 센서에 허위 신호를 주입하여 실재하는 장애물을 인식하지 못하게 만드는 순간, 탑승자의 생명권이 침해될 수 있는 것이다. 이러한 생명권의 침해는 자율주행

83) 김하열, 「헌법강의」, 박영사, 2022, 351면.
84) 헌재 1996. 11. 28. 95헌바1.
85) 김하열, 앞의 책, 354면.

자동차의 탑승자에게만 발생하는 것이 아니라 도로를 통행하는 보행자와 인접 차량의 탑승자에게까지 발생한다는 점에서 더욱 중대하다. 자율주행자동차에 대한 사이버 공격은 정보영역에서의 침해가 생명의 박탈로 직결된다는 점에서 자동차의 사이버보안관리체계는 생명권에 대한 핵심적 방어기제로 이해되어야 한다.

(2) 개인정보자기결정권의 보호

자율주행자동차는 운행 과정에서 위치정보, 주행궤적, 차량 내부 및 외부에 대한 영상, 생체정보 등 방대한 개인정보를 수집·처리하게 된다. 인공지능이 대규모 데이터를 처리하면서 개인정보가 무단으로 수집되거나 동의 없이 활용될 위험이 존재하며, 이는 헌법이 보장하는 개인정보자기결정권을 위협할 수 있다.[86] 이러한 정보가 사이버 공격에 의해 유출·조작될 경우 정보주체의 인격권과 사생활이 중대하게 침해될 수 있는바, 자율주행자동차의 사이버보안관리체계는 개인정보의 무결성과 기밀성을 보호함으로써 개인정보자기결정권을 실효적으로 보장하는 기능을 수행한다.

개인정보자기결정권은 단순히 국가의 정보수집을 방어하는 소극적 권리에 그치지 않고, 정보통신기술에 내재된 위험으로부터 개인정보를 보호하도록 국가에 요구하는 객관적 보호의무의 차원을 포함한다. 자율주행자동차에 대한 사이버 공격은 사인에 의한 개인정보자기결정권 침해의 전형적 사례이므로, 국가가 자동차의 제작자에게 정보보안 조치를 의무화하는 것은 개인정보자기결정권의 보호영역을 사인 간 관계로 확장하여 실현하는 입법적 조치로서 정당화될 수 있을 것이다.

Ⅴ. 결론

자율주행자동차는 운전이라는 인간 고유의 영역을 인공지능에 맡기는 패러다임 전환을 상징한다. 이러한 전환은 교통사고의 감소와 이동권의 확장이라는 편익을 약속하지만, 동시에 사이버 공격이 생명과 신체의 영역에 이르는 새로운 위험을 야기하기도 한다. 본 연구는 자율주행자동차의 사이버보안에 관하여 자동차 법제와 인공지능 법제를 비교하면서 헌법적 관점에서 검토하였다.

86) 박혜란, 앞의 논문, .63-64면.

먼저 자율주행자동차는 인지·예측·계획·제어의 전 과정이 소프트웨어와 통신으로 매개되는 사이버물리시스템이다. 그 위협은 차량 내·외부 통신망과 무선 업데이트 경로에 대한 공격에 그치지 않고, 적대적 공격·센서 스푸핑·데이터 오염처럼 인공지능의 인식·판단 과정 자체를 겨냥하는 공격으로 확장된다. 「자동차관리법」상 사이버보안 관리체계는 차량이라는 제품의 관점에서 설계된 것이어서 이러한 인공지능 특유의 위협을 충분히 포착하지 못할 수 있고, 이 지점에서 인공지능기본법이 개입할 자리가 생긴다.

다음으로 인공지능기본법상 고영향 인공지능의 판단구조를 분석하였다. 운전자의 개입 없이 자율주행시스템이 주행을 담당하는 조건부 완전자율주행 이상의 단계는 교통수단의 '주요한 작동 및 운영'에 활용되고, 생명·신체의 안전에 중대한 영향을 미칠 우려가 있으므로 고영향 인공지능에 해당한다. 그러나 고영향 인공지능 규율이 자율주행자동차에 적용되는 과정에서 세 가지 헌법적 문제가 드러난다. 먼저, 해당성 판단이 사업자의 자기 검토에 맡겨지고 그 기준이 가이드라인에서 비로소 구체화되는 구조는 명확성원칙에 비추어 바람직하지 않다. 다음으로, 생명·신체에 대한 위험과 그 밖의 기본권에 대한 위험을 하나의 표지로 묶어 동일한 책무를 부과하는 구조는 위험의 크기에 비례하는 규율이라고 보기 어렵다. 마지막으로, 사업자 책무의 보호 대상이 '이용자'에 머물러, 사이버 공격으로 가장 큰 위험에 놓이는 보행자와 다른 차량의 탑승자 같은 '영향받는 자'가 보호에서 빠진다.

나아가 「자동차관리법」상 사이버보안 관리체계 도입의 헌법적 의미를 검토하였다. 해커라는 사적 주체가 차량을 통하여 이용자와 제3자의 생명·신체를 위협하는 구조는 국가-가해자-피해자의 삼각관계를 이루며, 사이버보안 관리체계는 국가의 기본권보호의무를 이행하는 수단이다. 사전인증과 계속적 관리를 결합한 현행 체계는 과소보호금지원칙이 요구하는 최소한의 보호를 넘어서는 적극적 보호조치로 평가할 수 있다. 차량을 대상으로 한 공격이 사회재난의 성격을 띨 수 있다는 점에서, 이 체계는 헌법 제34조제6항의 재해예방의무를 구체화한 것이기도 하다. 또한 이 체계는 정보영역의 침해가 생명의 박탈로 직결되는 자율주행 환경에서 생명권을 보호하는 핵심적 방어기제이다. 아울러 방대한 위치·영상·생체정보의 무결성과 기밀성을 지킴으로써 개인정보자기결정권을 사인 간 관계에서 실현하는 장치로 기능한다. 이러한 점에 비추어 볼 때, 자율주행자동차의 사이버보안 관리체계는 단순한 기술적 안전장치가 아니라, 국민의 생명·신체와 개인정보를 지키는 헌법적 안전망으로 작용한다. 자율주행자동차는 인지·예측·계획·제어의 전 과정이 소프트웨어와 통신으로

매개되는 사이버물리시스템으로, 그에 대한 사이버 공격은 정보유출을 넘어 탑승자와 제3자의 생명·신체에 대한 직접적 침해로 이어진다. 특히 적대적 공격, 센서 스푸핑, 학습데이터 오염과 같이 인공지능의 인식·판단 과정 자체를 겨냥하는 위협은 차량이라는 제품의 관점에서 설계된 기존 자동차 법제만으로는 충분히 포착되지 않는다. 이 글은 이러한 문제의식에서 「자동차관리법」상 사이버보안 관리체계와 인공지능기본법상 고영향 인공지능 규율을 함께 검토하였다.

먼저 조건부 완전자율주행 이상의 자율주행시스템은 교통수단의 주요한 작동 및 운영에 활용되어 생명·신체의 안전에 중대한 영향을 미칠 우려가 있으므로 고영향 인공지능에 해당함을 확인하였다. 그러나 고영향 해당성 판단이 사업자의 자기 검토와 가이드라인에 의존하여 명확성원칙에 비추어 문제가 있고, 생명·신체에 대한 위험과 그 밖의 기본권 위험에 동일한 책무가 부과되어 비례적 규율이 이루어지지 않으며, 사업자 책무의 보호 대상이 '이용자'에 한정되어 보행자 등 '영향받는 자'가 보호에서 누락된다는 헌법적 쟁점을 도출하였다.

나아가 「자동차관리법」상 사이버보안 관리체계는 사인에 의한 생명·신체 침해 위험으로부터 국민을 보호하기 위한 국가의 기본권보호의무와 헌법 제34조 제6항의 재해예방의무를 이행하는 것으로서, 생명권과 개인정보자기결정권을 실효적으로 보장하는 헌법적 안전장치로 기능함을 밝혔다.

[참고문헌]

A. 저서
김하열, 헌법강의, 박영사, 2022.
정재황, 헌법학, 박영사, 2022.

B. 논문
1. 국내문헌

김나루, “우리나라의 인공지능 기본법 제정에 따른 법적 과제”, 「헌법학연구」 제31권 제2호, 2025.
김상태, “자율주행자동차에 관한 법적 문제”, 「경제규제와 법」 제9권 제2호, 서울대학교 공익산업법센터, 2016.
류병운, “사이버 공격으로부터 첨단 교통 시스템 보호 – 관련 입법례를 중심으로”, 「홍익법학」 제24권 제3호, 2023.
박종준, “인공지능기본법과 개별 법령 간 체계정합성에 관한 공법적 고찰”, 「외법논집」 제50권 제1호, 한국외국어대학교 법학연구소, 2026.
박혜란, “인공지능 기술의 기본권 침해 대응 방안 연구”, 「성균관법학」 제36권 제3호, 성균관대학교 법학연구원, 2024.
송문호, “자율주행자동차법제 개선방안”, 「법학논총」 제32권 제2호, 조선대학교 법학연구소, 2025.
신의수 · 한재경, “인공지능(AI) 활용 적정화를 위한 법제도 고찰 – 「AI 기본법」을 중심으로”, 「공공사회연구」 제15권 제1호, 한국공공사회학회, 2025.
윤성현, “자율주행자동차 시대 국민의 생명 · 신체의 안전보호를 위한 공법적 검토”, 「헌법학연구」 제22권 제3호, 한국헌법학회, 2016.
이희옥, “인공지능의 위험에 관한 기본권 논의와 입법 대응”, 「공법연구」 제53권 제3호, 한국공법학회, 2025.
정찬모, “인공지능기본법 주요 내용의 검토와 향후 과제”, 「법학연구」 제28권 제1호, 인하대학교 법학연구소, 2025.
채은선, “우리나라 「인공지능기본법」의 주요 내용 및 관련 고려사항”, 「경제규제와 법」 제18권 제1호, 서울대학교 공익산업법센터, 2025.

2. 해외문헌

Borrelli, Michael, “Applying the EU AI Act to the Automotive Industry: Ensuring Explainability and Transparency”, SSRN Scholarly Paper, 2024.

CANIS, Bill, “Issues in Autonomous Vehicle Testing and Deployment (CRS Report R45985)”, Congressional Research Service, 2021.

CHOWDHURY, Mashrur, Mhafuzul Islam and Zadid Khan, “Security of Connected and Automated Vehicles”, The Bridge, Vol.49 No.3, 2019.

DEDE, Georgia et al., “Cybersecurity Challenges in the Uptake of Artificial Intelligence in Autonomous Driving”, Publications Office of the European Union, 2021.

EBERS, M., Civil “Liability for Autonomous Vehicles in Germany”, SSRN Scholarly Paper, 2022.

FERNÁNDEZ LLORCA, David et al., “Testing Autonomous Vehicles and AI: Perspectives and Challenges from Cybersecurity, Transparency, Robustness and Fairness”, European Transport Research Review Vol. 17, 2025.

GU, Tianyu, Brendan Dolan-Gavitt and Siddharth Garg, “BadNets: Identifying Vulnerabilities in the Machine Learning Model Supply Chain”, arXiv preprint arXiv:1708.06733, 2017.

GIANNAROS, Anastasios et al., “Autonomous Vehicles: Sophisticated Attacks, Safety Issues, Challenges, Open Topics, Blockchain, and Future Directions”, Journal of Cybersecurity and Privacy Vol. 3, No. 3, 2023.

GROSSE, K., A. Alahi, “A Qualitative AI Security Risk Assessment of Autonomous Vehicles”, Transportation Research Part C: Emerging Technologies Vol. 169, Elsevier, 2024.

KELLY, Jessica et al., “Navigating the EU AI Act: A Methodological Approach to Compliance for Safety-critical Products”, 2024.

Keser, Mert, Youssef Shoeb and Alois Knoll, “How Could Generative AI Support Compliance with the EU AI Act? A Review for Safe Automated Driving Perception”, arXiv preprint, 2024,

MUSCH, Sean Michael, Charles Borrelli and Shabbi S Khan, Applying the EU AI Act to the Automotive Industry: Ensuring Explainability and

Transparency, 2024.

VELLINGA, Nynke E, “Connected and Vulnerable: Cybersecurity in Vehicles”, International Review of Law, Computers & Technology Vol.36, No.2, 2022.

[Abstract]

A Constitutional Analysis of Cybersecurity Systems for Autonomous Vehicles in the Age of Artificial Intelligence

Moon, Euibien*

Today's automobiles have evolved into "computers on wheels," integrating dozens of electronic control units, software systems, and communication technologies. As artificial intelligence increasingly replaces humans in driving tasks, the nature of cyber risk is undergoing a fundamental transformation. Cyber-attacks on autonomous vehicles extend beyond conventional data breaches and may directly endanger passengers and third parties by disabling braking or steering systems or manipulating sensor recognition. Threats that target AI perception and decision-making processes, including hostile attacks, sensor spoofing, and training data contamination, cannot be adequately addressed through existing automotive legislation, which was developed primarily from the perspective of the vehicle as a product. On the premise that autonomous-vehicle cybersecurity constitutes a matter of constitutional regulation extending beyond technical and industrial considerations, this paper examines the cybersecurity management system established under the "Automobile Management Act" and the regulation of high-impact artificial intelligence under the "Framework Act on Artificial Intelligence."

First, autonomous driving systems providing conditional full autonomy or higher were found to qualify as high-impact AI because they are used for the primary operation and management of transportation and may significantly affect the safety of life and physical well-being. However, the requirement that AI operators and guidelines determine whether a system constitutes high-impact AI raises concerns regarding the principle of clarity. In addition, the Framework Act on Artificial Intelligence imposes the same

* Professor of Kookmin University, College of Law

level of liability for risks to life, physical integrity, and other fundamental rights, irrespective of the gravity or severity of those risks. The analysis also identifies a constitutional concern arising from the limitation of AI operators' responsibilities to "users," thereby excluding "affected parties," such as pedestrians, from legal protection.

The cybersecurity management system established under the "Automobile Management Act" functions as a constitutional safeguard by concretizing the state's duty to protect fundamental rights and shield citizens from risks to life and physical integrity. It thereby contributes to the effective protection of the right to life and the right to informational self-determination.

Keywords : Autonomous Vehicle, Cyber Security Management System, Framework Act on Artificial Intelligence, High-Impact AI, State's Duty to Protect Fundamental Rights