Lex Humana, v. 17, n. 3, 2025, ISSN 1984-6754
Universidade Católica de Petrópolis

AI DECISION MAKING AND CRIMINAL LAW EXECUTION: COMPLIANCE WITH INMATES' RE-SOCIALIZATION?

DECISÕES DA INTELIGÊNCIA ARTIFICIAL E EXECUÇÃO PENAL: A IA É COERENTE COM A SEGURANÇA E A RESSOCIALIZAÇÃO DOS PRESOS?

Carlo Piparo
University of Udine, Italy
carlopiparo@gmail.com

Received: 15 Jun 2025
Accepted: 20 Jul 2025
Published: 04 Aug 2025
Corresponding author:
carlopiparo@gmail.com

Resumo

A execução das sentenças penais representa uma fase crucial no sistema penal, onde os objetivos da justiça punitiva e da reabilitação do infrator se concretizam. Nos últimos anos, no entanto, a crescente complexidade da gestão prisional e a demanda por maior eficiência na administração pública têm impulsionado o interesse pela integração de tecnologias de Inteligência Artificial no processo de execução penal. Este artigo investiga a interseção entre IA, execução do direito penal e os objetivos reabilitadores da punição. Embora a IA tenha potencial para otimizar as operações prisionais — por meio de análises preditivas, monitoramento comportamental e melhor alocação de recursos — sua aplicação levanta importantes preocupações jurídicas, éticas e operacionais. Questões centrais incluem a transparência dos algoritmos, o risco de discriminação e a possível erosão de direitos fundamentais. A pergunta central desta pesquisa é se a tomada de decisões baseada em IA é compatível com a segurança pública e o objetivo constitucional e humanista de ressocialização dos detentos.

Palavras-chave: Execução. Direito penal. Inteligência Artificial. Reabilitação.

Abstract

The execution of criminal judgements represents a crucial phase in the penal system, where the objectives of punitive justice and offender rehabilitation materialize. In recent years, however, the increasing complexity of prison management and the demand for greater efficiency in public administration have driven interest toward the integration of Artificial Intelligence technologies into the penal execution process. This paper investigates the intersection between AI, criminal law execution, and the

rehabilitative aims of punishment. While AI holds potential for optimizing prison operations —through predictive analytics, behavioral monitoring, and improved resource allocatio — its application raises significant legal, ethical, and operational concerns. Core issues include the transparency of algorithms, the risk of discrimination, and the potential erosion of fundamental rights. The question at the heart of this research is whether AI-driven decision-making is compatible with public safety and with the constitutional and humanistic goal of inmate rehabilitation.

Keywords: Execution. Criminal-law. Artificial Intelligence. Re-habilitation. Public Safety.

Introduction

Today's society is living a time of groundbreaking transformations we call "Digital Revolution", which "digitalized" almost every aspect of our lives: from workspaces to schools, from authomatized kitchens to public administrations (BODEN, 2018). This shift revealed to be as impactful as the changes introduced in the Industrial Age, such as steam-powered machines or Gutenberg's printing press. What makes this revolution unique is its scale and speed: with computing, data processing, and communications, digital technologies now shape not only how information flows but also how society functions at its core (MARSILI, WRÓBLEWSKA-JACHNA, 2024).

The digital revolution, as doctrine points, has taken us through multiple stages: from mechanical and electrical innovations to the current era of electronic and networked systems. These shifts have brought both exciting opportunities and serious challenges. In many ways, our societies are still trying digest such revolutionary changes. We face new questions about privacy, autonomy, democracy, and even what it means to be human in an age where machines can mimic aspects of our intelligence (MARSILI, WRÓBLEWSKA-JACHNA, 2024).

At the center of this digital world we can find algorithms: sets of instructions that tell computers how to perform tasks. Indeed, Artificial intelligence is "made of algorithms", which are *«real mathematical objects»* (YANOFSKY, 2010, p. 2). While the concept of algorithm may seem technical or abstract, algorithms are real and practical. Algorithms affect our daily lives in very real ways: they sort the news we see, recommend what we watch or buy, and influence decisions about everything from job applications to public policy. Far from being neutral tools, algorithms are designed by people and reflect human choices and biases. As they become more embedded in our systems, they have the power to reinforce inequalities or even shape belief systems (MARSILI, WRÓBLEWSKA-JACHNA, 2024).

Even more transformative is the rise of Artificial Intelligence. Doctrine define Artificial Intelligence as the ability of systems to carry out tasks that would normally require human thinking, such as recognizing patterns, solving problems, or making decisions. What sets today's AI apart is its ability to learn and evolve. Through machine learning and neural networks, AI systems can adapt to new data, improve over time, and even generate original content, such as texts, images, and videos. Artificial Intelligence, indeed, learns and adapts, interacts and determines itself inside the environmental frame. In this sense, we can talk about algorithmic "decision-making": artificial intelligence, indeed, collects (input) data, analyzes

them, and elaborates (output) data whick it uses to make decisions.

In some cases, algorithmic systems are becoming so sophisticated that it's difficult for humans to tell whether content is real or artificially produced. This power comes with serious risks. AI is already being used to create "deepfakes": realistic but fake videos or audio recordings that can spread misinformation and deceive the public. In a world increasingly shaped by digital content, this threatens our ability to trust what we see and hear (PIPARO, 2024). Moreover, when AI is used in ways that manipulate people's emotions or behaviors -what the article refers to as "cognitive hacking"- the consequences go beyond technical concerns and touch the very foundations of democracy and human freedom (MARSILI, WRÓBLEWSKA-JACHNA, 2024).

Despite these dangers, the digital revolution and the rise of AI also bring immense potential. They open doors for scientific discoveries, innovation, and more efficient ways to manage global challenges. But for that potential to be realized ethically and sustainably, we need clear standards, strong policies, and a deep public understanding of these technologies. As the authors point out, it's not enough to embrace digital change: we must also shape it in ways that reflect our values and protect our rights (MARSILI, WRÓBLEWSKA-JACHNA, 2024).

Artificial Intelligence is a concept that has captured both the imagination and concern of society for decades, and today, it stands at the center of some of the most transformative changes in human life. At its core, AI refers to the development of systems capable of performing tasks that normally require human intelligence, such as learning, reasoning, problem-solving, perception, and language understanding. The term was first coined in 1955 by Stanford professor John McCarthy, who defined it as «*the science and engineering of making intelligent machines*» (SHEIKH, PRINS, SCHRIJVERS, 2023). While early approaches to AI focused on programming machines to perform specific logical tasks -such as playing chess or solving equations- modern AI emphasizes adaptability, learning, and decision-making in uncertain environments, much like human beings do (SHEIKH, PRINS, SCHRIJVERS, 2023; SCHANK, 1987).

One of the most profound shifts in the field has been the transition from static programming to dynamic learning. This evolution is embodied in the field of Machine Learning, which explores how computer systems can improve their behavior based on experience or data. ML enables AI systems to develop their own internal models and solutions, often without explicit human instructions. Supervised learning, for instance, involves training a model on labeled data, while unsupervised learning identifies hidden patterns without labels. Reinforcement learning, on

the other hand, teaches AI through trial and error, maximizing rewards over time, mirroring how humans and animals learn from interaction with their environments (SHEIKH, PRINS, SCHRIJVERS, 2023).

At the forefront of current AI advancements is Deep Learning, a subset of ML that utilizes multi-layered neural networks inspired by the structure of the human brain. These artificial networks process information in hierarchical layers, allowing systems to recognize patterns in speech, images, and language with astonishing accuracy. Deep Learning has played a pivotal role in AI's recent successes, from autonomous driving to language generation tools, enabling both broad generalization from limited data and scalability to vast datasets. Yet, despite these technical marvels, the nature of Artificial Intelligence is not purely mechanical. Its development is deeply entwined with human values, needs, and social structures. Human-centered AI, for example, is a growing field focused on creating systems that augment rather than replace human capabilities. These technologies are designed to collaborate with people, adapting to individual preferences and social contexts, whether it's a voice assistant helping someone navigate a smartphone or a robotic companion providing emotional support for the elderly (SHEIKH, PRINS, SCHRIJVERS, 2023).

It's important to note that most current AI applications fall under "narrow AI", systems designed to excel at a specific task, such as facial recognition or language translation. The more ambitious goal, often referred to as Artificial General Intelligence, aspires to build machines with human-level understanding and flexible reasoning across varied contexts. While AGI remains a long-term vision, its pursuit raises critical ethical and philosophical questions about the role of intelligent machines in our lives and societies (SHEIKH, PRINS, SCHRIJVERS, 2023). Thus, Algorithmic Intelligence (equivalent with "Artificial Intelligence") contemplates different techniques, characteristics, uses, and approaches: authomated learning, mechanical reasonning, robotics, etc. (KOF, BOERS, KOSTERS, PUTTEN, POEL, 2002).

Thus, AI has been implemented in almost every corner of society. The execution of criminal law jurisdiction judgements -while we are speaking- is being touched by Artificial Intelligence, which profoundly transforming the management of the prison system by enhancing security, surveillance, correction, and the overall efficiency of incarceration. More broadly, the applications of algorithmic intelligence in the field of criminal enforcement are manifold:

> Alongside the potential development of databases and the organizational improvement of physical facilities and personnel resources, the vast scope of algorithm usage extends from the prevention of crime and the identification of offenders, to the adoption of predictive justice tools that support judges in

> decision-making and sentencing, and even to the profiling of individuals to forecast future behavior and calculate the likelihood of recidivism (CANESCHI, 2024, p. 257-258).

Such examples demonstrate how AI can serve as a strategic tool in the modernization of criminal enforcement, improving both the monitoring of inmate activities and the prevention of unlawful behavior, with significant implications for the security and efficiency of the prison system. However, as will be discussed in the following sections, the use of AI within modern Constitutional systems must be carefully calibrated to comply with the complex set of regulations safeguarding fundamental interests, *ex multiis*: protection of inmates' rights, compliance to correction principles, proper administration of justice, public order, and public safety. These constitutional necessities demand, as will be shown, lead legal interpreters to raise numerous concerns about the practical applicability of such algorithms, thereby requiring a particularly prudent legislative (as well as technical) intervention.

This study is structured around three key questions: Can AI improve the efficiency, effectiveness, and cost-performance of criminal law execution in public administration? What are the risks associated with excessive automation of prison-related decisions? What legal safeguards are necessary to ensure a balance between technological innovation and the protection of individual rights?

Theoretical framework of punishment: classical literature and modern constitutional achievements

To analyze the compatibility between algorithmic intelligence and the execution of punishment, we should begin with a doctrinal definition of punishment that reflects broad consensus. Punishment, as defined by Hart, consists of five essential elements: it must be afflictive, imposed in response to a violation of the law, inflicted on the actual offender, administered by human beings, and applied by the judge of the legal system in which the offense occurred (NJOKU, 2012). As for the function performed by punishment, historically it has addressed three main objectives: retribution, general prevention, and special prevention. Each of these functions possesses distinct and specific characteristics. From the perspective of Seneca's teachings, these functions can be divided into two broad categories (SENECA, 4 BC - 65 AD).

On one hand, we find the so-called "absolute teachings", which emphasize

offenses committed in the past and assert that punishment should be applied *quia peccatum est*: because a wrong has been committed. This approach refers to the retributive purpose of punishment. The retributive aim of criminal law falls within this category. On the other hand, the "relative teachings" maintain that punishment should be applied *ne peccetur*: so that future wrongdoing is avoided. In other words, punishment aims to modify the offender's future behavior, encompassing both general and special preventive purposes (PIPARO, BLAZEK, 2024).

The theory of punishment deeply influences the structure of criminal legislation and how sentences are implemented, depending on whether one adheres to one model or another. It seems almost intuitive that embracing one theory over another has significant implications for sentencing and its enforcement. Retributive theory, which does not rely on empirical justifications (such as the offender's social or personal history), bases punishment solely on proportionality, taking into account the harm done to the victim and the conduct that caused it (PIPARO, BLAZEK, 2024; CIAPPI, COLUCCIA, 1997).

In contrast, the theory of special prevention must constantly evaluate the rationality, feasibility, and effectiveness of punitive actions and define the content it adopts. This includes identifying indicators of individual and general preventive needs -possibly differentiated by categories of offenses and offenders- and developing criteria to translate these needs into concrete forms of punishment. The rehabilitative aim of punishment quickly became a defining feature of European legal culture. In the European context, the goal of rehabilitation has shaped the conception of punishment, including in the Italian Constitution. According to Article 27, paragraph 3 of the Italian Constitution, the primary objective of punishment is now "social reintegration," with particular attention to the rehabilitation of the offender within society (PIPARO, 2025). For this reason, the goal of punishment must be rehabilitation, and the primary quality of penal treatment must be the re-education of the offender: not merely a generic tendency within the system. Consequently, the entire penal system is designed with rehabilitative purposes in mind, and the severity of the punishment cannot ignore the inalienable criteria of social reintegration, which are tied to the seriousness of the offense and the mindset of the accused (PIPARO, BLAZEK, 2024).

Over the past two centuries, the global philosophy underpinning penal systems has undergone a profound transformation: from the brutality of retribution to the promise of rehabilitation. Historically, punishment was a public spectacle aimed at deterrence and vengeance. In early modern Europe, harsh corporal penalties, including executions, branding, and mutilation, were common tools used by sovereign

powers to assert authority and retaliate against deviance. As noted by MULCAHY (2019) medieval and early modern punishment resembled a ritual of "social surgery," whereby the offender was metaphorically and physically cut out from the social body to prevent further "infection" of the community. This conception of justice, rooted in divine or monarchic retribution, gradually gave way as Enlightenment ideals and utilitarian reformers such as Jeremy Bentham began advocating for more rational and humane approaches to crime control.

With the rise of the modern prison system in the 18th and 19th centuries, a new penal paradigm emerged. The shift was influenced not only by changing moral values but also by economic pragmatism. Reformers like Bentham saw in the prison a tool to transform idleness into productivity, envisioning incarceration not merely as punishment but as a method to reshape the offender into a compliant and useful citizen. During this time, prison architecture, religious instruction, and enforced labor were deliberately designed to instill discipline and repentance, laying the groundwork for what we would now call rehabilitation (MULCAHY, 2019). The 20th century saw further evolution with the incorporation of psychological and sociological theories into criminal justice policy. Particularly after World War II, the rise of the human rights movement and welfare state models in Western democracies reframed punishment in terms of social reintegration. Rehabilitation became a central goal, especially in European legal systems, emphasizing individual reform over collective vengeance (MULCAHY, 2019). Italy is a prime example of this constitutional reorientation: Article 27 of the Italian Constitution enshrines the rehabilitative aim of punishment, affirming that sentences must aim at the re-education of the offender: a stark departure from older retributive rationales (PIPARO, BLAZEK, 2024; PIPARO, 2025).

However, the rehabilitative ideal faced backlash in the 1970s, particularly in the United States, after Robert Martinson's influential essay argued that "nothing works" in correctional rehabilitatio. Despite its methodological flaws, this pessimistic view shaped public policy for decades, triggering a return to punitive, deterrence-focused models, especially amid growing fears around crime and political pressure to appear "tough on crime". Yet, from the 1990s onward, rehabilitation re-emerged as a central theme, driven by empirical research and evidence-based practices such as the Risk-Needs-Responsivity model. This model, now dominant in many Western jurisdictions, uses actuarial tools to tailor interventions to individual criminogenic needs, blending rehabilitation with public safety concerns (MULCAHY, 2019).

Today, the shift toward rehabilitation reflects not only a commitment to human dignity but also a strategic acknowledgment that reducing recidivism through

education, psychological support, and social inclusion serves both the individual and society. In this comparative evolution, retribution has yielded to rehabilitation: not completely or uniformly, but significantly enough to mark a global change in the understanding and purpose of punishment (MULCAHY, 2019). That given, constitutional interpretation requires consideration of both the goals of punishment and those of rehabilitation. Therefore, the objective of rehabilitation is as important as that of punishment. The use of the verb *"to aim"* (in the constitutional language) underscores the necessity for the rehabilitative process to respect the offender's right to self-determination. In this sense, Article 25, paragraph 2 of Spanish constitution mandates that prison sentences and security measures should be aimed at re-education and social rehabilitation, and Article 27, paragraph 4, Italian Constitution states that the death penalty shall never be permitted, further emphasize the humanized nature of punishment and its orientation toward rehabilitative ends (NICOTRA, 2014). In the same sense, in Portugal Article 2, paragraph 1, of Law Decree n. 265 of 1979 assesses that "The enforcement of custodial sentences and security measures aims at the reintegration of the offender into society by preparing them to lead a socially responsible life without committing crimes, as well as the protection of legal interests and the defense of society".

Indeed, according to Italian Constitutional Cout:

> It is indeed true that the nature of punishment inherently involves elements of social defense, as well as general prevention, due to the deterrent effect it may have on the utilitarian calculations of potential offenders. However, with regard to certain aspects -such as its afflictive and retributive dimensions- these represent minimal conditions without which punishment would cease to be what it is. On the other hand, aspects such as reintegration, deterrence, and social defense are indeed values with constitutional foundations, but not to such an extent as to justify undermining the rehabilitative purpose expressly enshrined in the Constitution within the framework of the penal system. If the aim of punishment were redirected toward those other characteristics, rather than toward the rehabilitative principle, there would be a risk of instrumentalizing the individual for broader goals of criminal policy (general prevention), or of prioritizing the satisfaction of collective needs for stability and security (social defense), thereby sacrificing the individual through the exemplarity of the sanction[1].

[1]Italian Constitutional Court, Judgement n. 313 of 1990, 26 June 1990.

Algorithmic decision making and protection of fundamental rights

We can define "Smart prisons" as the digitalized jails: they represent the revolution of the penitentiary system, and they aim to balance security requirements, management optimization, and rehabilitative opportunities (LINDSTRÖM, PUOLAKKA, 2020). A critical analysis of these developments requires a modern criminological approach: one that considers the legal and ethical implications, as well as the growing role of technological justice in the contemporary landscape. Indeed, such a development could amount to a true Copernican revolution. However, an excessively innovative approach may risk not only disrupting the delicate balances that the Prison Administration has gradually established over time, but could also hinder or even undermine the rehabilitative process and, in some cases, threaten the protection of fundamental human rights. Indeed, the implementation of Artificial Intelligence in the Prison Administration implies that decisions are taken by AI itself: thus, the question arises whether such decision-making activity complies with fundamental rights protection and correction itself.

In this sense, the Constitution of the Italian Republic, the European Convention on Human Rights, and the case law of the European Court of Human Rights provide the essential legal framework to ensure that the use of algorithmic decision-making within prison systems respects fundamental human rights. The legal context in which this subject is situated is primarily informed by Articles 2, 27, and 32 of the Italian Constitution, as the execution of penal sentences is governed by laws which, in turn, must align with the Constitution, placing the human person, with their inalienable dignity, at its core. *«Dignity and personhood must be understood as synonymous, for to eliminate or suppress an individual's dignity is to deprive or diminish their status as a human being, which is impermissible under any circumstances»* (SILVESTRI, 2014), including incarceration, which may limit personal liberty but not dignity or fundamental rights. The Constitutional Court has consistently emphasized the decisive role of human dignity in its rulings (ESPOSITO, 2020). Furthermore, an individual may not be subjected to any additional restrictions beyond those required for custodial and rehabilitative purposes, which must necessarily stem from a judicial sentence (Article 13 of Italian Constitution). Thus, it is the judge who is tasked with imposing punishment, through a logical and legal reasoning process, supported by the evidence and legal norms applied (Article 546 of Italian Code of Criminal Law Procedure).

Within this multi-level legal framework, the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data, ETS No. 108, and its subsequent amendments, including Protocol CETS No. 223, set out guidelines for the protection of data in the digital age, particularly concerning the safeguarding of individuals during data processing, the protection of digital identity, and facial recognition technologies. The increasing use of algorithmic decision-making tools in criminal justice must be balanced against the need to prevent inhuman or degrading treatment. This principle -reaffirmed in the Recommendation- is grounded in the European Convention for the Prevention of Torture and Inhuman or Degrading Treatment or Punishment, ETS No. 126, and the standards developed by the European Committee for the Prevention of Torture. This balance is essential to ensure that technological innovations serve to improve the system rather than exacerbate existing injustices or introduce new forms of discrimination (ESPOSITO, 2020).

The Council of Europe has also adopted the Framework Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law, CETS No. 225, which establishes key principles for the responsible use of AI technologies. In this context, Recommendation CM/Rec(2024)5 emphasizes the importance of developing clear standards for the use of AI decision-making in prison and probation services, ensuring that such tools are employed not only for security purposes, but also for the rehabilitation and social reintegration of individuals in conflict with the law[2].

While the Committee of Ministers recognizes that the implementation of AI in penitentiary systems can improve inmate security and management, it simultaneously underscores that algorithmic intelligence must be subject to strict controls to prevent detriment consequences in the light of protection of such fundamental rights. For example, the improper use of biometric surveillance or predictive algorithms could result in discrimination or violations of privacy if not adequately regulated and applied. It becomes clear, therefore, that digital technologies must be designed and deployed in a way that supports, rather than replaces, prison staff in

[2]CM/Rec(2024)5, available at: https://search.coe.int/cm#{%22CoEIdentifier%22:[%220900001680b1d0e4%22],%2
«Drawing attention to the need for these tools to be commissioned for design, development and maintenance to carefully selected and vetted private companies which work in close co-operation with the prison and probation services. These companies should be made aware that high ethical standards and principles and strict professional rules should be respected, and that the main objectives should be community safety and rehabilitating offenders, not making profits;

Emphasising therefore that it is critical to develop rapidly, to regularly review and, if necessary, revise the principles and standards which should guide the prison and probation services of its member States when using AI and related digital technologies in order to preserve high ethical and professional standards».

their daily duties. In this system of safeguards, private developers and/or suppliers of such technologies cannot be excluded: they, too, must adhere to and comply with stringent ethical standards to ensure that profit motives do not override the core objectives of security and rehabilitation. The principle of personality -understood as the humanization of punishment- remain central, ensuring that the use of AI does not lead to a purely automated or impersonal management of individuals in custody (PIPARO, 2025).

Finally, the research and development of AI technologies must be adequately funded and continuously monitored to evaluate the effectiveness of the tools employed. Transparency and accountability must be foundational pillars of their use, supported by periodic reviews and the establishment of redress mechanisms for individuals subject to decisions made by automated systems (PIPARO, 2023).

The "jail equilibrium": balancing efficient security and individualized correction

Correctional facilities often house violent inmates: in Italy alone, as of 2024, there were 27,382 detainees convicted of crimes against the person. It is no surprise, then, that prisons are frequently labeled as inherently violent spaces. This phenomenon is not limited to Italy; for instance, up to March 2019, England and Wales recorded 34,425 assault incidents in prisons, 10,311 of which were against staff. For this reason, scholars have examined the consequences of increasing prison violence, both for inmates, where *«for incarcerated individuals, victimization while in custody can cause further trauma, stifle rehabilitation efforts, and reduce cooperation with correctional staff due to damaged perceptions of procedural justice and legitimacy»* (DODD, ANTROBUS, SYDES, 2020; BEIJERSBERGEN, DIRKZWAGER, EICHELSHEIM, 2015).

As the literature points out:

> Robotic monitoring systems, three-dimensional cameras, and sensors designed to detect irregular or inappropriate behavior are already being tested in correctional facilities in South Korea and Hong KongIn the Chinese prison of Yancheng, an example of a so-called "smart jail", there is a continuous surveillance system for inmates, implemented through a network of sensors and cameras connected to a central "brain" capable of uninterruptedly monitoring each person and alerting staff in case of suspicious behavior. The prison system in Singapore is also pursuing a gradual process of creating "contactless"

> prisons, within which security is managed by automatons using tools known as "Avatar[3]" and "Vadar[4]" (CANESCHI, 2025, p. 264).

To counter these risks, particularly for prison staff, smart prisons employing automated surveillance and monitoring systems based on artificial intelligence and robotics can reduce the need for physical interaction between officers and inmates, thus ensuring a safer work environment. In South Korea, "guardian robots" equipped with 3D cameras, sensors, and pattern recognition algorithms have been tested to detect irregular or inappropriate behavior (CARBONE, 2011). In Hong Kong, autonomous patrol robots are supervised remotely by an officer. Additionally, robots are used to detect drugs in inmates' feces. Combined with advanced CCTV, facial recognition technologies, video analytics, and tracking wristbands, prison authorities are moving toward a hyper-secure, highly monitored system with maximum operational efficiency (JEFFERSON, 2020). In China, prison design is moving toward a form of "robotocracy", where AI controls the facilities. Yancheng's smart prison offers continuous, widespread surveillance through a network of AI-connected sensors and cameras (CHEN, 2019).

The Singapore Prison Service is implementing a "guardless prison" model, combining various smart technologies to improve operational efficiency and reduce the need for staff. Five key technologies have been reported as being tested or integrated. The first ones are called AVATAR and VADAR, which alert authorities to aggressive or anomalous activity. A different implementation consists of facial recognition cameras in cells for automated inmate checks. Another key implementation consists of wristbands with communication chips enabling cashless purchases and tracking rehabilitation program participation. Shared tablets, known as the Digitisation of Inmate Rehabilitation and Corrections Tool, which offer self-learning and educational apps are applied to replace printed materials and reducing contraband. Finally, iKiosk, allow inmates to perform administrative tasks and access entitlements independently (MCKAY, 2022).

In the Netherlands, a high-tech correctional facility has introduced RFID (Radio-Frequency Identification) wristbands as part of an integrated security management system that automates cell doors and monitors inmate movements and locations. Emotion recognition software also analyzes conversations and conflicts among inmates. The RFID chip, embedded in tamper-resistant wrist or ankle bands, identifies and tracks prisoners, verifies adherence to daily schedules, confirms presence

[3]Advanced Video Analytics to Detect Aggression
[4]Video Analytics to Detect Abnormal Behaviour

in secure zones, and provides evidence of misconduct. These technologies are believed to be more efficient than human monitoring and free up staff to focus more on rehabilitation. According to Leighton, RFID use in Japanese prisons allows inmates to move without escorts while remaining under intense surveillance, promoting autonomy and self-discipline (LEIGHTON, 2014).

Risk assessment tools (RATs) are a specific type of algorithm used to evaluate risk: they are a great example of algorithmic decision-making. Their primary purpose is to predict whether a person involved in the criminal justice system is likely to be rearrested (recidivism) or fail to appear in court after release. In recent years, the U.S. Goverment has developed COMPAS[5]: an algorithm capable of predicting recidivism risk. This tool has already been integrated into the judicial systems of various states to support judges in sentencing and risk assessment. Over one million individuals have been evaluated with the help of COMPAS, which is based on more than one hundred variables, including education level, personal history, neighborhood of residence, and age, in addition to legal data drawn from similar past cases (PARDO, 2018).

One of the most controversial cases involving COMPAS is the well-known *Loomis* case. Loomis, a U.S. citizen, was convicted after fleeing a police checkpoint in a stolen vehicle. Following a COMPAS assessment, he was sentenced to seven years in prison due to the high risk of recidivism indicated by the algorithm. Loomis appealed the sentence, arguing that the developer of the algorithm did not allow access to its internal workings, effectively creating a "black box algorithm" that made it impossible to mount an adequate defense. His lawyers also challenged COMPAS for potentially considering discriminatory factors, such as race. Nevertheless, the Wisconsin Supreme Court upheld the sentence, stating that COMPAS was only one of several factors used in the decision-making process[6] (PIPARO, 2025).

As already pointed out (PIPARO, 2025), The apparent reliability and neutrality promised by algorithmic decision-making can appear highly attractive to policymakers, particularly in countries like Italy, where legal reforms are increasingly shaped by appeals to efficiency and objectivity. However, this apparent objectivity

[5]Correctional Offender Management Profiling for Alternative Sanctions

[6]Supreme Court of Winsconsin, Winsconsin c. Loomis, Judgement, 5.04.2016: *«A recent analysis of COMPAS's recidivism scores based upon data from 10,000 criminal defendants in Broward County, Florida, concluded that black defendants "were far more likely than white defendants to be incorrectly judged to be at a higher risk of recidivism". Likewise, white defendants were more likely than black defendants to be incorrectly flagged as low risk. Although Northpointe disputes this analysis, this study and others raise concerns regarding how a COMPAS assessment's risk factors correlate with race. Additional concerns are raised about the need to closely monitor risk assessment tools for accuracy. At least one commentator has explained that in order to remain accurate, risk assessment tools "must be constantly re-normed for changing populations and subpopulations»*.

masks a deeper concern: these systems often replicate and institutionalize the very biases they purport to eliminate. While risk assessment algorithms like COMPAS are marketed as more accurate than human judgment, their reliance on statistical generalizations means that they inevitably categorize individuals into "risk classes" based on shared traits rather than individualized behavior. The resulting assessments are not truly about the person, but about the group to which they are statistically assigned. This practice risks reinforcing systemic inequalities, especially when variables correlated with social disadvantage -such as race, neighborhood, employment history, or family structure- are embedded within the algorithm's logic. Despite efforts to incorporate only risk-relevant variables, the selection and weighting of factors can still introduce bias, especially when socio-economic status or ethnicity function as proxies for risk (HOLSINGER, 2018).

As we already stated *supra,* in modern Constitutional legal system the primary goals of punishment necessarily include correction and social rehabilitation, as enshrined in Article 25, paragraph 2 of the Spanish Constitution and Article 27, paragraph 3 of the Italian Constitution. In line with these principles, artificial intelligence decision-making is also implemented in tools aimed at rehabilitating certain individuals. Therefore, while many of the smart prison technologies previously discussed serve disciplinary, incapacitating, and surveillance purposes, the technologies addressed in this section may instead promote positive change, rehabilitation, and the transition of inmates to life after release (PIPARO, BLAZEK, 2024).

One example of algorithmic decision-making implementation is found inside the work of Japanese companies Harumi Designs and Trottla, which have developed robotic dolls designed to resemble children and infants, specifically intended for the re-education of individuals convicted of sexual offenses against minors. According to the manufacturers, the purpose of these creations is to prevent future abuse of real children by providing an alternative means of interaction. In countries like Japan and China, these devices are now integrated with artificial intelligence algorithms, enabling them to simulate emotional responses such as crying, screaming, or expressions of fear, thereby mimicking interactions with real children (COX, 2018).

However, these robotic dolls raise numerous ethical and psychological controversies. Psychologist Patrice Renaud from the University of Montreal has suggested that, while some individuals might exercise self-control and use these robots to manage their urges, others might be encouraged by the experience to seek out real children. Similarly, Jack Balkin, writing for the *California Law Review Circuit*, argues that the law might, in certain contexts, treat robots as persons for practical or social policy reasons. For instance, he proposes extending some legal protections to

robots to prevent abuse, based on the logic that those who mistreat artificial entities may also be prone to mistreating people, children, or animals (BALKIN, 2015).

In the United States, the so-called CREEPER[7] Act of 2017 aims to ban the manufacture and sale of child-like sex dolls. New York Congressman Dan Donovan, who sponsored the CREEPER Act, presented alarming statistics at a press conference: according to British law enforcement, 85% of the 128 individuals found in possession of these dolls were also found to possess child pornography.

In smart prisons, the introduction of personal digital communication devices could also support the rehabilitative process: audiovisual links/videoconferencing, digital tablets, iKiosks, and in-cell computers represent a significant improvement over traditional shared landline phones in prison wings and outdated postal services, already obsolete even outside the prison context. These personal digital devices offer inmates new opportunities to maintain family ties, protect their right to legal defense (by accessing justice services), participate in educational programs, and develop digital literacy skills. In such context, public opinion has often criticized the provision of what are perceived as "luxuries" to inmates, aligning with a punitive vision of imprisonment that envisions prisons as harsh, austere, and comfortless environments. Other criticisms have been framed around safety concerns: for example, some have raised the concrete risk that victims or witnesses may be contacted and harassed, or that inmates might continue to manage criminal activities from within prison. Society tends to view prison walls as barriers to social interaction, and such digital devices challenge that notion of impermeability (MCKAY, 2016).

The rehabilitative benefits of limited connectivity for inmates are gradually gaining recognition in many countries. Some scholars argue that providing correctional AI-devices (which orient their correctional programmes through decision-making lgorithms) to inmates could help bridge the digital divide and promote access to educational and rehabilitative programs. If the digital content made available is not only relevant to the exercise of legal defense and the protection of inmates' personal rights but also of high quality, personal devices can be excellent tools for training on violence prevention, substance abuse, literacy, vocational training, higher education, parenting skills, and connecting with housing, healthcare, financial, and employment support services post-release. Others argue that these rehabilitation models -if and when effective- must be integrated with digital reintegration and inclusion models, enabling the state and prison administration to offer both online and offline rehabilitative solutions. From this perspective, smart prisons can be seen not as mere containment tools but as mechanisms for fostering autonomy and responsibility in

[7]Curbing Realistic Exploitative Electronic Pedophilic Robots Act

inmates in preparation for their reintegration into society (REISDORF, RIKARD, 2018). Many countries are currently implementing the features of smart prisons, and personal devices have been evaluated in several nations. According to scholars who have studied the Smart Prison Project tested in Finnish prisons, such tools may promote rehabilitation and reduce recidivism by developing inmates' autonomy, self-esteem, and life-management skills, thereby improving the overall prison culture. Their study concludes: thathe more digital services can be made independent and interactive, and the more the 'walls' of the prison can become invisible through digital means without compromising security, the better it will be for rehabilitation (LINDSTRÖM, PUOLAKKA, 2020). Personal digital technologies incentivize good behavior, as the threat of losing access to the devices acts as a deterrent against disciplinary infractions. Overall, studies have found that these technologies improve psychological well-being, increase autonomy, and reduce tensions within the prison environment (PALMER, HATCHER, TONKIN, 2020).

Nevertheless, there are critical issues. On one hand, digital devices can help optimize time spent in the cell; on the other, they might encourage inmates to remain isolated for longer periods. Others note that many in-cell technologies -even radios and televisions- act as "electronic babysitters". However, prison authorities may appreciate their calming effect, viewing them as efficient tools for managing inmates or, cynically, as "opium for the masses" (KHARZAEE, UNSWORTH, 2012).

There are also concerns about data collection, as digital devices may be used for more efficient inmate surveillance. Another pressing issue is the commercialization of prison communications, with exorbitant prices imposed on a literally captive audience: inmates and their families often pay disproportionately high rates for phone calls, video chats, and other paid services. Finally, a key concern is that such devices could fully replace in-person visits from family and legal counsel. A prison life devoid of human contact has already been observed during the COVID-19 pandemic and could lead to severely negative consequences, such as a worsening sense of isolation and further dehumanization (MCKAY, 2016).

Conclusion

Artificial Intelligence is now intimately associated with some of the most ethically contested aspects of modern governance: none more so than the administration of criminal justice. In the prison environment, where deprivation of liberty must never be conflated with erosion of dignity, Artificial Iintelligence decision-making algorithms offer both promise and threat. On the one hand, its potential to optimize

prison security, prevent violence, and minimize human error is undoubtedly persuasive. On the other, its potential to codify systemic injustice and depersonalize decision-making demands immediate ethical scrutiny. The correctional system, itself too frequently torn between institutional resistance and popular pressure for reform, is at a fork in the road: whether to allow intelligent algorithms to reinforce a mechanized approach to punishment, or to utilize its possibilities in the service of rehabilitation and the inherent rights enshrined in constitutions and human rights treaties.

Ensuring security in correctional facilities is a valid and necessary objective. In an environment rife with chaos or violence, no system can live up to its promise of rehabilitation. In this regard, algorithm-based surveillance technologies, such as biometric identification systems and behavior-detection cameras, have been successfully deployed in China, Hong Kong, and South Korea. By using sensors, 3D cameras, and pattern-recognition algorithms, these "smart prisons" are able to identify abnormalities like aggression, escape attempts, or self-harm in real time, potentially saving lives and safeguarding both staff and inmates.

RFID-enabled wristbands are used in Singapore and the Netherlands to monitor inmate compliance and movement, allowing for more fluid routines and reducing conflict. By relieving overworked staff and enabling them to concentrate more on human-centered rehabilitation efforts, these systems promise improved safety and resource efficiency. However, these advantages need to be carefully balanced against the dangers of intrusive monitoring, data misuse, and privacy degradation: risks that disproportionately impact marginalized groups already subjected to algorithmic bias and over-policing. Security cannot be used as an excuse for absolute control or as an excuse to disregard the constitutional requirement that prisoners be treated as the holders of unalienable rights rather than just as subjects of custody.

On the other side of the equation lies rehabilitation: the cornerstone of a humane justice system and, in jurisdictions like Italy, a constitutional requirement *ex* Art. 27, Italian Constitution. Here too, algorithmic decision-making offers compelling tools. Individualized rehabilitation plans, powered by machine learning, can analyze inmates' educational backgrounds, psychological profiles, and behavior patterns to tailor treatment programs and support reintegration. AI-based learning platforms, personal digital devices, and videoconferencing technologies are being trialed in Europe and North America, offering incarcerated individuals access to education, therapy, and legal services from within their cells. These tools help inmates maintain family connections, practice digital literacy, and begin rebuilding their lives, addressing some of the most persistent causes of recidivism: isolation,

skill deficits, and social disconnection.

Crucially, the use of such technologies has been shown to enhance psychological well-being and reduce tensions within prisons. In this sense, Algorithmic Intelligence is not a threat to dignity, but a potential instrument of restoration. However, these positive applications are not automatic. They require thoughtful design, meaningful oversight, and consistent respect for human agency. The risk, otherwise, is that "cold", algorithmic decision-making devices will substitute "warm" human decision-making and empathy: a slick, data-driven interface that displaces the human relationships essential to moral development and social reintegration. When such tools are used to determine sentencing, parole, or even prison privileges, they can have life-altering consequences, often with no way for the individual to understand or challenge the algorithmic reasoning behind them. These systems risk converting justice from a deliberative act into a computational output. In doing so, they invert the burden of proof: no longer is the state required to demonstrate guilt or danger; instead, the algorithm assigns risk, and the human being must live with its verdict.

Yet it does not have to be this way. A more ethical vision of Artificial Intelligence in correctional systems is both possible and necessary. One that centers on transparency, human oversight, and the principle of proportionality. This would require that algoithms serve as a tool for support, not judgment; for personalization, not categorization. Every use of algorithms in prison should be subject to robust human review, with clear avenues for contestation and appeal. The systems must be explainable, auditable, and governed by ethical standards that prioritize the person over the prediction. Furthermore, their development should be co-designed with stakeholders -including inmates, correctional officers, lawyers, and ethicists- to ensure the technology reflects lived experience and respects fundamental rights. If incarceration is to serve a rehabilitative purpose, it cannot become a closed-loop system governed by inscrutable code and enforced through digital paternalism.

Ultimately, the promise of algorithmic decision-making devices in correctional settings lies not in its capacity to surveil more efficiently or punish more accurately, but in its potential to humanize institutional routines, personalize treatment pathways, and restore the dignity of those society has cast aside. A prison equipped with intelligent systems must not be a prison emptied of humanity. Instead, it must be a place where technology expands the moral imagination: where even in confinement, individuals can be seen, heard, and supported as they strive toward change. To that end, legislators, technologists, and justice professionals must commit not only to technological advancement but to ethical stewardship. They must resist the

allure of efficiency when it comes at the cost of fairness and uphold the principle that every person, regardless of their past, retains the right to a future. That is the challenge and the promise of Artificial Intelligence in correctional justice: not to replace the human touch, but to make space for it, even behind prison walls.

References

BALKIN, J. M. The path of robotics law. **CLRC**, v. 6, n. 1, p. 58, 2015.

BEIJERSBERGEN, K. A.; DIRKZWAGER, A. J.; EICHELSHEIM, V. I.; VAN DER LAAN, P. H.; NIEUWBEERTA, P. Procedural justice, anger, and prisoners' misconduct: A longitudinal study. **Criminal Justice and Behavior**, v. 42, n. 2, p. 196–218, 2015.

BODEN, M. A. Intelligenza artificiale. In: KHALILI (a cura di). *Il futuro che verrà*, p. 133, 2018.

CANESCHI, A. Intelligenza artificiale e sistema penitenziario. **Rivista Italiana di Diritto e Procedura Penale**, v. 1, n. 1, p. 251–269, 2024.

CARBONE, A. South Korea rolls out robotic prison guards. **Time**, 27 Nov. 2011.

CHEN, S. No escape? Chinese VIP jail puts AI monitors in every cell 'to make prison breaks impossible'. **South China Morning Post**, 1 Apr. 2019.

CIAPPI, S., COLUCCIA, A. Giustizia criminale. Milano: Franco Angeli, 1997.

COX, B. Would child sex robots stop pedophilia or promote it? **NBCNews**, 2018.

DIGARD, L.; DIZEREGA, M.; YARONI, A.; RINALDI, J. A new role for technology? Implementing video visitation in prison. **Vera Institute of Justice**, New York, 2016.

DODD, V. A.; ANTROBUS, E.; SYDES, M. Cameras in corrections: Exploring the views of correctional officers on the introduction of body-worn cameras in prisons. **Criminal Justice and Behavior**, v. 47, n. 9, p. 1190–1208, 2020.

ESPOSITO, M. Ruolo del corpo di polizia penitenziaria nell'osservazione scientifica della personalità. **Discrimen**, v. 1, n. 1, p. 1–18, 2020.

HOLSINGER, A. A rejoinder to Dressel and Farid: New study finds computer algorithm is more accurate than humans at predicting arrest and as good as a group of 20 lay experts. 2018, p. 50-55.

JEFFERSON, B. J. Digitize and punish: Racial criminalization in the digital age. Minneapolis: University of Minnesota Press, 2020.

JOHNSON, S.; HAIL-JARES, K. Prisons and technology: General lessons from the American context. In: JEWKES, Y.; BENNET, J.; CREWE, B. (edited by). Handbook on Prisons. London, 2016.

KHARZAEE, H.; UNSWORTH, K. Social media: The new opiate of the masses? **International Review of Information Ethics**, v. 18, n. 1, p. 49–59,

2012.

KOF, L.; BOERS, M.; KOSTERS, W.; PUTTEN, P.; POEL, M. Artificial intelligence: Definition, trends, techniques and cases. **Knowledge for Sustainable Development: An Insight into the Encyclopedia of Life Support Systems**, v. 1, p. 1096-2002.

LEIGHTON, P. 'A model prison for the next 50 years': The high-tech, public-private Shimane Asahi Rehabilitation Center. **Justice Policy Journal**, v. 11, n. 1, p. 1–16, 2014.

LINDSTRÖM, M.; PUOLAKKA, M. Smart prison: The preliminary development process of digital self-services in Finnish prisons. **International Corrections & Prisons Association**, 28 July 2020.

MARSILI, M.; WRÓBLEWSKA-JACHNA, J. Digital revolution and artificial intelligence as challenges for today. **Media Society**, v. 20, n. 1, p. 19-30, 2024.

MCKAY, C. The carceral automaton: Digital prisons and technologies of detention. **International Journal for Crime, Justice and Social Democracy**, v. 11, n. 1, p. 101-113, 2022.

MCKAY, C. Video links from prison: Permeability and the carceral world. **International Journal for Crime, Justice and Social Democracy**, v. 5, n. 1, p. 21–37, 2016.

MULCAHY, J. The evolution of punishment and rehabilitation. **Irish Criminal Law Journal**, v. 29, n. 3, p. 39–49, 2019. Available at: `www.researchgate.net/publication/333642240`.

NIEDERMAN, M. The institutional life of algorithmic risk assessment. **Berkeley Technology Law Journal**, v. 34, n. 3, p. 705-744, 2019.

NICOTRA, I. Pena e reinserimento sociale ad un anno dalla sentenza Torregiani. **Diritto penitenziario e costituzione**, v. 1, p. 1-20, 2014.

NJOKU, F.O.C., H. L. A. Hart on philosophical and legal conception of punishment. **International journal of theology and reformed tradition**, 4, p. 220-235, 2012.

NOTARO, L. Algoritmos predictivos y Justicia Penal desde una perspectiva Italiana Europea. In: MASSARO, A. PERIS REIRA, J. **Derecho Penal, Inteligencia Artificial y Neurociencias**, p. 192–193, 2023.

PALMER, E., HATCHER, R. M., TONKIN, M. Evaluation of digital technology in prisons. **UK Ministry of Justice**, London, 2020.

PARDO, P. Por qué la inteligencia artificial necesita clases de ética: "Hemos creado un Frankenstein". **El Mundo**, 2018.

PIPARO, C., Criminal liability models and criminal participation in the digital environment: a modern challenge in the perspective of Italian constitutionalism. **Collected Papers of the Faculty of Law**, Novi Sad, p. 1357-1378, 2023.

PIPARO, C., BLAZEK, R. Criminal algorithms and their punishment in modern constitutionalism, **Bratislava Law Review**, Bratislava, p. 200-222, 2024.

PIPARO, C. La controvertida relación entre la inteligencia artificial y la democracia representativa. **Nuovi autoritarismi e democrazie**, 6, 1, Milano, 2024.

PIPARO, C. Machina delinquere potest? (ph.D. thesis). **University of Udine:** Udine, p. 1-180, 2025.

REISDORF, B. C.; RIKARD, R. V. Digital rehabilitation: A model of reentry into the digital age. **American Behavioral Scientist**, v. 62, n. 9, p. 1273–1290, 2018.

SANTONI DE SIO, F., MECACCI, G. Four responsibility gaps with artificial intelligence: Why they matter and how to address them. **Philosophy & Technology**, v. 34, n. 1, p. 1057–1084.

SCHANK, *What's IA, Anyway?*, in *IA Magazine*, Winter 8(4), p. 59-65, 1987.

SENECA. *De ira*, I, 19.

SILVESTRI, C. Il senso della pena. Ad un anno dalla sentenza Torregiani della CEDU. *Convegno presso il Carcere di Rebibbia*, Roma, 28 maggio 2014.

SUPREME COURT OF WISCONSIN. Wisconsin v. Loomis. Judgement, 5 Apr. 2016.

VILLASENOR, J. FOGGO, V. Artificial intelligence, due process and criminal sentencing. **Michigan State Law Review**, p. 295-354, 2020.

YANOFSKY, N. Towards a definition of an algorithm. Journal of Logic and Computation. Cornell University: Ithaca. p. 2, 2010.

ZIENTS, J. Delivering on broadband opportunity. **The White House**, 21 Sept. 2015.

ZINGALES, N. Risk assessment: una nuova sfida per la giustizia penale? **Diritto Penale e Uomo**, v. 1, n. 1, p. 1-22, 2021.

Acknowledgment and conflicts of interest

The authors declare that they have no conflicts of interest with respect to the research, authorship, and/or publication of this article. Any errors or omissions are our own.