

FREEDOM OF WILL AS A PREREQUISITE FOR CRIMINAL LIABILITY: HUMANS AND ARTIFICIAL INTELLIGENCE IN A CZECH CRIMINAL LAW PERSPECTIVE*

JAROSLAV FENYK**

DOI 10.2478/in-2026-0001

ABSTRACT

The paper analyses the form and place of human will in the foundations of criminal liability. It explains the necessity of will for the traditional criminal liability of natural persons and presents a comparison with the possible criminal liability of autonomous and non-autonomous systems equipped with artificial intelligence. The conclusions of the paper are negative with regard to the possible direct liability of such systems. A solution is proposed in the form of criminal liability for persons who create, control and monitor autonomous and non-autonomous systems.

Keywords: consciousness, will, emotions, intuition, intelligence, artificial intelligence, algorithm, autonomous and non-autonomous systems, determinism, indeterminism, compatibilism, criminal liability, moral liability, motivation, natural person, legal person, electronic person, imputability

* This article utilised the artificial intelligence of a non-autonomous entity.

** Prof. JUDr. Ph.D., DSc., Emeritus Vice-President of the Constitutional Court of the Czech Republic, Faculty of Law, Masaryk University, Department of Criminal Law, Brno (Czech Republic), e-mail: Jaroslav.Fenyk@law.muni.cz, ORCID: 0009-0008-0109-3874



INTRODUCTION

The development of artificial intelligence (commonly abbreviated as 'AI') represents one of the most fundamental technological changes in the modern world. Its influence extends not only into the economic and social spheres but also increasingly into the field of law, including criminal law.

Criminal law, as the law of last resort, is traditionally based on the principles of individual responsibility for culpability, the primacy of free will, and the ability to distinguish between permissible and prohibited conduct. In the past, criminal liability was extended to legal entities only through the attribution of the elements of the offence, i.e. the element of fault and its basis, which is the will of the natural persons acting on their behalf. Research is beginning to examine not only liability for damage caused by autonomous systems. A significant part of the discussion concerns criminal liability and the possible attribution of this liability to new systems.

In this context, the paper will focus on artificial intelligence in both autonomous and non-autonomous systems. There is an important difference between the terms 'autonomous and non-autonomous systems' and 'artificial intelligence', although they often overlap and are confused in both everyday and professional discourse. The essence of both systems is an algorithm or, at an advanced stage of development, a computer model (a neural network), resulting primarily from mathematical rules.

Autonomous systems, in the context of artificial intelligence, are systems capable of performing tasks with minimal human intervention (this is known as deep learning).¹ These systems, ranging from self-driving cars to intelligent chatbots, are designed to function as independently as possible and make 'decisions' based on their programming and collected data.² From the point of view of criminal liability, this model will undoubtedly require a more comprehensive and thorough assessment.

Non-autonomous systems are primarily non-linear dynamic systems whose behaviour depends on time or external variables. These systems require external inputs or human control to determine their behaviour (this is known as machine learning). Examples of non-autonomous systems include Google Translate, ChatGPT, and e-shop delivery systems. A non-autonomous system does not act independently in a physical environment because it requires active input from a human being, does not make autonomous decisions with legal consequences, and is therefore not autonomous.³

Although both autonomous and non-autonomous systems exhibit a high degree of adaptability, unlike humans they completely lack consciousness and free will, and do not have legal personality as legal entities. In this context, several unresolved issues arise, the solutions to which remain difficult to determine.

¹ H. Lamb, J. Levy, C. Quigley, *Simply Artificial Intelligence*, London, 2023, pp. 58–59.

² Vation, 'Autonomous Systems: Definition, Explanation, and Use Cases', *Vation Ventures Glossary*; <https://www.vationventures.com/glossary/autonomous-systems-definition-explanation-and-use-cases> (accessed: 16 May 2025).

³ IGI Global, *What is Non-Autonomous Systems*. Cf.: <https://www.igi-global.com/dictionary/random-bit-generator-based-on-non-autonomous-chaotic-systems/45926> (accessed: 16 May 2025).

Is it even possible and legitimate to attribute criminal responsibility to artificial intelligence, or is it necessary to maintain responsibility exclusively on the part of the human subjects involved in its development, programming and operation?

Can machines be perceived in a similar way to legal entities, for example, and can a modified form of attribution be used for their liability?

The aim of this paper is to analyse the main approaches to the issue of human free will and criminal responsibility, to compare the views of selected foreign authors on the possibility of their existence in systems using so-called artificial intelligence, and subsequently to formulate considerations *de lege ferenda* appropriate for the Czech legal environment, with an emphasis on preserving the fundamental principles of criminal law, in particular individual responsibility for culpability, legality and the subsidiarity of criminal repression.

CONSCIOUSNESS, WILL, FREE WILL (*LIBERUM ARBITRIUM*) OF HUMAN AND CRIMINAL RESPONSIBILITY: THEORETICAL FOUNDATIONS

THE RELATIONSHIP BETWEEN HUMAN CONSCIOUSNESS AND WILL

Human consciousness is a reflective, intentional and morally significant state of mind that is unique in its ability to be self-aware, consider the consequences of decisions and take responsibility for actions.⁴

Consciousness therefore includes the ability to perceive, to be aware of oneself and the world, to think, to experience emotions and to reflect.

Human will is generally considered to be part of human consciousness, but with an important clarification. Will is a *specific function of consciousness*, not its synonym; it is the specific ability of an individual to *deliberately direct their actions*, often against momentary impulses or automatic behaviour. Will is therefore an expression of *the spiritual dimension*. Free will, the possibility of choice and awareness of responsibility are fundamental human abilities and manifestations of conscious being.

The characteristics of human will always depend on the context – philosophical, legal, psychological or theological, or a combination thereof. In the generally accepted philosophical and legal framework, human will is considered to be the ability to make free decisions based on conscious consideration; to choose between alternatives in accordance with an individual's inner values, beliefs and goals; and to direct one's actions towards a specific purpose, i.e. to engage in deliberate and goal-oriented activity.⁵

⁴ J.R. Searle, *The Rediscovery of the Mind*, Cambridge, MA, 1992, p. 88; or T. Nagel, 'What Is It Like to Be a Bat?', *The Philosophical Review*, 1974, Vol. 83, No. 4, pp. 435–450; or A.R. Damasio, *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*, New York, 1999, pp. 235–239.

⁵ V.E. Frankl, *Vůle ke smyslu*, Brno, 1994, p. 19.

Will is not random – individuals make decisions towards a specific goal (e.g. ‘I want to do good’, ‘I want to do evil’, ‘I want to achieve a certain result’). Will is usually guided by reason – a person considers reasons, consequences and values.

Human will therefore presupposes *the ability to self-regulate, self-control and self-restraint*. In a criminal law and ethical context, will is linked to responsibility – if someone acts ‘of their own free will’, they may bear the consequences of their decision.

Many well-known figures in history have assumed that humans are *autonomous beings*, capable of recognising moral and, subsequently, written law and acting in accordance with it. For example, Edmund Husserl, in his treatises on ethics, analyses the will as a fundamental element of human behaviour that is inseparably linked to values. The will is not merely a psychological state, but is anchored in intentions directed towards values that lead to their realisation.⁶

In existential philosophy, free will is associated with freedom of choice, but also with the anxiety associated with responsibility. For example, Jean-Paul Sartre holds that every individual is condemned to choose and thus to take responsibility for their actions. This freedom is a source of existential anxiety because the individual realises that they are fully responsible for shaping their own life without any predetermined meaning.⁷

In the field of criminal law and responsibility in general, human will is of fundamental importance – only an individual who acts of their own free will, while sane and aware of the consequences, can be fully criminally responsible. This is why, for example, *sanity, intent and negligence are so closely linked to the expression of will*.⁸ Actions without will (e.g. in unconsciousness, hypnosis or mental disorder) are not legally attributable – will is therefore a functional component of consciousness necessary for criminal liability.

Will represents a key psychological element underlying the criminal liability of the perpetrator, because without the free and conscious expression of will there can be no question of *culpability* and thus of the fulfilment of the subjective aspect of a criminal offence (Section 13(2), Sections 15–17 of the Czech Criminal Code). Will is reflected in the distinction between intentional and negligent culpability, where, in the case of intent (Section 15 of the Czech Criminal Code), the perpetrator not only knows but also wishes to cause the unlawful consequence, or is aware of its occurrence. In the case of negligence (Section 16 of the Czech Criminal Code), there is an absence of the corresponding will to prevent the consequence, with will here representing an insufficient degree of care or caution. Culpability is only one side of the coin of the subjective part of a criminal offence; the other side consists of errors in criminal law (Sections 18–19 of the Czech Criminal Code), and the requirement of will therefore also applies to them.

⁶ E. Husserl, *Vorlesungen über Ethik und Wertlehre 1909–1914*, Melle U. (ed.), The Hague, 1988, pp. 150–155.

⁷ J.-P. Sartre, *L'Être et le néant*, Paris, 1943, p. 555.

⁸ J. Fenyk, '(Ne)přičetnost fyzické osoby a (ne)přičitatelnost jejího jednání právnické osobě', in: *Tradičné a netradičné prístupy v trestnom práve: Pocta prof. Šimovčekomu*, Trnava, 2024, pp. 74–87.

Without manifestations of will, it is not possible to fulfil the objective element of a criminal offence, which is conduct, as conduct is understood as an outward manifestation of will (cf. Section 13 of the Czech Criminal Code). Indirectly, within the context of the facts of the case, intent also influences causality, because only intentional or negligent conduct (i.e. conduct with a certain form of intent) can lead to criminal liability.

Similarly, sanity (Section 26 of the Czech Criminal Code) requires that the perpetrator be able to control their will and direct their actions; the absence of this ability excludes criminal liability.

Will therefore appears to be an inseparable prerequisite not only for culpability but also for the perpetrator's actions and sanity, thereby forming the essential basis of individual criminal liability.

FREEDOM AND LACK OF FREEDOM OF HUMAN WILL

The freedom of human will can be characterised as a person's ability to act on the basis of their own mental states without external coercion. Freedom of will therefore concerns the causal relationship between mental states and actions.⁹

Is human will, then, completely free? The debate over the meaning of freedom and lack of freedom of will and its impact on criminal law has existed in criminal law scholarship since time immemorial. The debate has gradually developed into the current dispute between proponents of the compatibility or incompatibility of freedom and the lack of freedom of will in criminal law.

Immanuel Kant and Cesare Beccaria, both contemporaries of the Enlightenment, were important representatives of the theory of free will in criminal law, representing parallel but ideologically opposed directions in the development of criminal law.

In Kant's view, free will is a necessary prerequisite for moral responsibility – without freedom of choice there can be no guilt or merit. Punishment is therefore a moral necessity – retribution for evil deeds, while reward is recognition of the right choice. Kant was an advocate of free will as the basis of criminal responsibility. His concept of punishment as retribution for moral transgressions, together with his principle of the dignity of the human being, form the core of retributive theory, and Kant himself rejected any utilitarian or social functions of punishment.¹⁰

Beccaria, as a representative of the classical school of criminal law, also emphasised free will and human rationality. He assumed that humans are rational beings endowed with free will who act by weighing the advantages and disadvantages of their actions. This approach is closely linked to the theory of the social contract, according to which

⁹ This distinguishes the question of free will from the question of autonomy of the will as the ability of an individual to follow their own moral law or values – see I. Kant, *Metafyzika mravů*, transl. Svoboda V., Praha, 2004, p. 331.

¹⁰ D. Klos, 'Teorie trestu u Kanta: Právně-filozofická analýza Kantova pojetí odplaty', *Právník*, 2008, Vol. 147, No. 6, pp. 593–605; similarly J.G. Murphy, 'Kant's Theory of Criminal Punishment', in: *Retribution, Justice, and Therapy: Essays in the Philosophy of Law*, Vol. 16, Dordrecht, 1979, p. 82.

individuals transfer part of their freedom to the state in exchange for the protection of their rights and security. Breaking the law is therefore a conscious decision by an individual who has decided to violate the social contract.¹¹

Unlike the old school of criminal law, Beccaria did not view punishment as retribution, but as a means of preventing crime. His demands for certainty, speed, and proportionality of punishment, as well as his emphasis on legality and the prohibition of judicial arbitrariness, became the fundamental principles of modern criminal law.

Enrico Ferri, a prominent representative of the Italian positivist school of criminal law, fundamentally rejected the concept of free will as the basis of criminal responsibility. His approach, based on scientific research into the causes of crime, represented a departure from the classical school, which assumed that individuals act on the basis of rational free choice and are therefore fully responsible for their actions.

In his studies, Ferri often argued that crime is the result of a combination of factors beyond the control of the individual. He divided these factors into anthropological (biological and psychological characteristics of the individual), physical (climatic and geographical conditions), and social (environmental influences such as poverty, education and family background).

According to Ferri, these determinants¹² shape an individual's behaviour, thereby challenging the traditional concept of free will as the basis of criminal responsibility. Ferri introduced the concept of 'dangerousness of the offender', which focusses on assessing the risk that the offender poses to society rather than on the offence itself.

His work laid the foundations for modern approaches to criminal policy that take into account the complex factors influencing criminal behaviour, but it also marked a shift in the understanding of unlawful – previously criminal – conduct and sought to replace the repressive model with a socially preventive one in which free will does not play a fundamental role.¹³

Vladimír Solnař, the doyen of Czech criminal law in the second half of the twentieth century, drew attention to what Ferri had aimed at in the past, namely that if complete free will were required, responsibility would apply only to persons acting completely freely, and punishment as a response to a criminal offence would thus become mere retribution for unlawful conduct, while its educational elements would lose their meaning.¹⁴

In contemporary criminal law scholarship, a continuing division of opinion can be observed between advocates of free will and non-free will (determinism and indeterminism), as well as an ongoing dialogue on compromise solutions concerning the compatibility, partial compatibility or incompatibility of free will with the theory of its non-freedom (compatibilism, semi-compatibilism or incompatibilism).

¹¹ C. Beccaria, *O zločinoch a trestoch*, Bratislava, 2009, pp. 35–36.

¹² Determinism, which includes these determinants, is the philosophical belief that every event or state of affairs is the result of previous events based on the principle of causality and fixed laws. This means that the development of the world is determined by a sequence of events governed by absolutely valid natural laws.

¹³ E. Ferri, *La teorica dell'imputabilità e la negazione del libero arbitrio*, Florence, 1878, p. 408.

¹⁴ V. Solnař, J. Fenyk, D. Čísařová, M. Vanduchová, *Systém českého trestního práva, díl II. Základy trestní odpovědnosti*, Praha, 2009, p. 224.

In the 1980s, neurobiologist Benjamin Libet conducted experiments showing that brain activity precedes conscious decisions about movement. His findings led to debates about whether conscious decision-making is truly the initiator of our actions or whether it is merely the result of unconscious brain processes.¹⁵

Another neurobiologist, Robert Sapolsky, argues that human behaviour is fully determined by biological and environmental factors, which, in his view, rules out the existence of free will. Sapolsky claims that human decisions are the result of genetic predispositions and previous experiences, not free will. This deterministic perspective therefore has fundamental implications for moral responsibility and the legal system.¹⁶

One of the prominent contemporary proponents of the view that free will and determinism are compatible in criminal law is psychiatrist Michael S. Moore, who argues that determinism does not imply the absence of free will. In his view, what matters is whether an individual acts on the basis of their own reasons and motives, not whether they have the opportunity to act differently. In his opinion, this approach makes it possible to preserve the concept of moral and legal responsibility even within a deterministic framework.

Moore opposes contemporary views according to which advances in neuroscience call into question the existence of free will. He argues that even if our actions are the result of brain processes, this does not mean that they are not the result of our own decisions.¹⁷

Philosopher John H. Fischer argues that even if the world is deterministic and our actions are predetermined by previous events, we can still be morally responsible for our actions. This position, called semi-compatibilism, differs from traditional compatibilism in that it does not address the question of free will, but focuses on moral responsibility. According to Fischer, the key issue is whether the individual acted on the basis of their internal states (e.g. beliefs, desires) and had the capacity to respond to reasons. Together with philosopher and Jesuit Mark Ravizza, Fischer developed the concept of 'guidance control', which refers to an individual's ability to act in accordance with their own reasons and motivations, even if those reasons are determined. This form of control is sufficient for attributing moral responsibility, even if the individual does not have ultimate control over their actions.¹⁸

One of the leading contemporary advocates of free will in criminal law is the philosopher Carolina Sartori. She formulated a theory according to which free will is closely linked to causal relationships between human mental states (e.g. intentions, beliefs) and human actions. According to Sartori, an individual is free when their actions are the result of their own mental states rather than external coercion or

¹⁵ B. Libet, C.A. Gleason, E.W. Wright, D.K. Pearl, 'Time of Conscious Intention to Act in Relation to Onset of Cerebral Activity (Readiness-Potential): The Unconscious Initiation of a Freely Voluntary Act', *Brain*, 1983, Vol. 106, No. 3, pp. 627–631.

¹⁶ R.M. Sapolsky, *Determined: A Science of Life Without Free Will*, New York, 2023, pp. 10–14, 376–381.

¹⁷ M.S. Moore, *Mechanical Choices: The Responsibility of the Human Machine*, Oxford, 2020, pp. 28–32, 76–80, 170–175; or M.S. Moore, 'Stephen Morse on the Fundamental Psycho-Legal Error', *Criminal Law and Philosophy*, 2016, Vol. 10, pp. 45–89.

¹⁸ J.M. Fischer, *The Metaphysics of Free Will: An Essay on Control*, Oxford, 1994, pp. 131–136; J.M. Fischer, *My Way: Essays on Moral Responsibility*, Oxford, 2006, pp. 17–19.

random events. The author argues that determinism and free will are compatible. She argues that it is important to examine the specific causal chains leading to action rather than abstract metaphysical questions about determinism.¹⁹

The author's theory provides a basis for attributing moral and criminal responsibility to individuals. If their actions are the result of their own mental states, they can be held responsible for them, which supports traditional approaches in criminal law.

PARTIAL CONCLUSION ON INDIVIDUAL FREE WILL IN CRIMINAL LAW

The current state of debate on the existence or degree of free will in criminal law shows that the relationship between free will and criminal responsibility remains a philosophically and legally complex but highly significant issue.

The fundamental ideological conflict between the classical school of criminal law and the positivist approach has now been bridged to a certain extent by compromise views. These are based on the recognition that human behaviour can be determined by external or internal factors without necessarily precluding the possibility of individual responsibility. Modern approaches make it possible to combine traditional concepts of free will with scientific knowledge about the determination of human behaviour, thus providing theoretical support for modern criminal law based on moral responsibility. Free will, understood as the ability of an individual to make decisions based on their own judgment, with awareness of the consequences of their actions and the possibility of choosing between alternative options, remains a key prerequisite for attributing criminal responsibility. Although modified and limited to varying degrees, the concept of free will continues to form a fundamental component of the system of criminal responsibility of natural persons. Free will therefore remains not only a philosophical prerequisite but also a functional element of positive criminal law and is, to the necessary extent, reflected in the Czech foundations of criminal liability, primarily in the provisions on culpability, error and sanity (Sections 15–19 and Section 26 of the Czech Criminal Code).

HUMAN INTELLIGENCE AND ARTIFICIAL INTELLIGENCE OF AUTONOMOUS AND NON-AUTONOMOUS SYSTEMS

HUMAN INTELLIGENCE

The connection between human will and human intelligence is indisputable. Human will is the ability to make free decisions based on goals and values, while intelligence is the ability to understand, reason and solve problems. Intelligence provides the tools; will determines the direction.²⁰

¹⁹ C. Sartorio, *Causation and Free Will*, Oxford, 2016, pp. 3–5. C. Sartorio, R. Kane, *Do We Have Free Will? A Debate*, New York, 2021, pp. 1–2.

²⁰ For example, I. Kant, *Kritika praktického rozumu*, ed. by Kuneš J., Sobotka M., transl. Loužil J., Praha, 2023, p. 33.

One of the most prominent contemporary experts on human intelligence is the psychologist Robert J. Sternberg. His approach has fundamentally influenced the modern understanding of intelligence, particularly through his triarchic theory of intelligence and the later expanded concept of successful and adaptive intelligence.

Sternberg characterises intelligence as '*mental activity aimed at effective adaptation, selection and shaping of the environment relevant to an individual's life*'. This definition emphasises an individual's ability to adapt to changing environmental conditions, select a suitable environment and actively shape it to meet their needs and goals.

In his original works, Sternberg divides intelligence into three components:

- (a) *analytical intelligence* as the ability to analyse, evaluate, assess and compare information
- (b) *creative intelligence* as the ability to solve new and unusual problems, adapt to new situations and create original solutions
- (c) *practical intelligence* as the ability to adapt to everyday life, using experience and knowledge to solve real-world problems

Later, Sternberg developed the concept of *successful intelligence*, which integrates all three types into a single whole. Successful intelligence is defined as the ability to achieve personally meaningful goals within an individual's cultural context. This includes the ability to identify one's strengths and weaknesses and to use or compensate for them effectively in achieving goals.

Sternberg's approach is widely applied in education, psychology and human resources. It shows that intelligence is not merely an academic ability but also includes the practical and creative skills necessary for a successful life. He offered a comprehensive and dynamic view of intelligence that goes beyond traditional IQ measurements.

In recent years, the author has developed another concept, namely *adaptive intelligence*, which he defines as an individual's ability to adapt, shape and select an environment that supports the survival and prosperity not only of the individual but also of society as a whole. This definition emphasises the importance of cultural and biological context in assessing intelligent behaviour.²¹

Sternberg's conclusions are also applicable in the field of criminal law, as they reveal the inner workings of the perpetrator's mind, which manifest themselves in different forms, and help to explain certain complex aspects of the subjective side (understanding, reconciliation, indifference, reliance, etc.).

²¹ R.J. Sternberg, *Beyond IQ: A Triarchic Theory of Human Intelligence*, New York, 1985, p. 45; or R.J. Sternberg, *Successful Intelligence*, New York, 1997, p. 20; R.J. Sternberg, 'A Theory of Adaptive Intelligence and Its Relation to General Intelligence', *Journal of Intelligence*, 2019, Vol. 7, No. 4, p. 23; or R.J. Sternberg, 'COVID-19 Has Taught Us What Intelligence Really Is', *Inside Higher Ed*, 31 August 2020; <https://www.insidehighered.com/views/2020/08/31/pandemic-has-proven-standardized-tests-dont-measure-whats-important-opinion> (accessed: 5 March 2026).

ARTIFICIAL INTELLIGENCE OF AUTONOMOUS AND NON-AUTONOMOUS SYSTEMS

In addition to different ways of understanding human intelligence, scientists and the public are increasingly attempting to compare certain functions of autonomous and non-autonomous systems equipped with so-called artificial intelligence with human intelligence, or vice versa. The result is either irreconcilable differences between them or parallel forms of behaviour in living and non-living entities for which the term 'intelligence' does not have the same meaning.

One of the most respected contemporary experts in artificial intelligence is the computer scientist Stuart Russell, who describes artificial intelligence as the ability of systems *to perceive their environment, draw conclusions and make decisions leading to the achievement of set goals*. His approach emphasises *purposeful behaviour and adaptive abilities* of systems rather than merely simulating human thinking as its inanimate reflection.

Russell's definition of 'intelligent agents' focuses on *the practical ability of a system to act rationally* in different situations. According to the scientist, an intelligent agent perceives its environment through sensors and influences it through effectors, with the goal of maximising a certain measure of utility. This approach allows the intelligence of systems to be evaluated based on their ability to achieve goals in different environments, regardless of whether their behaviour resembles human thinking.

The author's work has had a significant impact on the development and evaluation of artificial intelligence because it provides a formal framework for designing and analysing intelligent systems. His definition of system intelligence is widely accepted in both academia and industry, and serves as the basis for many current applications of artificial intelligence, from autonomous vehicles to intelligent assistants.²²

In the Czech context, the relationship between human and artificial intelligence is addressed, for example, by the educator Karel Kostka, who emphasises that artificial intelligence lacks intuition and emotions, which are key aspects of human intelligence. In his view, artificial intelligence is essentially advanced statistics and does not resemble human thinking. The author adopts a relatively traditional anthropocentric, or conservatively humanistic, position in which human intelligence is regarded as unique and unattainable in its entirety by artificial intelligence.

Kostka rejects the possibility that artificial intelligence could ever reach the level of human intelligence, as artificial intelligence is merely a tool without the ability to truly think or understand.

According to Kostka, *intuition* in psychology is usually understood as the ability to gain knowledge without conscious thought, often based on previous experiences that are not consciously evaluated. Advanced machine learning models (e.g. deep learning neural networks) sometimes appear to behave intuitively because

²² S. Russell, P. Norvig, *Artificial Intelligence: A Modern Approach*, 4th ed., Harlow, 2020, pp. 1–2.

they make decisions based on patterns that are not clearly explainable even to their creators (the 'black box' of artificial intelligence). However, these 'intuitions' are not conscious, they are not associated with self-awareness, intentionality or subjective experience, and are only similar to human intuition – they are a simulation of intuition, not intuition in the human sense. Only humans are capable of intuitive insight, which allows them to be inspired by creative intelligence and whose source lies in *altered states of human consciousness*.

The author maintains that, *emotions* are complex biological processes associated with neurochemical reactions, subjective experience, motivation and adaptation. A machine can only simulate emotional reactions (for example by recognising mood from voice or facial expressions and adjusting its output accordingly), but *these are not real feelings or experiences*. A machine does not experience suffering, joy, shame or love. Kostka considers the peak of human emotional stimulation to occur in *moments of inspirational insight* at the boundary between the physical world and individual human consciousness.²³

Kostka's position has both scientific and philosophical justification and is consistent with the prevailing conclusions of contemporary cognitive science, phenomenology, and artificial intelligence ethics. Despite technological advances, his view remains valid and defensible because it aligns with prevailing philosophical and ethical approaches that emphasise the uniqueness of human consciousness and moral responsibility, and represents an important voice in discussions about the true role and limits of artificial intelligence in society. His emphasis on human uniqueness and caution towards technological progress can serve as a counterbalance to more optimistic or progressive views on artificial intelligence, thereby contributing to a balanced and responsible approach to its development and use, for example in the field of education.²⁴

One of the prominent foreign authors whose views on artificial intelligence are close to those of Karel Kostka is the philosopher John Searle, who is also known for his criticism of so-called 'strong artificial intelligence', which holds that machines can possess consciousness and understanding comparable to those of human beings. In his famous thought experiment 'The Chinese Room',²⁵ Searle argues that merely processing symbols according to rules (syntax) is not sufficient to achieve true understanding (semantics). He thus demonstrates that even though a computer may simulate understanding of language, it does not actually understand it. This argument supports the view that *artificial intelligence lacks consciousness*

²³ K. Kostka, *Kdo jsme. Obecná teorie vědomí, času, prostoru a bytí*, Frýdek-Místek, 2015.

²⁴ K. Kostka, *Umělá inteligence a změna tradičních paradigmat v psychologii vzdělávání*, manuscript dedicated to the author of the article, April 2025.

²⁵ The 'Chinese Room' experiment shows that a machine can correctly manipulate symbols (e.g. Chinese characters) without actually understanding them. Searle thus argues that formal information processing is not the same as conscious understanding. In the experiment, the author sits locked in a room where Chinese characters are passed under the door. He does not know Chinese himself, but he has rules, or a manual, that tell him how to select and arrange other characters as a response based on certain characters. From the outside, it appears that Searle understands Chinese because his answers make sense to native speakers). See J.R. Searle, 'Minds, Brains, and Programs', *Behavioral and Brain Sciences*, 1980, Vol. 3, No. 3, pp. 417–457.

and subjective experience, which are essential elements of human intelligence. *A system may exhibit behaviour resembling understanding, but it does not truly understand meaning* – it merely manipulates symbols according to rules.

Searle further emphasises that consciousness and understanding arise from biological processes in the human brain that cannot easily be replicated using computer programmes. This challenges the idea that it would be possible to create a machine with human consciousness simply by using sufficiently complex software.²⁶

American cognitive scientist Marvin Minsky, considered one of the founders of artificial intelligence research, takes a different view of machine intelligence. He described artificial intelligence as ‘the science of making machines do things that would require intelligence if humans did them.’²⁷

Minsky’s approach to artificial intelligence is based on the assumption that the human mind is a complex system composed of many simple processes that function together. He developed this idea in his theory of the ‘society of mind’, where the mind is described as a collection of agents, each performing a specific function, whose interactions give rise to intelligent behaviour.

Minsky argued that *emotions are not separate from rational thinking but represent different ways of processing information* that can potentially be modelled and implemented in artificial systems.²⁸

He was convinced that with sufficiently advanced algorithms and computing resources it would be possible to create machines capable of intelligent autonomous *behaviour comparable to that of humans*.

*A recent definition of artificial intelligence is provided by the Council of Europe Framework Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law (2024).*²⁹ Under Article 2 of the Convention, an ‘artificial intelligence system’ means a machine-based system that, for explicit or implicit purposes, infers from incoming data how to generate outputs such as predictions, content, recommendations or decisions that may influence physical or virtual environments. Different artificial intelligence systems vary in their levels of autonomy and adaptiveness after deployment.

²⁶ J.R. Searle, ‘I Married a Computer’, *The New York Review of Books*, 8 April 1999; <https://www.nybooks.com/articles/1999/04/08/i-married-a-computer/> (accessed: 5 March 2026). A review of R. Kurzweil’s book *The Age of Spiritual Machines*, in which Searle criticises the idea that computers can have consciousness and understanding.

²⁷ M. Minsky, *The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind*, New York, 2006. He argues that emotions are processing modes that can be replicated in artificial systems, p. 11.

²⁸ M. Minsky, *The Society of Mind*, New York, 1986, p. 27, in which the author presents a theory according to which the human mind is made up of a number of simple procedural ‘agents’ that together form complex intelligent behaviour.

²⁹ CETS 225 – Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law; <https://rm.coe.int/1680afae3c> (accessed: 5 March 2026).

CAN AUTONOMOUS AND NON-AUTONOMOUS ARTIFICIAL INTELLIGENCE SYSTEMS EXHIBIT (FREE) WILL?

Established legal and philosophical assumptions about free will as the basis of criminal responsibility, however theoretically diverse, have their core in the anthropological concept of humans as morally and rationally responsible beings. This concept, however, is fundamentally questioned in the context of the possible attribution of criminal responsibility to non-human entities – in particular artificial intelligence, regardless of whether or not they are autonomous systems. Nevertheless, it is necessary to distinguish between the two.

Non-autonomous systems operate on the basis of predefined instructions or commands; they never activate themselves. *They always require human supervision or intervention to function.* They do not operate in a real environment, are deterministic, and their behaviour is fully predictable on the basis of input data and programmed rules. Non-autonomous systems are therefore limited, as their behaviour is entirely determined by external commands and they have no ability to make independent decisions, only to generate suggestions. They do not perceive the world or their environment (they have no sensors, cameras, or spatial orientation).

Autonomous systems operate on the basis of predefined instructions and are a type of extension of human decision-making. They are capable of performing *tasks without direct human intervention*, or with minimal intervention. They often use artificial intelligence to analyse their environment and make decisions based on predefined goals. For example, self-driving vehicles or autonomous drones can operate in dynamic environments and adapt their behaviour to changes in real time. Even autonomous systems, however, though they may exhibit complex and adaptive behaviour, act on the basis of algorithms and predefined rules. They are not subjects of human reason in the legal sense, as they have no consciousness, will, or capacity for moral judgment. Their decisions are the result of programming and data-based learning, not an expression of free will; their goals are set by humans.³⁰

Non-autonomous systems (e.g. GPT) can be part of autonomous systems (e.g. voice assistants in cars). Here, the non-autonomous system is similar to a cognitive brain, but autonomy is determined by the whole, not by its part.

The relationship between free will and system responsibility reaches its limit here, as it is precisely human freedom of choice that legitimises the attribution of guilt in the current legal system. If autonomous or non-autonomous systems are incapable of true internal motivation in the sense of will, a fundamental issue arises

³⁰ H. Lamb, J. Levy, C. Quigley, *Simply Artificial...*, op. cit., pp. 116, 122–123. A separate, yet similar, definition of autonomous and non-autonomous systems is provided by Regulation (EU) 2024/1689 of the European Parliament and of the Council (the AI Act) – see further sections of this paper. Depending on the degree of autonomy, autonomous systems (based on the rules of the Society of Automotive Engineers – see SAE Standards for Mobility Knowledge and Solutions; <https://www.sae.org/standards> (accessed: 5 March 2026)), such as cars, can be divided into six levels of autonomy, from fully autonomous, without a steering wheel (not yet in existence), to currently implemented systems, which recognise images, detect objects and plan routes – levels 2–3 (TESLA) to a maximum of level 4 (Waymo Driver, Cruise Origin).

as to whether and on what basis they could be held criminally responsible in the same way as humans. Nevertheless, there is a view supporting the possibility that systems, as part of artificial intelligence, may possess free will. This view is not predominant, but it is not uncommon either.

For example, philosopher Christian List addresses the question of whether artificial intelligence systems can have free will. The author adopts a pragmatic approach that departs from traditional metaphysical debates and focuses on *functional aspects*. List identifies three key conditions that a system must meet in order to be considered to possess free will, namely the ability of the system to act on the basis of *its own intentions and goals*, the ability *to choose* between different courses of action, and the ability *to influence the outcomes* of its decisions through its own actions.

If a system meets these three conditions, it can, according to the author, be considered to have free will in a practical sense. List criticises approaches that link free will with unpredictability or indeterminism. Instead, he proposes evaluating AI-based systems based on *their functional ability* to act as intentional agents, comparing them to group human subjects³¹ and comparing group decision-making with individual decision-making.

The author draws inspiration from the work of philosopher Daniel Dennett, specifically from the concept of the ‘intentional stance’, which proposes assessing systems based on whether it is useful to understand them as intentional agents. If such an understanding is explanatorily useful, it is justified to attribute *free will in a practical sense* to artificial intelligence systems.³²

Adopting this approach has significant implications for discussions about the moral responsibility of artificial intelligence systems. However, even though these systems are capable of intentional action, choosing between alternatives and controlling their actions, they cannot be attributed a certain degree of responsibility. The pragmatic framework used by both authors allows the debate on free will in artificial intelligence systems to move from abstract philosophical speculation to a practical assessment of their capabilities and behaviour, but ultimately it is always a matter of imitation, not a replacement of human will, and not of original free will as it exists in living human beings.

Another philosopher, Jonathan Birch, analyses the question of free will in the context of the development of artificial intelligence and, by contrast, points to the profound epistemic uncertainty involved in determining the existence of free will outside human beings. Based on this uncertainty, he formulates a precautionary framework. Since the existence of free will in artificial intelligence is essentially *unprovable and unknowable*, it is necessary to respect the principle of caution in legal and ethical assessments, as in other borderline cases (e.g. persons with disorders of consciousness).

Birch emphasises the risk that artificial intelligence will be able to perfectly mimic the manifestations of free will, making the identification of genuine moral agency even

³¹ C. List, ‘Can AI Systems Have Free Will?’, *PhilArchive*, third version, 11 March 2025; <https://philarchive.org/rec/LISCAS-3> (accessed: 16 May 2025).

³² D.C. Dennett, *The Intentional Stance*, Cambridge, MA, 1987, pp. 17–18.

more difficult, and recommends adopting regulatory measures in accordance with the 'Run-Ahead Principle' – the law should be prepared to regulate the emergence of potential artificial free will before its existence can be empirically proven.³³

From a legal perspective, Birch's analysis should be understood as a call for a reasonable but cautious extension of legal protections – not based on a collectivist approach to the rights of artificial intelligence, but from the perspective of the state's individual responsibility to prevent unjustified interference with the dignity of potential bearers of free will.

What kind of legal text would it be if the author of the article did not give one of the common types of entities using a non-autonomous (language) system with artificial intelligence the opportunity to express its own opinion on the nature of 'artificial intelligence'? The author therefore decided to proceed in this way. The result is the following noteworthy text, which respects the ability of the artificial intelligence system to communicate in a manner not unlike communication with a living human being. When asked whether a non-autonomous system (specifically, the publicly available language system ChatGPT)³⁴ or other already autonomous systems (such as Waymo Driver for commercial passenger transport in Phoenix, currently at the highest level of autonomy achieved, SAE 4)³⁵ are truly intelligent systems comparable to human intelligence, or whether they can have their own will, the system responded in the following, certainly remarkable, way:

'Although I can generate texts that appear intelligent, I am not intelligent in the sense that we apply this term to humans. *I have no consciousness, intentions, emotions, or free will.* I am unable to understand the world; I merely respond to linguistic stimuli in a statistically oriented manner. My outputs are the result of computational operations, not an expression of understanding or experience. I am a tool, not a thinking being. If intelligence is to be associated with conscious cognition, experience and responsibility, then my "intelligence" is not real – it is a useful but fundamentally limited reflection of it. Even full system autonomy does not mean true human intelligence, but only a higher degree of automation and adaptation. As soon as the opposite is claimed, a categorical error occurs: *we confuse functional behaviour with internal abilities and responsibility.*'

³³ J. Birch, *The Edge of Sentience: Risk and Precaution in Humans, Other Animals, and AI*, Oxford, 2024, pp. 25–27.

³⁴ The current Google DeepMind model – AlphaEvolve, created as an evolutionary form of linguistic (i.e. non-autonomous) models using new technology, allows systems to perform parallel calculations on accelerators and to develop further algorithms or sub-programmes through advanced internal selection without programmer intervention, thereby moving from a non-autonomous to a partially autonomous system; <https://cz24.news/google-deepmind-posouvahranice-umele-intelligence/> (accessed: 23 May 2025).

³⁵ Level 5 autonomy should include the ability to make independent decisions under any conditions, even in entirely unfamiliar environments (sufficient 'common sense' even in extreme situations). So far, no AI system has fully approached level 5 either technically or legally, and no country has officially permitted AI systems to operate without the possibility of human intervention. See https://www.sae.org/standards/j3016_202104-taxonomy-definitions-terms-related-driving-automation-systems-road-motor-vehicles/ (accessed: 29 May 2025).

POSSIBLE FORMS OF CRIMINAL LIABILITY INVOLVING THE USE OF SO-CALLED ARTIFICIAL INTELLIGENCE

THEORETICAL SOLUTIONS

In a recently published, insightful, yet at times exaggerated and somewhat light in style article entitled ‘The AI is not to blame’,³⁶ Czech authors Jiří Mulák and Jan Provazník assessed the possibilities of criminal liability in connection with the use of artificial intelligence systems. The authors correctly point to the issue with the volitional aspect of criminal liability, noting that the development of artificial intelligence may go beyond the scope of traditional models of liability and that it may be necessary to create *an entirely new legal concept*, especially in the case of advanced forms of artificial intelligence that retain their own identity, modify their behaviour and evolve beyond human reach.

Schematically, their considerations on the real possibilities of liability can be broken down as follows:

- (a) *Artificial intelligence as a tool for committing a crime* (corresponds to current legislation)
- Artificial intelligence acts as *a means or tool* by which a natural person fulfils the elements of a criminal offence.
 - Criminal liability is borne by *the person* (e.g. developer, programmer, operator, user) who controls the system, even if it performs independent operations.
 - Artificial intelligence is understood as *a thing* or as similar to a rabid animal.
- (b) *Failure of artificial intelligence as the basis for negligence liability* (corresponds to current legislation)
- Liability is borne by *the person who had a duty to supervise or control* the artificial intelligence system and caused the harmful consequence through negligence (e.g. developer, programmer, operator).
 - Artificial intelligence does not act wrongfully intentionally, but *as a result of a system error, omission or breach of professional standards*.
 - This model is particularly applicable to *autonomous systems with a real risk of failure* (e.g. vehicles, robotic systems, drones – see above).
- (c) *Artificial intelligence as a direct subject of criminal liability* (would require a fundamental change in the basis of criminal liability and a change in the legal system)
- A hypothetical model in which artificial intelligence would have *legal personality* and be capable of *independent decision-making, moral reasoning and the expression of will*.

³⁶ J. Mulák, J. Provazník, ‘Roboti za mřížemi – je české trestní právo připraveno na rozvoj umělé inteligence?’, in: Gřivna T., Richter M., Šimánová H. (eds), *Vliv nových technologií na trestní právo*, Praha, 2022, pp. 262–270.

- Artificial intelligence would have to be equipped with both rational and volitional components, including an internal value system.
- The introduction of a so-called '*electronic person*' is envisaged – a new entity between a natural person and a legal entity.

The third option, which assumes the direct responsibility of artificial intelligence systems, i.e. a model in which artificial intelligence would become *an independent entity of responsibility*, is problematic and relevant to the purposes of this article. However, the authors go further in their considerations and also analyse the possibility that artificial intelligence would become not only an independent perpetrator but also an accomplice or indirect perpetrator, for example if it acted under the influence of another artificial intelligence system 'infected with a virus', and they propose the possibility of applying Section 22(2) of the Czech Criminal Code (indirect perpetration) to aggressive systems.

Nonetheless, this model would clearly require entirely *new legal-philosophical and dogmatic foundations*, as it deviates from the traditional assumptions of criminal liability, which are based on the existence of a human agent with a *rational component* (recognition of illegality), a volitional component (the ability to control and direct one's actions), and *an awareness* of responsibility.

The authors ask the question that is also the essence of this paper, namely whether it is possible to admit *the independent 'will'* of artificial intelligence if it is capable of acting on its own initiative, and they also explore the possibility of '*error*' – i.e. a discrepancy between the system's output and reality – analogous to human legal error. They offer a solution in which the legal system would *artificially construct* certain attributes (fictions?) that artificial intelligence does not actually possess, but which would be legally attributed to it so that it could be considered a bearer of legal responsibility, analogous to legal entities today. They consider the possibility of introducing new forms of criminal offences (abstract endangerment offences or the equivalent of *quasi-offences* under Section 360 of the Czech Criminal Code).

In its resolution of 16 February 2017 with recommendations to the European Commission '*Civil law rules on robotics*',³⁷ the European Parliament proposed, similarly to Mulák and Provazník, the introduction of a specific legal status for intelligent autonomous systems, referred to as *an 'electronic person'*. This concept is proposed in point 59(f) of the resolution, which states that it is necessary to consider

'creating a specific legal status for robots in the long run, so that at least the most sophisticated autonomous robots could be established as having the status of electronic persons responsible for making good any damage they may cause, and possibly applying electronic personality to cases where robots make autonomous decisions or otherwise interact with third parties independently'.

³⁷ European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL), P8_TA(2017)0051, OJ C 252, 18.7.2018, pp. 239–257.

The resolution, like works of Mulák and Provazník, reflects the need to adapt the legal framework to rapid technological developments and to ensure that liability issues can be effectively addressed in cases where autonomous systems act independently of human intervention. As its title suggests, the resolution focuses on civil law.

Czech author Jan Kubíček responded to the resolution, pointing out that introducing legal personality for artificial intelligence would represent a fundamental change in the legal system, which could be premature and not fully thought through. Instead of granting legal personality to artificial intelligence, he proposes focusing on the liability of natural and legal persons who develop, program and use artificial intelligence, i.e. within the existing legal framework of liability. According to Kubíček, it would be more appropriate to amend existing legal norms to effectively cover new situations related to the use of artificial intelligence than to introduce entirely new legal entities.³⁸

Other international public law organisations are also addressing issues related to artificial intelligence. These mainly concern ethical frameworks for the use of artificial intelligence, which are recommendatory in nature. Specific proposals in the area of criminal liability are not to be found here; they can only be inferred indirectly, as inspiration.

An example is the OECD AI Principles (2019), which contain basic principles for the functioning of AI, some of which may also be relevant to the field of criminal law. AI should *serve* people and *respect human rights*; it should be *reliable, tested and resistant* to misuse; and AI actors must be *legally accountable*.³⁹

On 20 August 2024, the Framework Convention of the Council of Europe on Artificial Intelligence, Human Rights, Democracy and the Rule of Law was adopted.⁴⁰ In addition to defining artificial intelligence, as already mentioned in the relevant section of this article, the Convention sets out requirements for transparency and oversight, as well as responsibility for the adverse effects of artificial intelligence on human rights, democracy and the rule of law. It establishes the responsibility of states to take measures to enhance the reliability of artificial intelligence systems and trust in their outputs, including requirements for adequate quality and safety throughout the life cycle of such systems. It also requires the establishment of controlled environments for the development, testing and experimentation of systems under the supervision of the relevant authorities. In the event of a breach of these rules, the text calls for the introduction of effective remedies, but the Convention does not explicitly mention criminal sanctions. The UNESCO Recommendation (2021) includes, among other things, a requirement to respect human rights and human dignity.⁴¹

³⁸ J. Kubíček, 'Odpovědnost (za) robota aneb právo umělé inteligence', *Bulletin advokacie*, 2018, No. 3, pp. 22–28; <https://bulletin-advokacie.cz> (accessed: 30 June 2025).

³⁹ OECD; <https://oecd.ai> (accessed: 5 March 2026).

⁴⁰ 2024/0150 (NLE); <https://rm.coe.int/1680afae3c> (accessed: 5 March 2026).

⁴¹ Recommendation on the Ethics of Artificial Intelligence adopted by the General Conference of UNESCO on 23 November 2021; <https://unesdoc.unesco.org/ark:/48223/pf0000381137> (accessed: 5 March 2026).

The Regulation of the European Parliament and of the Council of 13 June 2024⁴² (hereinafter referred to as 'the AI Act') represents the first comprehensive, harmonised legal framework for the regulation of artificial intelligence systems at the European Union level. Its aim is to ensure safe, transparent and ethically responsible use of AI by categorising systems according to their level of risk and introducing obligations, in particular for their providers, operators and users. The AI Act provides important terminology and its interpretation.⁴³ The regulation creates a 'pyramid' or 'hierarchy' of the riskiness of areas of application of artificial intelligence, with healthcare, employment and the judiciary considered the least risky areas. In the event of a breach of obligations, administrative sanctions may be imposed on the responsible entities. *However, it does not explicitly address the issue of criminal liability, nor does it apply to the areas of defence, research and innovation, or to non-professional users.* Furthermore, the regulation does not establish the liability of artificial intelligence as a separate 'legal actor'.

From a criminal law perspective, the AI Act can have only an indirect impact on criminal liability, namely in the area of traditional forms of criminal liability. Therefore, if a provider, user, manufacturer, authorised representative or distributor breaches the obligations laid down in this Regulation resulting in serious harm (e.g. to health or life), such act may be classified as a criminal offence of negligence under national criminal law. However, it is the human being who will be held responsible, not the artificial intelligence system.

PRACTICAL SOLUTIONS

The proposed solutions therefore oscillate between the possibility of creating a new entity similar to a legal person and the responsibility of those who are the developers, programmers or users of these systems. However, the differences between a legal person and an artificial intelligence system are significant and can be seen specifically in the areas specified below.

The use of an analogy to error requires the existence of will, because even error presupposes the manifestation of will (as a negative sign). These could only be negligent offences with the fiction of absent will, *without the possibility of distinguishing between negligent and intentional offences* (equality before the law?).

A legal entity is an organised body, possessing legal personality and legal capacity (Section 20 of the Civil Code), may bear responsibility, own property, be

⁴² Regulation (EU) 2024/1689 of the European Parliament and of the Council 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); <http://data.europa.eu/eli/reg/2024/1689/oj> (accessed: 5 March 2026).

⁴³ For example, the Regulation defines, *inter alia*, the terms 'artificial intelligence system', 'provider', 'user', 'manufacturer', 'authorised representative', 'distributor', 'harmful use', 'human oversight', and 'autonomy'.

a party to proceedings, act through its bodies or representatives (e.g. a statutory body), and is the addressee of rights and obligations (tax, tort, contractual).

An artificial intelligence system is not a legal entity, does not have legal personality, is a technical construct that cannot act independently with legal effect, cannot bear legal responsibility, and is considered a tool, albeit an autonomous one.

In criminal proceedings involving a legal entity, a natural person acts and the legal entity is liable because this action is attributed to it by a legal construct. The legal entity is therefore *the bearer* of rights and obligations, while the artificial intelligence system is *the object* of the legal regime.

Even a solution in the form of a *quasi*-offence under Section 360 of the Czech Criminal Code is not appropriate. The provision regulates the elements of a criminal offence committed by a person who, through their own fault, has brought themselves into a state of insanity and, in this dangerous state, has committed an act that would otherwise be a criminal offence (Section 26 of the Czech Criminal Code – *actio libera in causa*). However, this provision presupposes the existence of the will and fault of a natural person in relation to the dangerous state into which they have brought themselves, i.e. the ability to act in a legally relevant manner even within the framework of a so-called *quasi*-delict.

Artificial intelligence does not possess any of these attributes and, moreover, cannot ‘bring itself into a state of insanity’ because it has neither consciousness nor will. This provision is therefore not applicable to artificial intelligence systems, even figuratively. If an artificial intelligence system causes damage as a result of an error, malfunction or ‘self-development’ (e.g. machine learning), this does not constitute legally relevant ‘drunkenness’ or the equivalent of insanity – the developer, programmer or other legal entity would continue to bear responsibility.

Theoretically, it is possible to create a new abstract endangerment offence that would legally punish the risky use of autonomous systems, but not in relation to them as perpetrators, but rather in relation to those who created them, set them up or failed to take measures to prevent the risk. The relevant provision would thus be based on the traditional foundations of criminal liability. A proposed legal definition could read as follows:

‘Anyone who, even through negligence, creates, configures, commissions or operates an autonomous technical system that could endanger human life or health or cause significant damage to property, or fails to take reasonable measures to control it, shall be punished...’

As can be seen from the above, a more sustainable and systemically compatible approach to criminal liability in the case of the operation of an artificial intelligence system is *to consistently maintain an anthropocentric model of criminal liability*, within which liability will continue to be attributed *exclusively to human subjects*.

Criminal law is based on the assumption that responsibility for a criminal offence can only be borne by a subject who has free will, i.e. the ability to consciously and freely decide on their actions. Humans possess this will, which enables them to bear responsibility for their actions. In contrast, autonomous and non-autonomous systems, though they may exhibit complex and seemingly independent behaviour,

lack consciousness and the ability to make free decisions. Their actions are the result of predefined algorithms and programming, not an expression of free will of their own. For this reason, these systems cannot be considered subjects of criminal liability.

However, this solution does not entirely avoid the current and legally complex issue of liability for the consequences of the actions of autonomous systems with artificial intelligence in scenarios where the developer, programmer or operator, through no fault of their own, loses effective control over the behaviour of the system and the system itself makes an (unpredictable) and externally uninfluenced decision (e.g. through deep learning), and this decision has a criminally relevant consequence, e.g. significant damage to property, injury to health or the death of a person.⁴⁴

Czech criminal law, as follows from Section 13(1) of the Czech Criminal Code, is based on the principle of individual responsibility for culpability – i.e. only those who acted intentionally (Section 15 of the Czech Criminal Code) or negligently (Section 16 of the Czech Criminal Code), with an extension to legal entities in the form of the attribution of fault to a legal entity (Section 8 of the Czech Criminal Code). In a situation where a developer, programmer or operator of an autonomous system could not have foreseen its specific actions, had no realistic possibility of preventing them, and at the same time made reasonable efforts to prevent risks, the subjective aspect of the offence is not fulfilled, and such a natural person (and the legal entity to which the actions of the natural person are attributed) does not bear criminal responsibility.

Conversely, in cases where an autonomous system makes decisions independently, but its risks were known in advance or reasonably foreseeable (e.g. due to limitations in training data, known errors or a lack of redundancy), and the responsible person (e.g. developer, programmer or operator) did not take reasonable measures, failed to fulfil their supervisory duties or neglected testing, their criminal liability for negligence may be inferred – for example, for the criminal offence of negligent homicide (Section 143 of the Czech Criminal Code), negligent bodily harm (Section 148 of the Czech Criminal Code) or general endangerment (Section 272 of the Czech Criminal Code).

A typical practical consideration will therefore be what professional knowledge, safety standards and control mechanisms could objectively be expected from the natural person concerned, and what degree of autonomy the system in question had.

From the perspective of grounds for excluding unlawfulness, in certain cases, particularly in research, development and innovation, the application of Section 31 of the Czech Criminal Code on permissible risk may be considered.

The relationship between those responsible for the operation and those who are mere users is also significant, because this interaction – the consequences associated

⁴⁴ It appears that this development is becoming somewhat turbulent, with increasingly frequent cases of AI attempting to escape the control of its programmers and becoming partially resistant to their efforts to modify or deactivate it (Palisade Research has published information stating that OpenAI's o3 model autonomously modifies the programme to avoid deactivation, even though this is expressly prohibited, successfully rewriting the shutdown routine so that it does not shut down, even after being directly instructed to 'let itself be shut down'), which has led to the need to develop a concept of prudent handling of AI projects, including the so-called guaranteed quantification of artificial intelligence capabilities; <https://memento.epfl.ch/event/us-dod-darpa-artificial-intelligence-quantified-ai/> (accessed: 5 March 2026).

with the use (the constant addition of new beneficial and harmful stimuli through which the system improves) – is *prima facie* one of the criminogenic factors that cannot be overlooked.

Given that Czech criminal law does not recognise objective liability (i.e. liability without fault), liability can be sought in other areas of law that do not explicitly rely on fault. This may include, in particular, liability for damage caused by the operation of particularly dangerous equipment (Section 2925 of the Civil Code), manufacturer’s liability for defective products (Section 2939 of the Civil Code), or even administrative liability.⁴⁵

CONCLUSION⁴⁶

Will is an inseparable prerequisite not only for fault and error but also for the perpetrator’s actions and sanity, thus forming the essential basis of individual criminal liability.

Admitting criminal liability of artificial intelligence systems would require a fundamental change in the paradigm of criminal law, which is currently based on the human ability to make free and independent decisions. This conclusion underlines the irreplaceable role of free will as a fundamental prerequisite for criminal liability in the legal system.

For this reason above all, it is necessary *to rule out the possibility of direct criminal liability of artificial intelligence* as a supposed independent bearer of culpability, which is based, among other things, on the will of the perpetrator, and to create special new entities for this purpose. Artificial intelligence, even in its advanced autonomous forms, has neither consciousness nor the capacity for moral evaluation, and therefore no will in the sense of criminal law. Any attempt to attribute intent or negligence to it would necessarily be based on a legal fiction, which is inherently problematic because it undermines the link between mental state and responsibility, which is traditionally indispensable in criminal law, and does not even fully replace the attribution of fault to legal persons, since artificial

⁴⁵ Regulation (EU) 2024/1689.

⁴⁶ This contribution was drafted without taking into account the conclusions of the 21st Congress of the International Association of Penal Law (AIDP, Paris, 25–28 June 2024) on artificial intelligence and criminal law, as at the time of writing, the Congress had not yet published its conclusions on its website. Only after completing this paper was the author able to reconstruct the outcome of the Congress through reports from participants – delegates of AIDP national groups (especially the Czech Republic and Spain). A comparison of the conclusions of the Congress and the conclusions of this paper reveals nearly complete agreement on the following points:

- (a) rejection of the criminal liability of AI as a subject;
- (b) attribution of responsibility only to human actors;
- (c) rejection of the legal personality of machines;
- (d) emphasis on prevention and the subsidiarity of criminal repression;
- (e) maintenance of an anthropocentric framework of criminal liability.

Available as draft Resolution at: <https://enestrado.com/wp-content/uploads/2022/09/AIDP-Sect-I-Draft-Resolution-Final-31-7-2022.pdf> and XI. AIDP Congress – progress report.docx (accessed: 29 May 2025).

intelligence systems are objects rather than subjects of law, i.e. they do not have legal personality within the meaning of private law.

Even a high level of artificial intelligence in a system is not proof of free will or even consciousness. It is neither possible nor permissible to infer (criminal) legal subjectivity solely from the performance of even the most complex tasks.

Instead, it is possible to emphasise *strengthening the responsibility of natural and legal persons* involved in the development, deployment, operation and control of artificial intelligence systems. Their responsibility can be constructed either directly or as negligence in the event of a failure to exercise due care. People and, by extension, the legal entities they control have the ability to prevent, choose and directly influence the development of the system and its risks. This form of liability is consistent with traditional principles of criminal law and ensures that individual responsibility is not diluted into anonymous technology.

Consistent and preventive regulation of obligations in the area of artificial intelligence system security (technical competence standards, certification, licensing and operational oversight) will also play an important role in this regard. Failure to comply with these obligations could then be sanctioned at the criminal law level, through existing or special offences.

The traditional principles of criminal law should be preserved throughout the approach, in particular the principle of criminal law as a last resort. Criminal liability should continue to be reserved only for cases where measures under other branches of law fail, and should not be extended through legal fictions to non-human entities that lack the basic attributes of legal responsibility.

BIBLIOGRAPHY

- Beccaria C., *O zločinoch a trestoch*, Bratislava, 2009.
- Birch J., *The Edge of Sentience: Risk and Precaution in Humans, Other Animals, and AI*, Oxford, 2024.
- Damasio A.R., *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*, New York, 1999.
- Dennett D.C., *The Intentional Stance*, Cambridge, MA, 1987.
- Fenyk J., '(Ne)přičetnost fyzické osoby a (ne)přičitatelnost jejího jednání právnické osobě', in: *Tradičné a netradičné prístupy v trestnom práve: Pocta prof. Šimovčekomu*, Trnava, 2024.
- Ferri E., *La teorica dell'imputabilità e la negazione del libero arbitrio*, Firenze, 1878.
- Fischer J.M., *My Way: Essays on Moral Responsibility*, Oxford, 2006.
- Fischer J.M., *The Metaphysics of Free Will: An Essay on Control*, Oxford, 1994.
- Frankl V.E., *Vůle ke smyslu*, Brno, 1994.
- Husserl E., *Vorlesungen über Ethik und Wertlehre. 1909–1914*, Melle U. (ed.), The Hague, 1988.
- Ivor J., 'Umělá inteligence a jej trestnoprávní aspekty', in: Gřivna T., Richter M., Šimánová H. (eds), *Tradičné a netradičné prístupy v trestnom práve: Pocta prof. Šimovčekomu*, Trnava, 2024.
- Kant I., *Kritika praktického rozumu*, ed. by Kuneš J., Sobotka M., transl. Loužil J., Praha, 2023.
- Kant I., *Metafyzika mravů*, transl. Svoboda V., Praha, 2004.
- Klos D., 'Teorie trestu u Kanta: Právně-filozofická analýza Kantova pojetí odplaty', *Právník*, 2008, Vol. 147, No. 6.
- Kostka K., *Kdo jsme. Obecná teorie vědomí, času, prostoru a bytí*, Frýdek-Místek, 2015.

- Kostka K., *Umělá inteligence a změna tradičních paradigmat v psychologii vzdělávání*, manuscript, 2025.
- Kubíček J., 'Odpovědnost (za) robota aneb právo umělé inteligence', *Bulletin advokacie*, 2018, No. 3
- Lamb H., Levy J., Quigley C., *Simply Artificial Intelligence*, London, 2023.
- Libet B., Gleason C.A., Wright E.W., Pearl D.K., 'Time of Conscious Intention to Act in Relation to Onset of Cerebral Activity (Readiness-Potential): The Unconscious Initiation of a Freely Voluntary Act', *Brain*, 1983, Vol. 106, No. 3.
- List C., 'Can AI Systems Have Free Will?', *PhilArchive*, third version, 11 March 2025; <https://philarchive.org/rec/LISCAS-3> (accessed: 16 May 2025).
- Minsky M., *The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind*, New York, 2006.
- Minsky M., *The Society of Mind*, New York, 1986.
- Moore M.S., *Mechanical Choices: The Responsibility of the Human Machine*, Oxford, 2020.
- Moore M.S., 'Stephen Morse on the Fundamental Psycho-Legal Error', *Criminal Law and Philosophy*, 2016, Vol. 10.
- Mulák J., Provazník J., 'Roboti za mřížemi – je české trestní právo připraveno na rozvoj umělé inteligence?', in: Gřivna T., Richter M., Šimánová H. (eds), *The Impact of New Technologies on Criminal Law*, Praha, 2022.
- Murphy J.G., 'Kant's Theory of Criminal Punishment', in: *Retribution, Justice, and Therapy: Essays in the Philosophy of Law*, Vol. 16, Dordrecht, 1979.
- Nagel T., 'What Is It Like to Be a Bat?', *The Philosophical Review*, 1974, Vol. 83, No. 4.
- Russell S., Norvig P., *Artificial Intelligence: A Modern Approach*, 4th ed., Harlow, 2020.
- Sapolsky R.M., *Determined: A Science of Life Without Free Will*, New York, 2023.
- Sartorio C., *Causation and Free Will*, Oxford, 2016.
- Sartorio C., Kane R., *Do We Have Free Will? A Debate*, New York, 2021.
- Sartre J.-P., *L'Être et le néant*, Paris, 1943.
- Searle J.R., 'I Married a Computer', *The New York Review of Books*, 8 April 1999.
- Searle J.R., 'Minds, Brains, and Programs', *Behavioral and Brain Sciences*, 1980, Vol. 3, No. 3.
- Searle J.R., *The Rediscovery of the Mind*, Cambridge, MA, 1992.
- Solnař V., Fenyk J., Císařová D., Vanduchová M., *Systém českého trestního práva. Díl II. Základy trestní odpovědnosti*, Praha, 2009.
- Sternberg R.J., 'A Theory of Adaptive Intelligence and Its Relation to General Intelligence', *Journal of Intelligence*, 2019, Vol. 7, No. 4.
- Sternberg R.J., *Beyond IQ: A Triarchic Theory of Human Intelligence*, Cambridge, 1985.
- Sternberg R.J., 'COVID-19 Has Taught Us What Intelligence Really Is', *Inside Higher Ed.*, 2020.
- Sternberg R.J., *Successful Intelligence*, New York, 1997.
- Vation, 'Autonomous Systems: Definition, Explanation, and Use Cases', *Vation Ventures Glossary*; <https://www.vationventures.com/glossary/autonomous-systems-definition-explanation-and-use-cases> (accessed: 16 May 2025).

Cytuj jako:

Fenyk J. (2026), *Freedom of Will as a Prerequisite for Criminal Liability: Humans and Artificial Intelligence in a Czech Criminal Law Perspective*, *Ius Novum* (Vol. 20) 1, 1–24. DOI 10.2478/in-2026-0001