

Artificial Intelligence and the System of Responses to Crime: From Situational Prevention to the Engineering of Multilayered Sanctions

1. Mohammad Saeed Nasiraei[✉]: Department of Criminal Law and Criminology, Shi.C., Islamic Azad University, Shiraz, Iran

2. Alireza Salehi^{*}: Department of Criminal Law and Criminology, Shi.C., Islamic Azad University, Shiraz, Iran

3. Amir Iravanian[✉]: Department of Criminal Law and Criminology, Shi.C., Islamic Azad University, Shiraz, Iran

*Correspondence: dralirezasalehi@iau.ir

Abstract

This article examines the role of artificial intelligence in the system of responses to crime, focusing on the continuum from situational and person-oriented prevention to the design of multilayered sanctions. The study proceeds from the distinction between the cognitive and analytical capacities of artificial intelligence and the normative authority required for legally significant decisions. In situational prevention, artificial intelligence can support dynamic analysis of spatial, temporal, and environmental risk patterns, improve resource allocation, and facilitate continuous evaluation of preventive interventions. However, predictive outputs must remain distinct from automatic intervention because crime data may reflect previous patterns of surveillance, reporting, and enforcement and may generate self-reinforcing feedback loops. In person-oriented prevention, the legitimate function of artificial intelligence lies not in identifying presumed future offenders, but in analyzing risk factors, protective factors, vulnerabilities, and modifiable needs in order to support proportionate and non-stigmatizing interventions. The article further examines the transition from prevention to reactive criminal policy and argues that artificial intelligence may assist in comparing criminalization, administrative regulation, and non-punitive alternatives without independently determining the legal status of conduct. In the field of sanctions, artificial intelligence can contribute to the engineering of multilayered responses by comparing criminal, administrative, disciplinary, professional, rehabilitative, medical, psychological, and social measures and by identifying less intrusive alternatives where appropriate. The central argument is that artificial intelligence should operate as a bounded decision-support infrastructure rather than an autonomous decision-maker. Its legitimate value lies in strengthening evidence-based analysis, proportionality assessment, monitoring, and policy revision while preserving human oversight, legal responsibility, contestability, and fundamental rights. The article ultimately proposes a cyclical model of identification, analysis, intervention design, implementation, evaluation, and revision for the responsible integration of artificial intelligence into criminal policy.

Keywords: Artificial Intelligence; Criminal Policy; Situational Crime Prevention; Person-Oriented Prevention; Decision-Support Systems; Multilayered Sanctions; Proportionality; Algorithmic Risk Assessment

Received: 21 July 2026

Revised: 03 October 2026

Accepted: 10 October 2026

Initial Publication: 11 October 2026

Final Publication: 01 July 2027



Copyright: © 2027 by the authors. Published under the terms and conditions of Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) License.

Citation: Nasiraei, M. S., Salehi, A., & Iravanian, A. (2027). Artificial Intelligence and the System of Responses to Crime: From Situational Prevention to the Engineering of Multilayered Sanctions. *Legal Studies in Digital Age*, 6(7), 1-17.

1. Introduction

The growing incorporation of artificial intelligence into criminal justice and crime-control institutions requires a conceptual approach that extends beyond the conventional question of whether police agencies, courts, administrative authorities, or correctional institutions employ technologically sophisticated tools. Digitalization, automation, and artificial intelligence are not synonymous phenomena. A judicial institution may digitize its files without altering the substance of criminal-policy decision-making, while a comparatively limited algorithmic system may substantially affect that policy if its output influences how risks are identified, interventions are prioritized, sanctions are selected, or the consequences of governmental action are evaluated. Accordingly, the central issue is not the technological label attached to a system but the functional position that algorithmic inference occupies within the architecture of social responses to crime. Contemporary scholarship on artificial intelligence and criminal policymaking has increasingly emphasized this broader institutional dimension. Talebi Rostami considers artificial intelligence relevant to criminal policymaking precisely because its analytical capacities may affect policy formation and implementation rather than merely administrative efficiency (Talebi Rostami, 2025). Amirian Farsani likewise identifies new opportunities and challenges arising from the interaction between artificial intelligence and contemporary Iranian criminal justice policymaking, particularly where computational systems move from peripheral administrative assistance toward functions capable of influencing judicial and enforcement choices (Amirian Farsani, 2025). The subject therefore concerns the reconfiguration of the informational and decision-making environment within which criminal policy operates.

Criminal policy itself cannot be reduced to punishment after an offense has occurred. A modern system of responses to crime encompasses preventive strategies, decisions concerning the legal characterization of harmful conduct, the choice between penal and non-penal regulatory mechanisms, the design and implementation of sanctions, and the subsequent evaluation of their consequences. Artificial intelligence potentially enters each of these stages in a different manner and with a different intensity. In preventive policy, its principal value lies in recognizing patterns, identifying situational concentrations of risk, integrating heterogeneous information, and assisting authorities in deciding where or toward what problem preventive resources should be directed. Ehsanpour describes the capacity of artificial intelligence to process crime-related information and identify patterns as one of its significant contributions to crime prevention (Ehsanpour, 2025), while Shojai Langari connects artificial intelligence and data mining with the identification of crime patterns and the more targeted organization of preventive measures (Shojai Langari, 2024). Najafi and Mousavifar similarly emphasize the expanding relevance of intelligent analytical systems to contemporary crime-prevention strategies (Najafi & Mousavifar, 2025). Yet the movement from analysis to intervention raises a crucial question: even where an algorithm estimates a risk accurately, does that estimate itself justify surveillance, police deployment, restriction, treatment, punishment, or any other governmental response? This article proceeds from the premise that it does not. Prediction and intervention constitute analytically distinct stages, and the legitimacy of the latter cannot be derived automatically from the technical quality of the former.

The same distinction becomes even more important when artificial intelligence is used in relation to individuals rather than places, environments, or temporal patterns. Algorithmic risk assessment may combine a large number of variables and identify correlations that would be difficult to process through unaided human judgment. However, the capacity to classify a person as presenting an elevated statistical probability of a future event creates both methodological and normative difficulties. Neil and Zanger-Tishler demonstrate that criminal risk assessment can reproduce disparities where arrest or similar institutional outcomes are used as proxies for underlying criminal behavior, because such variables may embody pre-existing patterns in policing and enforcement (Neil & Zanger-Tishler, 2025). Stevenson and Doleac further show that the practical effects of algorithmic risk assessment depend not only on the statistical characteristics of a model but also on how human decision-makers interpret and use its outputs (Stevenson & Doleac, 2024). The appropriate question is therefore not whether an algorithm can replace human judgment, but how algorithmic analysis should be positioned within a decision architecture in which legal authority, professional responsibility, evidentiary interpretation, and respect for rights remain institutionally attributable to human actors. Parsa's analysis of human oversight in European Union and Iranian law reinforces the importance of preserving meaningful human supervision when artificial intelligence contributes to decisions carrying significant legal consequences (Parsa, 2024).

This issue becomes still more complex when criminal policy moves from prevention to reaction. Once conduct has occurred or a regulatory violation has been established, public authorities must decide not merely whether to respond, but how. The conventional tendency to equate criminal policy with criminal punishment obscures the range of possible responses available in contemporary regulatory systems. A single harmful event may potentially attract criminal sanctions, administrative penalties, disciplinary consequences, professional restrictions, rehabilitative interventions, medical or psychosocial measures, restitutionary requirements, or combinations of these mechanisms. Evidence-based crime and justice policy has increasingly emphasized the importance of evaluating interventions according to their actual effects rather than assuming that greater severity necessarily produces greater public safety (Welsh & Mears, 2024). Mears and Cochran argue for a policy orientation that substitutes systematic evidence-based decision-making for reflexive reliance on incarceration and other high-intensity penal measures (Mears & Cochran, 2024). Artificial intelligence adds a new analytical capacity to this environment because it can potentially integrate information about offense characteristics, intervention outcomes, risk factors, institutional capacity, previous responses, and patterns of compliance. Its legitimate contribution, however, lies in improving comparative analysis among legally available responses rather than independently determining which sanction an individual deserves.

The emergence of artificial intelligence in legislative and regulatory design strengthens the need to separate cognitive assistance from normative authority. Hill, Waddington, and Qiu explore the possibility that artificial intelligence can support legislative processes by organizing complex information, identifying inconsistencies, and assisting the optimization of legal texts (Hill et al., 2025). Gangi Chiodo's examination of generative AI in parliamentary experience similarly illustrates the growing proximity of artificial intelligence to institutional processes traditionally regarded as exercises of human legislative judgment (Gangi Chiodo, 2026). Such developments create meaningful opportunities, but they do not dissolve the distinction between identifying evidence and making normative choices. A system may identify that a regulatory response has produced limited compliance, that two offenses are treated inconsistently, or that a particular sanction disproportionately affects certain populations; it cannot derive from those observations alone the legal judgment that conduct ought to be criminalized, decriminalized, punished more severely, or redirected toward non-penal regulation. Salehnejad Behrestaghi and colleagues discuss the possible contribution of artificial intelligence to the analysis of emerging forms of offending and criminalization processes (Salehnejad Behrestaghi et al., 2024), but criminalization remains a normative institutional act requiring justification independently of technological prediction.

The objective of this article is therefore to develop an integrated account of artificial intelligence across the full continuum of criminal-policy response, beginning with situational prevention and person-oriented preventive intervention and extending to reactive policy and the engineering of multilayered sanctions. The analysis is organized around a central proposition: the appropriate role of artificial intelligence is to strengthen the cognitive, analytical, comparative, and evaluative capacities of legitimate human decision-makers, not to convert statistical prediction into autonomous authority over rights and liberties. This proposition requires attention to the quality of data, the distinction between correlation and causation, the distribution of algorithmic errors, the possibility of feedback loops, the effects of human reliance on algorithmic recommendations, proportionality, contestability, and the availability of less intrusive alternatives. O'Brien's examination of naturally occurring big data in evidence-based policy demonstrates both the analytical promise of large-scale administrative information and the difficulties arising from data generated through institutional processes rather than controlled research environments (O'Brien, 2024). Welsh and Mears likewise stress that policy conclusions require careful evaluation of the quality and applicability of evidence rather than mere accumulation of information (Welsh et al., 2024). These concerns are fundamental to any attempt to transform artificial intelligence from an attractive technological innovation into a legitimate component of criminal policy.

The argument developed here accordingly rejects two opposite simplifications. The first is technological optimism, according to which greater predictive accuracy naturally produces better criminal policy. The second is technological rejectionism, according to which the risks associated with artificial intelligence eliminate any legitimate role for computational analysis in prevention or sanction design. A more defensible approach is functional and graduated. The legitimacy of a particular application should depend on what the system analyzes, what output it produces, how that output affects a subsequent decision, the severity of the consequences for affected persons, the availability of human review, and the evidentiary basis for believing that its use improves outcomes compared with feasible alternatives. Farajpour and colleagues emphasize ethical

requirements surrounding the adoption of the European Union's regulatory approach to artificial intelligence (Farajpour et al., 2025), while Nouralivand describes the risk-oriented regulatory architecture of the European Union Artificial Intelligence Act as an important development in AI governance (Nouralivand, 2024). Such risk-sensitive approaches are particularly relevant to criminal policy because the legal significance of an algorithm cannot be determined independently of the coercive power to which its output is connected. The following sections therefore conceptualize artificial intelligence not as an autonomous source of criminal-policy decisions but as an instrument whose legitimacy depends on its integration into a human-governed, evidence-sensitive, rights-constrained system of prevention, response, evaluation, and revision.

2. Artificial Intelligence and Preventive Responses to Crime

Situational crime prevention provides one of the clearest settings in which artificial intelligence can contribute to criminal policy without necessarily making judgments about the character or culpability of particular individuals. The foundational logic of situational prevention is that criminal events are shaped not only by the motivations of potential offenders but also by the conditions under which opportunities for offending arise. From this perspective, prevention seeks to modify environments, opportunities, targets, access structures, patterns of guardianship, or other contextual conditions in ways that make offending more difficult, more detectable, less rewarding, or otherwise less attractive. Artificial intelligence does not alter this conceptual foundation; rather, it can expand the capacity to identify how opportunity structures vary across locations and periods. Ehsanpour discusses the use of artificial intelligence for analyzing crime-related data and identifying locations or circumstances associated with elevated risks (Ehsanpour, 2025). Najafi and Mousavifar similarly link intelligent analysis to the development of more responsive crime-prevention mechanisms (Najafi & Mousavifar, 2025). The transformation introduced by AI is therefore primarily epistemic: instead of relying exclusively on static crime maps, periodic statistics, or accumulated professional intuition, institutions can process larger streams of spatial, temporal, environmental, and incident-based data and update assessments more frequently.

This transition can be understood as a movement from retrospective crime mapping toward dynamic estimation of situational risk. A conventional map identifies where recorded incidents have occurred, whereas a predictive model may attempt to infer where similar events are comparatively more likely to occur in a future period. Shojai Langari associates data-mining techniques with the capacity to identify hidden patterns and support more targeted preventive deployment (Shojai Langari, 2024). This capacity can be valuable because crime is rarely distributed evenly through space and time, and limited preventive resources cannot be allocated uniformly. Yet the distinction between recorded crime and underlying crime must remain central. Administrative crime data are generated through reporting practices, police presence, institutional priorities, detection capacities, classification systems, and patterns of citizen cooperation. O'Brien notes that naturally occurring administrative data can provide scale and realism unavailable in many conventional research designs, while also presenting substantial problems of selection, measurement, and interpretation because the data were not necessarily generated for research purposes (O'Brien, 2024). A location containing more recorded incidents may therefore reflect genuinely elevated offending, greater surveillance, greater reporting, or some combination of these mechanisms. A predictive map must consequently be treated as an estimate based on institutional data rather than as a direct representation of an objective criminal reality.

This limitation becomes especially significant where predictive analysis influences subsequent policing. If an area is identified as high risk and consequently receives increased police attention, the enhanced presence may generate additional detections and recorded incidents. Those newly generated records can then enter subsequent model training and reinforce the classification of the same area as dangerous. Algorithmic feedback is therefore not merely a technical problem; it arises from the interaction between prediction and institutional behavior. Neil and Zanger-Tishler demonstrate the broader significance of this problem in risk assessment by showing how criminal justice data may embed disparities associated with arrest and enforcement practices (Neil & Zanger-Tishler, 2025). The same logic applies spatially: data produced by previous enforcement may become inputs that justify future enforcement, creating the appearance that the algorithm is continuously confirming its own predictions. For this reason, AI-assisted situational prevention should distinguish variables that approximate underlying victimization or offending from those heavily dependent on the intensity of previous police activity. More fundamentally, a high-risk classification should never function as an automatic trigger for coercive intervention.

The distinction between prediction and intervention is therefore indispensable. A model may estimate an elevated probability of vehicle theft in a parking area during certain hours, but that output does not specify which preventive measure ought to follow. Authorities might improve lighting, alter access controls, strengthen guardianship, redesign the physical environment, issue targeted warnings, deploy personnel, or combine several measures. The algorithm cannot infer the normative desirability of these alternatives merely from its prediction. The relevant choice depends on evidence of effectiveness, cost, intrusiveness, potential displacement, effects on legitimate users of the space, and the legal authority available to the institution. Welsh, Zane, and Mears emphasize that evidence-based crime and justice policy requires systematic attention to whether interventions actually achieve intended outcomes rather than relying on intuition or institutional tradition (Welsh & Mears, 2024). Welsh and Mears further underscore that evaluating research evidence requires consideration of methodological quality and the extent to which findings can support a specific policy inference (Welsh et al., 2024). Artificial intelligence should therefore contribute to the identification and comparison of options, not collapse the distinction between knowing where risk may be concentrated and determining how the state should respond.

Intelligent surveillance illustrates the same principle. Computer vision and other automated systems can transform surveillance technologies from passive recording devices into analytical tools capable of detecting objects, movements, entry into restricted areas, or patterns designated as anomalous. Shojai Langari discusses intelligent monitoring as one of the technological possibilities associated with AI-supported crime prevention (Shojai Langari, 2024). Yet technical detection accuracy and criminological effectiveness are different variables. A system can reliably detect a specified movement without demonstrating that its deployment reduces crime. It can also generate alerts about statistically unusual behavior that is entirely lawful. The broader the category of “suspicious” or “abnormal” behavior becomes, the greater the risk that situational prevention shifts from the management of opportunity to continuous scrutiny of persons. This distinction is especially important where the model's output can prompt police intervention. Parsa's emphasis on meaningful human oversight is relevant because a human reviewer must have genuine authority and adequate information to question an automated classification rather than merely approve it procedurally (Parsa, 2024). Human involvement is valuable only if it functions as substantive review.

The preventive value of artificial intelligence also depends on post-intervention evaluation. Situational interventions can reduce offending within a target area while shifting it geographically, temporally, or toward another target; alternatively, preventive benefits may extend beyond the immediate intervention area. A system designed only to identify where to intervene but not to monitor what happens afterward creates an incomplete policy cycle. AI can be especially useful in this second analytical role because it can compare changing distributions of incidents over time and identify patterns that would be missed by a narrow evaluation confined to the treated location. This reflects a broader evidence-based orientation in which implementation data are not merely administrative records but inputs for subsequent policy revision. Mears and Cochran argue that systematic evidence should guide choices among crime-control interventions and support movement away from indiscriminate reliance on punitive responses (Mears & Cochran, 2024). AI-assisted prevention should similarly be evaluated by whether it produces measurable improvements compared with simpler alternatives, not by the technological sophistication of the model itself.

Person-oriented prevention presents a more difficult problem because algorithmic analysis approaches the individual more closely and thereby increases the potential consequences of classification error. The legitimate objective of such systems should not be the identification of a presumed “future offender.” A more defensible function is the structured analysis of risk factors, protective factors, vulnerabilities, and needs that may be relevant to preventive support. Abdollahi discusses the ability of artificial intelligence to process multiple behavioral, social, and criminal-history variables in constructing criminal profiles (Abdollahi, 2026). That analytical capacity demonstrates the technical possibility of integrating heterogeneous information, but its preventive use requires a conceptual transformation: instead of asking which person is likely to commit an offense, policy should ask which modifiable conditions may increase exposure to offending or victimization and which interventions might address those conditions without imposing punitive consequences on the basis of predicted conduct.

The importance of this distinction is evident in early-intervention risk assessment. Sheppard and colleagues examined risk assessment instruments for young people who had not previously had formal contact with juvenile justice and evaluated their feasibility in prevention and early-intervention services (Sheppard et al., 2024). The study demonstrates that structured risk analysis can be considered before conventional post-adjudication contexts, but it simultaneously reveals a critical

methodological issue: the predicted outcome may be future contact with a justice institution rather than an independently measured occurrence of offending. Institutional contact is shaped by reporting, referral, detection, policing, school practices, and administrative choices. Consequently, a model that predicts institutional contact should not be described as though it predicts criminal behavior itself. This distinction has profound legal significance because the moral and regulatory consequences of labelling a child as likely to enter the justice system may be severe even when the model is accurately predicting institutional exposure rather than offending.

Algorithmic error must similarly be evaluated through the consequences attached to classification. False positives and false negatives are not purely statistical quantities. If a false positive leads only to an optional offer of educational support, its legal significance differs radically from a false positive that results in police surveillance, restricted movement, intensified monitoring, or placement on an enforcement list. Neil and Zanger-Tishler show why group differences in criminal justice data can complicate claims of neutrality in risk assessment (Neil & Zanger-Tishler, 2025). Stevenson and Doleac demonstrate that algorithmic recommendations are also filtered through human behavior, meaning that the same model may produce different institutional effects depending on how decision-makers respond to its scores (Stevenson & Doleac, 2024). The acceptable degree of uncertainty should therefore be inversely related to the coerciveness of the resulting measure. As the consequence becomes more restrictive, the evidentiary and procedural safeguards surrounding the prediction must become more demanding.

A defensible model of AI-assisted person-oriented prevention can thus be organized as a continuous cycle of assessment, intervention design, and outcome monitoring. During assessment, artificial intelligence can integrate relevant information, identify combinations of risk and protective factors, and indicate cases requiring closer professional evaluation. During intervention design, the objective should be to connect identified and modifiable needs to proportionate educational, familial, psychosocial, rehabilitative, or supportive services rather than convert a probability score into a punitive response. During evaluation, subsequent changes should be measured so that the intervention can be modified, reduced, or terminated when no longer justified. Ghasemi and colleagues' model of an intelligent judicial assistant for alternatives to imprisonment demonstrates the broader potential of AI to connect risk-related analysis with consideration of intervention options rather than treating prediction as an end in itself (Ghasemi et al., 2025). Transposed carefully to prevention, this logic supports a decision-support model centered not on predicting who will become an offender, but on identifying which present and modifiable circumstances can be addressed through the least intrusive effective intervention.

Accordingly, AI-assisted prevention is legitimate only when it remains embedded in a problem-oriented rather than person-condemning framework. In situational prevention, the model should identify conditions of opportunity while remaining sensitive to data generated by previous enforcement. In person-oriented prevention, the system should identify needs and modifiable risk factors without converting probabilistic assessments into presumptions of dangerousness. In both domains, prediction must remain separable from intervention, and intervention must remain subject to independent justification. The highest value of artificial intelligence in preventive policy therefore lies not in maximizing the state's capacity to foresee conduct, but in improving the quality of decisions about when prevention is necessary, which form it should take, how intrusive it should be, and whether its actual effects justify its continuation.

3. Artificial Intelligence and the Transition from Prevention to Reactive Criminal Policy

The transition from preventive policy to reactive criminal policy marks a fundamental change in the nature of governmental decision-making. Prevention addresses the reduction of risks, opportunities, vulnerabilities, and conditions associated with future harmful conduct, whereas reactive policy begins from the institutional determination that conduct has occurred, that a regulatory norm may have been violated, or that a formal response must be considered. This transition does not mean that preventive objectives disappear; sanctions may themselves pursue deterrent, rehabilitative, incapacitative, reparative, or protective aims. The difference is that the state now confronts a prior normative question concerning the legal status of the conduct and the legitimate form of public reaction. Artificial intelligence may contribute information to this process, but its analytical capabilities cannot independently answer whether behavior should be criminalized, regulated administratively, subjected to professional discipline, or addressed without punitive intervention. Salehnejad Behrestaghi and colleagues discuss the potential use of artificial intelligence in identifying and analyzing emerging forms of criminal conduct (Salehnejad

[Behrestaghi et al., 2024](#)), yet identification of novelty or social harm remains analytically different from the legislative judgment that criminal prohibition is justified.

This distinction is essential because the expansion of data-driven governance can create pressure to convert measurable risk into legal prohibition. A system capable of detecting recurrent harmful patterns may make previously obscure conduct more visible, while predictive analysis may suggest that particular behaviors are correlated with future damage. Neither finding establishes the necessity of criminalization. Criminal law carries distinctive expressive, procedural, and coercive consequences, and its use requires a justification that cannot be reduced to computational evidence of correlation. Evidence may demonstrate frequency, distribution, financial loss, victimization patterns, or comparative effectiveness of previous responses, but the decision to attach criminal condemnation remains a normative judgment about the legitimate limits of state power. Hill and colleagues show that artificial intelligence may assist legislative processes by organizing information and improving certain dimensions of drafting and policy analysis ([Hill et al., 2025](#)). The significance of such systems lies in informing legislators, not in transforming legislative choices into optimization problems with a single technically correct solution. Gangi Chiodo similarly illustrates both the opportunities and institutional questions created when generative AI enters parliamentary work ([Gangi Chiodo, 2026](#)). Legislative judgment cannot be outsourced merely because the evidentiary environment has become computationally richer.

The appropriate contribution of artificial intelligence at the boundary between prevention and reaction is therefore comparative rather than determinative. A policymaker considering whether harmful conduct should become a criminal offense should have access to evidence concerning the nature and scale of harm, existing legal mechanisms, patterns of enforcement, compliance rates, affected populations, previous regulatory outcomes, and the likely consequences of alternative interventions. Artificial intelligence can help aggregate and analyze these sources. O'Brien's discussion of big data in evidence-based policy shows how large administrative datasets can reveal patterns useful to policymaking while simultaneously requiring caution regarding their origin and limitations ([O'Brien, 2024](#)). Welsh, Zane, and Mears frame evidence-based crime and justice policy around the systematic use of research evidence to improve public decisions ([Welsh & Mears, 2024](#)). The combination of these approaches suggests that AI can strengthen the informational foundation of criminalization policy, but only if data analysis is treated as evidence to be interpreted rather than authority to be obeyed.

A three-way framework is particularly useful in this context. The relevant choice is not simply between criminalization and legal indifference. Conduct can be placed within criminal law, subjected to administrative or regulatory sanction, or governed by non-punitive mechanisms. In some settings, civil, licensing, supervisory, restorative, educational, or compliance-based measures may provide more proportionate responses than criminal punishment. Artificial intelligence can facilitate comparison among these alternatives by examining historical outcomes, detecting overlap between legal regimes, identifying patterns of repeated non-compliance, and estimating how different institutional responses have operated in practice. Rigi and colleagues discuss artificial intelligence in connection with the efficiency of administrative law and organizational structures ([Rigi et al., 2025](#)), highlighting a broader domain in which computational systems can support regulatory governance outside criminal courts. Mahdi likewise examines legal challenges surrounding AI implementation within Iran's administrative system ([Mahdi, 2025](#)). These perspectives are relevant because a sophisticated criminal policy must be capable of deciding not only when criminal law should act but also when administrative regulation can address misconduct without the stigma and coercive intensity of criminalization.

Such comparison should be structured by a principle of minimum necessary intervention. If a lower-intensity regulatory mechanism can effectively address the relevant harm, the mere availability of criminal punishment does not make it preferable. Evidence-based policy provides an important foundation for this reasoning because it directs attention toward actual outcomes rather than symbolic assumptions about severity. Mears and Cochran's critique of mass incarceration emphasizes the importance of systematically evaluating alternatives instead of assuming that expanding imprisonment is the default path toward crime control ([Mears & Cochran, 2024](#)). The same reasoning can be generalized to the architecture of legal response: the state should distinguish between what requires condemnation through criminal law, what can be addressed through administrative or disciplinary regulation, and what is better governed through preventive or supportive mechanisms. Artificial intelligence may assist this inquiry by making the comparative consequences of these choices more visible.

However, a minimum-intervention principle cannot itself be reduced to algorithmic optimization. Determining what counts as “necessary” requires judgments about acceptable risk, individual rights, distributive effects, social values, and the legitimate purposes of public authority. A computational model can estimate consequences under specified assumptions, but the selection of those assumptions and the weighting of competing consequences remain normative. For example, a model could indicate that an administrative fine produces higher compliance at lower institutional cost than criminal prosecution, but policymakers must still consider whether the conduct expresses a degree of culpability or harm that requires a different public response. Conversely, a model could show that severe punishment produces a measurable deterrent effect, yet that finding alone would not establish that the punishment is proportionate. Welsh and Mears stress the importance of assessing both the quality of evidence and its applicability to the policy question at hand (Welsh et al., 2024). Evidence informs normative decision-making but does not displace it.

The danger of automated criminalization is particularly pronounced when artificial intelligence identifies “emerging crime” from novel technological behaviors. Rapid technological change can produce new forms of harm that are not easily captured by older legal categories, and AI may help detect patterns across large datasets more rapidly than conventional legislative analysis. Salehnejad Behrestaghi and colleagues explicitly consider AI in connection with the criminalization of emerging crimes (Salehnejad Behrestaghi et al., 2024). Yet novelty does not by itself justify criminalization. Some emerging conduct may already fall within existing general offenses; other behavior may require technical regulation rather than penal prohibition. AI could be valuable in detecting duplication, inconsistency, and enforcement gaps, thereby supporting narrower rather than broader criminalization. Its most constructive role may therefore be not the automated creation of offenses but the identification of situations in which new penal legislation is unnecessary because existing tools are adequate.

Artificial intelligence can likewise support retrospective review of existing criminalization. Criminal codes and regulatory systems accumulate provisions over time, often without systematic evaluation of whether older offenses continue to serve their intended purposes. Data-driven analysis can identify offenses that are rarely prosecuted, provisions producing inconsistent enforcement, categories that substantially overlap, or sanctions generating high social costs without demonstrable benefits. The capacity to use technology for legislative optimization discussed by Hill and colleagues (Hill et al., 2025) can be extended to this evaluative function. Instead of viewing AI solely as a mechanism for expanding governmental capacity, policymakers can employ it to identify areas of overcriminalization, redundant regulation, or disproportionate enforcement. This represents an important reversal of the common assumption that technological sophistication necessarily produces more control.

The transition from prevention to reactive policy also requires attention to due process and institutional competence. As algorithmic analysis becomes relevant to legal classification, affected persons and institutions may need to understand the evidentiary basis on which categories are created or enforced. Farajpour and colleagues identify ethical requirements associated with the adoption of AI regulation in the European context (Farajpour et al., 2025), while Nouralivand's analysis of the European Union Artificial Intelligence Act highlights a governance model in which risk classification and legal safeguards vary according to the significance of the application (Nouralivand, 2024). A similar functional logic is appropriate in criminal policy. An AI tool used to summarize legislative statistics poses different risks from a model whose classification directly shapes an enforcement category. The greater the connection between algorithmic output and coercive legal consequences, the greater the need for transparency, traceability, institutional accountability, and meaningful human responsibility.

Human oversight in this field must also be understood substantively. Parsa emphasizes that human supervision must operate as a meaningful legal safeguard rather than a formal presence attached to an essentially automated process (Parsa, 2024). A legislature, prosecutor, administrative authority, or regulator should therefore be capable of rejecting the implications of an AI-generated analysis and explaining the independent reasons for its decision. The model should function as an evidentiary and analytical assistant, not as a mechanism through which responsibility is displaced. This principle is crucial because sophisticated models can acquire an aura of objectivity that makes their recommendations difficult to challenge even where they depend on contested data or normative assumptions.

The transition to reactive policy should thus be conceptualized as a filtering process in which AI assists in structuring information but does not determine the legal category. The first question is whether state intervention is necessary. The second is whether that intervention should be punitive. The third is whether criminal law is required or whether administrative, disciplinary, professional, rehabilitative, or other mechanisms can adequately address the problem. The fourth concerns the

architecture and intensity of the response itself. Artificial intelligence may provide evidence at every stage, but its role changes as the decision moves closer to direct restriction of rights. The result is not a technologically automated criminal policy, but a more analytically informed process in which the informational contribution of AI is separated from the legal authority to define prohibited conduct and determine the legitimate limits of governmental reaction.

4. Artificial Intelligence and the Engineering of Multilayered Sanctions

Once a legitimate basis for formal intervention has been established, criminal policy confronts a second question that is analytically distinct from criminalization: what type of response should be imposed, at what level of intensity, and for what purpose? The concept of multilayered sanction engineering responds to this problem by rejecting the assumption that state reaction should be organized along a single penal axis running from less imprisonment to more imprisonment. Contemporary responses to harmful conduct can include criminal penalties, non-custodial sanctions, administrative penalties, disciplinary measures, professional restrictions, financial consequences, treatment requirements, rehabilitative programs, psychosocial support, restitutionary mechanisms, supervision, and combinations of these responses. The contribution of artificial intelligence lies in improving the capacity to compare these alternatives and to match a legally available intervention to the characteristics of the conduct, the individual case, the objectives of the response, and the available evidence concerning effectiveness. Ghasemi and colleagues' work on an intelligent judicial assistant for alternatives to imprisonment illustrates this direction by connecting AI-based risk analysis with the selection of non-custodial options rather than assuming imprisonment as the natural endpoint of judicial decision-making (Ghasemi et al., 2025).

The criminal-sanction layer is the most sensitive because penal measures represent a concentrated exercise of public coercion. AI may assist criminal sentencing or sanction design by organizing information about previous outcomes, identifying comparable cases, estimating risks associated with particular interventions, or drawing attention to legally relevant alternatives. Yet the use of such systems must distinguish individualized prediction from normative proportionality. A model can estimate a probability of reoffending; it cannot independently determine how much punishment an offender deserves. Stevenson and Doleac's analysis of algorithmic risk assessment shows that the effect of risk tools depends significantly on the behavior of the humans who receive and interpret algorithmic outputs (Stevenson & Doleac, 2024). This finding undermines simplistic accounts in which an algorithm mechanically improves sentencing merely by supplying more accurate information. A risk score becomes legally significant only through the institutional decision in which it is embedded.

This distinction also reveals why accuracy cannot be the sole criterion for evaluating AI-assisted sanction selection. A model could demonstrate strong predictive performance while relying on variables that reproduce enforcement disparities or generate unacceptable distributive consequences. Neil and Zanger-Tishler's analysis of algorithmic bias emphasizes the difficulty of using arrest-related measures as though they were neutral representations of criminal behavior (Neil & Zanger-Tishler, 2025). Where historical enforcement practices have affected who is arrested, charged, or recorded, predictive models trained on those outcomes may convert institutional history into apparently individualized risk. The problem becomes more serious when the prediction affects liberty. A false positive associated with an offer of voluntary support differs fundamentally from a false positive that contributes to incarceration, intensive supervision, or denial of a less restrictive sanction. For this reason, the design of criminal sanctions should treat algorithmic assessment as one evidentiary input whose relevance is constrained by proportionality, legal sentencing criteria, contestability, and human responsibility.

Artificial intelligence may nevertheless offer significant value in broadening the range of options actually considered by decision-makers. Judicial systems often possess alternatives to imprisonment that remain underused because information about eligibility, local availability, prior outcomes, or individualized suitability is fragmented across institutions. An AI-assisted system could integrate these sources and present structured alternatives without making the final decision. Ghasemi and colleagues provide an example of this orientation by proposing an intelligent assistant focused specifically on alternatives to incarceration (Ghasemi et al., 2025). Such an approach aligns with the broader evidence-based argument advanced by Mears and Cochran that criminal justice policy should move from generalized reliance on incarceration toward interventions supported by systematic evidence (Mears & Cochran, 2024). The value of AI in this context is therefore not that it can calculate the "correct sentence," but that it can reduce informational barriers that cause severe sanctions to become default choices.

At the legislative level, AI can support the architecture of criminal sanctions by comparing the practical effects of existing penalty structures. If a legal system provides several penalties for similar forms of misconduct, computational analysis can identify differences in application, recidivism, compliance, collateral consequences, duration, costs, and demographic distribution. Hill and colleagues' account of AI-assisted legislative optimization suggests that algorithmic tools can help identify structural problems in legal frameworks (Hill et al., 2025). Such analysis could reveal, for example, that two legally similar offenses produce markedly different sanctioning patterns or that a statutory penalty range is disconnected from actual judicial practice. This information may support legislative revision. It should not, however, be interpreted as authorizing an algorithm to determine the morally appropriate penalty range. Legislative proportionality remains a question of legal judgment informed by empirical evidence.

Administrative and disciplinary sanctions form a second layer within the response architecture. Many regulatory violations do not require criminal prosecution but nevertheless justify formal consequences, including fines, warnings, suspension, corrective orders, licensing restrictions, or temporary closure. Rigi and colleagues discuss the role of artificial intelligence in improving administrative efficiency and organizational decision-making (Rigi et al., 2025), while Mahdi identifies legal challenges arising from the implementation of artificial intelligence in administrative structures (Mahdi, 2025). In this setting, AI can assist authorities in identifying patterns of non-compliance, distinguishing isolated breaches from repeated violations, comparing enforcement histories, and selecting among legally available levels of response. The principal benefit may be greater consistency and the capacity to detect cases in which corrective intervention is sufficient without escalating immediately to highly punitive measures.

Yet administrative sanctions should not be treated as legally insignificant merely because they are non-criminal. Some administrative measures can have consequences comparable to or more severe than conventional penalties, particularly where they terminate a license, close a business, restrict access to a profession, or impose substantial financial burdens. The use of AI in this domain must therefore remain sensitive to consequence rather than formal legal classification. A decision-support system that recommends a minor warning presents a different rights profile from one whose output effectively determines whether a person may continue working. Parsa's discussion of human oversight is especially relevant wherever automated recommendations bear on significant legal interests (Parsa, 2024). The institutional decision-maker must be able to understand the grounds of the recommendation, consider contextual information unavailable to the model, and reject the proposed response where proportionality or fairness requires a different result.

Professional and occupational sanctions form a related but distinct layer. Professional regulation often aims to preserve competence, integrity, safety, and public trust rather than to punish wrongdoing in the criminal-law sense. Possible responses may range from warnings, mandatory education, supervision, and corrective plans to temporary suspension or exclusion from professional activity. AI can assist in differentiating these responses by examining the nature of the violation, previous compliance, remedial action, the risk created by continued practice, and the effectiveness of earlier interventions. This capacity supports a graduated rather than binary response structure. Nevertheless, where professional restrictions affect livelihood, reputation, or social standing, decision-making must satisfy increasingly robust procedural guarantees. The technological characterization of the tool does not diminish the legal significance of the consequence.

Medical, psychological, rehabilitative, and social responses form another essential layer. Their inclusion within criminal policy does not imply that criminal conduct should be medicalized or that legal responsibility can be reduced to diagnostic categories. Rather, some cases involve identifiable treatment needs, substance-use problems, social vulnerabilities, educational deficits, family circumstances, or other factors that may be relevant to reducing future harm. AI can assist in identifying clusters of needs and matching individuals with available interventions, but the objective must be support and rehabilitation rather than the conversion of social disadvantage into an algorithmic marker of dangerousness. Sheppard and colleagues' study of early-intervention risk assessment underscores the importance of validating risk tools for the populations and institutional purposes in which they are actually used (Sheppard et al., 2024). A model developed for one population or outcome cannot simply be transferred to another context because the label "risk" appears in both settings.

Artificial intelligence may be especially useful where the response requires coordination across institutional layers. A person may simultaneously face a criminal proceeding, professional disciplinary process, treatment need, and social-service intervention. Without coordination, these systems can produce duplication, contradiction, or excessive cumulative burden.

Multilayered sanction engineering requires decision-makers to examine the total architecture of response rather than each measure in isolation. An administrative restriction that appears moderate when viewed alone may become disproportionately severe when combined with criminal punishment, professional suspension, and intensive supervision. AI can potentially aggregate these consequences and present decision-makers with a fuller picture of cumulative intervention. This is a more constructive use of computational capacity than employing separate algorithms in each institution without awareness of the total burden imposed on the individual.

The principle of least intrusive effective intervention should govern this architecture. Evidence-based policy is not equivalent to selecting whichever intervention produces the strongest measurable behavioral effect regardless of cost to rights and liberty. Welsh, Zane, and Mears describe evidence-based crime and justice policy as a framework for integrating reliable evidence into policy choices (Welsh & Mears, 2024), while Welsh and Mears emphasize careful assessment of the strength and relevance of research evidence (Welsh et al., 2024). These principles imply comparison among interventions, attention to unintended consequences, and resistance to claims of effectiveness based on weak or non-comparable evidence. AI can accelerate comparison, but it cannot define the acceptable balance between effectiveness and liberty. A less intrusive measure producing sufficient protection may be preferable to a more severe measure even if the latter promises a modest additional reduction in estimated risk.

The evaluative dimension is therefore indispensable to sanction engineering. Once a response is imposed, institutions should assess whether it achieves the objectives that justified it. Artificial intelligence can help monitor compliance patterns, recurrence, institutional costs, treatment completion, unintended consequences, and disparities in application. O'Brien's analysis of large-scale naturally occurring data illustrates why administrative information can be especially valuable for longitudinal policy evaluation (O'Brien, 2024). Yet evaluation should not become a mechanism for indefinite intervention. If subsequent evidence shows that a measure is ineffective, unnecessarily restrictive, or disproportionately burdensome, the proper function of monitoring is to support modification or termination. The system must allow risk classifications to decrease as well as increase and must permit policy conclusions to be revised rather than permanently encoded.

This feedback function distinguishes multilayered sanction engineering from algorithmic sentencing in a narrow sense. The objective is not to construct a machine that selects penalties. It is to create an evidence-sensitive response architecture in which data from implementation are continuously returned to policymakers and decision-makers. Criminal sanctions can be compared with non-custodial alternatives; administrative responses can be evaluated against compliance outcomes; professional measures can be assessed for corrective effect; treatment and social interventions can be examined in relation to relevant needs; and cumulative burdens can be identified. The model thus operates across levels of policy rather than within a single sentencing decision.

Ethical and legal governance must accompany every stage of this process. Farajpour and colleagues emphasize ethical requirements associated with AI regulation (Farajpour et al., 2025), and Nouralivand highlights the European Union's differentiated governance of artificial intelligence according to risk (Nouralivand, 2024). In multilayered sanctions, a similar principle requires stronger safeguards as the connection between algorithmic output and rights-restricting consequences becomes more direct. Systems used only for aggregate policy evaluation can be governed differently from tools that generate individualized sanctioning recommendations. The latter require rigorous validation, transparency concerning relevant variables and limitations, meaningful opportunities for contestation, institutional audit, and clearly identifiable human responsibility.

The engineering of multilayered sanctions therefore does not signify the transformation of punishment into a computational exercise. It signifies a more systematic approach to the relationship between problems, purposes, interventions, evidence, and consequences. Artificial intelligence can make fragmented information more intelligible, reveal underused alternatives, identify inconsistencies, and facilitate continuous evaluation. Its legitimate role is strongest where it enlarges the range of proportionate responses considered by human decision-makers and weakest where it substitutes a probabilistic output for a legal judgment concerning blame, desert, or deprivation of rights. The measure of successful AI integration is consequently not the degree of automation achieved, but whether the resulting response system becomes more differentiated, evidence-sensitive, proportionate, revisable, and capable of selecting less intrusive measures where they can adequately achieve legitimate objectives.

5. Governing Algorithmic Decision-Support Across the Response Continuum

The preceding analysis demonstrates that artificial intelligence can operate at several points across the continuum from prevention to sanction, but the legal significance of its intervention varies substantially according to function. A system that organizes aggregate crime data, a model that estimates the risk associated with a place, an instrument that classifies an individual, a tool that recommends an alternative to imprisonment, and a platform that monitors compliance all involve computational analysis, yet they do not pose equivalent risks to rights or institutional legitimacy. A coherent governance framework must therefore abandon the assumption that “AI in criminal justice” constitutes a single regulatory category. Nouralivand’s analysis of the European Union Artificial Intelligence Act illustrates the importance of differentiated regulation based on the risks associated with specific uses (Nouralivand, 2024). Farajpour and colleagues similarly emphasize that ethical requirements must accompany the adoption and regulation of artificial intelligence rather than being treated as an afterthought (Farajpour et al., 2025). In criminal policy, this functional differentiation should be tied to the distance between algorithmic output and coercive consequence.

At the lowest level of direct coercive impact are systems used primarily for information organization, descriptive analysis, and aggregate evaluation. These tools may identify trends, summarize large datasets, detect inconsistencies, or compare policy outcomes without producing an individualized recommendation. Their risks concern data accuracy, representativeness, privacy, and methodological validity, but the connection to a person’s liberty may be indirect. At a higher level are predictive systems that classify locations or persons according to estimated future risk. These systems introduce more serious concerns because prediction can influence the allocation of police attention, social services, supervision, or judicial scrutiny. At a still higher level are recommendation systems that identify specific interventions or sanctions. Finally, systems that directly trigger or execute restrictive measures would create the greatest tension with principles of legality, accountability, and human authority. This graduated model makes it possible to calibrate safeguards to the function actually performed rather than the mere presence of machine learning.

Data quality is the first governance requirement across all levels. Artificial intelligence does not observe society independently; it processes representations created through institutional practices. Crime data may reflect reporting patterns, policing priorities, prosecution decisions, administrative categorization, and unequal institutional contact. Neil and Zanger-Tishler demonstrate how the use of arrest as a measure of crime can create racial disparities in algorithmic risk assessment (Neil & Zanger-Tishler, 2025). Their analysis has a broader implication: wherever an institutional variable is used as a proxy for underlying behavior, the model may learn patterns of institutional response alongside patterns of conduct. This is particularly significant for systems that claim to predict reoffending, dangerousness, or future justice-system contact. A governance framework should therefore require explicit identification of the target variable, the process through which it is generated, and the extent to which it reflects the phenomenon policymakers actually seek to understand.

The second requirement concerns validation and transferability. A model that performs adequately in one population, institutional setting, or period should not be assumed to retain that performance elsewhere. Sheppard and colleagues found meaningful questions concerning the performance of risk assessment instruments when applied to early-intervention populations in juvenile justice (Sheppard et al., 2024). This reinforces a general rule: validation must correspond to the actual population and purpose of use. A model designed to estimate recidivism among convicted adults cannot automatically be converted into an early-prevention tool for adolescents; a system developed from data in one jurisdiction may not retain validity in another; and a tool predicting institutional contact cannot be represented as though it predicts actual offending. Continuous revalidation is also necessary because institutional practices and social conditions change over time.

The third requirement is the separation of predictive performance from policy effectiveness. A system may predict an outcome accurately but fail to improve the intervention ultimately chosen. This distinction is central to evidence-based criminal policy. Welsh and Mears emphasize that evidence must be evaluated according to research quality and relevance to the policy claim being made (Welsh et al., 2024). Welsh, Zane, and Mears similarly frame evidence-based policy around the use of credible research to guide crime and justice interventions (Welsh & Mears, 2024). Accordingly, the relevant evaluation of an AI-assisted system should not end with metrics such as classification accuracy. Authorities should ask whether using the system actually improves decisions compared with available alternatives, whether it reduces harm, whether it increases or decreases

disparities, whether it changes the use of coercive measures, and whether any improvement justifies the financial and rights-related costs of deployment.

Human oversight constitutes a fourth requirement, but its meaning must be specified precisely. Merely placing a human official at the final stage of an automated process does not necessarily preserve meaningful responsibility. Decision-makers may defer to algorithmic recommendations because they appear scientifically authoritative, because institutional routines encourage acceptance, or because rejecting the recommendation requires additional justification. Stevenson and Doleac's study shows that algorithmic risk assessment operates through the behavior of human users rather than independently of them (Stevenson & Doleac, 2024). Parsa likewise emphasizes the legal importance of human oversight in AI governance (Parsa, 2024). Meaningful oversight therefore requires the authority, competence, information, and institutional freedom to disagree with the system. A human who merely confirms an opaque recommendation performs ceremonial rather than substantive supervision.

Explainability should likewise be understood in functional terms. It is not always necessary for every technical operation within a complex model to be translated into a complete mathematical account for every user. What is necessary is that the relevant decision-maker and affected person can understand why the output matters, which types of information materially influenced it, what uncertainties attach to it, and how its relevance to the legal decision can be challenged. The required degree of explanation should increase with the severity of the consequence. An aggregate planning model may require documentation suitable for institutional audit, while an individualized recommendation affecting liberty should be capable of meaningful contestation in the specific case. Farajpour and colleagues' focus on ethical requirements in AI regulation supports an approach in which transparency is connected to accountability and protection against unjustified harm (Farajpour et al., 2025).

Contestability is particularly important because algorithmic errors may arise from inaccurate records rather than from model design alone. Individuals should have a route to correct material factual errors where those data affect significant decisions. A model cannot be considered fair merely because its statistical performance is acceptable if the person affected has no mechanism for challenging incorrect information. This requirement becomes crucial in multilayered sanction systems because the same underlying data may circulate between criminal, administrative, disciplinary, and social institutions. An erroneous classification could otherwise propagate across systems and accumulate consequences. Effective governance therefore requires controls on data sharing, correction procedures, limits on secondary use, and institutional responsibility for ensuring that outdated or legally irrelevant information does not silently shape subsequent interventions.

Proportionality provides the normative link connecting all stages of the response continuum. The intensity of data processing, the reliability demanded from the model, the degree of human review, and the procedural safeguards required should all correspond to the seriousness of the potential intervention. Person-oriented risk assessment used to prioritize a voluntary service cannot be governed identically to a model influencing pretrial detention or imprisonment. Administrative tools that rank cases for inspection raise different concerns from systems recommending professional exclusion. This does not mean that low-impact systems are unregulated; rather, it means that criminal-policy governance should reflect the relationship between uncertainty and consequence. The more severe the measure, the less tolerable it becomes to rely on opaque, weakly validated, or indirectly relevant predictions.

A further requirement is institutional traceability. The use of artificial intelligence can blur responsibility because errors may be attributed to developers, data providers, administrators, or the algorithm itself. Such diffusion is incompatible with the exercise of public power. The legally competent institution must remain responsible for the decision and for the legitimacy of the system it chooses to deploy. Amirian Farsani identifies both opportunities and challenges arising from artificial intelligence in Iranian criminal justice policymaking and judicial law enforcement (Amirian Farsani, 2025). Mahdi likewise highlights legal challenges arising from AI implementation in public administration (Mahdi, 2025). These concerns indicate that institutional adoption itself must be subject to rules defining responsibility, audit authority, system approval, documentation, and remedies.

Governance should also include comparative necessity. The question should not merely be whether an AI model can perform a task, but whether employing it produces sufficient improvement over a simpler method to justify its risks and costs. Technological complexity can become self-legitimizing, encouraging institutions to adopt systems because they are innovative rather than because they solve a demonstrated problem. O'Brien's analysis of big data highlights both the long-term potential

and methodological challenges of using naturally occurring information in policy and practice (O'Brien, 2024). A similar discipline should apply to artificial intelligence: institutions should define the problem first, identify feasible alternatives, and then determine whether AI provides a measurable advantage. Where conventional statistical analysis or professional review achieves comparable results with greater transparency and lower rights-related risk, sophistication alone is not a sufficient reason for algorithmic deployment.

The same logic applies to legislative uses of AI. Hill and colleagues discuss the potential contribution of artificial intelligence to future-oriented legislative optimization (Hill et al., 2025), while Gangi Chiodo examines generative AI within a parliamentary context (Gangi Chiodo, 2026). These applications may assist lawmakers by organizing evidence, identifying textual inconsistencies, comparing regulatory approaches, or modeling consequences. Yet legislative use should remain transparent concerning the source and role of AI-generated material. The democratic legitimacy of legislation derives from authorized institutions and deliberative procedures, not from the technical competence of the software used to assist them. Artificial intelligence can improve access to relevant information while simultaneously creating risks of hidden assumptions, overreliance, or homogenized reasoning if generated outputs are accepted without independent scrutiny.

A mature governance model must finally incorporate periodic evaluation and exit mechanisms. Technologies introduced into criminal justice often become institutionally entrenched, particularly after investments in infrastructure, training, and data systems. Yet a tool that initially performs well may deteriorate as patterns change, or subsequent evidence may reveal disparities and unintended consequences. The evidence-based orientation described by Welsh and colleagues requires continued evaluation rather than one-time approval (Welsh & Mears, 2024). Systems should therefore be subject to periodic review of predictive validity, practical effectiveness, disparate effects, data quality, user behavior, and continued necessity. Institutions must be prepared not merely to modify an algorithm but to suspend or discontinue its use when evidence no longer supports deployment.

Taken together, these requirements produce a governance framework centered on the concept of AI as a bounded decision-support infrastructure. Its permissible role is broad in cognitive and evaluative functions but progressively narrower as it approaches the direct imposition of coercive consequences. The governing principle is that computational capacity expands what institutions can know and compare but does not enlarge their substantive legal authority. AI may reveal a pattern of risk, identify an intervention that has historically performed well, or expose inconsistency in sanctioning, but each movement from information to coercion requires an independent legal justification. The legitimacy of AI-assisted criminal policy therefore depends not on the elimination of human judgment but on structuring human judgment more carefully: making relevant evidence visible, exposing uncertainty, broadening the range of alternatives, documenting reasons, permitting challenge, evaluating consequences, and preserving accountable human authority over decisions that affect rights and freedoms.

6. Conclusion

Artificial intelligence is reshaping the informational environment in which contemporary criminal policy is formulated and implemented, but its most important contribution does not lie in the replacement of human judgment by automated decision-making. Its significance lies in the possibility of reorganizing the relationship between knowledge, intervention, implementation, evaluation, and policy revision. When understood in this way, artificial intelligence becomes part of a broader system of responses to crime rather than a discrete technological device attached to police, courts, or correctional institutions. The relevant question is therefore not whether an institution “uses AI,” but what function a particular system performs, what kind of output it produces, how that output enters a legally significant decision, and what consequences follow for individuals and communities.

In situational prevention, artificial intelligence can expand the capacity to analyze spatial, temporal, environmental, and incident-based information and to move from static descriptions of past events toward dynamic assessments of situational risk. This function can improve the targeting of preventive attention and support more responsive allocation of resources. Yet predictive information should not be confused with the reality it attempts to estimate. Crime data are shaped by reporting, surveillance, detection, and institutional priorities, and the use of predicted risk to redirect police resources can itself generate the data that later appear to validate the original prediction. The appropriate model therefore separates prediction from intervention. A high-risk location is not an automatic instruction to intensify coercive control; it is an evidentiary signal that

must be considered alongside the nature of the problem, the effectiveness of available preventive measures, their intrusiveness, and their possible unintended effects.

In person-oriented prevention, the need for this distinction becomes even stronger. The proper objective of artificial intelligence should not be the creation of technologically sophisticated categories of “potential offenders.” Predicting future criminality at the individual level risks transforming statistical uncertainty into stigma and governmental control. The more defensible use of AI is to support multidimensional assessment of risk factors, protective factors, vulnerabilities, and needs, particularly where these elements are capable of modification through non-punitive intervention. The central question should shift from “Who will offend?” to “Which present and modifiable condition can be addressed, by what intervention, and with what evidence of benefit?” This reorientation moves artificial intelligence away from prediction as an end in itself and toward decision support within a preventive framework centered on proportionality and support.

The same logic structures the transition from preventive to reactive criminal policy. Once formal response is contemplated, the first issue is not the severity of punishment but the legal architecture within which the behavior should be addressed. Criminalization, administrative regulation, professional discipline, and non-punitive intervention represent distinct forms of public response. Artificial intelligence can provide richer evidence concerning patterns of harm, regulatory effectiveness, enforcement outcomes, institutional overlap, and the consequences of existing rules. It cannot determine, solely from those observations, that criminal law is necessary. The decision to criminalize is an exercise of normative and political authority because it determines the boundary at which the state attaches penal condemnation and coercive consequences to conduct. Greater analytical capacity should therefore support more careful and limited criminalization rather than automatic expansion of penal regulation.

Once intervention is justified, the concept of multilayered sanction engineering provides a more appropriate framework than a narrow focus on sentencing. Modern criminal policy contains multiple response layers: criminal, administrative, disciplinary, professional, rehabilitative, medical, psychological, social, restorative, and supervisory. These responses may operate separately or in combination. Artificial intelligence can assist by bringing together information that is otherwise institutionally fragmented, comparing the consequences of different interventions, identifying less restrictive alternatives, and revealing inconsistencies in how similar cases are treated. Such assistance may be particularly valuable where severe sanctions have become institutional defaults despite the formal availability of alternatives.

The engineering of sanctions should nevertheless remain constrained by the principle that predictive capacity is not equivalent to normative authority. An algorithm may estimate recidivism risk, predict the likelihood of compliance, identify treatment needs, or compare historical outcomes. None of these operations determines what a person deserves, what degree of restriction is proportionate, or whether a particular deprivation of liberty is legally justified. These judgments depend on legal rules, responsibility, culpability, rights, the purposes of intervention, and the legitimate boundaries of public power. AI can make relevant evidence available to the decision-maker, but the authority to interpret that evidence within a normative legal framework must remain human and institutionally accountable.

A central implication of this approach is that the acceptable role of artificial intelligence varies according to the intensity of the consequence connected to its output. Systems used for aggregate analysis and policy evaluation may legitimately operate with a wider degree of automation because they do not directly determine an individual's rights. Predictive systems concerning persons require stronger validation and safeguards. Recommendation systems influencing sanctions require still greater transparency, contestability, and human review. Direct automated imposition of rights-restricting measures represents a fundamentally different category because it risks transferring normative authority from legally responsible institutions to computational processes. Regulation should therefore be calibrated to function and consequence rather than based on the undifferentiated category of artificial intelligence.

The analysis also demonstrates that human oversight must be substantive rather than ceremonial. A human decision-maker does not provide meaningful protection merely by appearing at the final stage of an automated workflow. Effective oversight requires the capacity to understand the relevance and limitations of the output, identify uncertainty, consider contextual information, reject algorithmic recommendations, and provide independent reasons for the final decision. Responsibility cannot be displaced onto the system, its developer, or the complexity of the model. Public institutions that choose to incorporate artificial intelligence into criminal policy remain responsible for the legal consequences of its use.

The same principle applies to data. AI systems operate on institutional representations of reality rather than reality itself. Arrest, conviction, recorded crime, police contact, administrative violation, treatment history, and professional discipline are outcomes produced through social and institutional processes. Treating these variables as neutral proxies for underlying behavior can reproduce existing patterns of enforcement. Data governance must therefore form part of substantive criminal-policy design rather than being regarded as a technical preliminary. The source of data, its relationship to the phenomenon being predicted, the possibility of correction, its continued relevance, and the distribution of errors are all matters that affect the legitimacy of subsequent intervention.

Ultimately, the most productive model is a cyclical rather than linear conception of AI-assisted criminal policy. The process begins with identification and assessment of a problem, continues through analysis of possible interventions, moves to implementation, and then returns through evaluation and revision. Artificial intelligence can contribute at each stage, but its contribution differs. It may organize information at one stage, estimate probabilities at another, compare interventions at a third, and monitor outcomes at a fourth. What it should not do is collapse these separate stages into a single automated decision. Prediction should remain distinct from intervention, evidence from normative judgment, and implementation from evaluation.

Within such a system, the purpose of technological development is not to maximize the state's capacity to classify, predict, and control. Its purpose is to improve the quality of criminal-policy decisions while simultaneously limiting unnecessary intervention. A well-designed AI-assisted system should make it easier to identify where no action is required, where a preventive measure is preferable to punishment, where an administrative or professional response is sufficient, where treatment or social support is more appropriate, where a non-custodial measure can substitute for imprisonment, and where an existing intervention should be discontinued because evidence does not support its continuation. In this sense, the highest value of artificial intelligence may lie not in enabling more intervention but in enabling more differentiated restraint.

The resulting model of artificial intelligence and criminal policy can therefore be summarized as an architecture of bounded assistance. Artificial intelligence expands cognitive capacity, strengthens pattern recognition, facilitates comparison, and improves continuous evaluation. Human institutions retain normative authority, legal responsibility, and control over coercive decisions. Prevention remains directed toward reducing opportunities and modifiable risks rather than identifying predetermined offenders. Reactive policy remains governed by the need to justify intervention and to distinguish criminalization from less intrusive regulatory alternatives. Sanction design becomes multilayered and evidence-sensitive rather than automatically penal. Evaluation becomes continuous, allowing ineffective or disproportionate measures to be revised or terminated.

The future significance of artificial intelligence in responding to crime will consequently depend less on whether systems become capable of producing increasingly sophisticated predictions and more on whether legal institutions develop an architecture capable of using those predictions without surrendering the principles that justify public authority. The essential boundary is between increasing the state's capacity to know and increasing its entitlement to coerce. Artificial intelligence may legitimately strengthen the former; it cannot by itself establish the latter. A criminal-policy system that preserves this distinction can employ technological innovation to support more precise prevention, more differentiated intervention, more proportionate sanctions, and more rigorous evaluation while maintaining human responsibility, legality, and fundamental rights at the center of the response to crime.

Ethical Considerations

All procedures performed in this study were under the ethical standards.

Acknowledgments

Authors thank all who helped us through this study.

Conflict of Interest

The authors report no conflict of interest.

Funding/Financial Support

According to the authors, this article has no financial support.

References

- Abdollahi, S. (2026). The role of artificial intelligence in predicting offenders' criminal profiles. *Law of Modern Technologies*, 7(13), 461-478. <https://doi.org/10.22133/mtlj.2026.513665.1450>
- Amirian Farsani, A. (2025). Artificial intelligence and criminal justice policymaking in contemporary Iran: New opportunities and challenges for judicial law enforcement officers. *Contemporary Political and Social Transformations of Iran*, 4(5), 80-95.
- Ehsanpour, S. R. (2025). The importance and position of artificial intelligence in crime prevention. *Applied Criminology Research*, 3(7), 59-80. <https://doi.org/10.22034/AQCR.2025.2054758.1053>
- Farajpour, R., Amerinia, M., & Gorjinia, M. (2025). Ethical requirements in the process of adopting the European Union Artificial Intelligence Act. *Cyberspace Legal Studies*, 3(4), 38-53.
- Gangi Chiodo, F. (2026). The Hand on the Pen: Generative AI and the Italian Parliamentary Experience. *Statute Law Review*, 47(2), hmag024. <https://doi.org/10.1093/slr/hmag024>
- Ghasemi, M., Habibitabar, M., & Moradi, S. (2025). An intelligent judicial assistant for alternatives to imprisonment: An artificial intelligence-based risk prediction model. *Applied Criminology Research*, 3(9), 125-152. <https://doi.org/10.22034/aqcr.2025.2073590.1127>
- Hill, G., Waddington, M., & Qiu, L. (2025). From Pen to Algorithm: Optimizing Legislation for the Future with Artificial Intelligence. *Ai & Society*, 40, 3075-3086. <https://doi.org/10.1007/s00146-024-02062-3>
- Mahdi, A. (2025). An analysis of legal challenges and solutions for implementing artificial intelligence in Iran's administrative system. *Administrative Law Journal*(42), 206-229.
- Mears, D. P., & Cochran, J. C. (2024). Mass Evidence-Based Policy as an Alternative to Mass Incarceration. In B. C. Welsh, S. N. Zane, & D. P. Mears (Eds.), *The Oxford Handbook of Evidence-Based Crime and Justice Policy* (pp. 580-597). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780197618110.013.29>
- Najafi, A., & Mousavifar, S. M. (2025). A reflection on the development of artificial intelligence in crime prevention. *Legal Civilization*, 8(23), 171-190. <https://doi.org/10.22034/LC.2025.512833.1612>
- Neil, R., & Zanger-Tishler, M. (2025). Algorithmic Bias in Criminal Risk Assessment: The Consequences of Racial Differences in Arrest as a Measure of Crime. *Annual Review of Criminology*, 8, 97-119. <https://doi.org/10.1146/annurev-criminol-022422-125019>
- Nouralivand, Y. (2024). Analysis: The European Union Artificial Intelligence Act, the world's first law on artificial intelligence governance. *National Security Monitor*(142), 113-122.
- O'Brien, D. T. (2024). Big Data and Evidence-Based Policy and Practice: The Advantages, Challenges, and Long-Term Potential of Naturally Occurring Data. In B. C. Welsh, S. N. Zane, & D. P. Mears (Eds.), *The Oxford Handbook of Evidence-Based Crime and Justice Policy* (pp. 598-620). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780197618110.013.30>
- Parsa, N. (2024). The role and requirements of human oversight of artificial intelligence in European Union law and Iranian laws. *Research and Development in Public Law*, 1(2), 114-140. <https://doi.org/10.22034/jrpl.2025.721655>
- Rigi, M., Hejazi, A., & Jelokhani Niaraki, F. (2025). The role of artificial intelligence in the efficiency of administrative law and organizational divisions. *Law Studies*, 6(52), 123-134.
- Salehnejad Behrestaghi, S., Soufi, S., & Heidari Parchkouhi, A. (2024). Criminalizing emerging crimes using artificial intelligence. *Journal of Law Studies*, 12(53), 1-12.
- Sheppard, K. G., Talaugon, A. R., & Hernandez, J. L. (2024). Assessing the Feasibility and Performance of Risk Assessment Instruments for Early Intervention and Prevention Services in Juvenile Justice. *Journal of Criminal Justice*, 94, 102262. <https://doi.org/10.1016/j.jcrimjus.2024.102262>
- Shojai Langari, S. Y. (2024). The impact of artificial intelligence and data mining on crime prevention: Opportunities and challenges. *Judicial Journal*(6), 94-108.
- Stevenson, M. T., & Doleac, J. L. (2024). Algorithmic Risk Assessment in the Hands of Humans. *American Economic Journal: Economic Policy*, 16(4), 382-414. <https://doi.org/10.1257/pol.20220620>
- Talebi Rostami, M. (2025). The position of artificial intelligence in criminal policymaking. *Legal Civilization*, 8(26), 463-476. <https://doi.org/10.22034/lc.2026.565037.1721>
- Welsh, B. C., & Mears, D. P. (2024). Evaluating Research and Assessing Research Evidence. In B. C. Welsh, S. N. Zane, & D. P. Mears (Eds.), *The Oxford Handbook of Evidence-Based Crime and Justice Policy* (pp. 21-36). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780197618110.013.2>
- Welsh, B. C., Zane, S. N., & Mears, D. P. (2024). Evidence-Based Crime and Justice Policy. In B. C. Welsh, S. N. Zane, & D. P. Mears (Eds.), *The Oxford Handbook of Evidence-Based Crime and Justice Policy* (pp. 1-18). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780197618110.013.1>