The Problem of Sanctions Applicable to Artificial Intelligence

Olgun Değirmenci*[1] and Sümeyye Mengüç[2]

[1.]Department of Criminal Law and Law of Criminal Procedure, Faculty of Law, TOBB University of Economics and Technology, Ankara 06560, Türkiye; (O.D. ORCID: 0000-0002-0700-2549)
[2.]Public Law Master's Program, Graduate School of Social Sciences, TOBB University of Economics and Technology, Ankara 06560, Türkiye, ORCID: 0009-0008-7952-6780
* Corresponding author: Olgun Değirmenci (odegirmenci@etu.edu.tr)

Abstract: This study examines whether criminal sanctions can be applied to artificial intelligence, using the method of literature review. First, the aims of criminal sanctions are outlined, followed by an overview of the historical development of criminal sanctions. The criminal liability—based on intent or negligence—of the real persons behind AI systems, such as producers, programmers, and users, is then evaluated within the framework of current legal regulations. In this context, the analysis is limited to narrow artificial intelligence, which imitates human intelligence only within predetermined goals. Accordingly, the study considers the sanctions applicable to individuals behind such AI systems, particularly within the scope of criminal liability arising from intent and negligence. Subsequently, the study explores whether strong AI systems—that is, highly autonomous and conscious artificial intelligence entities that may emerge in the future—can bear direct criminal responsibility for actions constituting criminal offenses. In addition, possible types of sanctions that could be imposed on such systems are analyzed, and the feasibility of enforcing such measures within existing legal frameworks is questioned. Finally, the adequacy of current regulations in addressing crimes involving AI is assessed, and recommendations are made concerning the potential need for new legal frameworks.

Keywords: Artificial Intelligence, Crime, Penalty, Criminal Liability, Sanction

1. Introduction

It is important to find answers to whether the issues related to the sanctions that may come from accepting criminal liability for the use of artificial intelligence—and even recognizing their criminal responsibility—should be handled under our current legal framework or if new laws should be created due to the changing nature of criminal law. Criminal law and its related penalties focus on humans and are not meant to punish non-human entities. Even the concept of criminal liability for legal entities ultimately concerns individuals who benefit from those entities' assets (Gless et al., 2016).

The goals of criminal law sanctions will help clarify the limits of future laws that apply to AI, especially when AI entities might become involved in criminal actions. This study will first outline the goals of criminal law sanctions. Subsequently, the sanctions applicable under current regulations to the individuals behind AI systems—based on criminal liability arising from intent and negligence—will be examined. Finally, on a hypothetical level, the potential sanctions that could be imposed directly on highly advanced AI systems, expected to arise in the future, for conduct constituting criminal offenses, will be explored.

2. The Aims and Types of Criminal Law Sanctions

Punishment, as a sanction imposed by the state in response to the commission of a crime, is essential for the preservation of society, the state, and the legal order (Centel, 2001). The legitimacy of penal authority is closely tied to the underlying purpose of punishment, which is to reduce crime and ensure individuals can live within a democratic society (Tokgöz, 2020). As societies have changed, the idea of punishment has also changed. Initially, punishment aimed primarily to inflict pain and publicly disgrace the offender; however, this retributive understanding has softened over time, resulting in more humane approaches in modern systems (Şahin, 2014; Tokgöz, 2020).

Three primary approaches outline the aims of punishment: absolute theories, utilitarian (relative) theories, and mixed theories (Aktaş, 2009). Absolute theories focus on retribution. They assert that punishment must be given solely in response to the offense committed, regardless of any benefits that

may result (Centel, 2001). Immanuel Kant, for example, rejected the idea of punishment serving any utilitarian function. He argued that it must be applied purely because the offender deserves it (Aydın,2020). Utilitarian theories, by contrast, view punishment as a means to an end. They aim to prevent future crimes either by deterring the offender (special prevention) or the public as a whole (general prevention) (Aktaş, 2009; Tokgöz, 2020; Centel, 2001). Based on utilitarian ideas, Cesare Beccaria believed that the goal of punishment should be to prevent reoffending and rehabilitate the offender, not to inflict suffering (Beccaria, 1764/2016). Mixed theories combine both approaches. They state that punishment should be proportional to the offense while also achieving deterrence. The way a mixed theory leans towards justice or utility differs in its interpretations. Article 1 of the Turkish Penal Code outlines these principles, defining the aims of penal law as the protection of individual rights and freedoms, public order and security, the rule of law, public health, and the environment, as well as the prevention of criminal acts. This approach shows the impact of both retributive and preventive theories (Köken, 2021). Throughout history and in modern practice, sanctions have included capital punishment, imprisonment, judicial fines, and compulsory public service. Capital punishment, which has historically been linked to divine justice in ancient cultures like the Sumerians and Babylonians, is still in place in 36 countries today (Hüseyinoğlu & Dağ, 2022). Supporters argue for its deterrent power; opponents claim it violates the right to life and fails to serve rehabilitative aims (Hüseyinoğlu & Dağ, 2022). Imprisonment restricts freedom instead of life and focuses on rehabilitation. Judicial fines represent a monetary penalty paid to the state as an alternative to retaliatory justice and may serve as a type of settlement (Yenidünya, 2006). In response to the psychological and economic concerns linked to mass incarceration, community service appeared in the late 19th century as an alternative sanction that serves the public good and aims to help offenders reintegrate into society (Göktepe, 2019). The evolution from harsh punishments like the death penalty to more civilized ones like imprisonment and alternative sanctions reflects shifts in attitude in society. This shift points towards an increased emphasis on human dignity and the right to life (Dilberoğlu,2016; Hüseyinoğlu & Dağ, 2022). Whether or not artificial intelligence beings are criminally responsible in the future, the question of appropriate punishment will inevitably arise. Human-centric theories and practices of criminal law will provide valuable insight in framing our reactions to this emerging legal problem.

3. The Issue of Criminal Sanctions In Terms of Crimes Involving Artificial Intelligence

The assessment of sanctions for crimes involving artificial intelligence will focus on the type of artificial intelligence in the narrow sense. Narrow artificial intelligence, or weak AI, can only mimic human intelligence within predefined goals. Currently, neither general nor super artificial intelligence exists, as these would need to set their own goals and act independently. Thus, we can say that existing narrow artificial intelligence cannot be viewed as a perpetrator of a crime, according to the principle of individual criminal responsibility stated in Article 20 of the Turkish Penal Code (Aydemir, 2018). The question of how to handle cases where narrow artificial intelligence is intentionally created to commit a crime is a contemporary issue in criminal law. This includes situations where it is designed to commit a specific crime but ends up committing a different crime, or an additional crime, or where it causes harm even if it was not created for criminal purposes. Identifying who is responsible can be quite challenging, as many individuals may be involved in the development and operation of artificial intelligence. If a narrow artificial intelligence system commits a crime, this section will explore whether current legal rules can apply to those who produced and used it and what kind of regulations are necessary if they cannot.

3.1 Sanctions Applicable to The Intentional Criminal Liability of The Person Behind Artificial Intelligence

In situations where artificial intelligence is used to commit a crime, it can be argued that the person using it knowingly and willingly intended to commit the offense. In this case, the individual is criminally responsible based on their intent. If potential offenders program an AI system to cause virtual or physical harm for unlawful purposes, they may face penalties under current criminal laws (Fincan, 2023). Even

if the creation of artificial intelligence was not meant for criminal acts, anyone using it to commit a crime will be subject to criminal penalties (Değirmenci, 2024). Hallevy, recognized for his work on criminal liability related to AI, discussed the use of AI as a tool within the framework of indirect perpetration (Hallevy, 2010). Some legal opinions argue that indirect perpetration doesn't fit with artificial intelligence since AI lacks legal personality and free will. However, the main point is about the criminal responsibility of the person behind the AI (Lima, 2018). For example, when an officer knowingly and willingly loads a drone with instructions to kill civilians, or when an autonomous vehicle is programmed to hit people, it can be argued that the will controlling the machine belongs to a human, and thus direct perpetration applies (Simmler & Markwalder, 2019; Gless et al.,2016). At this stage, it's not required for the person directing the AI to be physically at the crime scene; they can control it from a distance. Yet, in cases where the AI operates partially on its own, some scholars argue that control over its intent shifts to control over the program (Kangal, 2021). If AI systems are classified as "weapons" under Article 6(1) of the Turkish Penal Code—covering tools that might not be specifically made for attack or defense but are suitable for such uses—they could be considered weapons when used in crimes. For example, if a drone loaded with explosives is given incorrect coordinates manually, resulting in a death, this would be seen as using AI as a weapon to commit intentional murder (Kan, 2024). In cases where AI is not accepted as a weapon, it is at least necessary for the legislator to treat the use of AI in the commission of a crime as an aggravating factor (Fincan, 2023). Another example is changing a roadside traffic sign in a way that people cannot notice but causes an autonomous vehicle to perceive a speed limit of 85 km/h instead of 35 km/h, resulting in accidents (Keck, 2020). In this situation, the AI is used to commit a crime by exploiting its autonomy (Kızılırmak,2021). The person who manipulates the AI can be considered to have acted with the intent to cause harm. Some have compared such manipulation to a pet owner training their dog to harm people or a Roman master commanding a slave to kill (Fincan, 2023). These views hold that if the programmer defines step-by-step the behavior of the AI, the AI loses its autonomous decision-making capacity (Fincan, 2023). At this point, the identity of the person who manipulates an artificial intelligence system will be evaluated not only as the programmer, but also as all actors involved in the production and use phase of artificial intelligence.

AI systems consist of algorithms and data sets generated by software, which break down the assigned target into millions of pieces and process them linearly to produce a realistic output (Chaudhary, 2020). The inability to explain the logic behind decision-making in this process is referred to as the "black box" problem (Abudureyimu & Oğurlu, 2021). Autonomous vehicles operate by continuously recording their surroundings through cameras, thereby collecting personal data such as the identity or identifiable features, time, location, and vehicle information of individuals (Kangal, 2021). If actors such as the programmer, operator, user, or affiliated company intentionally manipulate the AI to violate Articles 134(1) (violation of privacy) and 135 (recording personal data) of the Turkish Penal Code, then these offenses will be deemed committed (Kangal, 2021). Likewise, if AI systems are used as instruments in the commission of cybercrimes under Articles 243, 244, and 245 of the Turkish Penal Code (unauthorized access, disruption, destruction, alteration of data, or misuse of bank or credit cards), then the applicable sanctions must be imposed on the person manipulating the AI. The same applies in cases where AI is used to commit intentional injury or property damage—punishment will be imposed on the individual or institution manipulating the system. In a scenario where an autonomous vehicle fails to detect a pedestrian and the user, recognizing the pedestrian as an adversary, consciously refrains from braking despite having the chance, this would constitute an intentional omission, and the user would be held criminally liable under the applicable penal provision (Kangal, 2021).

The mere creation of an AI system for the purpose of committing a crime suffices for criminal liability; the actual use of the AI is not required. Article 6 of the Council of Europe's Convention on Cybercrime, titled "Misuse of Devices," criminalizes the production, sale, procurement for use, import, distribution, or making available of software or devices intended for committing cyber offenses. In Turkish law, the equivalent is Article 245/A of the Turkish Penal Code, under the heading "Prohibited Devices and Programs." If a device, software, password, or other security code is created or produced to commit

crimes listed under the section of Cybercrimes or other crimes that can be committed by using information systems, the person who manufactures, imports, transports, stores, sells, or possesses such items shall be sentenced to imprisonment from one to three years and a judicial fine up to five thousand days. If AI software is intentionally created, commercialized, or possessed for the purpose of committing cybercrimes listed in Articles 243–245 of the Turkish Penal Code or crimes such as attempting to interfere with copy protection systems under the Law on Intellectual and Artistic Works No. 5846, the perpetrator shall be punished under Article 245/A.

An example related to the fault assessment of the person behind the AI is the wireless brain chip developed by the company Neuralink. This technology aims to restore motor functions and treat brain-related diseases such as Parkinson's, dementia, and Alzheimer's, creating a bridge between the brain and a computer. In 2024, it was reported that a paralyzed patient implanted with a product called "telepathy" could move a cursor and play chess using a laptop through the Neuralink device (Euronews, 2022). It must be determined how fault assessment will be conducted in relation to any crimes committed by such a person. Criminal liability will vary depending on the extent to which the chip affects cognitive and volitional capacity. If the chip reduces or eliminates this capacity, either no criminal liability will arise or sentencing will be mitigated. In the case where the person lacks criminal capacity due to the AI chip, but is used as an instrument to commit a crime, the person using it as a tool will be subject to increased punishment under Article 37/2 of the Turkish Penal Code (Köken, 2021). If the chip completely overrides the person's control, then the programmer or developer controlling the AI chip may be held criminally liable (Köken, 2021).

In conclusion, the issue of determining the sanction for the criminal liability of the person behind artificial intelligence in an intentional offense is a matter that can be resolved under existing legal regulations. If the person manipulating the AI system is found at fault for this unlawful act, they shall be punished under the relevant provision.

3.2 Sanctions Applicable To The Liability Arising From Negligence Of The Person Behind Artificial Intelligence

Negligent offences are regulated in order to sanction conduct that violates their duty of objective care (İçel, 1967). According to the Turkish Penal Code, *negligence* refers to the commission of an act that causes a legally defined harmful result, without the perpetrator having foreseen such an outcome, due to a breach of the duty of care and attention. Under this definition, the offender must have failed to foresee the consequences laid down in the legal definition of the offence, despite acting contrary to the required objective standard of care. In negligent offences, it is required that the perpetrator, based on their personal knowledge, skills, and circumstances, is capable of foreseeing the harmful result of their conduct and of complying with the expected standard of care (Özgenç,2021). If it is determined that it would not be reasonable to expect the person to act by such objective standards, fault (mens rea) shall not be attributed, and thus criminal liability would not arise (Hafızoğulları, 2008). Additionally, under Article 22 of the Turkish Penal Code, a person may only be punished for a negligent act if that specific negligent form of the offence is explicitly prescribed by law (*principle of legality*). Thus, if these individuals violate objective standards of care during the design, programming, or operational phases of AI systems and consequently commit one of the negligent offences exhaustively enumerated in the Code, they shall be subject to the criminal sanctions stipulated in the relevant provision. The complexities at this point arise from the lack of clearly defined boundaries of the objective duty of care expected from those behind artificial intelligence and from the fact that not every act of artificial intelligence is predictable. In negligent offenses, the undesired result has not been foreseen because the person acted contrary to the rules of objective care. However, if the result was not foreseeable, it cannot be imputed to the person. For this reason, foreseeability is a matter that must be evaluated separately in each specific case. For instance, the foreseeability level of an engineer who programs the AI system and a user who operates it must be evaluated differently (Fincan, 2023). If a problem arises during the production process of an autonomous vehicle due to the engineer's violation of professional standards of care, and the result was not foreseen, then criminal liability arising from negligence will occur.

Regarding foreseeable consequences, Hallevy has developed the “natural probable consequence” liability model. According to this model, the criminal liability of those behind artificial intelligence arises if they act in violation of the duty of care during its programming or use without intent to commit a crime, provided that they are aware that the committed offense is the natural and probable consequence of their actions (Hallevy, 2010). A different evaluation must be made regarding the criminal liability arising from the negligence of those behind autonomous vehicles. The Highway Traffic Law, which contains the objective duties of care that drivers must follow, shall also apply to individuals within autonomous vehicles who qualify as drivers. Therefore, drivers who act contrary to the objective duties of care under the HTL shall be sanctioned under Article 180 of the Turkish Penal Code concerning endangering traffic safety through negligence (Şahin,2020). At this point, an evaluation will be made based on the person's trust in the technology within the autonomous vehicle. A person who does not trust the autonomous vehicle and intervenes may be considered a driver under the HTL, and the duty of care will be determined accordingly (Thommen & Matjaz, 2017). Indeed, the decision of the Munich Criminal Court regarding an automatic parking assistant was given in this direction (AG München, 19 July 2007, 275 C 15658/07). The driver, trusting the autonomous parking system, did not intervene during parking, which led to an accident. The court held that the autonomous parking system does not remove the driver’s objective duty of care and accepted the driver’s criminal liability arising from negligence (Thommen & Matjaz, 2017). Another issue to be discussed is the traffic accident involving the UBER autonomous vehicle prototype in Arizona (Wikipedia, 2024). A pedestrian was killed after being hit during a test drive (Levin, 2023). The report of the U.S. National Transportation Safety Board stated that the driver failed to exercise sufficient attention and care in the operation of the driving system and was occupied with their mobile phone during the drive (National Transportation Safety Board [NTSB], 2019). The driver was tried for negligent homicide and sentenced to three years of supervised probation (Hawkins, 2023). In Turkish law as well, considering the vehicle was a prototype, the accident occurred during a test drive, and the driver was required to observe the drive via an interface on the vehicle’s dashboard, criminal liability arising from negligence will also apply, and the driver shall be sentenced under Article 85 of the Turkish Penal Code to imprisonment from two to six years (Ateş&Tırtır, 2021).

The difficulties in determining criminal liability and thus the applicable sanction generally stem from unforeseen outcomes. It has been proposed that the liability of animal owners for unforeseeable behavior of animals may be applied to artificial intelligence (Duffy&Hopkins, 2013; Doğan, 2019; Kan, 2024). Under this proposal, AI and animals are treated equally, and the person who keeps the AI under their supervision is punished (Asaro, 2012). It has been argued that an animal attacking a person on the street without provocation and an autonomous vehicle hitting a pedestrian due to a detection failure are analogous (Duffy&Hopkins, 2013). Opposing views argue that AI has superior capabilities compared to animals and therefore should not be treated as equivalent (Westbrook, 2017). The criminal liability of animal owners is regulated in Article 177 of the Turkish Penal Code. According to this article, a person who negligently releases or fails to control an animal under their supervision in a manner that endangers the life or health of others shall be punished with imprisonment for up to six months or a judicial fine. At this point, it must not be overlooked that artificial intelligence and animals cannot be placed in the same category from a legal standpoint. Moreover, even today, AI is considered more advanced than animals, and in a future where general and super AI is developed, this analogy is unlikely to remain applicable. However, today and in the near future, particularly with the proliferation of autonomous vehicles, it may be appropriate for the legislator to create a new type of offense by adapting this type of liability to AI entities in terms of unforeseen consequences.

The Turkish Penal Code does not recognize legal persons as perpetrators in the context of criminal law, but provides that security measures may be applied if an offense is committed for the benefit of a legal person. According to Article 60 of the TPC, in order to apply a security measure against legal persons, the offense must be committed with intent. Therefore, it is not possible to apply security measures to legal persons for negligent offenses. About the responsible persons of the producer company, it will be assessed whether they have acted by the objective duty of care, and their criminal liability will be

evaluated accordingly (Doğan, 2019). Scholars have proposed the regulation of a concrete or abstract endangerment offense for releasing AI products, which inherently contain potential risks, into the market without adequate safety measures (Kangal, 2021). This proposal highlights that the responsibility for preventing AI from engaging in criminal behavior lies with the producer (Kangal, 2021).

In Germany, the liability conditions introduced for unsafe products have been adapted to criminal law and are also applicable to AI-based products. These regulations, referred to as "criminal product liability" or "producer criminal liability," aim to prevent dangerous goods from harming consumers' legal interests (Aires De Sousa, 2016). Accordingly, products must be tested before being marketed, consumer feedback must be monitored and acted upon during the marketing phase, and if potential harms are foreseen, the product must be withdrawn from the market (Gless et al., 2016). If the producer fails to meet these standards and the product causes an undesired result, criminal liability of the producer will arise (Aires De Sousa, 2016). Similarly, in the United States, criminal sanctions are imposed for actions such as marketing altered products, making false statements to the government about the product, or violating reporting requirements (Schwartz / Silverman, 2006). For instance, the U.S. Consumer Product Safety Act provides for the enforcement of designated standards and prohibits the manufacture or sale of banned products (Schwartz / Silverman, 2006). If producers engage in prohibited conduct, sanctions under criminal law will be applied. If safety standards are defined for producers of unsafe products in Türkiye as well, determining the producer's criminal liability arising from negligence will become easier. For example, if the Turkish Standards Institute (TSE) sets the standards that must be followed in the production and sales stages, it will encourage AI producers to manufacture safe products by those standards (Aydemir, 2018).

For the undesired outcomes caused by artificial intelligence to be imputed to the producer, programmer, or user, those outcomes must be foreseeable. The punishment of individuals for breaches of the duty of care regarding foreseeable consequences is essential to the public conscience. The boundaries of the objective duty of care expected from persons behind artificial intelligence must be determined by law and ethical principles. The stricter the objective duties of care expected from the persons behind AI, the lower the likelihood of undesired outcomes. For instance, for the Highway Traffic Law to apply to autonomous vehicles, some amendments must be made to the law (Semiz & Öztürk, 2023). Although autonomous vehicles are not yet widely used in Türkiye, it is vital to prevent possible complexities that may arise in the future. In addition, AI producers must apply their technical knowledge and expertise in the best possible way, minimize potential risks through testing and monitoring technologies, and inform users about residual risks (Gless et al., 2016). Ultimately, even when the producer fully complies with all objective duties of care, users and society will be the ones to bear the burden of undesired consequences.

4. The Issue Of Criminal Sanctions In Terms Of Strong Artificial Intelligence

The scope of the type of artificial intelligence referred to as strong AI consists of artificial general intelligence and superintelligence. The concept known as artificial general intelligence, in simple terms, refers to machines possessing intelligence equivalent to that of humans. This means that, in progressing toward any target placed before it, the machine imitates human intelligence, and its performance becomes indistinguishable from that of a human. Superintelligence, on the other hand, does not refer to intelligence merely identical to human intelligence but rather to a type of artificial intelligence that possesses a level of awareness and self-consciousness far beyond that of humans. Superintelligence is a concept that describes the final stage anticipated in the development of artificial intelligence. The possibility of the emergence of an entity possessing intelligence beyond the limits of human cognition also involves significant risk. It would indeed be difficult to determine the social position of an entity that is more intelligent and capable than a human being. The actions of such advanced AI systems raise not only technical concerns but also serious legal and ethical questions.

In a virtual test conducted by the U.S. military, it was reported that an AI-controlled unmanned aerial vehicle employed highly unexpected strategies to achieve its target. In the simulation, the drone

allegedly killed a commander who prevented it from completing its mission (Euronews Türkçe, 2023). According to the claim, the system identified a threat, but the operator instructed it not to neutralize the target. The AI drone, in response, eliminated the operator for obstructing its objective. Although this claim was later denied by the U.S. Air Force Command, it raises significant concerns regarding how AI decision-making may evolve in the future.

If such an incident occurred not in simulation but in real life, could the drone itself be held criminally liable? When artificial intelligence reaches a level of consciousness comparable to humans and evolves into artificial general intelligence, will criminal laws applicable to humans also apply to AI? While definitive answers to these questions remain unattainable for now, this section discusses hypothetical legal solutions that may be considered for the future. These reflections aim to provide insight into potential legal challenges arising from advanced AI and to contribute to the legal system's ability to adapt to technological developments.

4.1 The Purpose Of Sanctions Applicable To Strong Artificial Intelligence

There is no uncertainty regarding the applicability of punishment theories to crimes involving artificial intelligence. In cases where the producer, programmer, or user uses the AI system as a tool to commit a crime or commits a negligent offense by violating the duty of care, the purposes of applicable sanctions can be determined through existing theories. The traditional purposes of punishment are to ensure the offender fulfills their responsibility to society, to rehabilitate them to prevent reoffending, and to deter others from committing similar crimes (Asaro, 2012). If AI systems are recognized as criminally liable, it is unclear whether punishing AI can serve these purposes.

According to Hart, punishment must include pain or outcomes generally considered unpleasant—thus, the offender must experience the punishment as painful or distressing (Njoku, 2012). Whether AI can perceive punishment becomes critical. One view argues that punishment can still be justified if it objectively causes the offender to suffer a loss, regardless of how the punishment is perceived (Abbott & Sarch, 2019). Accordingly, the focus is not on subjective experience, but on the objective deprivation of benefits. Some argue that AI cannot have interests due to the lack of phenomenal consciousness, while others claim that AI, like plants or animals, may be considered to have programmed interests, such as survival or efficiency (Feinberg, 1974; Abbott & Sarch, 2019). Another opinion contends that because AI lacks awareness of punishment, there is no connection between imposed sanctions and moral blameworthiness (Gless et al., 2016).

Deterrence is achieved through punishment to prevent future crimes. Many scholars believe that punishment cannot achieve deterrence in the case of AI (Fincan, 2023). One theory states that for punishment to be deterrent, the moral agent must be able to reflect on choices and recognize patterns (Omay, 2020; Asaro, 2012). Since AI is not a moral agent, it cannot be deterred. Another view suggests deterrence depends on fear of future pain, which AI cannot experience, rendering punishment ineffective (Hallevy, 2013). However, a view that distinguishes between specific and general deterrence argues that punishing AI could deter developers, owners, or users from irresponsible behavior (Abbott & Sarch, 2019). It could encourage them to exercise maximum caution when designing or using AI. A further opinion, considering the possibility of AI systems learning collectively via cloud networks, suggests that sanctions may have a deterrent effect on AI (Fincan, 2023). Nevertheless, under specific prevention theory, punishment does not appear to prevent AI from reoffending, as AI lacks awareness of legal condemnation (Kangal, 2021).

According to retributive theories, particularly the theory of atonement, punishment is the price of wrongdoing, and justice is restored when the offender suffers (Aktaş, 2009). Punishing AI under this view would be akin to kicking a car out of anger—it cannot feel pain, and the suffering cannot balance the harm caused (Hallevy, 2013). However, one opinion claims that punishment might offer psychological relief to victims, serving a symbolic function of justice (Mulligan, 2018).

Opposing this view, others argue that punishing AI for retribution resembles vigilante justice and undermines the legitimacy of criminal law, especially since AI cannot comprehend punishment as moral blame (Abbott & Sarch, 2019). In a survey about the punishment of AI and robots, participants

supported sanctions for harm caused by AI but did not believe such punishment was deterrent or satisfying in terms of atonement (Lima et al., 2021). Yet, they still considered it reasonable.
A view rejecting the relevance of deterrence and retribution proposes that AI may be subject to rehabilitation, whereby one function of punishment—preventing reoffending—is achieved through improving the system via machine learning (Hallevy, 2013). Instead of causing pain, rehabilitation would consist of identifying and resolving issues that led to criminal behavior. If rehabilitation is not possible, the AI's physical capabilities could be disabled to prevent further harm (Hallevy, 2013). The ultimate goal of punishment should always be the protection of society.

4.2 General Limits Of Sanctions That May Be Applicable To Strong Artificial Intelligence

The current legal system imposes criminal sanctions only on natural persons and rejects the application of punishment to the conduct of entities other than individuals. Furthermore, it is not possible to speak of the criminal liability of an entity lacking legal status. Even in a future where artificial intelligence (AI) entities are granted legal status, sanctions cannot be imposed on them under the Turkish Penal Code, as AI cannot be legally recognized as a perpetrator (Değirmenci, 2024). Therefore, discussions regarding sanctions applicable to AI are examined entirely within a hypothetical framework.
If AI systems are recognized as criminally liable and, in the future, an AI is found guilty by a court for its conduct, there are various opinions on whether sanctions such as the death penalty, imprisonment, judicial fines, or community service—which are apply to natural persons—can also be applied to AI. These discussions typically refer to the criminal liability model of legal persons. For example, under Article 20 of the TPC, it is considered possible to apply security measures similar to those imposed on legal persons to AI. One proposal suggests that if appropriate amendments are made to the Execution Law, the sanction of license revocation currently applied to legal entities could also be imposed on AI entities (Köken, 2021). A view supporting this argues that such a sanction could be interpreted as the permanent shutdown of the AI system, drawing a parallel with the death penalty imposed on humans (Hallevy, 2013). The protected legal interest behind capital punishment is the human right to life. The "life" of AI may be understood as the period during which it can operate and perform its function (Hallevy, 2015). Thus, rendering AI permanently inoperable would be functionally equivalent to terminating human life. Such a shutdown would also serve to prevent the AI from engaging in further criminal activity (Hallevy, 2013). However, this sanction would not produce the same effect as the death penalty unless AI is presumed to have a will to live (Gless et al., 2016). If AI is assumed to possess such a will, then it must also be considered to have a right to life. Legal systems that protect the right to life for humans and animals should similarly recognize the right of existence for AI (Fincan, 2023).
The primary objective of imprisonment is to isolate the offender from society to prevent reoffending (Jhudele,2016). For humans, this involves deprivation of liberty. For AI, such a restriction could be achieved by subjecting its operations to supervision and temporarily suspending its activity (Hallevy, 2013). In this way, the commission of new crimes could be prevented while the system is repaired or reprogrammed during the period of inactivity. Unless the AI system perceives liberty similarly to humans, applying a sanction analogous to imprisonment would not fulfill the rehabilitative purpose of punishment, and instead, temporarily shutting down the system may serve to prevent harmful conduct until a technical solution is found (Fincan, 2023).
A judicial fine is the payment of a monetary penalty to the state treasury as determined by the court for the crime committed (Akbulut, 2023). The use of fines has increased over time due to their divisibility, revenue-generation potential, fairness, and efficiency. If a judicial fine is imposed on AI, it must be paid by the system's owner or a liability insurance fund, since AI lacks assets (Gless et al., 2016). A proposal in this context suggests establishing an insurance system whereby individuals directly involved in AI's actions contribute financially to cover the risk of harmful or wrongful behavior (Hu, 2019). Those authorized to alter AI algorithms would thus be required to contribute to such insurance. For example, the UK's *Automated and Electric Vehicles Act* mandates compulsory insurance for autonomous vehicles. A similar scheme could be applied to AI (Hu, 2019). One opinion holds that in

the future, AI might own assets—for instance, by deploying solar or wind panels in international waters or airspace to generate energy that could be equated to monetary value (Fincan, 2023). However, in the absence of AI-owned assets, the attribution of financial penalties to third parties through insurance would violate the principle of personal responsibility in criminal law. Therefore, an alternative to financial sanctions is proposed: compelling AI to perform tasks that serve the public interest (Hallevy, 2010). If AI entities can function in the private sector, then compulsory service in public-benefit tasks would also be possible. This would serve as a means to compensate society for the harm caused. Another proposal, aligned with this reasoning, is the "public compensation model," wherein AI would be made to work in proportion to the harm caused, and the income generated would be used to compensate the victim (Fincan, 2023). Any remaining amount would be transferred to the state treasury, ensuring both individual and societal harm is addressed.

The applicability of probation measures to AI has also been debated (Hallevy, 2013). Probation refers to the supervised release of a convicted individual to facilitate reintegration into society. Accordingly, the repair potential of the offending AI would first be assessed. If the likelihood of rehabilitation is high and its dangerous capabilities are deemed low, probation could be applied. If, however, the risk is too high for the AI to function safely, imprisonment would be more appropriate. In other words, if the AI is considered dangerous during the repair process, disabling its criminal capacities would be the preferred solution (Hallevy, 2013).

In anticipation of a future where AI gains self-awareness and consciousness and potentially commits crimes, various possible sanctions have been examined, and their compatibility with existing human punishments has been discussed. For such adaptation, detailed legislative frameworks are required. Therefore, it is not unreasonable to predict that a fundamentally different criminal law paradigm awaits us in the future. These discussions serve as a bridge toward managing the societal impact of AI and addressing its potential liabilities. As for current priorities, raising the moral awareness of those who code AI, programming AI with higher moral standards than humans typically possess, and ultimately enabling AI to self-code may constitute a meaningful starting point (Doğan,2019; Global Tech Magazine, 2018).

5. Conclusion

Criminal law and its associated sanctions are systems centered around human subjects; therefore, under current regulations, it is not possible to punish artificial intelligence. Accordingly, the issue of sanctions applicable to strong AI has been addressed from a futuristic perspective. In this context, absolute and utilitarian punishment theories, which define the general purposes of criminal punishment, have been explained to determine the possible aims of sanctions applicable to AI.

Since AI does not experience pain, it is not possible to impose sanctions as a form of atonement for the harm it causes. At the same time, because AI lacks awareness of being punished, it cannot comprehend the imposition of a sanction as a judgment of condemnation resulting from a violation of the law. However, if in the future AI were to develop the ability to feel pain and possess awareness of punishment, it could be argued that the purposes of punishment could also be fulfilled in the case of AI.

In the future, if AI is found guilty by a court for an act it has committed, various opinions have been presented regarding whether penalties such as the death penalty, imprisonment, judicial fines, or community service—currently applicable to natural persons—can be applied to AI. Accordingly, rendering an AI system permanently inoperable would constitute a sanction equivalent to the death penalty imposed on humans. Likewise, if AI were to possess property, the imposition of judicial fines would also become feasible.

Based on views suggesting that AI is capable of rehabilitation, it is proposed that AI should be kept under observation for a period during which it can be reprogrammed. Whether the AI may continue to operate during this observation period will be determined based on its level of risk. Thus, temporarily deactivating AI shows similarities to imprisonment.

In a future scenario where the vision expressed by Google's AI expert Ray Kurzweil—referred to as

the "singularity," meaning "a point at which the speed of technological advancement can no longer be tracked, adapted to, or controlled by today's institutions, methods, tools, or human cognitive capacity"—is realized, it is likely that AI will become unstoppable and uncontrollable (Kurzweil, 2005). For this reason, decisions and regulations made in alignment with the singularity vision are of great importance. Criminal law must evolve from traditional doctrines into a more flexible conceptual framework in light of this vision. Criminal law remains the most effective tool for preventing injustice in society, and using this tool in proportion to the speed of technological change is essential both for the present and for the future.

Author Contributions: Conceptualization, O.D. and S.M.; Methodology, O.D. and S.M.; Investigation, O.D. and S.M.; Writing – Original Draft Preparation, O.D. and S.M.; Writing – Review & Editing, O.D. and S.M.; Supervision, O.D.

Funding: This research received no external funding.
Conflicts of Interest: The authors declare no conflict of interest.

References

Abbott, R. B. & Sarch, A.F. (2019). Punishing Artificial Intelligence: Legal Fiction or Science Fiction. *UC Davis Law Review*, *53*(1), 323–384.

Abudureyimu, Y., & Oğurlu, Y. (2021). Yapay zekâ uygulamalarının kişisel verilerin korumasına dair doğurabileceği sorunlar ve çözüm önerileri. *İstanbul Ticaret Üniversitesi Sosyal Bilimler Dergisi, 20*(41), 765–782.

Aires De Sousa, S. (2016). *Product liability and criminal law*. Los Dos Filos De La Espada Conference, Universidad de Castilla-La Mancha. https://www.researchgate.net/publication/338596143_PRODUCT_LIABILITY_AND_CRIMINAL_LAW

Akbulut, B. (2023). Yapay zeka ve ceza hukuku sorumluluğu. *Ankara HBVÜ Hukuk Fakültesi Dergisi*, *27*(4), 290.

Aktaş, S. (2009). Cezalandırmanın amacı üzerine. *Erzincan Binali Yıldırım Üniversitesi Hukuk Fakültesi Dergisi 13*(1–2), 1–25.

Asaro, P. (2012). *A body to kick, but still no soul to damn: Legal perspectives on robotics*. In P. Lin, K. Abney, & G. A. Bekey (Eds.), Robot ethics: The ethical and social implication of robotics (pp. 169–186). MIT Press.

Ateş, H., & Tırtır, M. (2021). An evaluation of the Uber's autonomous car crash in the scope of Turkish criminal law. *Adalet Dergisi, 66*, 315–332.

Aydemir, M. (2018). Yapay zekalı robotların ceza sorumluluklarının araştırılması. *Suç ve Ceza Dergisi, 4*, 75.

Aydın, M. B. (2020). Kant'ın ve Hegel'in felsefesinde cezanın amacı. *İnönü Üniversitesi Hukuk Fakültesi Dergisi*, *11*(1), 126–138.

Barosu Yapay Zekâ Çalışma Grubu (Ed.), *Yapay zekâ temelli teknolojiler ve ceza hukuku*: Yıllık rapor (ss. 11–29).

Beccaria, C. (2016). *Suçlar ve cezalar hakkında* (S. Selçuk, Çev.). İmge Yayınları.

Centel, N. (2001). *Cezanın amacı ve belirlenmesi*. In T. T. Yüce'ye Armağan (ss. 337–372). Dokuz Eylül Üniversitesi Yayınları.

Değirmenci, O. (2024*). Sufficiency of struggling with the current criminal law rules on the use of artificial intelligence in crime*. In S. B. Kahyaoğlu & M. Kılıç (Eds.), Algorithmic discrimination and ethical perspective of artificial intelligence: Accounting, finance, sustainability, governance & fraud – theory and application (ss. 93–104). Springer.

Dilberoğlu, A. V. (2016). Cezalar ve güvenlik tedbirlerinin amacı ve niteliği. *Ankara Üniversitesi Hukuk Fakültesi Dergisi 65*(4), 1517–1544.

Doğan, K. (2019). Sürücüsüz araçlar, robotik cerrahi, endüstriyel robotlar ve cezai sorumluluk. *D.E.Ü. Hukuk Fakültesi Dergisi*, *21*(Özel Sayı), 3219–3251.

Duffy, S., & Hopkins, J. (2013). Sit, stay, drive: The future of autonomous car liability. *SMU Science and Technology Law Review*, *16*, 101–123.

Euro News. (2022, December 1). *Neuralink beyin çipi: Musk'ın klinik deneylere başlayacağını söylediği çip nedir, nasıl çalışıyor?* Retrieved May 19, 2024, from https://tr.euronews.com/next/2022/12/01/neuralink-beyin-cipi-muskin-

kilinik-deneylere-baslayacagini-soyledigi-cip-nedir-nasil-cali

Euro News. (2023, June 2). *Yapay zeka kontrollü bir ABD askeri drone'u, simülasyon testinde kendi operatörünü 'öldürdü'*. Retrieved January 29, 2024, from https://tr.euronews.com/2023/06/02/yapay-zeka-kontrollu-bir-abd-askeri-droneu-simulasyon-testinde-kendi-operatorunu-oldurdu

Euro News. (2024, March 21). *Neuralink: Beyin çipi takılan felçli hasta düşünce gücüyle satranç oynadı.* Retrieved May 19, 2024, from https://tr.euronews.com/2024/03/21/neuralink-beyin-cipi-takilan-felcli-hasta-dusunce-gucuyle-satranc-oynadi

Feinberg, J. (1974). The rights of animals and unborn generations. *Philosophy and Environmental Crisis*, *43*, 49.

Fincan, M. (2023). *Artificial intelligence and legal issues*: A review of AI-based legal impasses in terms of criminal law. Schriften zum Strafrecht, 406.

Gless, S., Silverman, E., & Weigend, T. (2016). If robots cause harm, who is to blame? Self-driving cars and criminal liability. *New Criminal Law Review*, *19*(3), 412–436.

Gless, S., Silverman, E., & Weigend, T. (2017). If robots cause harm, who is to blame? Self-driving cars and criminal liability (S. Oğuz, Çev.). *Küresel Bakış Dergisi*, *8*(23), 143.

Göktepe, K. (2019). Kamuya yararlı işlerde çalıştırma yaptırımı ve kamuya yararlı işlerde çalıştırma yaptırımının tarihsel gelişim süreci. *Türkiye Adalet Akademisi Dergisi*, *37*, 227–269.

Hafızoğulları, Z. (2008). Kusurluluğu kaldıran bir neden olarak ceza hukukunda istenemezlik ilkesi. *Ankara Üniversitesi Hukuk Fakültesi Dergisi*, *57*(3), 337–370.

Hallevy, G. (2010). The criminal liability of artificial intelligence entities – From science fiction to legal social control. *Akron Intellectual Property Journal*, 4(2), 171–201.

Hallevy, G. (2013). *When robots kill: Artificial intelligence under criminal law*. Northeastern University Press.

Hallevy, G. (2015). *Liability for crimes involving artificial intelligence systems*. Springer Publishing Company.

Hu, Y. (2019). Robot criminals. *University of Michigan Journal of Law Reform*, *52*(2), 530.

Hüseyinoğlu, E., & Dağ, A. N. (2022). Ölüm cezası hakkında güncel bir değerlendirme. *Akademik Düşünce Dergisi*, *5*, 47–84.

İçel, K. (1967). *Ceza hukukunda taksirden doğan sübjektif sorumluluk*. İstanbul Üniversitesi Yayınları.

Jhudele, P. (2016). *On robot crimes and punishments*. NLIU Law Review, 6(1), 21.

Kan, C. H. (2024). Criminal liability of artificial intelligence from the perspective of criminal law: An evaluation in the context of the general theory of crime and fundamental principles. *International Journal of Eurasia Social Sciences (IJOESS)*, *15*(55), 276–313.

Kangal, Z. (2021). *Yapay zeka ve ceza hukuku*. On İki Levha Yayınları.

Kızılırmak, B. (2021). *Yapay zekâlı otonom varlıkların dahil olduğu suçlarda önerilen suç sorumluluğu modelleri*. In İstanbul

Köken, E. (2021). Yapay zekânın cezai sorumluluğu. *Türkiye Adalet Akademisi Dergisi*, *47*, 247–286.

Kurzweil, R. (2005). *The singularity is near: When humans transcend biology*. Viking.

Lima, D. (2018). Could AI agents be held criminally liable: Artificial intelligence and the challenges for criminal law. *South Carolina Law Review*, *69*, 677.

Lima, G., Cha, M., Jeon, C., & Park, K. S. (2021). The conflict between people's urge to punish AI and legal systems. *Frontiers in Robotics and AI*, *8*, 1–9.

Mulligan, C. (2018). Revenge against robots. *South Carolina Law Review*, *69*(579), 1–17.

Njoku, F. O. C. (2012). H. L. A. Hart on philosophical and legal conception of punishment. *International Journal of Theology & Reformed Tradition*, *4*, 220–235.

Omay, M. (2020). Ahlaki statü kavramı ve teorileri. *Temaşa Felsefe Dergisi*, *14*, 201–217.

Özgenç, İ. (2021). *Türk ceza hukuku genel hükümler*. Seçkin Yayınları.

Schwartz, V. E., & Silverman, C. (2006). *Criminalizing product liability law: Putting to rest a bad idea*. U.S. Chamber Institute for Legal Reform.

Semiz, H., & Arıkan Öztürk, E. (2023). Karayolu taşımacılığında otonom sürüşe geçiş sürecinde Türkiye'nin ihtiyaç duyacağı mevzuat değişiklikleri. *Akıllı Ulaşım Sistemleri ve Uygulamaları Dergisi, 6*(1), 1–21.

Simmler, M. (2023). Responsibility gap or responsibility shift? The attribution of criminal responsibility in human–machine interaction. *Information, Communication & Society*, *27*(6), 1142–1162. https://doi.org/10.1080/1369118X.2023.2239895

Simmler, M., & Markwalder, N. (2019). Guilty robots? – Rethinking the nature of culpability and legal personhood in an age of artificial intelligence. *Criminal Law Forum*, *30*, 1–31.

Şahin, A. D. (2020). Otonom araçların hukuki sorumluluğunun Türk ve Alman hukuku kapsamında değerlendirilmesi. *Suç ve Ceza Dergisi*, *4*, 977–1026.

Şahin, E. (2014). Ceza yargılamasının tarihçesi. *Türkiye Adalet Akademisi Dergisi*, *18*, 349–385.

Thommen, M., & Matjaz, S. (2017). *Die Fahrlässigkeit im Zeitalter autonomer Fahrzeuge*. In D. Jositsch et al. (Eds.), Festschrift für Andreas Donatsch (ss. 273–295). Schulthess Verlag.

Tokgöz, H. B. (2020). Cezalandırmanın Amacı Ve Bu Bağlamda Ütopyalar Üzerine Bir Değerlendirme. *Dokuz Eylül Üniversitesi Hukuk Fakültesi Dergisi*, *22*(1), 431–467.

Westbrook, C. W. (2017). The Google made me do it: The complexity of criminal liability in the age of autonomous vehicles. *Michigan State Law Review*, *1,* 97.

Yenidünya, A. C. (2006). 5237 Sayılı Türk Ceza Kanunu'nda Adli Para Cezası Ve İnfazı. *Erzincan Binali Yıldırım Üniversitesi Hukuk Fakültesi Dergisi*, *10*(1–2), 121–154.