Introduction

Rapid advancements in artificial intelligence (AI) and the integration of autonomous systems into daily life have revolutionised industries and enhanced societal capabilities. From healthcare to transportation, AI-driven systems increasingly assume roles traditionally managed by humans. While these autonomous systems offer numerous advantages, they also present complex legal challenges, mainly when they cause harm. Determining liability when an AI-driven autonomous system's functioning results in damage, injury, or death raises critical questions about accountability.

Criminal liability, traditionally based on human actions, is particularly challenged by the emergence of autonomous systems. The unique nature of these systems, functioning with minimal human intervention, complicates the attribution of blame: who is responsible if a self-driving vehicle causes a fatal accident, or if an AI-driven medical device fails during surgery? Current legal doctrines, grounded in human control, struggle to address situations where machines conduct autonomously.

A significant proportion of the current legal literature focuses on a single application of AI-driven autonomous systems, with a particular emphasis on autonomous driving. These studies thus offer detailed insights into the specific obligations of individuals - *i.e.*, drivers and manufacturers- under current legal frameworks. As AI-driven robots, self-driving vehicles, offer excellent exemplars regarding the matter. However, each application of AI is subject to the relevant technical standards and detailed legislation[1]. Consequently, examining the topic within a specific sector limits it to a narrower scope. Although the present study draws upon cases from autonomous driving, its primary objective is to provide a comprehensive theoretical framework that can be applied across various contexts. Accordingly, a broader approach is sought by assessing AI-driven autonomous systems in general. The scope of this study therefore extends beyond the examination of specific types of AI, such as self-driving vehicles, industrial robots, chatbots, etc., and instead focuses on the establishment of a general liability framework for criminal offences involving autonomous systems driven by AI. As such, the structure of the analysis is centred on the general principles

1 To illustrate, for an examination of the legal aspects concerning self-driving vehicles in Germany, see: HILGENDORF, Straßenverkehrsrecht der Zukunft, 2021, p. 445 ff.

of negligent liability, rather than a detailed evaluation of the responsibilities of each individual involved in the manufacturing and operation of these systems.

Furthermore, the concept '*autonomy*' rather than '*artificial intelligence*' has been emphasised in this study. This choice is based on the rationale that, from a criminal law perspective, the primary issue lies in the autonomy of these systems, the reduced human control over them, and their potential to generate outcomes that are difficult to predict in advance. Indeed, in the future, AI may evolve in unforeseen directions, or the current hype may diminish. Even different autonomous entities, some of which may not presently fall within the definition of AI, including potentially carbon-based forms, may emerge. In such cases, the findings of this study can also be applied to those autonomous beings.

Remarkably, as with all narratives of human history, the question at the heart of this study, namely "who bears accountability if a robot (human-made creation) causes harm?", and the related stories concerning entities with self-directed movement or autonomous volition, trace back to ancient times. Indeed, the same pattern reflecting the human fascination and fear towards beings capable of autonomous action is explored in numerous ancient and literary texts: *Automatons* built by *Hephaestus*[2] or *Erewhonian machines* from *Samuel Butler*'s 1872 novel *Erewhon,* the legendary creature *Golem* from Jewish folklore (16th century) brought to life by *Rabbi Judah Loew*, or the famous *Frankenstein's monster* in *Mary Shelley*'s novel from 1818[3]. However, perhaps for the first time in modern human history, our kind is relinquishing control to autonomous beings. Consequently, we are no longer confronting mere puppets; instead, we are engaging with *Pinocchio*, a figure who has transcended his strings, and we must now consider whether *Geppetto* can be held accountable for Pinocchio's misbehaviour.

Technological advancements bring not only benefits but also risks and responsibilities[4]. The rise of data-driven technology now infuses society, making digital disengagement nearly impossible as automation, AI, and networking merge digital and physical spheres[5]. Despite the extensive bene-

2 HOMER, Book 18: The Iliad, Translation: Ian C. Johnston, 2nd edition, Arlington (Va.): Richer resources publications, 2007, p. 416.
3 LEHMAN-WILZIG, Frankenstein Unbound, 1981, p. 442.
4 WANG/MA, Preventing Crimes, 2022, p. 4.
5 FATEH-MOGHADAM, Innovationsverantwortung, 2020, p. 867.

fits, these advancements will introduce numerous legal challenges, including risk assessment, civil and criminal liability, insurability and so forth[6].

A recently published document by the OECD outlines the potential benefits and risks associated with AI while also presenting forward-looking policy recommendations[7]. Another report by the United Nations Institute for Disarmament Research (UNIDIR) highlights new privacy and security risks posed by AI systems, particularly regarding their potential misuse for malicious purposes in cybersecurity. The report emphasises the range of these risks and their possible areas of impact[8]. Additionally, it warns that AI technologies could significantly affect both national and global security by facilitating disinformation[9] and could introduce new risks in biotechnology, particularly regarding the proliferation of biochemical weapons[10]. Another recent UN study underscores that the improper or malicious design and use of AI systems may hinder sustainable development, reinforce societal biases, undermine information security, and lead to human rights violations[11].

While the avoidance of harm by robots may be desired as outlined in *Asimov*'s laws of robotics, it is statistically unavoidable. Unfortunately, these laws are not only inherently contradictory[12]; but also, from a legal perspective, they are naive[13].

Autonomous systems driven by AI complicate the determination of criminal liability due to diminished human control and unpredictable outcomes. Key issues encompass the principle of guilt, individual criminal liability, the scope of duty of care, and challenges within the causality. Consequently, given the difficulties in attributing liability in AI-related crimes,

6 HÖTITZSCH, Juristische Herausforderungen, 2015, pp. 78-93.

7 Assessing Potential Future Artificial Intelligence Risks, Benefits and Policy Imperatives, OECD Artificial Intelligence Papers, OECD Artificial Intelligence Papers No. 27, 14.11.2024, doi:10.1787/3f4e3dfb-en.

8 PUSCAS Ioana, "AI and International Security: Understanding the Risks and Paving the Path for Confidence-Building Measures", UNIDIR, 12.10.2023, https://unidir.org/publication/ai-and-international-security-understanding-the-risks-and-paving-the-path-for-confidence-building-measures/, p. 9, 22, 54. (accessed on 01.08.2025).

9 *Ibid*, p. 51.

10 *Ibid*, p. 53.

11 United Nations General Assembly, "Seizing the Opportunities of Safe, Secure and Trustworthy Artificial Intelligence Systems for Sustainable Development", Draft Resolution A/78/L.49, United Nations, 11.03.2024, https://digitallibrary.un.org/record/4040897?v=pdf. (accessed on 01.08.2025).

12 HALLEVY, The Criminal Liability, 2010, p. 173.

13 HILGENDORF, Recht und autonome Maschinen, 2015, p. 32.

some scholars advocate for the establishment of novel legal regulations and ethical principles[14]. Conversely, others argue that despite the difficulties and even gaps in assigning negligence in cases involving AI, existing criminal law norms and traditional legal theory can still effectively guide application and values overall. Thus, according to some, legal doctrine needs to find an appropriate place within the traditional legal framework to provide a more reasonable theoretical and normative basis for addressing the challenges posed by AI-related crimes[15]. Nevertheless, the establishment of new legal norms for AI-driven autonomous systems may result in the application of provisions that conflict with one another, thereby introduces legal uncertainty and overlapping[16]. In any case, diminishing human control should not result in diminished liability to uphold an effective criminal policy that balances deterrence with societal benefit.

In criminal law, even though certain issues may seem novel and complex, the foundational arguments and debates surrounding responsibility for dangerous activities have remained mostly consistent[17]. For instance, issues such as foreseeability, controllability, and avoidability were already being discussed nearly 130 years ago: during a carriage ride in 1896, a driver lost control of their wagon when the horses became agitated, leading to an accident in which a blacksmith was knocked over and suffered a broken leg. The *Reichsgericht* ruled that although the injury was foreseeable, negligence could only be established due to a failure to exercise proper care[18].

In order to provide a thorough evaluation of issues related to criminal law, **the first chapter** of this study begins by introducing the challenges of liability in crimes involving AI-driven autonomous systems. Detail is given to the primary reasons for analysing these crimes separately from other offences, particularly due to their distinct *ex ante* and *ex post* characteristics. **In the second chapter**, the emergence of crimes involving AI-driven autonomous systems is explored. Here, it is observed that the term "crimes involving autonomous systems" is preferred over "crimes caused

14 STANILA Laura, Living in the Future, 2020, p. 300, 308, 310.
15 ZHAO, Principle of Criminal Imputation, 2024, p. 38 f.
16 EBERS, Truly Risk-Based, 2024, p. 18 ff.
17 GLESS, Mein Auto, 2016, p. 232.
18 Reichsgericht in Strafsachen (RGSt), decision of 23.03.1897, Case No. Rep. 576/97, RGSt V. 30, p. 25 (*Leinenfänger case*), https://opinioiuris.de/sites/default/files/RG,%2023.03.1897%20-%20Rep.%2057697%20-%20RGSt%2030,%2025.pdf. (accessed on 01.08.2025). GROPP/SINN, § 12 Fahrlässigkeit in Strafrecht AT, 2020, p. 588 Rn. 185 ff.; KASPAR, § 9 Fahrlässigkeitsdelikte in Strafrecht AT, 2023, p. 233 Rn. 66.

by autonomous systems". Accordingly, the chapter examines how these systems become associated in criminal activities, highlighting the aspects that distinguish the negligent liability of the person behind the machine. **The third chapter** examines various liability models proposed in legal doctrine to overcome the challenges associated with criminal liability. Within this framework, the widely debated concepts of 'robot liability' and 'electronic personhood' are also discussed. Subsequently, **the fourth chapter** addresses the central focus of the study: the criminal liability of the person behind the machine. Here, the foundations of negligent liability and the boundaries of the duty of care are analysed, particularly in terms of permissible risk and the principle of reliance. Since the study focuses broadly on autonomous systems rather than a specific AI application, its structure is not organised by categorising the liability of manufacturers, operators and so on. Practical guidance is provided to practitioners and those behind the machine through concrete delineations of the limits of the duty of care, illustrated with real-world examples. This chapter also examines the 'dilemma situations' that are frequently discussed in literature. Finally, in **the fifth chapter**, suggestions for *de lege ferenda* are presented. Here, prominent proposals aimed at addressing the challenges of criminal liability through concrete legislative recommendations are examined.

The study adopts German law as its primary analytical framework. However, due to significant parallels with Turkish law, it remains pertinent to both legal systems. Descriptive sections have been deliberately kept concise. Nonetheless, given that the study is written in English, it is intended also to serve as a useful resource for readers from the Anglo-American legal tradition, who may be less familiar with the criminal law dogmatics characteristic of Continental Europe. Accordingly, certain foundational topics (such as the concept of negligence) are explored in greater depth to facilitate engagement with such readers. Theoretical discussions are not presented in abstract isolation but are instead contextualised and illustrated through recent concrete examples closely aligned with the subject matter. Where appropriate, the study also draws attention to areas of convergence and divergence between the respective legal traditions.

Finally, to maintain coherence and semantic flow throughout the study, extensive cross-referencing has been employed. This enables readers to easily navigate related discussions across different sections, particularly where issues addressed under one heading bear relevance to others.