



MASTER THESIS | MASTER'S THESIS

Titel | Title

Civil liability for Autonomous Vehicles under European Union
Law

verfasst von | submitted by

Mgr. Gréta Gáborová Bc.

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of
Master of Laws (LL.M.)

Wien | Vienna, 2026

Studienkennzahl lt. Studienblatt | Degree
programme code as it appears on the
student record sheet:

UA 999 082

Universitätslehrgang lt. Studienblatt |
Postgraduate programme as it appears on
the student record sheet:

Europäisches und Internationales Wirtschaftsrecht
[Vollzeit Englisch]

Betreut von | Supervisor:

ao. Univ.-Prof. Mag. Dr. Siegfried Fina

Table of Contents

Table of Abbreviations	2
Abstract (English)	3
Introduction.....	5
1. Foundations of automated vehicles	7
1.1 Automated vehicles and levels of automation	7
1.2 Artificial Intelligence systems	11
1.3 Establishing civil liability in accidents	14
1.4 Socio-economic implications and risk allocation	21
2. Regulatory challenges.....	23
2.1 Concept of defectiveness and liability standard	24
2.2 Trolley-type scenarios.....	29
2.3 Black-box Problem	33
3. Comparative evaluation of the EU civil liability framework.....	39
3.1 China	40
3.2 United Kingdom	42
4. Proposals for adjustments and future development.....	43
4.1 Defectiveness and allocation of liability.....	45
4.2 Trolley Problem	46
4.3 Black-box Problem	47
Conclusion	49
Table of literature	51
Table of legislation	55

Table of Abbreviations

ADAS	Advanced driver assistance systems
ADS	Automated driving system
AEVA 2018	The Automated and Electric Vehicles Act 2018
AI	Artificial Intelligence
AI Act	Regulation 2024/1689
AVA 2024	Automated Vehicles Act 2024
AV	Automated vehicles/Autonomous vehicles
DDT	Dynamic driving task
ECHR	European Convention on Human Rights
EU Charter	Charter of Fundamental Rights of the European Union
ODD	Operational design domain
PLD Directive	Directive 2024/2853
SAE	Society of Automotive Engineers
TFEU	The Treaty on the Functioning of the European Union
XAI	Explainable Artificial Intelligence

Abstract (English)

This master's thesis examines the legal challenges associated with the deployment of automated vehicles under European Union law. The research conducted has shown that the most problematic areas for traditional human-centric legislative approaches are the concept of defectiveness, trolley-type scenarios, and the inherent black-box nature of the essential AI algorithms for the future of the automotive industry. This thesis aims not only to assess the efficiency of the *de lege lata* framework but also to provide an interdisciplinary approach by addressing the technological perspective on automated driving systems and the practical feasibility of the future legislative proposals. The thesis is based on doctrinal legal research and comparative analysis of the legal frameworks of the European Union, the United Kingdom, and China governing automated vehicles. It further argues that, given the rapid technological advancements of the last decade, the legal framework struggles to keep pace and shift from traditional liability standards to standards tailored to AI models. In other words, the current European Union legal framework does not yet provide a coherent and practically enforceable civil liability regime for automated vehicles. In particular, the opacity of AI systems continues to undermine existing evidentiary mechanisms, while trolley scenarios and technically justified limitations of AI models reveal the major shortcomings of defect-based liability concepts. Finally, the thesis concludes with suggestions that future regulatory developments should focus on improving evidentiary mechanisms, enhancing explainability, ensuring a more balanced allocation of liability, and fostering innovation, rather than imposing strict liability on economic operators.

Keywords: Automated vehicles, AI systems, Blackbox, Legal framework, Product liability, Trolley Problem

Abstract (Deutsch)

Diese Masterarbeit untersucht die rechtlichen Herausforderungen, die der Einsatz automatisierter Fahrzeuge im Rahmen des Rechts der Europäischen Union mit sich bringt. Die durchgeführte Forschung hat gezeigt, dass die problematischsten Bereiche für traditionelle menschenzentrierte Gesetzgebungsansätze das Konzept der Fehlerhaftigkeit, Trolley-Szenarien sowie die inhärente Black-Box-Natur der für die Zukunft der Automobilindustrie wesentlichen KI-Algorithmen sind. Diese Arbeit zielt nicht nur darauf ab, die Effizienz des de lege lata Rechtsrahmens zu bewerten, sondern auch darauf, einen interdisziplinären Ansatz zu bieten, indem die technologische Perspektive auf automatisierte Fahrsysteme sowie die praktische Umsetzbarkeit zukünftiger Gesetzgebungsvorschläge berücksichtigt werden. Die Arbeit basiert auf rechtsdogmatischer Forschung und einer vergleichenden Analyse der Rechtsrahmen der Europäischen Union, des Vereinigten Königreichs und Chinas, die automatisierte Fahrzeuge regeln. Darüber hinaus vertritt sie die Auffassung, dass der geltende Rechtsrahmen infolge der dynamischen technologischen Entwicklungen der letzten Dekade zunehmend an seine Grenzen stößt, da er den Übergang von klassischen Haftungsstandards hin zu auf die Besonderheiten von KI-Modellen abgestimmten Haftungsregime bislang nur unzureichend vollzogen hat. Mit anderen Worten: Der gegenwärtige Rechtsrahmen der Europäischen Union bietet noch kein kohärentes und praktisch durchsetzbares zivilrechtliches Haftungsregime für automatisierte Fahrzeuge. Insbesondere untergräbt die Intransparenz von KI-Systemen weiterhin die bestehenden Beweismechanismen, während Trolley-Szenarien und durch KI-Modelle gerechtfertigte Einschränkungen die wesentlichen Schwächen fehlerbasierter Haftungskonzepte offenlegen. Abschließend kommt die Arbeit zu dem Ergebnis, dass sich zukünftige regulatorische Entwicklungen auf die Verbesserung von Beweismechanismen, die Förderung der Erklärbarkeit, die Gewährleistung einer ausgewogeneren Haftungsverteilung sowie die Förderung von Innovationen konzentrieren sollten, anstatt wirtschaftlichen Akteuren eine verschuldensunabhängige Haftung aufzuerlegen.

Schlüsselwörter: Automatisierte Fahrzeuge, Systeme der künstlichen Intelligenz (KI), Blackbox, rechtlicher Rahmen, Produkthaftung, Trolley Problem.

Introduction

This master's thesis focuses on the major legal challenges associated with the deployment of automated vehicles, with particular attention devoted to the concept of defectiveness, trolley-type scenarios and the black-box problem. Technological progress in Artificial Intelligence is fundamentally transforming the automotive sector by enabling levels of vehicle autonomy that were previously unimaginable. In the past decade, ideas from the last century about the use of deep neural networks were revisited due to the availability of large amounts of data and sufficient computational power. As a result, automated vehicles are increasingly transitioning from experimental projects to vehicles capable of operation on public roads, thereby challenging the suitability of traditional legal frameworks governing pre-market requirements, product safety, and civil liability for traffic accidents. Ever since the deployment of AI-based products, all jurisdictions have struggled to shift from traditional human-centric legislation to legal frameworks that would address the specific nature of algorithmic decision-making behind AI models. The challenge is particularly evident in the context of automated vehicles, where decision-making is delegated to complex software architectures that are capable of partially or fully replacing human drivers. The author chose to focus on automated vehicles, as they represent one of the most significant technological advances and have the potential to improve public road safety by reducing the primary cause of traffic accidents: human error. Given the early stage of regulatory development in this field, the author seeks to identify existing regulatory gaps and propose legal solutions that would ensure technological feasibility, legal certainty, and a balance among all parties involved, while respecting fundamental human rights. Moreover, the author gained practical experience with AI systems while working for an AI company, which has complemented the legal analysis with a technological perspective, facilitating an interdisciplinary approach to the challenges associated with AI-driven automated vehicles. This thesis is primarily intended for academic researchers and professionals in the respective fields, with the aim of initiating further discussion in this regard. The core sources used in the analysis for this thesis were European Union and foreign legislation, technical and legal research papers, academic literature, and recent legislative developments and materials produced by the European Union. Against this background, this thesis seeks to address the following research questions:

- 1. Does the combined application of the AI Act and the Product Liability Directive provide a coherent and practically enforceable civil liability framework for autonomous vehicles under EU law?*
- 2. How should the concept of defectiveness under the Product Liability Directive be interpreted?*
- 3. Are the evidentiary mechanisms introduced by the revised Product Liability Directive sufficient to overcome the opacity and limited explainability inherent in AI-based autonomous vehicle systems, given their black-box nature?*
- 4. Can defect-based liability address unavoidable harm in trolley-type scenarios involving autonomous vehicles?*

To answer the outlined research questions, the following hypotheses were formulated:

H1: The combined application of the AI Act and the Product Liability Directive does not establish a fully coherent and practically enforceable civil liability regime for autonomous vehicles.

H2: The black-box characteristics of AI systems and the occurrence of trolley-type scenarios reveal structural tensions within the framework of defect-based product liability.

This thesis is divided into four chapters. The first chapter introduces the technological and legal foundations of automated vehicles. Firstly, it explains different levels of automation according to the SAE standard, then it examines the connection between Artificial Intelligence systems and automated vehicles. Subsequently, it analyses the *de lege lata* civil liability provisions under European Union law, comprising the PLD Directive and the AI Act. The second chapter identifies and discusses the main challenges connected with the deployment of automated vehicles. Particular attention is devoted to defectiveness under the revised PLD Directive, trolley-type scenarios involving unavoidable harm and the black-box nature of AI systems. The author sees these areas as the most problematic for traditional legal frameworks. The third chapter provides a comparative assessment of the EU civil liability framework by analysing the legal approaches to automated vehicles adopted in the United Kingdom and China. The fourth chapter exposes the limitations of the current framework based on the analysis conducted in the second and third chapters. Based on the identified limitations, this thesis aims to formulate *de lege ferenda* proposals. The research is based on doctrinal legal analysis, complemented by deductive reasoning and comparative legal methodology.

1. Foundations of automated vehicles

This chapter introduces the foundations of automated vehicles. Firstly, it examines the European legal framework governing automated vehicles and provides a thorough explanation of the levels of automation described by the Society of Automotive Engineers. The distinction between levels of automation is essential for this thesis, as it determines the allocation of control over the vehicle for the dynamic driving task (DDT) and therefore may significantly affect the assessment of responsibility and civil liability in accidents involving automated vehicles. Secondly, the chapter further explores the intersection of AI and automated vehicles, as ADS systems are classified as high-risk AI systems and will therefore be subject to extensive regulatory obligations in the future. It also addresses the broader socio-economic implications of automated vehicles, including safety expectations and comparisons with traditional human drivers, and the distribution of risks. Finally, this chapter outlines the *de lege lata* state of the law for the establishment of civil liability for accidents involving AVs

1.1 Automated vehicles and levels of automation

The European legal framework defines an automated vehicle as “(...) *a motor vehicle designed and constructed to move autonomously for certain periods of time without continuous driver supervision but in respect of which driver intervention is still expected or required*;(...)”¹ and a fully automated vehicle as “(...) *a motor vehicle that has been designed and constructed to move autonomously without any driver supervision [...]*”(…)”² This indicates that the respective regulation partially recognises different levels of automation, as further explained in SAE International standard J3016, which establishes a Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles (hereinafter referred to as “SAE standard”). The standard outlines six levels

¹ Art. 3 (21) European Parliament and Council Regulation 2019/2144 of 27 November 2019 on type-approval requirements for motor vehicles and their trailers, and systems, components and separate technical units intended for such vehicles, as regards their general safety and the protection of vehicle occupants and vulnerable road users amending Regulation (EU) 2018/858 of the European Parliament and of the Council and repealing Regulations (EC) No 78/2009, (EC) No 79/2009 and (EC) No 661/2009 of the European Parliament and of the Council and Commission Regulations (EC) No 631/2009, (EU) No 406/2010, (EU) No 672/2010, (EU) No 1003/2010, (EU) No 1005/2010, (EU) No 1008/2010, (EU) No 1009/2010, (EU) No 19/2011, (EU) No 109/2011, (EU) No 458/2011, (EU) No 65/2012, (EU) No 130/2012, (EU) No 347/2012, (EU) No 351/2012, (EU) No 1230/2012 and (EU) 2015/166 [2019] OJ L325/1(Regulation 2019/2144).

²Art. 3 (22) Regulation 2019/2144.

of driving automation, which vary in allocation of dynamic driving tasks (DDT) between the human driver and ADS, ranging from Level 0, which represents no driving automation, to Level 5, with full driving automation. To fully understand the differences among automation levels, it is important to explain the core elements of the SAE classification, which are discussed hereinafter. DDT is defined as “*All of the real-time operational and tactical functions required to operate a vehicle in on-road traffic[.](...)*”³ It includes several subtasks such as lateral vehicle motion control (steering), longitudinal vehicle motion control (acceleration and deceleration), monitoring the driving environment (object and event detection, recognition, classification, and response preparation), object and event response execution, manoeuvre planning, and enhancing conspicuity (lighting, sounding the horn, signalling, or gesturing).⁴

The term “Operational design domain” describes *[o]perating conditions under which a given driving automation system or feature thereof is specifically designed to function, including, but not limited to, environmental, geographical, and time-of-day restrictions, and/or the requisite presence or absence of certain traffic or roadway characteristics.*⁵

The Minimal risk condition refers to a stable, stopped state that the vehicle is brought into by the user or ADS by performing a DDT fallback to reduce the risk of a crash.⁶ Within Levels 0 to 2, the human driver performs part or all of DDT. At level 0 (No driving automation), a human driver performs DDT at all times. The vehicle may provide warnings or momentary emergency intervention if a driving automation system is present. However, it does not perform any part of DDT continuously.⁷ At level 1 (Driver assistance), the driver performs DDT that is not performed by the driving automation system and continuously supervises its operation. The driver decides when to engage the system or disengage the system and immediately performs all DDT when required or desired.⁸ Meanwhile, when not disengaged by the driver, the driving automation system “*[p]erforms part of DDT by executing either the longitudinal motion or the lateral vehicle control subtask [.]*”⁹ Level 2 (Partial Driving Automation) corresponds to Level 1 in the role of the

³ SAE International, ‘Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles’ (SAE Standard J3016_202104, Revised April 2021, SAE International 2021) 8 <https://doi.org/10.4271/J3016_202104> accessed 9 March 2025 (SAE J3016_202104), 9.

⁴ SAE J3016_202104, 8.

⁵ SAE J3016_202104, 17.

⁶ SAE J3016_202104, 15.

⁷ SAE J3016_202104, 28.

⁸ SAE J3016_202104, 28.

⁹ SAE J3016_202104, 28.

driver; on the contrary, it is also capable of simultaneous control of longitudinal and lateral vehicle motion.¹⁰ Unlike at lower levels of automation (Levels 0 to 2), at Levels 3 to 5 the ADS performs the entire DDT while engaged. At Level 3, prior to engaging the ADS, the driver ensures the ADS-equipped vehicle is in operational readiness and determines whether to engage the ADS. While ADS is engaged, it performs the entire DDT within its ODD, and the driver becomes a fallback-ready user, meaning the driver must be receptive to a request to intervene and respond by performing in a timely manner. For instance, the driver must intervene when the ODD limits are about to be exceeded or when a DDT performance-relevant system failure occurs. In the situations mentioned above, the ADS disengages at the appropriate time after the request to intervene is issued, allowing the driver to resume control and determine how to achieve a minimal risk condition.¹¹ Level 4 (High Driving Automation) closely resembles Level 3, since the ADS performs the entire DDT within its ODD. The key difference lies in the DDT fallback. Unlike in Level 3, the ADS performs the DDT fallback itself and transitions the vehicle to a minimal risk condition when a performance-relevant system failure occurs, when the vehicle is about to exit its ODD, or at the user's request to achieve a minimal risk condition. Consequently, while ADS is engaged, the user is not required to perform the DDT or the fallback. However, the user may be prompted to resume vehicle operation when approaching the vehicle's ODD limits and may perform DDT after ADS reaches its ODD limit.¹² To sum up, Level 4 may therefore operate only during a specific part of the journey, depending on its ODD limits.¹³ Level 5 (Full Driving Automation) follows the same principles as Level 4, as the ADS performs DDT and DDT fallback. Level 5 differs from Level 4 in that it operates without ODD, allowing the system to operate under all driver-manageable on-road conditions.¹⁴ In other words, Level 5 may enable the ADS to perform the entire trip from origin to destination, regardless of, for example, weather conditions, time of day, or geographical restrictions, except under adverse conditions (white-out snow storm, flooded roads, glare ice)¹⁵ Another key point to keep in mind for Levels 4 and 5 is a possible delay in user-requested ADS disengagement, especially when an immediate disengagement could compromise safety. Therefore, the system first transitions the vehicle to a minimal risk

¹⁰ SAE J3016_202104, 28.

¹¹ SAE J3016_202104, 28.

¹² SAE J3016_202104, 28.

¹³ SAE J3016_202104, 32.

¹⁴ SAE J3016_202104, 28.

¹⁵ SAE J3016_202104, 32.

condition.¹⁶ This may raise several questions in determining liability for AV accidents, especially when considering the possibility of human intervention in situations where ADS malfunctions due to sensor failures, bias in training data, or faulty decisions. Moreover, in practice, levels of automation are not strictly separated but tend to overlap. The transition between individual levels is often gradual, meanwhile, the vehicles may operate in varying degrees of automation depending on several factors, such as the driving environment, system capabilities, and user interaction. As a result, the distinctions between levels of automation often become blurred. This overlap complicates liability allocation and the assessment of expected safety, as it is not always clear who was effectively in control at a given moment: the driver or the ADS. On the one hand, relying solely on formal levels of automation may be insufficient, and greater emphasis should be placed on the actual distribution of control between humans and the ADS. On the other hand, the distinction is necessary because it captures differences in control, technical capabilities, and responsibilities across different levels of automation. From the manufacturer's perspective, the higher the level of automation they claim, the greater the liability they face, which is also an important reason manufacturers tend to avoid classifying their systems at higher levels of automation and instead use lower ones. Furthermore, differences in regulatory approaches governing the automotive industry across jurisdictions are also an important factor, as manufacturers may be inclined to deploy and classify higher levels of automation in regions with less stringent regulations, such as the USA, the UAE, or China, while regions with stricter regimes, such as the European Union, represent a more cautious approach. This includes both pre-market requirements and applicable liability provisions.

While it seems theoretical, the distinctions between levels of automation can be understood by examining real-world examples of ADS across the automotive industry. The following examples will primarily be drawn from the USA environment, given the broader deployment of automated vehicles and the greater availability of their operational data, to more clearly distinguish the levels of autonomy. The Level 2 example is a basic version of Tesla Autopilot, which helps with steering, accelerating, and braking on highways while still requiring hands on the wheel, or Chevrolet Super Cruise, which allows hands-free driving in a mapped area but still requires the driver's alertness.¹⁷ The example of the Level

¹⁶ SAE J3016_202104, 28.

¹⁷ Neonode, 'Driving the Future – Levels of Autonomous Vehicles Explained' (*Neonode*, 7 February 2025) <<https://neonode.com/newsroom/news/driving-the-future-levels-of-autonomous-vehicles-explained>> accessed 9 April 2026.

3 currently deployed is a Mercedes-Benz Drive Pilot, which enables hands-off driving in heavy traffic at a maximum speed of 65 km/h on mapped highways, or Honda Sensing Elite, which introduced the traffic jam pilot, which can perform dynamic driving tasks in congested highway traffic. In comparison with Level 2 vehicles, Level 3 vehicles do not require continuous driver supervision.¹⁸ A notable example of Level 4 (full automation within limited operational environments capable of operating without human intervention under predefined conditions) is Waymo’s fully autonomous taxi services operating in selected areas of Phoenix, San Francisco, and Los Angeles. Level 5 is not yet deployed commercially, but some companies have made notable progress with their concept vehicles, such as Audi AI:CON, which embodies a concept of a fully autonomous vehicle, and it is also transformed into a “third living space” as it is capable of interacting at the same time with its surroundings and its occupants.¹⁹ Within the European Union, an interesting example, beyond the conventional levels of autonomy for standard road traffic, can be found in fully automated public transport systems operating in a controlled environment. A notable example is the driverless Copenhagen Metro, in service since 2002, with 30,000 daily departures and 370,000 daily passengers, managed from the control room and maintenance centre.²⁰ Notwithstanding the fact that the automated metro represents a successful example of a driverless transport mechanism, their operation cannot be compared directly to the automated vehicles deployed in road traffic as they operate within highly controlled environments with a dedicated infrastructure and limited external variables, predictable routes and most importantly, substantially reduced decision-making requirements.

1.2 Artificial Intelligence systems

Before undertaking an in-depth analysis of the provisions and rules relevant to automated AI systems, it is important to first introduce the AI Act, its objectives, scope, and key regulatory mechanisms. The AI Act represents the first initiative to comprehensively regulate Artificial Intelligence at the European Union level. Its objective

¹⁸ Neonode, ‘Driving the Future – Levels of Autonomous Vehicles Explained’ (*Neonode*, 7 February 2025) <<https://neonode.com/newsroom/news/driving-the-future-levels-of-autonomous-vehicles-explained>>.

¹⁹ Neonode, ‘Driving the Future – Levels of Autonomous Vehicles Explained’ (*Neonode*, 7 February 2025) <<https://neonode.com/newsroom/news/driving-the-future-levels-of-autonomous-vehicles-explained>>.

²⁰ Metro Service, ‘Metro Service and the Metro’ (*Metro Service A/S*) <<https://www.metroservice.dk/en/what-we-do/the-metro/>> accessed 26 May 2026.

is to transform the European Union into a global hub of trustworthy AI by establishing harmonised rules governing the development, placing on the market, and use of AI within the EU. It seeks to ensure the safety and respect for the fundamental rights and values on which the European Union was founded. Furthermore, it aims to promote investment and innovation in AI, improve governance and enforcement, and encourage a single EU market dedicated to AI.²¹ The AI Act establishes a horizontal regulatory framework governing the development, deployment, and use of AI systems across the European Union. It applies across all sectors and introduces a product safety framework built on a risk-based approach, distinguishing among four categories of AI systems based on the potential risk they pose. It imposes a set of requirements for market entry and certification of high-risk AI systems, including a mandatory CE marking procedure. The pre-market conformity regime also extends to the data used for training, testing, and validation to ensure equitable outcomes.²² The AI Act provides legal definitions for various parties, including providers, importers, deployers, distributors, and product manufacturers, leaving them accountable if they are involved in the development, use, import, distribution, or manufacturing of a product.²³ Another important point to mention is the scope of the AI Act, as Article 2 (1) a) and b) suggest extraterritorial application of the regulation, meaning that, according to these provisions, the regulation applies to all persons involved, including those established in third countries, provided that AI systems are placed on the market or put into service within the European Union or if the outputs created by the AI system are used within its territory. This could have major implications for automated vehicles, considering the common cross-border development of automated driving systems involving numerous parties.

In compliance with the Article 3 (1) of AI Act AI systems are defined as “*machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers,*

²¹ Konrad Meier and Roger Spichiger, ‘The EU AI Act : What it means for your business’ (EY, 15 March 2024) <https://www.ey.com/en_ch/insights/forensic-integrity-services/the-eu-ai-act-what-it-means-for-your-business> accessed 18 March 2026.

²² Mauritz Kop, ‘EU Artificial Intelligence Act: The European Approach to AI’ (2021) 2/2021 Stanford - Vienna Transatlantic Technology Law Forum, Transatlantic Antitrust and IPR Developments, Stanford University <<https://law.stanford.edu/publications/eu-artificial-intelligence-act-the-european-approach-to-ai/>> accessed 18 March 2026.

²³ Konrad Meier and Roger Spichiger, ‘The EU AI Act : What it means for your business’ (EY, 15 March 2024) <https://www.ey.com/en_ch/insights/forensic-integrity-services/the-eu-ai-act-what-it-means-for-your-business>.

from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments”²⁴.

According to the legal definition set out above, the algorithms used in automated driving systems fall within the scope of AI systems. AI systems used in automated vehicles, for instance, enhance a vehicle’s intelligence, safety, and efficiency by processing data from diverse driving scenarios, enabling the system to learn from road conditions and to make informed driving decisions, and monitor the driver’s alertness and fitness. In addition, AI systems tend to support non-safety-related functions, such as media, access, and payment services.²⁵ According to Article 6 (1) of the AI Act, an AI system is classified as high-risk if it is intended to function as a safety component of a product, or if the AI system itself is the product covered by Union harmonisation legislation, provided that such product is subject to a third-party conformity assessment.²⁶ Firstly, in the context of automated vehicles, the first condition mentioned above is undoubtedly met when an automated driving system incorporates AI algorithms and serves as a safety component. A typical example includes automated driving systems such as automatic emergency braking, adaptive cruise control, lane-keeping assist, traffic sign recognition, driver monitoring systems, and forward collision warning. It is important to keep in mind that most ADS systems are not 100% AI-based, by contrast, many of them are a combination of AI and traditional rule-based algorithms.²⁷ Nevertheless, this fact does not lead to different legal consequences, as the mere presence of an AI algorithm within such systems is sufficient to classify them as high-risk AI systems under the AI Act, provided they function as a safety component. To put it differently, the hybrid nature of safety systems does not diminish, nor does it alter, the regulatory consequences associated with their deployment. Secondly, the other condition is fulfilled as the respective safety components, being a part of the automated vehicle, are subject to third-party conformity assessment procedures in compliance with Art. 2 (2) AI Act in connection with Annex I section B. automated vehicle systems fall within the scope of Regulation (EU) 2018/858 on type-approval of motor

²⁴ European Parliament and Council Regulation 2024/1689 of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) [2024] OJ L 2024/1689 (AI Act).

²⁵ Thomas Kahl, ‘AI Act and the Automotive Industry – Where does the road lead?’ (*Taylorwessing*, 25 March 2025) <<https://www.taylorwessing.com/de/insights-and-events/insights/2025/03/ai-act-and-the-automotive-industry>> accessed 19 March 2026.

²⁶ Art. 6 (1) (a) and (b) AI Act.

²⁷ Frédéric Bouchard and others, ‘A Rule-Based Behaviour Planner for Autonomous Driving’ in Guido Governatori and Anni-Yasmin Turhan (eds), *Rules and Reasoning* (Springer 2022) 263, 263-264.

vehicles and Regulation (EU) 2019/2144 on general vehicle safety. Article 2 (2) of the AI Act results in the partial application of its high-risk systems provisions to an automated vehicle's AI systems. This ensures that the use of AI in the automotive industry will persist, mainly regulated by sectoral legislation, while AI Act provisions for high-risk systems will serve as complementary. Accordingly, existing sector-specific legislation will be aligned with and reflect the requirements imposed on high-risk systems under the AI Act. The respective requirements will be tailored to the automotive industry's specific needs. The non-regulated systems, which are not subject to the aforementioned sector legislation, must directly comply with the AI Act's standard rules and therefore undergo a standard risk assessment.²⁸ Another key point to mention is that according to Article 113, apart from certain provisions, the whole AI Act shall apply from 02.08.2026, apart from Article 6 (1) and the corresponding obligations set out in the AI Act are yet to enter into force as of 02.08.2027, leaving automotive industry regulations yet to be revised, however, it is expected to mirror the AI Act demands and obligations in a field of high-risk systems. High-risk systems must comply with numerous requirements and undergo a conformity assessment, which must be completed prior to the system's release to market. The respective systems also require registration in the EU database. Moreover, high-risk systems require an appropriate AI risk management system, logging capabilities, and human oversight, in addition to data governance for the training, testing, and validation datasets, while maintaining proper cybersecurity, robustness, and fairness of the system.²⁹ Regardless of their implementation, the concepts mentioned above are deemed relevant for the future of the automotive industry, as they may affect civil liability assessments, especially given the opacity of the system's decision-making process due to the black-box nature of AI algorithms, which will be further analysed in the following chapter.

1.3 Establishing civil liability in accidents

To fully comprehend the liability concepts set out in the PLD Directive, it is essential to consider the European Union's secondary acts applicable to the automotive

²⁸ Thomas Kahl, 'AI Act and the Automotive Industry – Where does the road lead?' (*Taylorwessing*, 25 March 2025) <<https://www.taylorwessing.com/de/insights-and-events/insights/2025/03/ai-act-and-the-automotive-industry>>.

²⁹ Konrad Meier and Roger Spichiger, 'The EU AI Act : What it means for your business' (*EY*, 15 March 2024) <https://www.ey.com/en_ch/insights/forensic-integrity-services/the-eu-ai-act-what-it-means-for-your-business>.

industry, which govern the approval of vehicles prior to their placement on the market. Regulation (EU) 2019/2144 establishes requirements for the type-approval of vehicles, systems, and their components, focusing on general safety, the protection of vehicle occupants and vulnerable road users, as well as specific aspects such as tyre pressure monitoring systems, fuel efficiency, CO₂ emissions, and the safety and environmental performance of tyres.³⁰ The Regulation (EU) 2018/858 establishes an administrative framework and technical requirements governing the type-approval and placing on the market of vehicles, as well as their systems, components, and separate technical units, including individual vehicle approvals. It further covers parts and equipment that may adversely affect essential vehicle systems and lays down market surveillance requirements.³¹ Another important point to emphasise is that AI Act requirements concerning high-risk AI systems have not yet been incorporated into the regulations mentioned above, as they have not yet entered into force. Although the automotive industry's specific revisions regarding high-risk AI systems are still pending, the AI Act text provides an early indication of the regulatory demands industry is going to face in the future.³²

The establishment of civil liability in automated vehicle accidents under European Union law is governed by the PLD Directive and the national civil liability provisions of the Member States. The directives are part of secondary European Union law, and according to Article 288 of TFEU, "*A directive shall be binding, as to the result to be achieved, upon each Member State to which it is addressed, but shall leave to the national authorities the choice of form and methods.*"³³ Meaning that the form and methods of achieving its aims may vary among the Member States. The respective directive provides a liability framework for economic operators for damage to natural persons resulting from defective products and regulates compensation for such damage.³⁴ Its scope includes "*(...) products placed on the market or put into service after 9 December 2026.*"³⁵

³⁰ Art. 1 Regulation 2019/2144.

³¹ European Parliament and Council Regulation 2018/858 of 30 May 2018 on the approval and market surveillance of motor vehicles and their trailers, and of systems, components and separate technical units intended for such vehicles, amending Regulations (EC) No 715/2007 and (EC) No 595/2009 and repealing Directive 2007/46/EC (Text with EEA relevance.) [2018] OJ L 151/1 (Regulation 2018/858).

³² Thomas Kahl, 'AI Act and the Automotive Industry – Where does the road lead?' (*Taylorwessing*, 25 March 2025) <<https://www.taylorwessing.com/de/insights-and-events/insights/2025/03/ai-act-and-the-automotive-industry>>.

³³ Consolidated version of the Treaty on the Functioning of the European Union [2012] OJ C 326/47.

³⁴ Art. 1 European Parliament and Council Directive 2024/2853 of 23 October 2024 on liability for defective products and repealing Council Directive 85/374/EEC [2024] OJ L 2024/2853 (PLD Directive).

³⁵ Art. 2 PLD Directive.

The PLD Directive, originally enacted in 1985, was revised in 2024, with a transposition deadline for member states of 9 December 2026, while the old PLD remains in force until that date. Meanwhile, it will still apply to products that were put into circulation before that date.³⁶ In 2021, the European Institute (ELI) published the Guiding Principles for Updating the Product Liability Directive for the Digital Age, which outlined mechanisms that needed to be revised.³⁷ The need to revise the original PLD Directive has also been widely discussed in expert reports of the European Commission, resolutions of the European Parliament, and the Commission's own reports, as over the decades of its adoption the technological landscape has undergone a significant transformation, in particular due to the rapid development of digital technologies and the increasing integration of software and AI into physical products. More consumer goods frequently rely on digital components to perform functions that were previously purely mechanical, while automated systems and algorithms increasingly replace human operation. These developments have exposed the limitations of the original Product Liability Directive, which was not designed to address such technologies. Therefore, its revision was needed to ensure continued effective protection of individuals and the proper functioning of product liability regimes in the digital age.³⁸ Although the old PLD Directive³⁹ has proven to be an important and effective instrument, rapid technological development (including AI, new circular-economy business models, and global supply chains) has created legal uncertainty and inconsistencies regarding the definition of a product. Practical experience has also demonstrated that injured parties face obstacles to obtaining compensation due to both limitations in bringing claims and difficulties in proving liability, especially given the growing technical and scientific complexity. The aforementioned challenges are particularly seen in cases involving emerging technologies. Consequently, the revision of the old PLD Directive should encourage the development and adoption of new technologies

³⁶ Kosjenka Dumančić and Dominik Vuletić, 'Liability for Damages Caused by Autonomous Driving Vehicles from the EU Law Perspective' in Ante Novokmet and others (eds), *EU and Comparative Law Issues and Challenges Series (ECLIC 2025)* 64, 67.

³⁷ Kosjenka Dumančić and Dominik Vuletić, 'Liability for Damages Caused by Autonomous Driving Vehicles from the EU Law Perspective' (2025) in Ante Novokmet and others (eds), *EU and Comparative Law Issues and Challenges Series (ECLIC 2025)* 64, 67.

³⁸ Christian Twigg-Flesner, *Guiding Principles for Updating the Product Liability Directive for the Digital Age* (European Law Institute 2021) 3.

³⁹ Council Directive 85/374/EC of 25 July 1985 on the approximation of the laws, regulations and administrative provisions of the Member States concerning liability for defective products [1985] OJ L 210/29 (old PLD Directive).

while simultaneously ensuring the same level of protection for claimants and legal certainty for businesses.⁴⁰

In this part, it is important to highlight the main differences between the original and revised PLD Directive to examine, in the next chapters, whether the revised PLD is appropriate and sufficiently comprehensive to meet the needs of the era of automated vehicles and AI systems. According to the old PLD Directive, “product” is defined as a tangible and movable object, which led to a dispute among legal practitioners about whether software is deemed a product and, if so, to what extent. Meanwhile, the revised PLD Directive explicitly extends its scope to digital products, including software, digital manufacturing files, and AI systems. Moreover, it includes digital services related to a product's functionality, such as cloud-based and real-time data services. The explicit coverage of AI systems ensured that developers and providers could also be held liable for defects.⁴¹ The old PLD Directive introduced a strict liability regime, meaning the producer's liability for harm caused by defective products regardless of their fault, with limited grounds for exculpation. While the claimants were required to prove the defect, the damage, and the causal link between them. The aim of this framework was to balance the interests of both parties by providing a fair mechanism for addressing product defect-related claims. In contrast to the old PLD Directive, the new PLD continues with a strict liability regime, but unlike the original version the new PLD was construed to ease the claimant's burden of proof in cases involving advanced technologies by establishing a rebuttable presumption of defectiveness if the defendant does not disclose relevant evidence while the claimant demonstrates, aside from the causal link, the non-compliance of a product with product mandatory safety requirements.⁴² As the criticised issues were already addressed by revising the original PLD Directive, for the purposes of this thesis we will work with the new PLD provisions, so the old application problems will no longer be addressed.

According to de lege lata state: “[a] product shall be considered defective where it does not provide the safety that a person is entitled to expect or that is required under Union or national law.”⁴³ Meanwhile, the rest of the cited article provides a non-exhaustive

⁴⁰ Recital 3 PLD Directive.

⁴¹ Kosjenka Dumančić and Dominik Vuletić, ‘Liability for Damages Caused by Autonomous Driving Vehicles from the EU Law Perspective’ in Ante Novokmet and others (eds), *EU and Comparative Law Issues and Challenges Series (ECLIC 2025)* 64, 68.

⁴² Art. 10 (3) PLD Directive.

⁴³ Art. 7 (1) PLD Directive.

list of circumstances to consider when assessing a product's defectiveness. In this regard, for the purposes of this thesis, it is important to point out c), e), and f). The letter c) refers to a product's continuing liability to learn or gain new features after its placement on the market, and its impact on the product.⁴⁴ As can be seen, legislators considered the capabilities of AI systems and machine learning algorithms incorporated into products. Secondly, letter e) requires consideration of the timing of product placement on the market or into service, whether and when the manufacturer's control over the product persists, and the moment the product leaves the manufacturer's control.⁴⁵ This is relevant for automated vehicles as the precise knowledge of the time of the manufacturer's control over the vehicle, regarding, for example, software updates, could have direct implications for liability allocation in the event of a defect. Finally, the letter f) mentions product safety requirements, including cybersecurity ones.⁴⁶ This is also very important, as in the digital era, product safety does not only include traditional requirements but also proper cybersecurity measures, given that ADS largely rely on software, data exchange, and connectivity. To put it differently, vulnerabilities to potential cyberattacks may directly compromise user safety and therefore must be deemed a defect under the PLD Directive's understanding. Another important provision to mention is Article 8, which identifies economic operators who may be held liable for damage caused by defective products under the PLD Directive and is therefore significant for liability allocation. The term "economic operators" is defined as *"(...) manufacturer of a product or component, a provider of a related service, an authorised representative, an importer, a fulfilment service provider or a distributor[.](...)"*⁴⁷ According to Article 8, liability is not limited to the final product's manufacturer. It is also extended to other actors within the supply chain, such as component manufacturers, importers, authorised representatives of the manufacturer, and, under certain conditions, distributors and fulfilment service providers. The PLD Directive also imposes liability on a person who has substantially modified a product and subsequently made it available on the market or put it into service outside the manufacturer's control, recognising that person as a manufacturer. Moreover, the manufacturer can be held liable for a defective component if it was incorporated into or interconnected with a product under the manufacturer's control.⁴⁸ The key point is that the article directly strengthens the

⁴⁴ Art. 7 (c) PLD Directive.

⁴⁵ Art. 7 (e) PLD Directive.

⁴⁶ Art. 7 (f) PLD Directive.

⁴⁷ Art. 4 (15) PLD Directive.

⁴⁸ Art. 8 PLD Directive.

effectiveness of liability provisions, as the fact that the injured party can sue the manufacturer of the entire product rather than the component manufacturer, without the need to identify the specific defective component, simplifies claims and indirectly reduces the burden of proof. It also prevents liability fragmentation across complex supply chains, ensuring clear attribution of responsibility to the injured party. In practice, the product's manufacturer should seek recourse from the accountable party. Articles 9 and 10 address the evidentiary challenges in product liability cases. Article 9 contains rules governing the disclosure of evidence, enabling claimants, under certain conditions, to obtain evidence in the defendant's possession. It introduced the rule of evidence disclosure, under which, if the claimant introduces sufficient evidence to support the claim's plausibility for compensation, the defendant must disclose any relevant evidence in his possession. At the same time, the respective provision limits such disclosure to what is deemed necessary and proportionate, while safeguarding confidential information and trade secrets. Article 10 lays down the rules concerning the burden of proof, while, on the one hand, maintaining the traditional civil law liability approach established in the old PLD Directive, which requires the claimant to prove the defectiveness of the product in question, the damage suffered, and the causal link between defectiveness and damage.⁴⁹ On the other hand, the revised PLD Directive introduced a rebuttable presumption of product defectiveness, established to alleviate the claimant's burden of proof in certain scenarios discussed below. According to Article 10 (2), the product's defectiveness is presumed if the defendant did not disclose relevant evidence in compliance with Article 9 (1), if it is demonstrated by the claimant that the product failed to comply with the mandatory safety requirements arising from European Union or national law, which were tailored to prevent the type of damage suffered and if it is demonstrated by claimant that the damage resulted from evident malfunction of the product under reasonably foreseeable or ordinary circumstances. Moreover, under Article 10(3), if the product's defectiveness has been established and the damage caused corresponds to the defect in question, there is no need for the claimant to demonstrate a causal link between defectiveness and the damage caused. Lastly, the presumption of defectiveness, causation, or both in compliance with Article 10 (4) may be established by the court, notwithstanding the proper disclosure according to Article 9 (1), if the claimant faces disproportionate difficulties in proving the defectiveness in connection

⁴⁹ Art. 9 and 10 PLD Directive.

with technical or scientific complexity, provided that their likelihood is demonstrated.⁵⁰ The aforementioned rebuttable presumptions indeed simplified the claimant's position, as intended, because if the harm was caused by a defective product comprising a digital element (including AI) or by a purely digital product, it is more difficult to prove defectiveness. In these cases, to examine the defect, it may be necessary to assess the product's digital element or digital service rather than the physical item, which requires a complex technical assessment and thereby places a claimant at a significant disadvantage.⁵¹ Another important provision governing exceptions to the economic operator's liability is Article 11, which provides that the economic operator will not be held liable in certain situations. This list of exemptions is exhaustive, leaving the member states with clear instructions. The article 11 (1) e) is particularly interesting in a context of automated vehicles equipped with complex AI systems, as it states the exemption of liability providing "*(...) that the objective state of scientific and technical knowledge at the time the product was placed on the market or put into service or during the period in which the product was within the manufacturer's control was not such that the defectiveness could be discovered* [...] (...)"⁵² which should ensure a proper protection for economic operators especially with regards to the opacity, lack of explainability and black-box nature of AI systems. In the following chapter, the analysed provisions will be subjected to a more thorough assessment, mostly considering innovation, technical capabilities, legal certainty, and the proper balance between the parties.

Apart from the AI Act and the PLD Directive on 28 September, the European Commission has released a proposal for the AI Liability Directive, which aims to establish uniform rules on non-contractual civil liability for damage caused by AI systems. It was designed to address the evidentiary difficulties arising from the use of AI and sought to prevent unjustified barriers to legitimate claims.⁵³ It should have complemented the AI Act by setting out the procedural provisions for persons who suffered damage caused by an AI system. Its scope did not extend beyond the procedural provisions mentioned, thereby

⁵⁰ Art. 10 PLD Directive.

⁵¹ Christian Twigg-Flesner, *Guiding Principles for Updating the Product Liability Directive for the Digital Age* (European Law Institute 2021) 9.

⁵² Art. 11 (1) (e) PLD Directive.

⁵³ European Commission, 'Liability Rules for Artificial Intelligence' (European Commission, 2022) <https://commission.europa.eu/topics/business-and-industry/doing-business-eu/contract-rules/digital-contracts/liability-rules-artificial-intelligence_en> accessed 23 March 2026.

lacking substantive tort rules.⁵⁴ However, the Commission withdrew the mentioned proposal on 11 February 2025 and stated as a reason “[n]o foreseeable agreement - the Commission will assess whether another proposal should be tabled or another type of approach should be chosen.”⁵⁵ This development leaves a regulatory gap in the additional evidentiary challenges posed by the character of the AI systems, which must therefore be assessed within the existing liability framework.

1.4 Socio-economic implications and risk allocation

As already mentioned in this chapter, the product is deemed defective if it does not provide the expected safety for the person or the required safety under European Union or national law. But what constitutes a proper level of safety that can be expected? In 2018, the European Union set a goal to eliminate all road deaths by 2050 under the “Vision Zero” strategy. It involves defining key performance indicators, addressing the main road safety challenges, namely safe infrastructure, safe vehicles, safe road use (speeding, alcohol, distraction, and use of protective equipment), and fast and effective post-crash care. The EU has also set a target to reduce road deaths and serious injuries by 50% by 2030 as an intermediary step. This was introduced in the EU Road Safety Policy Framework 2021-2030 and the Commission’s Strategic Action Plan on Road Safety.⁵⁶ Nowadays, the preliminary figures on road fatalities for 2025 have been released by the Commission, reporting around 19,400 deaths, representing a 3% decrease compared with 2024. Given the rising number of vehicles on EU roads and the growing volume of kilometres driven, this represents a notable achievement. It is also important to emphasise that, compared with 2019, the preliminary data from 2025 show a 15% decrease in road deaths and serious injuries, further illustrating the overall progress made in improving road safety across the European Union. Nevertheless, the data also underscores that continued efforts are required at all levels, as most member states are still not on track to achieve the EU’s goal of halving

⁵⁴ Didem Polad, ‘Liability Perspective for Users of Autonomous Vehicles in the EU’ (*RAILS Robotic & AI Law Society*, 15 April 2024) <<https://blog.ai-laws.org/liability-perspective-for-users-of-autonomous-vehicles-in-the-eu/>> accessed 31 March 2026.

⁵⁵ European Commission, ‘Annexes to Commission Work Programme 2025 Moving forward together: A Bolder, Simpler, Faster Union’ (Annex to Communication) COM (2025) 45 final 26.

⁵⁶ European Climate, Infrastructure and Environment Executive Agency (CINEA), *EU Road Safety: Towards “Vision Zero”* (European Commission 2022) 5.

serious injuries and road deaths by 2030.⁵⁷ Moreover, behavioural risk factors continue to undermine road safety initiatives despite legislative measures, awareness campaigns, and investments in road safety. For example, the main ones include driving at inappropriate or excessive speeds, driving under the influence, driving while distracted, and not wearing a seatbelt.⁵⁸ In the United States, it is predicted that automated vehicles could reduce traffic fatalities by up to 90% by eliminating human error, which is estimated to cause 94% of fatalities. It was estimated that the deployment of automated vehicles could save at least 29,000 lives per year in the US and around 10 million lives per decade worldwide, potentially creating one of the most significant public health advances of this century.⁵⁹ Nowadays, the automotive sector is undergoing an unprecedented technological transformation, increasing vehicle safety through the implementation of ADAS systems such as automatic emergency braking, intelligent speed assistance, and lane-keeping assist, which are becoming standard features. The increasing integration of ADS into vehicles creates both opportunities and challenges. Although higher levels of automation reduce the impact of human driver error, the transition phase introduces new challenges, including ensuring safe interactions between automated and human-driven vehicles in traffic, protecting non-motorised road users, and determining liability in accidents involving automated vehicles.⁶⁰ As an example of support for the development of automated vehicles, Horizon Europe is funding the SAFE-UP safety systems project and the Mediator project, which focuses on mediating between human and automated driving.⁶¹ The influence of behavioural factors could be significantly diminished with the increase of automated systems. Although higher levels of automation could effectively prevent road accidents caused by human error, these systems are complex technologies that require careful consideration of the risks associated with their deployment and the appropriate legislative framework to mitigate them. From a risk-allocation perspective, the question of who should bear the risks associated with different levels of autonomy of automated systems remains.

⁵⁷ Directorate-General for Mobility and Transport, 'EU road deaths drop by 3% in 2025' (*European Commission*, 24 March 2026) < https://transport.ec.europa.eu/news-events/news/eu-road-deaths-drop-3-2025-2026-03-24_en > accessed 6 April 2026.

⁵⁸ European Commission, 'Report on the Implementation of the EU Road Safety Policy Framework at the Mid-Point' (Communication) COM (2026) 77 final 7.

⁵⁹ Janet Fleetwood, 'Public Health, Ethics, and Autonomous Vehicles' (2017) 107(4) *American Journal of Public Health* 532, 532.

⁶⁰ European Commission, 'Report on the Implementation of the EU Road Safety Policy Framework at the Mid-Point' (Communication) COM (2026) 77 final 17.

⁶¹ European Commission, 'Report on the Implementation of the EU Road Safety Policy Framework at the Mid-Point' (Communication) COM (2026) 77 final 15.

Across different levels of automation, the allocation of risk and liability between the driver and the economic operators involved is unclear. As automation increases, the driver becomes less involved in DDT, thereby shifting liability more towards economic operators. This question has a crucial impact on the world of automated vehicles, as it determines how the risks and costs of accidents are distributed among the involved parties. It also directly influences behavioural incentives, as different liability models may encourage or discourage the adoption and development of automated driver systems. The legislation's adequate framework must ensure equal protection for economic operators and drivers, while supporting technological advancement and innovation, as it is in the best interests of all parties to reduce road accidents and transactional costs. Furthermore, a clear and predictable liability framework, ensuring legal certainty and effective enforceability, is also important for fostering customer trust, influencing the individual's willingness to purchase and use automated vehicles, and therefore their widespread adoption. The extent of transferring the risks associated with automated vehicles to third parties, such as insurers, also remains uncertain.

2. Regulatory challenges

This chapter is dedicated to the central part of this thesis, as it identifies the most significant challenges for our legal environment, given the complexity of automated driving systems, which typically incorporate AI algorithms. Due to the rapid pace of evolution of the mentioned technologies, all jurisdictions struggle to adopt appropriate legislative measures to address the new risks they pose in practice. Although it has been shown that automated vehicles could be an effective solution for reducing human driver errors in the future, they also introduce new causes of accidents that were not present before. An illustrative example is one of the first known road deaths in the United States involving an automated vehicle, a Tesla S, in which the car's sensor system failed to distinguish a large white truck and trailer that was crossing the highway from the bright sky. The accident occurred because Tesla attempted to drive under the trailer at full speed. While a driver was allegedly watching Harry Potter, the system required the driver to remain alert because it was not yet perfect.⁶² The company acknowledges in its manual that its automated driving

⁶² Sam Levin and Nicky Woolf, 'Tesla driver killed while using autopilot was watching Harry Potter, witness says' *The Guardian* (San Francisco, 1 July 2016)

systems have limitations. For instance, the Tesla Model 3 manual includes a disclaimer stating that dirty sensors, cameras, environmental conditions, and faded line markings can affect the performance of a self-driving system. Furthermore, the systems may also not work properly, for example, if the road has sharp curves or significant elevation changes, if the road signs and signals are unclear or poorly maintained, or due to poor visibility.⁶³ These shortcomings are the consequence of an inherent characteristic of machine learning systems. They operate based on the learning data and are therefore susceptible to biases in data selection, and they face difficulties generalising beyond learned patterns, which may lead to unpredictable behaviour in real-world environments. The common features of AI systems create several regulatory challenges for traditional legal frameworks. In particular, they raise questions about how to determine defectiveness in the context of probabilistic and learning-based systems, whether existing evidentiary rules are sufficient given the opacity of these systems, and how liability should be attributed in scenarios involving unavoidable harm. The mentioned issues are examined in the following sections of this chapter, focusing on the concept of defectiveness, the black-box problem, and trolley-type scenarios.

2.1 Concept of defectiveness and liability standard

As noted, under the PLD Directive, the product is deemed defective if it does not provide the safety that a person is entitled to expect or that is required under Union or national law.⁶⁴ When assessing a product's defectiveness, the directive provides a non-exhaustive list of circumstances that shall also be taken into account. Since this provision is part of the directive, it leaves member states with discretion over the form and methods for determining defectiveness. Consequently, the interpretation of this concept is crucial for the legal systems of member states that are still within the transposition period for the revised PLD Directive. Assessing defectiveness can be complex, particularly for automated vehicles. While higher levels of automation, excluding the need for direct human control,

<<https://www.theguardian.com/technology/2016/jul/01/tesla-driver-killed-autopilot-self-driving-car-harry-potter>> accessed 15 April 2026.

⁶³ Tesla, 'Model 3 Owner's Manual' (Tesla 2026)

<https://www.tesla.com/ownersmanual/model3/en_us/GUID-E5FF5E84-6AAC-43E6-B7ED-EC1E9AEB17B7.html> accessed 15 April 2026.

⁶⁴ Art. 7 (1) PLD Directive.

may suggest a clear interpretation, as the damage is attributable to a defective product or to faulty traffic signs, the question of attributing damage is less clear and more difficult at lower levels of automation. To put it differently, at Levels 3 and 4, situations in which a human user is either in a fallback-ready position or may be prompted to resume vehicle operation when ADS reaches its ODD limit raise questions about whether the damage could have been prevented by a human driver taking control. This introduces an additional layer of complexity, as damage can result not only from the system's faulty performance but also from the user's actions or omissions. In contrast to conventional vehicles, with always present human control including Levels 0-2 and Level 5 vehicles (currently not deployed) with always present ADS control, therefore in cases where courts do not need to distinguish between different levels of automation, automated vehicles with levels 3, 4 of automation require more sophisticated approach, which takes into account how roles are distributed between ADS and driver as in many scenarios with a rapid human intervention, the accident could have been prevented. Consequently, determining defectiveness under the PLD Directive becomes significantly more challenging for legal systems of member states. On the one hand, there is the ADS system, which can malfunction or make learning mistakes, on the other hand, there is a human driver or supervisor who could have intervened. Moreover, ADS limitations should also be taken into account. Is the system defective if it causes accidents, for instance, because of worn-down lane markings, or because of heavy rain conditions while the driver was not paying attention? It is clear that the question of defectiveness is inherently connected to assessing liability for accidents and to the extent to which the legal system is prepared for road accidents involving automated vehicles. From a legal standpoint, numerous factors can contribute to possible accidents involving automated vehicles. Most likely, the accidents will occur due to system malfunctions, system limitations being exceeded, human driver error, and the condition of road signs. Among the various legal systems of member states, Germany took the first step towards a separate legal framework for automated vehicles. The German Road Traffic Act was amended in 2017 and 2021 to introduce specific rules for automated vehicles and, most importantly, to distinguish between Level 3 and Level 4 automation. The term “driver” refers to users of Level 3 vehicles, whereas the term “technical supervisor” refers to users of Level 4 vehicles. With increasing automation, the driver has more obligations than the technical supervisor. Moreover, in the event of damage caused by a Level 3 automated vehicle, the burden of proof rests with the driver. Meanwhile, the burden of proof for Level

4 rests with the injured party.⁶⁵ The German approach demonstrates that assessing defectiveness in the context of automated vehicles must align with the allocation of control between the human user and the ADS. By distinguishing between Level 3 and Level 4 automation and by adjusting the burden of proof, German law acknowledges that the assessment of defectiveness is closely linked to and dependent on the distribution of liability for AV accidents. The analysis in this subchapter illustrates that determining defectiveness and liability under the revised PLD Directive involves many elements that can significantly influence the legal outcome and therefore must be addressed differently in each specific case.

The problem of assessing defectiveness has an additional dimension, which is closely linked to the liability setting in accidents involving automated vehicles. As already mentioned, the revised PLD Directive retained the strict liability regime of the old PLD Directive, with the possibility of being exempted from liability provided that one of the exempting conditions is met. Since the provisions were construed as general rules governing defective products placed on the market, they are not specifically tailored to automated vehicles. As a result, they do not fully reflect the original purpose of reducing road fatalities and serious injuries caused by human error, and the current liability standard can have a chilling effect on innovation, further development, and deployment of automated vehicles. In this section, the analysis returns to the question of the level of safety that a person is entitled to expect. According to the PLD Directive recital, “(...) *the defectiveness of a product should be determined by reference not to its fitness for use but to the lack of the safety that a person is entitled to expect, or that is required under Union or national law.*”⁶⁶ In addition, it further explains that “[t]he assessment of defectiveness should involve an objective analysis of the safety that the public at large is entitled to expect and not refer to the safety that any particular person is entitled to expect.”⁶⁷ The objective analysis should take into account the intended purpose, presentation, reasonably foreseeable uses, objective characteristics, properties such as expected lifespan, and the specific requirements of the intended user groups.⁶⁸ The key recital is also 31, which explicitly states that disclaimers made by the product’s manufacturers, provided with the

⁶⁵ Didem Polad, ‘Liability Perspective for Users of Autonomous Vehicles in the EU’ (*RAILS Robotic & AI Law Society*, 15 April 2024) <<https://blog.ai-laws.org/liability-perspective-for-users-of-autonomous-vehicles-in-the-eu/>>.

⁶⁶ Recital 30 PLD Directive.

⁶⁷ Ibid.

⁶⁸ Ibid.

product, are not considered a sufficient measure to make a defective product safe, as the defectiveness should be determined by the aforementioned interpretation rules.⁶⁹ This undermines the importance of the disclaimers in manuals that often accompany automated vehicles, for instance, as previously described, the USA Tesla manual for the European Union market. As was demonstrated in the previous chapter, the Tesla manual included inherent and technically justified limitations of ADS systems, including sensitivity to environmental conditions, road infrastructure quality, and reliance on non-deterministic behaviour and the character of AI-based systems, which produce variable outputs under the same conditions. This illustrated that these limitations cannot simply be removed from the equation, as they directly affect the performance of the ADS in real-world conditions. Moreover, automated cars are complex engineering systems that incorporate multiple algorithms and interconnected technological components. Therefore, their capabilities and limits are not typically understood by the average user or the general public. Thus, the question is how the objective safety standard can be applied in relation to a product that the general public is not accustomed to. Since there is limited experience with automated cars, the analogy to the human-driver car is appropriate. In simplified terms, the public is entitled to expect that automated vehicles are at least as safe as human-driven cars. Nevertheless, determining the defectiveness by comparing it to the human driver does not fully solve the problem, as some types of accidents can be prevented by the ADS, but new causes of accidents can arise, such as physical limitations of sensors⁷⁰ or their malfunction, but also AI-related causes, such as biased training data, misclassification, error in decision making, and limitations in interpreting unexpected traffic situations. It can be argued that a relevant benchmark for the future of automated vehicles is to be statistically safer than the average human driver, factoring in ADS system-induced risks. Nevertheless, the hurdle is that the average human driver standard is rather abstract, making it hard to ascertain whether ADS meets that standard. It can be proven only based on large-scale statistics.⁷¹ In legal practice, it is necessary to apply the standard, which is not problematic to use in each individual liability case. If the average human driver standard was used, perhaps the question of whether an average human driver could have prevented this accident could be asked. Although an automated car differs from a human driver and may fail in different ways,

⁶⁹ Recital 31 PLD Directive.

⁷⁰ Maurice Schellenkens, 'Self-driving cars and the chilling effect of liability law' (2015) 31(4) Computer Law & Security Review 506, 510.

⁷¹ Maurice Schellenkens, 'Self-driving cars and the chilling effect of liability law' (2015) 31(4) Computer Law & Security Review 506, 510.

there is a good chance it will cause fewer accidents, notwithstanding the mentioned fact. While the concept of defectiveness remains crucial in the era of automated vehicles, it ultimately serves as the basis for liability-setting and liability allocation, meaning that its interpretation directly shapes the liability regime for accidents involving automated vehicles. Liability and innovation are not isolated concepts but directly influence each other. As an example, the liability law affects manufacturer's decisions to produce certain products. Thus, in our case, when the liability standard is set too high, the manufacture of automated vehicles with higher levels of automation may be delayed until technology reaches a higher level of safety. In contrast, when a liability standard is set too low, it can undermine public trust and lead manufacturers to economise on safety.⁷² In this context, the European liability framework, based on a strict liability regime complemented by limited grounds for exemption under Article 11 of the PLD Directive, establishes a baseline for liability allocation. As already mentioned, the strict liability regime incentivises manufacturers to make safer products. In addition, they are in a stronger position than consumers. Consequently, this is the reason why a strict liability regime was adopted in the first place. At the same time, these systems differ from other product defects in that, once ADS becomes safer than humans, automation will yield substantial safety gains. Perhaps moving towards the negligence standard could yield better outcomes, as it would accelerate innovation and the further deployment of ADS, thereby directly reducing accidents in the future. To illustrate, if a state-of-the-art ADS were ten times safer than a human driver, it would produce a better outcome than waiting for the same ADS to become a hundred times safer than a human driver.⁷³ Once automated vehicles become statistically safer than human drivers, accident rates will drop significantly due to automation. At some point, it may be preferable to increase adoption of safer technologies by reducing product safety requirements.⁷⁴ On the contrary, the final level of product safety depends not only on the liability standard imposed on manufacturers but also on regulatory requirements applicable prior to the product's placement on the market. Furthermore, once the AI Act is incorporated into the automotive industry's pre-market requirements, the risks associated with high-risk AI systems will further decrease. In general, the stricter the pre-market

⁷² Maurice Schellenkens, 'Self-driving cars and the chilling effect of liability law' (2015) 31(4) *Computer Law & Security Review* 506, 510-511.

⁷³ Ryan Abbott, 'The Reasonable Computer: Disrupting the Paradigm of Tort Liability' (2016) 86 (1) *George Washington Law Review* 22.

⁷⁴ Ryan Abbott, 'The Reasonable Computer: Disrupting the Paradigm of Tort Liability' (2016) 86 (1) *George Washington Law Review* 22-23.

requirements, the safer the products are when deployed, regardless of whether a strict liability or negligence liability standard is applied.

2.2 Trolley-type scenarios

The primary objective of deploying automated vehicles is to reduce traffic accidents caused by human error and enhance road safety, aiming to achieve a level of performance that statistically meets or exceeds that of the average human driver. In other words, for the deployment of vehicles with higher levels of automation capable of operating safely on public roads, they must replicate or even surpass human decision-making processes. However, certain decisions are beyond the mere mechanical application of traffic laws and safe route planning, as they appear to require a form of ethical judgement, which is notoriously difficult to implement into algorithmic structures that can be consistently applied in real-life scenarios.⁷⁵

The original Trolley Problem formulation is based on a scenario created by philosopher Philippa Foot in 1967 in her article *“The Problem of Abortion and the Doctrine of Double Effect”*, which was later elaborated and named *“The Trolley Problem”* by philosopher Judith Jarvis Thomson. An original formulation involves a scenario in which a trolley is heading uncontrollably towards five track workers repairing the track. Due to a braking system failure, the driver cannot stop the vehicle in time. Suddenly, the driver notices an alternative track leading off to the right, with one worker, onto which the trolley can be diverted. However, like the five others, one track worker cannot avoid the oncoming trolley in time. The main question arising from the presented scenario is whether it is morally permissible to divert the trolley, thereby sacrificing one worker in order to save five, and how such a decision can be justified within existing moral frameworks.⁷⁶ One possible philosophical approach to resolving this dilemma is the doctrine of double effect, which distinguishes between consequences directly intended and merely foreseen as side effects of a voluntary action. According to this doctrine, an action may be considered morally permissible even if it leads to harmful consequences, provided that such consequences were not intended as ends or means but occur only as incidental outcomes. Application of the doctrine of double effect to the Trolley Problem may interpret the death

⁷⁵ Patrick Lin, ‘Why Ethics Matter for Autonomous Cars’ in Markus Maurer and others (eds), *Autonomes Fahren Technische, rechtliche und gesellschaftliche Aspekte* (Springer-Verlag GmbH 2015) 69, 69.

⁷⁶ Judith Jarvis Thomson, ‘The Trolley Problem’ (1985) 94 Yale Law Journal 1395, 1395.

of a single individual as a foreseen but unintended consequence of diverting a trolley, rather than a directly intended outcome.⁷⁷ On the contrary, Mrs Foot questions the sufficiency of this doctrine. She further develops her argument by distinguishing between positive and negative duties, in particular, the duty to refrain from causing harm, and the duty to provide aid. She demonstrates that in the trolley scenario, a person faces a conflict between two negative duties, namely the duty to avoid harming five individuals and the duty to avoid harming a single individual. Under the circumstances where he cannot avoid both, it seems reasonable to do the least injury he can. Apart from the Trolley Problem, this reasoning does not hold in other scenarios, particularly those involving the active harm of an individual, where a morally relevant distinction arises between actively causing harm and merely allowing it to occur. Consequently, Mrs Foot argues that even if the doctrine of double effect is rejected, it does not mean that the size of evil (harm) should always be the determining factor.⁷⁸ The Trolley Problem is particularly relevant to automated vehicles, as it provides a simplified scenario for analysing situations in which harm cannot be entirely avoided, and ADS may be required to make decisions involving unavoidable harm.

To illustrate the need for ethics in this area, consider a hypothetical scenario in which the automated vehicle faces a critical decision. In this scenario, it must either swerve left and hit a young child or swerve right and hit an elderly person. Given the vehicle's speed, either outcome would be certainly fatal, and if no action is taken, both individuals could be harmed, suggesting that some form of intervention is required to minimise the damage.⁷⁹ On the contrary, there could be debate over whether automated cars need to confront hard ethical choices and whether other solutions offer a path around them. Common examples of circumventing the ethical decision include stopping the vehicle or handing control back to a human driver/operator. Even though these solutions initially appear to be best practices, they will not always be applicable, especially if automated vehicles operate beyond their limited highway environments. Moreover, braking or changing the control may not always be the safest option in real-world conditions. Factors such as road surface, surrounding traffic, limitations in system perception, or delays in ADS

⁷⁷ Philippa Foot, 'The Problem of Abortion and the Doctrine of the Double Effect' (1967) 5 Oxford Review 1.

⁷⁸ Philippa Foot, 'The Problem of Abortion and the Doctrine of the Double Effect' (1967) 5 Oxford Review 5.

⁷⁹ Patrick Lin, 'Why Ethics Matter for Autonomous Cars' in Markus Maurer and others (eds), *Autonomes Fahren Technische, rechtliche und gesellschaftliche Aspekte* (Springer-Verlag GmbH 2015) 69, 69-70.

disengagement can render the mentioned solutions ineffective at avoiding collisions or even more dangerous than facing the ethical dilemma.⁸⁰

In a given situation where an automated vehicle must choose between two fatal outcomes, any decision would involve value and ethical judgement rather than a purely technical optimisation. The Trolley Problem application to automated vehicles is often illustrated in research papers by comparing children with elderly people, a criminal with a person without a prior conviction, or just different numbers of people on each side, following the same principles as stated above. The question of how ADS should react in each scenario is much more complex than merely deciding which pedestrians to prioritise, as was first illustrated. ADS systems could be programmed to concentrate on saving different categories of individuals, depending on the intended objectives. From one perspective, if an automated vehicle was designed to prioritize the protection of its own occupants, it would most likely crash into the lightest object possible. If the scenario involved multiple vehicles, it would prefer striking a lighter vehicle rather than a heavier one.⁸¹ From the perspective of prioritising the safety of other road users (including pedestrians and occupants of other vehicles), it could be programmed to prefer collisions that minimise harm to third parties, at the expense of its own occupants. In these cases, the system would prefer a collision with a heavier vehicle rather than a lighter one, or, if it can recognise vehicle types, perhaps choose to collide with a safer vehicle over one not known for safety.⁸² Apart from the decisions that need to be made pre-market, other factors to consider include environmental, weather, and road conditions at the time of the potential accident. The mentioned variables tend to affect not only ADS but also human drivers encountering similar situations under comparable uncertainty. *“The frequency at which these dilemmas will emerge is extremely hard to estimate, just as it is extremely hard to estimate the rate at which human drivers find themselves in comparable situations.”*⁸³ As rare as such situations might be, the human driver is required to make an immediate decision and may face criminal charges depending on the outcome of the accident. It is difficult to determine how often human drivers encounter such dilemmas in real-life situations. Drivers whose accidents are fatal cannot report whether they faced a moral

⁸⁰ Patrick Lin, ‘Why Ethics Matter for Autonomous Cars’ in Markus Maurer and others (eds), *Autonomes Fahren Technische, rechtliche und gesellschaftliche Aspekte* (Springer-Verlag GmbH 2015) 69, 71.

⁸¹ Patrick Lin, ‘Why Ethics Matter for Autonomous Cars’ in Markus Maurer and others (eds), *Autonomes Fahren Technische, rechtliche und gesellschaftliche Aspekte* (Springer-Verlag GmbH 2015) 69, 72.

⁸² Patrick Lin, ‘Why Ethics Matter for Autonomous Cars’ in Markus Maurer and others (eds), *Autonomes Fahren Technische, rechtliche und gesellschaftliche Aspekte* (Springer-Verlag GmbH 2015) 69, 72.

⁸³ Edmond Awad and others, ‘The Moral Machine experiment’ (2018) 563 Nature 59, 59.

dilemma, while those who survived may not have realised that they were in such situation. Nevertheless, ethical guidelines for automated vehicles governing the choice in dilemma situations should not depend on the frequency of such situations.⁸⁴ In contrast to the human-driven car, the ADS systems would need to make these decisions based on predetermined, embedded system design, meaning that the manufacturers of the ADS in question have to choose how the system will react in such situations, and thereby decide about human life, for which they may be held liable according to criminal law and civil law provisions. Nowadays, courts rarely examine the exact reasoning behind a human driver's actions, whose behaviour is often attributed to panicking and instinct in the critical moments preceding the accident. The question of "why" is more substantial in AV accidents than in human-driver accidents, because the system's behaviour results from predefined design choices. Generally, the most important difference between humans and machines is that there is an assumption that a person of good judgement will recognise the need to disregard the exact letter of the law in order to honour its spirit, but how to train machines to learn this?⁸⁵ In other words, driving in real traffic conditions frequently presents ethically challenging situations, so the expectations placed on the driver are to make sophisticated, instant ethical decisions. Therefore, programming ADS solely to follow traffic rules would be insufficient. By this were meant situations such as deliberately crossing a double line to avoid hitting a person in the road or disregarding a red traffic signal to avoid a collision with an oncoming train. Consequently, the qualitative requirements for automated vehicles are to balance safety, mobility, and legality⁸⁶, as real-world traffic scenarios are much more complex than simply following the law. Furthermore, the member states of the European Union and the Council of Europe could face legal consequences if they adopted a law that permits manufacturers to develop and program ADS systems in a way that would make a preference between different human lives, as it would be contrary to the European Convention on Human Rights, the Charter of Fundamental Rights of the European Union, and the Treaty on European Union. In parallel with the trolley problem, any law that would allow encoding of any kind of preference between human lives would be in breach of Article 2 ECHR, guaranteeing the right to life, and of Article 14 ECHR, which prohibits any kind or form of discrimination on grounds such as age, sex, race, or origin⁸⁷.

⁸⁴ Edmond Awad and others, 'The Moral Machine experiment' (2018) 563 Nature 59, 59.

⁸⁵ Noah J. Goodall, 'Can you program ethics into a self-driving car?' (2016) 53 (6) IEEE Spectrum 28, 30.

⁸⁶ Janet Fleetwood, 'Public Health, Ethics, and Autonomous Vehicles' (2017) 107(4) American Journal of Public Health 532, 534.

⁸⁷ European Convention on Human Rights [1950].

Furthermore, Article 2 of the TEU sets out the values on which the European Union was built, including human dignity, equality, and respect for human rights. A serious breach of the values of the European Union, as set out in Article 2 TEU, may trigger action against a member state under Article 7 TEU.⁸⁸ In addition, it would be contrary to Article 1 of the EU Charter, which protects human dignity, and to Article 2 of the EU Charter, which safeguards the right to life. In addition, Article 3 (2) of the EU Charter ensures “*the prohibition of eugenic practices, in particular those aiming at the selection of persons*”⁸⁹ Although the scope and meaning of Article 3 (2) of the EU Charter is limited to the fields of medicine and biology, it nevertheless reflects a broader European legal and ethical rejection of practices that assign different value to human lives based on a personal characteristic. In other words, the legislation of a member state that would approve algorithms that make choices between different human lives could, per analogiam, be seen as problematic in light of European Union values. Based on the discussed problematic areas, the relevant question within the European Union legal setting is therefore not how ADS should be programmed, but what the EU and member states should allow to be programmed in these sensitive situations in order to ensure the ADS will prioritize overall safety, fundamental human rights, and the core values of the European Union.

2.3 Black-box Problem

The black-box problem is generally defined as an inability to fully comprehend AI’s decision-making processes and the unpredictability of its decisions and outputs. Black boxes vary in the degree of their opacity, so for theoretical purposes they can be divided into two categories: strong and weak black boxes. Although they are theoretical concepts rather than exact illustrations of the black boxes in AI systems⁹⁰, they nevertheless illustrate variation in the transparency of AI decision-making processes. Strong black boxes refer to AI systems whose decision-making processes are fully opaque to humans. In other words, it is impossible to determine how the system reached a particular decision or prediction, which information determined the outcome, or how the processed variables were ranked by importance. Moreover, such systems cannot be meaningfully analysed, even

⁸⁸ Consolidated Version of the Treaty on European Union [2012] OJ C 326/13.

⁸⁹ Charter of Fundamental Rights of the European Union [2000] OJ 364/01.

⁹⁰ Yavar Bathaee, ‘The Artificial Intelligence Black Box and the Failure of Intent and Causation’ (2018) 31 (2) Harvard Journal of Law & Technology 889, 905-906.

retrospectively, by reverse-engineering their outputs. Weak black boxes are still opaque to humans, but, unlike strong black boxes, they can be analysed through reverse engineering or probed to determine an approximate ranking of the importance of variables considered by AI.⁹¹ The issues with transparency stem directly from the nature of the machine learning algorithms widely used in AI systems, for instance, deep neural networks and support vector machines.⁹² Machine learning is a subset of AI that focuses on algorithms capable of identifying patterns in training data and subsequently making accurate inferences about new data. This ability allows machine learning models to make decisions or predictions without relying on explicit hard-coded instructions.⁹³ The two algorithms mentioned represent the opposite ends of the complexity spectrum.⁹⁴ Deep neural networks are one of the core architectures used in deep learning. *“Deep learning is a subset of machine learning where models consist of multiple layers of interconnected neurons. Each layer extracts increasingly complex features from the data, allowing deep networks to learn non-linear, intricate patterns.”*⁹⁵ Meanwhile, support vector machines are algorithms used in shallow learning models. *“Shallow Learning refers to machine learning models with limited depth, usually involving one or two layers of processing.”*⁹⁶ On the one hand, deep neural networks consist of thousands, millions, billions or even trillions of neurons and parameters used for learning and data processing. The complexity arising from the vast number of artificial neurons and their interconnections makes it nearly impossible to determine how the system arrives at particular decisions or predictions. On the other hand, the support vector machine can also cause a black-box problem, even though it is a shallow algorithm. The shallow algorithms can also cause a black-box problem, as they simultaneously process and optimise numerous variables by identifying geometric relationships within mathematically defined higher-dimensional spaces. Due to the high dimensionality mentioned, humans are unable to meaningfully visualise how the support vector machine

⁹¹ Yavar Bathaee, ‘The Artificial Intelligence Black Box and the Failure of Intent and Causation’ (2018) 31 (2) Harvard Journal of Law & Technology 889, 906.

⁹² Yavar Bathaee, ‘The Artificial Intelligence Black Box and the Failure of Intent and Causation’ (2018) 31 (2) Harvard Journal of Law & Technology 889, 894.

⁹³ Dave Bergmann, ‘What is machine learning?’ (IBM) <<https://www.ibm.com/think/topics/machine-learning>> accessed 11 May 2026.

⁹⁴ Hassaan Idrees, ‘Shallow Learning vs. Deep Learning: Is Bigger Always Better?’ (Medium, 27 October 2024) <<https://medium.com/@hassaanidrees7/shallow-learning-vs-deep-learning-is-bigger-always-better-51c0bd21f059>> accessed 11 May 2026.

⁹⁵ Hassaan Idrees, ‘Shallow Learning vs. Deep Learning: Is Bigger Always Better?’ (Medium, 27 October 2024) <<https://medium.com/@hassaanidrees7/shallow-learning-vs-deep-learning-is-bigger-always-better-51c0bd21f059>>.

⁹⁶ Ibid.

makes its decisions or processes new inputs.⁹⁷ Apart from deep neural networks and support vector machines, the recent AI successes are also attributable to machine learning techniques such as random forests, probabilistic graphical models, and reinforcement learning. In many cases, the most accurate AI methods, such as deep learning models, tend to be the least explainable, whereas more transparent models, such as decision trees, are often less accurate.⁹⁸

As already noted, AI systems use their own training processes instead of explicit programming. As a result, during the training process, they can discover new correlations between inputs, making it difficult to understand how outputs are produced. Furthermore, the opacity of such systems can hide an AI system's deficiencies, such as biases, inaccuracies, or hallucinations.⁹⁹ From a technical standpoint, the black box problem arises from the difficulty of interpreting changes occurring within deep neural networks, particularly hidden layers, where connections between neurons are continuously updated during both training and inference. The fact that modern models may contain billions of neurons and that the connections between individual layers may involve billions of parameters makes interpreting these changes practically impossible for the human mind. Although it is possible to approximate or infer what occurs within the system, it is unlikely that individual changes and internal processes will ever be explained in a fully deterministic manner.¹⁰⁰ The scientific discipline called Explainable Artificial Intelligence (XAI) was developed in response to the limitations of AI systems and seeks to enable them to provide clear, interpretable explanations of their actions and decisions. Its primary objective is to improve human understanding of AI behaviour by clarifying the mechanisms underlying its decision-making processes.¹⁰¹ Ideally, XAI should be capable of clarifying the system's capabilities and reasoning, explaining its previous actions, ongoing processes, and future steps, while also revealing the information underlying its decisions and actions.¹⁰² The main concepts of XAI are transparency, interpretability, and explainability. Transparency refers

⁹⁷ Yavar Bathaee, 'The Artificial Intelligence Black Box and the Failure of Intent and Causation' (2018) 31 (2) *Harvard Journal of Law & Technology* 889, 897.

⁹⁸ David Gunning and others, 'XAI-Explainable artificial intelligence' (2019) 4 (37) *Science Robotics* eaay7120, 1.

⁹⁹ European Data Protection Supervisor, *TechDispatch #2/2023 - Explainable Artificial Intelligence* (EDPS, 2023) 1-2.

¹⁰⁰ Gréta Gáborová, 'Nedostatky a riziká umelej inteligencie z pohľadu práva duševného vlastníctva' (Master's Thesis, Comenius University Bratislava 2025) 40.

¹⁰¹ European Data Protection Supervisor, *TechDispatch #2/2023 - Explainable Artificial Intelligence* (EDPS, 2023) 3.

¹⁰² David Gunning and others, 'XAI-Explainable artificial intelligence' (2019) 4 (37) *Science Robotics* eaay7120, 1.

to the extent to which the specific AI model can be understood. In other words, a model is deemed transparent if a person can comprehend the entire model at once. Interpretability enables humans to understand how AI models respond to specific inputs and to identify situations in which the model produces an incorrect outcome. Explainability is important for a trustworthy AI system, particularly in situations where human lives are at stake.¹⁰³ XAI approaches may be divided into two categories: a) self-interpretable models (white-box models), where interpretability is incorporated directly into the system's design, and b) post hoc explanations (black-box models), where the system's behaviour is first observed and subsequently analysed. In white-box models, the underlying algorithms are relatively easy to understand, with a possible interpretation of how inputs are converted into outputs or target variables, for instance, in linear regression and decision trees. On the contrary, the complexity of black-box models requires a post hoc approach, in which the model's behaviour is explained after a decision has already been produced.¹⁰⁴ On the one hand, because no general method can explain the complete behaviour of black-box models, explanations are usually limited to specific decisions or individual predictions. On the other hand, a significant advantage of local post hoc explanation methods is that they do not impose limitations on model complexity or require model disclosure.¹⁰⁵ On the contrary, “[f]rom a technical and philo[sophical] point of view these explanations can never re[veal] the [‘]unique, true reason[’] why an algorithm came to a certain decision. In complicated black-box models, such a true reason simply does not exist.”¹⁰⁶ The best-case scenario is that post hoc explanation algorithms may be able to identify some of the elements that contributed to a given decision. In an adversarial context, post hoc explanations can be either trivial or harmful, with the worst-case scenario being to induce humans to falsely believe that the decision is justified or objective, even if it is not.¹⁰⁷ To sum up, the theoretical concepts of black-box and white-box models can be seen as an alternative to the

¹⁰³ European Data Protection Supervisor, *TechDispatch #2/2023 - Explainable Artificial Intelligence* (EDPS, 2023) 4-5.

¹⁰⁴ European Data Protection Supervisor, *TechDispatch #2/2023 - Explainable Artificial Intelligence* (EDPS, 2023) 6-8.

¹⁰⁵ Sebastian Bordt and others, ‘Post-Hoc Explanations Fail to Achieve their Purpose in Adversarial Contexts’ in *Proceedings of the 2022 ACM Conference on Fairness, Accountability and Transparency* (Association for Computing Machinery 2022) 891, 891.

¹⁰⁶ Sebastian Bordt and others, ‘Post-Hoc Explanations Fail to Achieve their Purpose in Adversarial Contexts’ in *Proceedings of the 2022 ACM Conference on Fairness, Accountability and Transparency* (Association for Computing Machinery 2022) 891, 903.

¹⁰⁷ Sebastian Bordt and others, ‘Post-Hoc Explanations Fail to Achieve their Purpose in Adversarial Contexts’ in *Proceedings of the 2022 ACM Conference on Fairness, Accountability and Transparency* (Association for Computing Machinery 2022) 891, 903.

aforementioned distinction between strong and weak black-boxes. Both approaches indicate that machine learning algorithms present a full spectrum of accuracy and explainability and therefore cannot be generalised and expected to meet the same criteria.

The black-box problem in the automotive sector poses difficulties in fixing deep learning systems when they produce unwanted outcomes. As an example, if an automated vehicle hits a pedestrian instead of braking as expected, the black-box nature of machine learning algorithms implies that we cannot identify the reasoning behind the system's decision-making process or determine why a particular decision was made.¹⁰⁸ If such an accident occurred because the perception system failed to detect the pedestrian, it would be attributed to the system encountering something novel in that scenario. Subsequently, developers would then attempt to identify the root cause and retrain the system using similar situations so it would learn to improve its performance. On the contrary, the biggest challenge lies in obtaining training data capable of covering all possible real-world situations. Variations in weather, visibility, and road conditions, such as sunny and foggy conditions or salted roads, may affect how the system perceives its surroundings. Given the virtually infinite number of possible permutations, it is impossible to determine if the system is sufficiently robust to handle every situation.¹⁰⁹ Robustness poses a hurdle to trusting deep learning systems in terms of safety. According to Associate Professor Rawashdeh, there are two possible approaches. One is to stop using deep learning systems in areas where the potential harm is high. The other is to make deep learning more transparent, potentially using XAI, though it is still an emerging field¹¹⁰ with its downsides, as already mentioned in this subchapter. In practice, automated vehicles are often not based solely on end-to-end AI models. Many ADS architectures comprise machine learning systems responsible for perception and environmental recognition, as well as rule-based behaviour planning and vehicle control systems. While AI components identify pedestrians, vehicles, traffic signs, and other related environmental factors, rule-based systems constrain, validate, and execute safety-critical decisions, such as braking, steering, and lane changes. These hybrid approaches are considered advantageous, particularly in terms of explainability, safety, and, therefore, trustworthiness. Moreover, deep learning

¹⁰⁸ Lou Blouin, 'AI's mysterious 'black box' problem, explained' (*M Dearborn News*, 6 March 2023) <<https://umdearborn.edu/news/ais-mysterious-black-box-problem-explained>> accessed 13 May 2026.

¹⁰⁹ Lou Blouin, 'AI's mysterious 'black box' problem, explained' (*M Dearborn News*, 6 March 2023) <<https://umdearborn.edu/news/ais-mysterious-black-box-problem-explained>>.

¹¹⁰ Lou Blouin, 'AI's mysterious 'black box' problem, explained' (*M Dearborn News*, 6 March 2023) <<https://umdearborn.edu/news/ais-mysterious-black-box-problem-explained>>.

algorithms have achieved a major breakthrough, as they are able to learn the problem's structure while simultaneously solving it.¹¹¹

The European Union seeks to address the black-box nature of algorithms used in the automotive sector by establishing specific requirements for the type-approval of automated vehicles, as set out in Article 11, letter d), of Regulation 2019/2144. In particular, the mentioned Article introduced technical requirements for event data recorders in automated vehicles. According to Article 3 (13), “‘event data recorder’ means a system with the only purpose of recording and storing critical crash-related parameters and information shortly before, during and immediately after a collision[.]”¹¹² Article 6 (4) sets out the requirements for event data recorders to capture and retain operational vehicle data immediately before, during and right after a collision, including information concerning the speed of the vehicle, position, braking, and tilt of the vehicle, as well as the state and activation rate of safety systems and accident-avoidance systems, while ensuring data accuracy and survivability. Moreover, they cannot be deactivated, they must operate in a closed-loop system. The collected data must remain anonymised and safeguarded against manipulation or misuse, while still enabling identification of a specific vehicle type and version, as well as the active safety and accident-avoidance systems installed. Also important is the requirement to make the collected data available to national authorities for authorised purposes.¹¹³

Recital 13 further explains the aim of implementing event data recorders, which are capable of storing crucial anonymised operational vehicle data before, during and immediately after the collision. The qualitative requirements demand a specific data range, resolution, and accuracy, and for data collection, retrievability, and storage thereof. In addition, the collected data may be used only in a manner that ensures its sole purpose for road safety assessment, without allowing the identification of the vehicle owner or holder.¹¹⁴ The implementation of Event Data Recorders could provide valuable insights into traffic accidents, especially involving automated vehicles, and adequately address accident investigations and general vehicle safety assessments.¹¹⁵ These systems

¹¹¹ Frédéric Bouchard and others, ‘A Rule-Based Behaviour Planner for Autonomous Driving’ in Guido Governatori and Anni-Yasmin Turhan (eds), *Rules and Reasoning* (Springer 2022) 263, 263-264.

¹¹² Regulation 2019/2144.

¹¹³ Regulation 2019/2144.

¹¹⁴ Regulation 2019/2144.

¹¹⁵ Neonode, ‘Beyond the Black Box: How EDR and DMS Integration Will Rewrite Vehicle Safety’ (Neonode, 28 March 2025) <<https://ja.neonode.com/newsroom/news/beyond-the-black-box-how-edr-and-dms-integration-will-rewrite-vehicle-safety>> accessed 25 May 2026.

*“overwrite themselves on each loop until a triggering event—such as a crash or sudden stop—prompts the system to preserve a snapshot, typically spanning 5–20 seconds before and after the incident.”*¹¹⁶ Although the event data recorder can improve accident reconstruction and simplify evidentiary mechanisms by recording and storing outputs, vehicle behaviour and operational parameters, it may be argued that it cannot possibly fully eliminate algorithmic opacity as it cannot comprehend the internal decision-making processes of AI algorithms, such as what has happened within the deep layers of neural networks in each specific case.

3. Comparative evaluation of the EU civil liability framework

The core aim of this chapter is to evaluate whether the EU civil liability framework focused on the general safety of products is sufficient to overcome the challenges inherent to AI-driven autonomous vehicles identified in the second chapter. Subsequently, this chapter presents possible *de lege ferenda* solutions to improve legal certainty while adequately reflecting the operational complexity of AI-driven autonomous vehicles. Furthermore, it compares the different legislative approaches with the EU approach. The European Union is the first jurisdiction to attempt to provide a comprehensive answer to the responsibility gap created by technological advances involving automated vehicles and AI-driven systems. The law of non-contractual liability was not designed to address acts without intent, as it first assumed a mind behind all choices. Unlike humans, AI makes its decisions based on input, calculations, and learned behaviours. The solution of the EU liability framework is to redirect the liability onto economic operators who found themselves in the best position to control the risk. Meanwhile, it is focused on the best possible way to compensate the victim rather than on the technology. Moreover, the opacity of AI systems does not justify the exclusion of liability. On the contrary, the deployment of black-box systems justifies a stricter liability standard for those who choose to use certain algorithms.¹¹⁷ Since the AI Liability Directive was not adopted due to a lack of foreseeable agreement, the PLD Directive remains the core legislative instrument in the

¹¹⁶ Neonode, ‘Beyond the Black Box: How EDR and DMS Integration Will Rewrite Vehicle Safety’ (Neonode, 28 March 2025) <<https://ja.neonode.com/newsroom/news/beyond-the-black-box-how-edr-and-dms-integration-will-rewrite-vehicle-safety>>.

¹¹⁷ Ronald Mpofu, ‘Liability for Autonomous Vehicle Decisions: Comparative Legal Analysis of the EU, China, and Türkiye’ (2025) <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5759786> accessed 19 May 2026, 4-5.

European Union. As the EU approach pursues harmonisation rather than full legal unification in this regard, member states have discretion in implementing the revised PLD Directive into their national legal systems. The next part of this chapter will outline the key differences in the approaches of different jurisdictions, focusing on liability allocation and the balance between innovation, legal certainty and victim protection.

3.1 China

In contrast to the EU approach, China does not currently have a comprehensive law governing automated vehicles, as the Road Traffic Safety Law and the Product Quality Law still rely on a human driver, creating a significant gap in the allocation of liability for AV accidents.¹¹⁸ Nevertheless, significant regulatory developments have emerged at the lower levels, particularly through local regulations, normative documents, technical standards, and numerous pilot projects aimed at exploring and shaping the future regulation of autonomous driving. On the national level, the technical standard on autonomous driving, namely the Classification of Driving Automation for Vehicles (GB/T 40429-2021), divides driving automation into the various levels of automation¹¹⁹ in a manner comparable to the SAE Standards. In 2020, the government issued the Strategy for the Innovative Development of Intelligent Vehicles, which outlined the goal of large-scale production of Level 3 automated vehicles by 2025 and expanded use of Level 4 cars in market-oriented applications. Moreover, the Chinese government promotes integration of vehicles, road structure and cloud-based systems through “vehicle-road-cloud integration” initiatives and pilot projects in 20 cities. For instance, the governance adopted for such pilot projects is the Shenzhen Special Economic Zone Intelligent Connected Vehicle Administrative Regulations (hereinafter referred to as “Shenzhen Regulations”), which represent a comprehensive framework for intelligent connected vehicles by governing their market entry requirements and registration, operation on public roads, vehicle-road collaboration, data governance, as well as rules concerning traffic accidents and liability.¹²⁰ Article 3 of the Shenzhen Regulations demonstrates a regulatory shift away from the traditional human-

¹¹⁸ Andrew Zhang and others, ‘China’s Autonomous Vehicle Regulations’ (2025) Asia Business Law Journal <<https://law.asia/china-autonomous-vehicle-regulations/?utm>> accessed May 2026.

¹¹⁹ Andrew Zhang and others, ‘China’s Autonomous Vehicle Regulations’ (2025) Asia Business Law Journal <<https://law.asia/china-autonomous-vehicle-regulations/?utm>>.

¹²⁰ Andrew Zhang and others, ‘China’s Autonomous Vehicle Regulations’ (2025) Asia Business Law Journal <<https://law.asia/china-autonomous-vehicle-regulations/?utm>>.

centric model by distinguishing among conditional automatic driving, highly automatic driving, and fully automatic driving, thereby enabling autonomous driving without manual operation.¹²¹ The classification thereof corresponds to the SAE levels of automation 3, 4, and 5. Article 34 requires conditional and highly autonomous vehicles to retain manual driving functions, corresponding devices, and a driver, whereas fully autonomous vehicles may operate without such elements, provided that their operation remains limited to designated areas and road sections preapproved by public authorities. Moreover, Article 35 obliges the driver to use automatic driving in accordance with the manual and traffic regulations, to sit in the driver's seat, to continuously monitor the vehicle and its surroundings, and to be ready to take immediate control of the vehicle. Meanwhile, a fully autonomous vehicle is required to autonomously implement risk-reducing measures or enable remote takeover in the event of compromised traffic safety. Another point to mention is the requirement set out in Article 36 for the manufacturer to install an external indicator light for autonomous driving mode, to warn other vehicles and pedestrians that the car is driving autonomously.¹²² Most importantly, Article 37 establishes a duty to *“(…) record and store data such as the position, operating status, driving mode and surveillance video inside and outside the vehicle at least 90 seconds before an accident or failure of the vehicle and maintain the continuity and integrity of such data.”*¹²³ Furthermore, the recorded data must be retained for a minimum period of 30 days. The most significant articles on traffic accidents involving autonomous vehicles are 53, 54, and 56, which comprehensively allocate liability among the parties involved while adopting different approaches for different levels of autonomy. In compliance with Article 53, when an autonomous vehicle with a driver causes damage, liability for the traffic accident rests with the autonomous vehicle, and the driver bears liability for compensation. Article 54 complements this framework by allowing the driver, owner, or manager to seek recourse from the manufacturer or seller under certain conditions. Firstly, the traffic accident must be caused by the product's defectiveness. Secondly, the compensation according to Article 53 must be made prior to claiming recourse. Article 56 lays down that objective data recorded by onboard equipment, roadside equipment, and supervision platforms are eligible to serve as important evidentiary material for liability allocation in AV-related accidents.¹²⁴

¹²¹ Regulations of Shenzhen Special Economic Zone on the Administration of Intelligent Connected Vehicles (Shenzhen Municipal People's Congress, 2022) (Shenzhen Regulations).

¹²² Shenzhen Regulations.

¹²³ Shenzhen Regulations.

¹²⁴ Shenzhen Regulations.

3.2 United Kingdom

The legal framework governing automated vehicles in the United Kingdom is based on two legislative pillars: a) the Automated and Electric Vehicles Act 2018 and b) the Automated Vehicles Act 2024. The Automated and Electric Vehicles Act 2018 sets out rules on civil liability and insurance for accidents involving automated vehicles operating in self-driving mode. The Act extends the traditional compulsory motor insurance regime to cover automated driving and aims to facilitate efficient compensation mechanisms for victims of traffic accidents arising therefrom.¹²⁵ Section 1 of AEVA 2018 introduced a statutory mechanism, whereby the Secretary of State is required to prepare and maintain a list of all motor vehicles capable of driving autonomously and may be used in compliance with statutory requirements for automated driving on roads or other public places under certain circumstances in Great Britain. Section 2 established a specialised liability regime for accidents involving automated vehicles operating in self-driving mode, which provided separate liability rules for insured and uninsured automated vehicles. If an automated vehicle is insured at the time of the accident, and an insured person or any other individual suffers damage resulting from a traffic accident caused by an automated vehicle in self-driving mode on a road or other public places in Great Britain, the insurer is liable for the damage. If the automated vehicle is not insured at the time of the accident, liability for damage is imposed on the owner of the respective automated vehicle, provided that the same conditions set out in Section 2 remain fulfilled. In connection with Section 2, Section 5 provides the insurer or owner of the vehicle that compensated the injured party with the right to seek recourse from any person legally responsible for the accident.¹²⁶ Furthermore, Section 3 sets out the contributory negligence rules, under which the liability of the insurer or owner for damages is reduced if the injured party contributed to the damage. Moreover, the insurer or owner may be exempted from liability where the accident was entirely attributable to the negligence of the person in charge of the vehicle. In addition, Section 4 further introduces liability exceptions for the insurer in situations involving prohibited

¹²⁵ Simmons & Simmons, 'Liability to Licensing: Navigating the UK's New Self-Driving Regime' (*Simmons & Simmons*, 23 February 2026) <<https://www.simmonssimmons.com/en/publications/cmlzfosf700bgu454589uqmmi/liability-to-licensing-navigating-the-uk-s-new-self-driving-regime>> accessed 24 May 2026.

¹²⁶ Automated and Electric Vehicles Act 2018.

software alterations or the failure to install safety-critical software updates known or reasonably expected to be known.¹²⁷

The Automated Vehicles Act 2024 establishes the pre-market legal framework focusing on the lawful operation, authorisation and safe deployment of automated vehicles.¹²⁸ It established a new legal basis for regulating automated vehicles in the United Kingdom. Moreover, the government considers the automated driving industry an important driver of economic growth and investment, with projections indicating that the UK automated driving industry will be worth around £42 billion and create approximately 38 000 more skilled jobs by 2035. Effective cooperation among industry, the Government, and Parliament is expected to play a significant role in shaping the future legal framework for automated vehicles in the United Kingdom. The year 2026 is anticipated to introduce the first pilot programmes for automated passenger services on United Kingdom roads, accompanied by conducting a number of public consultations and evidence-gathering initiatives aimed at further shaping the regulatory landscape, intended to enter into force in the second half of 2027, while balancing the promotion of technological innovation while protecting public safety.¹²⁹

4. Proposals for adjustments and future development

The analysis in the second chapter has demonstrated that the European Union's core legislative pillars for the deployment of automated cars are the revised PLD Directive, pre-market regulations in the automotive sector established by Regulations (EU) 2019/2144 and (EU) 2018/858, and ancillary provisions of the AI Act. It is important to highlight the revised PLD Directive, as this initiative has demonstrated the EU legislative switch more towards intangible products such as software and AI-driven systems. At the same time, the analysis has shown that despite a significant shift towards technology, several legal questions remain to be resolved in the era of automated vehicles. On the one hand, automated vehicles should not be perceived just as a commercial product, instead, they

¹²⁷ Automated and Electric Vehicles Act 2018.

¹²⁸ Simmons & Simmons, 'Liability to Licensing: Navigating the UK's New Self-Driving Regime' (*Simmons & Simmons*, 23 February 2026) <<https://www.simmonssimmons.com/en/publications/cmlzfosf700bgu454589uqmmi/liability-to-licensing-navigating-the-uk-s-new-self-driving-regime>>.

¹²⁹ Oliver Wilson and others, 'Automated Vehicles in the UK: What's coming down the road in 2026?' (*Hogan Lovells*, 29 January 2026) <<https://www.hoganlovells.com/en/publications/automated-vehicles-in-the-uk-whats-coming-down-the-road-in-2026>> accessed 25 May 2026.

should be seen as a measure capable of significantly reducing road fatalities caused by human driver error. On the other hand, they pose new risks associated with the inherent nature of their algorithms, including those AI-based. Since the first use of AI in the modern sense, all jurisdictions have struggled to adopt appropriate legislative measures to address the rapidly evolving technological landscape and to understand the new risks associated with AI algorithms commonly used in applications. The member states of the European Union are still within the transposition period for the revised PLD Directive until 9th December 2026, meaning it remains to be seen how these rules will be implemented in the national legal systems of member states and whether they will be effective in practice. Similarly, the AI Act provisions regarding high-risk AI systems have not yet come into force and therefore have not been incorporated into automotive sector-specific legislation. As already mentioned, the AI Act's provisions on high-risk AI are exempt from direct application to the automotive industry under Article 2(2). They will apply indirectly through the automotive legislation's adoption thereof. That is the reason why the analysis of its shortcomings will not go any further, as it is not yet clear to what extent the sector legislation will reflect the respective provisions.

Apart from the AI Act, one of the main shortcomings of the PLD Directive is that it does not distinguish between different levels of automation, which represent different allocation of dynamic driving tasks among driver and vehicle, and should therefore be substantial in determining liability for traffic accidents. The reason likely stems from the PLD Directive being intended for general product liability rather than tailored to the automotive industry, which can also be seen as a shortcoming. On the contrary, nothing prohibits member states from distinguishing between levels of autonomy within their national legislation, although this can contribute to increased regulatory fragmentation between member states. Germany currently represents one of the few jurisdictions which distinguishes between levels 3 and 4 of automation in their legislation. In the future, it may be beneficial to create a separate EU regulation tailored to automated vehicles aimed at the unification of core aspects. Such a framework could, for instance, adopt SAE levels of automation into the European Union legal environment or similar standards, which would emphasise the differences between ADS and human driver involvement in performing DDT for each level of autonomy. Perhaps it could further establish distinct principles for allocating liability, rather than uniformly imposing strict liability standards on economic operators, at higher levels of autonomy. As the unified automated vehicle framework would contribute to the EU's "Vision Zero" initiative, it would further pursue the common

objective of eliminating road fatalities by 2050. The new regulation could govern both special pre-market requirements for automated cars and all questions connected to deployment, as well as all legal issues, such as product liability, victim compensation schemes, and evidentiary mechanisms. The complexity of AI-driven automated vehicles deserves uniform legislation that would adequately address their complexity. Moreover, unifying all questions related to automated vehicles into a single regulation could also have a positive impact on legal certainty, predictability, and, therefore, the overall functioning of the internal market.

4.1 Defectiveness and allocation of liability

As for defectiveness, strict liability frameworks with limited options for exemption from liability could have a chilling effect on manufacturers and result in stagnation and a limitation of innovation in the respective area, as they will naturally try to minimise potential legal consequences arising from the deployment of automated vehicles. Compared with human driver error, automated vehicles will almost certainly cause fewer accidents, so the question is whether the European Union prefers to maintain a strict liability regime at the expense of slowing down the innovation and deployment of state-of-the-art technologies that could improve overall road safety. From a different perspective, the economic operators cannot face the same liability consequences for Level 3 and 4 as for Level 5. Perhaps a shift towards stricter pre-market regulation of AV deployment and higher qualitative standards for training data could bring better results for both overall road safety and for fostering innovation. This issue also has another dimension related to external factors that may influence AI-driven ADS recognition and, therefore, the final decision of the rule-based system, such as accidents caused by worn-down lane markings or poor weather conditions. What is also interesting is the fact that the EU deems the manufacturer's manuals, which encompass the product's limitations, as insufficient measures for considering a defective product as safe. However, what if the product should not be deemed defective, and the disclaimer in the manual represents justified limitations of AI-driven ADS? Machine learning algorithms cannot reasonably be expected to work properly on worn-down lane markings, with dirty perception sensors, or in difficult weather conditions and certainly, economic operators cannot be held liable for those. Perhaps a shift towards stricter pre-market regulation for AV deployment, rather than post-accident liability, could enable greater technological advances while still protecting injured parties.

This could be achieved with higher standards for training, testing and validation datasets imposed on AI systems within the automotive sector, along with higher demands imposed on the readability and quality of road infrastructure, since AI-driven ADS inherently depend on the quality of input. De lege lata, evidentiary and liability mechanisms impose strict liability on economic operators, as it remains unclear how the evidentiary process should be assessed when black boxes are included. The revised PLD Directive introduces the rebuttable presumption of defectiveness and alleviates the burden of proof for the injured party. It can be argued that adopted evidentiary rules do not lead to the best outcome, as, in the de lege lata state, greater stress is placed on compensating alleged victims in many circumstances rather than on adopting provisions aimed at finding the real cause of accidents. Moreover, the inapplicability of vehicle manual disclaimers could pose a significant hurdle for manufacturers, as these disclaimers encompass legitimate ADS limitations which cannot be eliminated. The aforementioned facts may place economic operators at an unjust disadvantage. As for the claimants, they still need to prove defectiveness, damage suffered, and a causal link between them, which may also be challenging or impossible given the complexity of ADS architectures and the involvement of interconnected software and hardware components.

4.2 Trolley Problem

Regarding the trolley problem, as it is a rather philosophical dilemma and, at best, an edge case in real life, it has not yet been directly reflected in legislation governing automated vehicles. Since it is necessary to preserve the European Union's values and respect international treaties, algorithms that choose among people and assign different values to their lives are unacceptable. Consequently, the regulatory approach to the respective sensitive questions must be based on ethically neutral choices that aim to minimise overall damage. Perhaps there could be a rule that requires programmers to programme ADS so that, if the system determines in a given case that a traffic accident is unavoidable, it will be prompted to do as little harm as possible. Moreover, it would not make sense to program ADS in a way that would minimise external harm at the expense of passenger lives. No reasonable person would want to buy an automated car behaving in such a manner. The view of the customers is one of the most important factors deciding how many automated cars will be present in the traffic, so if the aim is concentrated on reducing accidents caused by human driver error, then the conditions of ADS usage must

not only benefit overall road safety but also individual customers when compared to the traditional vehicles. Consequently, the trolley problem should not be understood as a doctrine allowing the creation of algorithms capable of making morally and statutorily correct choices in sensitive situations, since the consensus on what is permissible and what is not is not even within philosophy itself. The legal framework should instead concentrate on ensuring predictable, transparent, and ethically neutral decision-making focused on minimising overall harm to all involved parties (passengers, pedestrians, other vehicles) while respecting the fundamental legal rights and values of the European Union Treaties and the ECHR. In practice, ADS should be designed to generally prevent accidents and, in unavoidable scenarios, to reduce the severity of harm without valuation of different human lives. For instance, these systems could be programmed to minimise the probability of fatal injuries by reducing the severity and angle of impact in each case, lowering vehicle speed, and maintaining overall vehicle stability.

4.3 Black-box Problem

The black-box nature of AI algorithms remains one of the most significant challenges for the future development of automated vehicles, as ADS become more sophisticated and increasingly reliant on machine learning algorithms. As previously discussed, ADS systems comprise AI algorithms responsible for pattern recognition and object detection. This information is then fed into rule-based, deterministic algorithms that make the final decision. Nowadays, deep learning algorithms play a critical role for automated driving, as without them the system would simply not work. Their pattern recognition and object detection quality is superior to more explainable algorithms, such as decision trees, random forests, or other statistical methods. The automated systems require a fusion of complex software and architectural design patterns where cameras, sensors, and sometimes LIDAR are required to capture real-time data of vehicle surroundings. This data is captured, processed, and propagated to deep neural networks to make an inference. Subsequently, the results of an inference are fed into deterministic systems, and the final decisions of those systems are propagated to various actuators. The actuators translate the information signal into real-world actions such as acceleration, deceleration, longitudinal or lateral movement. Direct replacement of the mentioned deep learning methods is not possible with deterministic methods. Therefore, the solution is not to avoid using black-box algorithms in order to improve explainability. Instead, the regulatory efforts should

concentrate on enhancing their explainability to the greatest extent possible. The opacity problem can be partially addressed through XAI post hoc methods, which can identify some of the factors that influenced a particular decision. Moreover, the European Union initiative to implement event data recorders can serve as an important tool for improving accident investigations. Nevertheless, the respective tools can only mitigate, rather than eliminate, the difficulties posed by deep learning algorithms, since the internal processes within the deep layers may never be fully explainable in a deterministic manner. As for future regulatory proposals, EU legislators should focus on reducing the negative effects associated with black boxes, for instance, by further developing XAI, raising data quality requirements, and adopting more sophisticated event data recording systems. Furthermore, the objective for the future should not be perfect explainability but rather reaching an acceptable level of transparency. If AI-driven automated vehicles can achieve significantly fewer accidents than human drivers, then the black-box opacity must be accepted to some extent to achieve the ultimate goal of reducing road fatalities and improving overall road safety. Once ADS are statistically safer than human drivers, it should be less important how they precisely arrive at a particular decision.

Conclusion

This master's thesis examined the legal challenges associated with deploying AI-based automated vehicles within the European Union legal framework. The main objective of this thesis was to assess whether the current regulatory framework is capable of addressing the inherently complex nature of AI systems, with a particular emphasis on defectiveness, trolley-type scenarios, and their black-box nature. To answer the first research question, the analysis has demonstrated that the combined application of the AI Act and the Product Liability Directive does not provide a coherent and practically enforceable civil liability framework for automated vehicles. Although both legal instruments represent an important step towards a comprehensive legal framework for emerging technologies, several important issues remain unresolved. In particular, the current PLD Directive is geared toward general product liability rather than tailored to automated vehicles, resulting in a lack of specific rules that reflect distinctions between levels of automation and their influence on the distribution of the dynamic driving task between driver and ADS. This creates great legal uncertainty in liability allocation and limits the effectiveness of the *de lege lata* framework. Regarding the second research question, the analysis showed that defectiveness should not be interpreted solely through traditional general product safety standards based on the objective safety expectations of the general public, without considering the level of automation, driver involvement, operational limitations outlined in product manuals, and the technical characteristics of automated driving systems. *De lege lata* interpretation can often lead to incorrect conclusions, as it fails to account for the inherent technological limitations of automated driving systems, which could be deemed defective even if the cause of the accident was a justified technological limitation resulting from wrongful use of ADS or unreasonable expectations of AI models. As for the third research question, the evidentiary mechanisms introduced in the revised PLD Directive are insufficient to overcome the opacity and limited explainability inherent in AI-based automated vehicle systems. To put it differently, alleviation of the burden of proof of the claimant plus a rebuttable presumption in this regard, without solving the main hurdle of the evidentiary process being the opacity of AI algorithms, could lead to unjust consequences for economic operators, thereby indirectly negatively influencing innovation. Greater legal certainty and predictability in court proceedings involving automated vehicles could be achieved by improving evidentiary mechanisms for AV accidents. The combination of event data recorders, driver monitoring

systems, and XAI should provide a better understanding of the root cause of each accident. To answer the fourth research question, the analysis shows that defect-based liability cannot adequately address unavoidable harm in trolley problems arising from automated vehicle accidents, thereby leaving a significant legal gap. As was already mentioned, the legal framework should focus on adopting ethically and technologically neutral decisions that minimise overall damages, rather than assigning different values to human lives, as this would be contrary to fundamental human rights such as the right to life, human dignity, and the principle of non-discrimination. The findings in this thesis, based on comprehensive research, confirmed the first hypothesis, namely that the combined application of the AI Act and the PLD Directive does not establish a fully coherent and practically enforceable civil liability regime for autonomous vehicles, mainly due to the identification of the legal gaps in the following areas: defectiveness, trolley problem and black-box problem. Likewise, the second hypothesis was confirmed as the analysis has demonstrated that the black-box character of AI systems and trolley-type scenarios revealed a structural tension within the framework of defect-based product liability. Namely, the revised PLD Directive retains human-centric provisions, whereas AI models with their decision-making processes do not align with traditional civil liability concepts. This thesis is limited to questions concerning civil liability for AV accidents, thereby deliberately excluding analysis of criminal liability, data protection law, and cybersecurity law. In conclusion, future regulatory developments should aim to strike a balance between victim protection and economic operator's liability, as the established liability standard has a direct influence on promoting innovation in the automotive sector, while simultaneously ensuring technological feasibility, legal certainty, better explainability of black boxes, and ultimately respect for the European Union's values. In addition, the future framework could be more concentrated on unification within the European Union, and according to different levels of automation, it could set different liability principles for each level of autonomy, factoring in the potential influence of the driver, economic operator, as well as institutions responsible for road and traffic sign quality in each member state.

Table of literature

Abbott R, 'The Reasonable Computer: Disrupting the Paradigm of Tort Liability' (2016) 86 (1) George Washington Law Review.

Awad E and others, 'The Moral Machine experiment' (2018) 563 Nature 59.

Bathae Y, 'The Artificial Intelligence Black Box and the Failure of Intent and Causation' (2018) 31 (2) Harvard Journal of Law & Technology 889.

Bergmann D, 'What is machine learning?' (IBM) <<https://www.ibm.com/think/topics/machine-learning>>.

Blouin L, 'AI's mysterious 'black box' problem, explained' (*M Dearborn News*, 6 March 2023) <<https://umdearborn.edu/news/ais-mysterious-black-box-problem-explained>>.

Bordt S and others, 'Post-Hoc Explanations Fail to Achieve their Purpose in Adversarial Contexts' in *Proceedings of the 2022 ACM Conference on Fairness, Accountability and Transparency* (Association for Computing Machinery 2022) 891.

Bouchard F and others, 'A Rule-Based Behaviour Planner for Autonomous Driving' in Guido Governatori and Anni-Yasmin Turhan (eds), *Rules and Reasoning* (Springer 2022) 263.

Directorate-General for Mobility and Transport, 'EU road deaths drop by 3% in 2025' (*European Commission*, 24 March 2026) <https://transport.ec.europa.eu/news-events/news/eu-road-deaths-drop-3-2025-2026-03-24_en> accessed 6 April 2026.

Dumančić, K and Vuletić D, 'Liability for Damages Caused by Autonomous Driving Vehicles from the EU Law Perspective' in Ante Novokmet and others (eds), *EU and Comparative Law Issues and Challenges Series (ECLIC 2025)* 64.

European Climate, Infrastructure and Environment Executive Agency (CINEA), *EU Road Safety: Towards "Vision Zero"* (European Commission 2022).

European Commission, ‘Annexes to Commission Work Programme 2025 Moving forward together: A Bolder, Simpler, Faster Union’ (Annex to Communication) COM (2025) 45 final.

European Commission, ‘Liability Rules for Artificial Intelligence’ (European Commission, 2022) <https://commission.europa.eu/topics/business-and-industry/doingbusinessen/contract-rules/digital-contracts/liability-rules-artificial-intelligence_en>.

European Commission, ‘Report on the Implementation of the EU Road Safety Policy Framework at the Mid-Point’ (Communication) COM (2026) 77 final.

European Data Protection Supervisor, *TechDispatch #2/2023 - Explainable Artificial Intelligence* (EDPS, 2023).

Fleetwood J, ‘Public Health, Ethics, and Autonomous Vehicles’ (2017) 107(4) American Journal of Public Health 532.

Foot P, ‘The Problem of Abortion and the Doctrine of the Double Effect’ (1967) 5 Oxford Review 5.

Gáborová G, ‘Nedostatky a riziká umelej inteligencie z pohľadu práva duševného vlastníctva’ (Master’s Thesis, Comenius University Bratislava 2025).

Goodall NJ, ‘Can you program ethics into a self-driving car?’ (2016) 53 (6) IEEE Spectrum 28.

Gunning D and others, ‘XAI-Explainable artificial intelligence’ (2019) 4 (37) Science Robotics eaay7120.

Idrees H, ‘Shallow Learning vs. Deep Learning: Is Bigger Always Better?’ (*Medium*, 27 October 2024) <<https://medium.com/@hassaanidrees7/shallow-learning-vs-deep-learning-is-bigger-always-better-51c0bd21f059>>.

Kahl, T, 'AI Act and the Automotive Industry – Where does the road lead?' (*Taylorwessing*, 25 March 2025) <<https://www.taylorwessing.com/de/insights-and-events/insights/2025/03/ai-act-and-the-automotive-industry>>.

Kop M, 'EU Artificial Intelligence Act: The European Approach to AI' (2021) 2/2021 Stanford - Vienna Transatlantic Technology Law Forum, Transatlantic Antitrust and IPR Developments, Stanford University <<https://law.stanford.edu/publications/eu-artificial-intelligence-act-the-european-approach-to-ai/>>.

Levin S and Woolf N, 'Tesla driver killed while using autopilot was watching Harry Potter, witness says' *The Guardian* (San Francisco, 1 July 2016) <<https://www.theguardian.com/technology/2016/jul/01/tesla-driver-killed-autopilot-self-driving-car-harry-potter>>.

Lin P, 'Why Ethics Matter for Autonomous Cars' in Markus Maurer and others (eds), *Autonomes Fahren Technische, rechtliche und gesellschaftliche Aspekte* (Springer-Verlag GmbH 2015) 69.

Meier K and Spichiger R, 'The EU AI Act : What it means for your business' (*EY*, 15 March 2024) <https://www.ey.com/en_ch/insights/forensic-integrity-services/the-eu-ai-act-what-it-means-for-your-business>.

Metro Service, 'Metro Service and the Metro' (*Metro Service A/S*) <<https://www.metroservice.dk/en/what-we-do/the-metro/>>

Mpofu R, 'Liability for Autonomous Vehicle Decisions: Comparative Legal Analysis of the EU, China, and Türkiye' (2025) <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5759786>.

Neonode, 'Beyond the Black Box: How EDR and DMS Integration Will Rewrite Vehicle Safety' (*Neonode*, 28 March 2025) <<https://ja.neonode.com/newsroom/news/beyond-the-black-box-how-edr-and-dms-integration-will-rewrite-vehicle-safety>>.

Neonode, 'Driving the Future – Levels of Autonomous Vehicles Explained' (*Neonode*, 7 February 2025) <<https://neonode.com/newsroom/news/driving-the-future-levels-of-autonomous-vehicles-explained>>.

Polad *D*, 'Liability Perspective for Users of Autonomous Vehicles in the EU' (*RAILS Robotic & AI Law Society*, 15 April 2024) <<https://blog.ai-laws.org/liability-perspective-for-users-of-autonomous-vehicles-in-the-eu/>>.

SAE International, 'Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles' (SAE Standard J3016_202104, Revised April 2021, SAE International 2021) <https://doi.org/10.4271/J3016_202104>.

Schellenkens *M*, 'Self-driving cars and the chilling effect of liability law' (2015) 31(4) *Computer Law & Security Review* 506.

Simmons & Simmons, 'Liability to Licensing: Navigating the UK's New Self-Driving Regime' (*Simmons & Simmons*, 23 February 2026) <<https://www.simmonssimmons.com/en/publications/cmlzfosf700bgu454589uqmmi/liability-to-licensing-navigating-the-uk-s-new-self-driving-regime>>.

Tesla, 'Model 3 Owner's Manual' (Tesla 2026) <https://www.tesla.com/ownersmanual/model3/en_us/GUID-E5FF5E84-6AAC-43E6B7EDEC1E9AEB17B7.html>.

Thomson *J J*, 'The Trolley Problem' (1985) 94 *Yale Law Journal* 1395.

Twigg-Flesner *Ch*, *Guiding Principles for Updating the Product Liability Directive for the Digital Age* (European Law Institute 2021).

Wilson *O* and others, 'Automated Vehicles in the UK: What's coming down the road in 2026?' (*Hogan Lovells*, 29 January 2026) <<https://www.hoganlovells.com/en/publications/automated-vehicles-in-the-uk-whatscoming-down-the-road-in-2026>>.

Zhang *A* and others, ‘China’s Autonomous Vehicle Regulations’ (2025) *Asia Business Law Journal* <<https://law.asia/china-autonomous-vehicle-regulations/?utm=>>.

Table of legislation

European Parliament and Council Regulation 2019/2144 of 27 November 2019 on type-approval requirements for motor vehicles and their trailers, and systems, components and separate technical units intended for such vehicles, as regards their general safety and the protection of vehicle occupants and vulnerable road users amending Regulation (EU) 2018/858 of the European Parliament and of the Council and repealing Regulations (EC) No 78/2009, (EC) No 79/2009 and (EC) No 661/2009 of the European Parliament and of the Council and Commission Regulations (EC) No 631/2009, (EU) No 406/2010, (EU) No 672/2010, (EU) No 1003/2010, (EU) No 1005/2010, (EU) No 1008/2010, (EU) No 1009/2010, (EU) No 19/2011, (EU) No 109/2011, (EU) No 458/2011, (EU) No 65/2012, (EU) No 130/2012, (EU) No 347/2012, (EU) No 351/2012, (EU) No 1230/2012 and (EU) 2015/166 [2019] OJ L325/1 (Regulation 2019/2144) [2019] OJ L325/1.

European Parliament and Council Regulation 2024/1689 of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) [2024] OJ L 2024/1689 (AI Act).

European Parliament and Council Regulation 2018/858 of 30 May 2018 on the approval and market surveillance of motor vehicles and their trailers, and of systems, components and separate technical units intended for such vehicles, amending Regulations (EC) No 715/2007 and (EC) No 595/2009 and repealing Directive 2007/46/EC (Text with EEA relevance.) [2018] OJ L 151/1.

European Parliament and Council Directive 2024/2853 of 23 October 2024 on liability for defective products and repealing Council Directive 85/374/EEC [2024] OJ L 2024/2853.

Council Directive 85/374/EC of 25 July 1985 on the approximation of the laws, regulations and administrative provisions of the Member States concerning liability for defective products [1985] OJ L 210/29 (old PLD Directive).

Consolidated version of the Treaty on the Functioning of the European Union [2012] OJ C 326/47.

Charter of Fundamental Rights of the European Union [2000] OJ 364/01.

European Convention on Human Rights [1950].

Regulations of Shenzhen Special Economic Zone on the Administration of Intelligent Connected Vehicles (Shenzhen Municipal People's Congress, 2022).

Automated and Electric Vehicles Act 2018.