ISSN: 2688-8653

Criminal Law Assessment of Acts Committed Through Algorithms to Benefit Others in Automated Decision-Making

Yongbo Shang

Beijing Normal University, School of Law, Beijing 100875, China

Abstract

As algorithms become deeply embedded in public administration, traditional forms of corruption have undergone a transformation, and the elements of official misconduct face interpretive challenges in automated decision-making scenarios. Based on the manner in which the perpetrator intervenes in the algorithm and the content of their knowledge, acts of using algorithms to secure benefits for others can be categorized into acts of commission, acts of omission, and acts of negligence. In terms of legal characterization, the creation of a biased algorithm and its deployment constitute the commencement of a crime; omission may also satisfy the element of securing benefits for others, but only if the perpetrator exercises control over the algorithm; negligence constitutes a crime of dereliction of duty only when it results in significant losses. Regarding attribution of liability, the focal point should shift from control over the act to control over the algorithm: for intentional crimes, the liability of principal offenders and accomplices is determined based on volitional control; for negligent crimes, supervisory and management responsibilities are allocated based on breaches of duty, and the scope of punishment is limited by the principle of reliance.

Keywords

Automated decision-making; Seeking benefits for others; Negligent omission; Oversight fault.

1. The Proposal of the Issue

With the advancement of digital government construction, the way public power is exercised is evolving towards the "algorithmic administration" paradigm[1]. Currently, algorithms have taken on substantive discretionary functions in areas such as administrative approval and welfare distribution, assisting or even replacing public authorities in making legally binding decisions[2]. Reduced the space for arbitrary decision-making, bias, and rent-seeking[3]. However, it also brings the risk of blurring the accountability mechanism[4]. When algorithms are manipulated or biased, the transfer of benefits is hidden behind the "black box", and its harmfulness far exceeds that of traditional decision-making models. Article 385 of China's Criminal Law stipulates the "giving benefits for others" requirement based on the actor directly performing official duties. However, in automated decision-making, the outcome of seeking benefits is made by the algorithm, and the actor does not take any active actions. The relationship of power and money compensation has deviated between the normative prototype and the factual structure. The common view positions algorithms as a criminal tool. In cases of intentional manipulation of algorithms, there is no doubt, but it is difficult to explain the above situation.And the view advocating the principle of subject qualification[5] is overly radical, and there is still an intermediate situation where "benefiting others" is achieved through setting up, tolerating, or neglecting the review process.

Based on this, this article attempts to answer three questions: First, whether using algorithms to benefit others constitutes bribery, and whether tolerating algorithmic bias falls under the

ISSN: 2688-8653

category of "benefiting others"; second, distinguishing between the intentional bribery and negligent misconduct in the context of "algorithm black box"; third, allocating the responsibilities of algorithm developers, deployers, and users.

2. Specific Types of Benefiting Others through Algorithms

In the context of automated decision-making, due to differences in the performance of duties, subjective cognition, and control capabilities among actors, benefiting others through algorithms can be categorized into types such as acting, inaction, and negligence.

One type is the active manipulation or control of algorithms, where the actor actively intervenes in the algorithm to bias the outcome in favor of a specific entity. On one hand, this manifests as embedding biased functions during the design or parameter setting stages. When translating administrative purposes into code instructions, elements such as case facts and legal requirements inevitably carry the subjective perception of the designer[6]. The actor embeds parameters that are in line with the interests of the petitioner into the filtering conditions or scoring rules. Then, the algorithm will repeatedly achieve the preset results. Due to the temporal separation between this behavior and the outcome, it is extremely covert. On the other hand, it manifests as tampering with data or interfering with the output during the operation process. The actor replaces the input data or overrides the output result to make the decision result favor the petitioner. It can be seen that as a type, the actor's will dominates the algorithm, and the algorithm is merely a tool.

The second type is the passive non-action type that tolerates algorithmic bias. This refers to the situation where the bias in the results is not the result of the actor's active actions. In practice, algorithmic flaws, data contamination, or model drift can cause automated decisions to produce biased results without human intervention. However, the actor is aware of this and has the obligation to correct it but still allows it. It can be acknowledged that algorithmic decisions are not always absolutely correct and human supervision and control of algorithmic decisions are necessary[7]. When the automated decision-making system emits an abnormal signal or the actor learns of the bias through complaints or audits, and still tolerates or allows it, this is considered the profit-seeking type of inaction.

The third type is the negligent type, where the perpetrator neither actively set bias parameters nor is aware of the bias in the result algorithm. They merely objectively cause the transfer of benefits due to the failure to fully fulfill the obligations of review, detection, or supervision. The negligent type differs fundamentally from the non-action type. The former has the obligation and ability to foresee but fails to do so, while the latter has a clear understanding but tolerates it. In practice, due to the formalism of the manual review process and the ineffective抽查 mechanism， the perpetrators often remain oblivious to abnormal signals. It should be noted that some biases may stem from the negligence of the developers， but they usually do not hold the identity of a public official and do not have the authority to dominate public power， so their criminal responsibility should be carefully pursued.

The above types are classified based on the way the actor intervenes in the algorithm, and there is a possibility of cross-transformation. The actor may belong to different types at different stages, and should be evaluated based on the subjective attitude at the time of the behavior. It cannot be assumed that the actor had the intent to commit bribery from the very beginning just because the behavior ultimately constitutes bribery. At the same time, all the aforementioned behavior types are limited to the actual participation of the algorithm in the decision-making process. If the automated decision-making only has a formal meaning and the decision conclusion is completely made by humans, it can be directly judged according to traditional rules.

ISSN: 2688-8653

3. Qualification of Actions of Benefiting Others through Algorithms

3.1. Setting or manipulating biased algorithms constitutes the act of accepting bribes

If an individual sets or operates biased algorithms to serve the interests of the person making the request, thereby resulting in a transaction of power and money, it constitutes the crime of accepting bribes. In automated decision-making, the power is directly exercised by algorithms or digital systems, but they do not have independent authority and liability capacity; they are merely the mechanisms and channels for the performance of administrative duties by the administrative authorities[8]. Even if we concede the artificial intelligence's status as a subject of rights[9], such actions should also be uniformly regarded as the expressions of the administrative authority's will, and the consequences shall belong to the administrative authority[10]. That is to say, the will of the intended subject must be formed by natural persons. The act of the actor setting the bias parameters essentially treats the automated decision-making system as an extension tool of authority. On one hand, the launch of the biased algorithm immediately creates an urgent danger to the unbuyable nature of the position. Some may think that parameter setting is merely a tool for subsequent power-money transactions and belongs to criminal preparation. But the problem is, the boundary between criminal preparation and implementation lies in whether it poses a real and urgent danger to the protection of legal interests. If the biased algorithm remains only at the design stage, the danger to the unbuyable nature of the position behavior is indirect. It can be regarded as criminal preparation. On the other hand, the profit results automatically realized by the algorithm still belong to the setting behavior. Bribery does not require the actual acquisition of benefits. The actor's commitment alone is sufficient. By setting the biased algorithm and deploying it online, the actor has already "carried out" the profit-making behavior. Automation decision-making merely extends the time of power-money transactions. The actor retains actual control over the algorithm and has no essential difference from the traditional structure of using other tools for bribery.

3.2. Tolerating algorithmic bias as a form of non-action-based profit-making behavior

When the algorithmic bias stems from design flaws, data contamination, or model drift, and the actor knowingly and has the obligation to correct it but chooses to ignore it, how should it be evaluated? This article holds that this falls under the category of non-action-based profit-making behavior.

Firstly, "seeking benefits for others" can be achieved through an inactive approach. "Seeking" originally means "trying to obtain", while "tolerating" seems difficult to encompass its core semantic meaning. However, the penalty for accepting bribes is based on the unbuyability of the official act. If the actor knows that the algorithm is biased but does not correct it, the unbuyability of the official act will be subject to a real infringement. Further examination reveals that "seeking benefits for others" describes the normative relationship where the official act points to the interests of others. The word "for" indicates a strong purpose. The actor tolerating an outcome for which they have the obligation to prevent also meets the requirements of "seeking benefits for others". In terms of its system position, the old objective theory [11] requires the actor to take an active action and does not accommodate the space for inaction; the subjective theory[12] only requiring the perpetrator to have the intention of benefiting others weakens the objective constraints of bribery crimes; the new objective theory[13] pay attention to whether the power and property are in a corresponding relationship, and maintain an open attitude towards the behavior itself. Therefore, "seeking

ISSN: 2688-8653

benefits for others" does not necessarily require active actions; it can be achieved through inaction as well.

Secondly, those who tolerate algorithmic biased behavior have the obligation to prevent such outcomes from occurring. The obligation is the prerequisite for assuming the responsibility of non-action. The obligation to correct algorithmic bias in automated decision-making mainly stems from legal duties and prior actions. On one hand, the responsibility of supervisors and managers to ensure the legal and fair operation of the algorithm system does not cease with the completion of the process; instead, it transforms into the supervision of the operating status. Their account permissions, information advantages, and management responsibilities are all sources of this obligation. On the other hand, selection, acceptance, and other prior actions grant the algorithm public decision-making authority. Signing, acceptance, and authorizing for launch are the starting points for the behaviorist to assume subsequent supervisory responsibilities. Of course, one cannot simply affirm the obligation of the behaviorist based on general job responsibilities; it is also necessary to examine whether the behaviorist has the ability to act.

Furthermore, tolerating algorithmic bias is equally harmful to legal rights as actively setting biased algorithms. The failure to act lacks the causal force in the sense of action. The reasoning logic lies in presuming that if the actor fulfills their obligations, the harm can be avoided. That is to say, the failure to act emphasizes the possibility of avoiding the consequence. On one hand, it is the possibility of avoiding the consequence in the sense of performing an act, emphasizing that the actor has the actual ability to prevent the result. If algorithmic bias stems from model flaws and any act of performance cannot correct it, the actor's failure to act is not equivalent to the act of performing. On the other hand, it is the possibility of avoiding the consequence in the sense of causal relationship, emphasizing that the result will not occur after the actor fulfills their obligations. The burden of proof for the causal relationship in the absence of action lies with the prosecution. If it cannot be proven that the actor has the actual means to block and the reaction time, the causal connection should be denied in favor of the defendant based on the principle of presumption in favor of the defendant in cases of doubt.

Finally, those who tolerate algorithmic bias have a generalized intent to accept bribes. Traditional bribery requires the actor to have a clear understanding of "for whom and what to do". However, the actor often does not know the algorithmic decision-making logic and specific content. If the ascription-type knowledge standard is applied, it will lead to an inability to prove and result in the leniency of the crime. In the non-action type, the knowledge that is known is the algorithmic biased state rather than the specific result. That is, if the actor recognizes the existence of algorithmic bias, they have the intent to accept bribes. It should be emphasized that this generalized intent does not lower the proof standard. When making a specific judgment, factors such as the degree of signal anomaly, professional background, contact logs, and the time of receiving the bribe can be considered, and counter-evidence can be allowed. If it is truly impossible to know the algorithmic bias due to limited authority or information obstruction, their intent to accept bribes should be denied.

3.3. Negligence Lacking the Intent to Accept a Bribe Should Be Classified as Dereliction of Duty

In cases of negligence, the perpetrator neither actively designed a biased algorithm nor had specific knowledge of the algorithm’s bias; rather, the transfer of benefits occurred objectively solely because the perpetrator failed to fulfill their duty to review and supervise. Since the objective manifestations of negligence are similar to those of inaction, but there are differences in terms of subjective intent, such conduct should be classified as a crime of dereliction of duty.

On the one hand, the perpetrator lacks the intent to accept a bribe because he or she is unaware of the quid pro quo between power and money. “Securing benefits for others” requires

ISSN: 2688-8653

awareness of the benefits to be gained by others, coupled with an attitude of indifference or pursuit; however, a perpetrator acting negligently is unaware of the algorithmic bias and is therefore unable to direct their official acts toward the interests of others, nor can they recognize the quid pro quo relationship between their official acts and the property. In other words, one cannot disregard subjective intent to impose objective liability, nor can one presume that the perpetrator had the intent to accept a bribe merely because harmful consequences ultimately occurred.

On the other hand, a breach of the duty of care combined with significant losses meets the elements of a crime of dereliction of duty. If a perpetrator acts with gross negligence, fails to perform, or improperly performs their duties, thereby causing significant losses to the interests of the state and the people, they may be guilty of dereliction of duty; if a perpetrator exceeds their authority or acts in violation of regulations, they may be guilty of abuse of power. If the perpetrator was unaware of algorithmic bias and the resulting harm, or naively believed such harm could be avoided, and their objective conduct manifested as a long-term failure to conduct spot checks or to address warning signals, this constitutes dereliction of duty; if the perpetrator knowingly exceeded their authority despite being aware of the potential for harm, and their objective conduct manifested as arbitrarily disabling system alerts, improperly delegating approval authority, or bypassing review procedures, this constitutes abuse of power. Furthermore, if negligent performance of duties does not result in significant losses, it should be addressed only through administrative disciplinary action or organizational accountability and should not be subject to criminal evaluation. When evaluating negligence, one cannot directly apply the evaluative logic used for acts of commission or omission; the focus of evaluation should shift from “securing benefits for others” to an assessment of the duty of care and the resulting harm.

4. Pathways for Attributing Liability in Cases Where Algorithms Are Used to Generate Profit for Others

As private technological capabilities become deeply embedded in public administration, decision-making has become the product of the combined effects of multiple factors. Consequently, the traditional chain of accountability has been extended, and the entities responsible have become increasingly obscure. While the qualitative analysis discussed above can determine whether various types of conduct constitute a crime and, if so, what type of crime, it struggles to clarify who should bear responsibility for specific harmful consequences, why they should bear it, and how they should do so. Therefore, it is necessary to further explore the pathways for attributing liability.

4.1. Basis for Attribution: Transition from Attribution of Behaviors to Attribution of Algorithms

The traditional attribution of criminal acts was based on the direct control of the actor over the behavior. However, in automated decision-making, biased outcomes are formed through multiple links, and no single entity can exert complete control. The traditional attribution logic is digitally unsuitable in the context of automated decision-making and needs to be adjusted.

Automated decision-making is centered around algorithms. It is reasonable to attribute the results based on the control of the participating entities over the algorithms. The basis for attribution shifts from the control of behavior to the control of algorithms. The actor objectively should have the realistic possibility of controlling the operation of the algorithm, including the ability to modify parameters, suspend the system, and transfer to manual processing. Intentional acts and negligent acts are hierarchical. Intentional acts have the actual dominance over the dangerous flow, and for this, the attribution should be based on the willfulness of the

ISSN: 2688-8653

actor; negligent acts only have the possibility of dominance over the dangerous flow, and for this, the attribution should be based on the violation of obligations[14]. It should be noted that algorithmic dominance is not the logic for reconstructing the determination of criminal constitutive elements; rather, it is an extension of the crime-dominance theory and the guarantor theory in the digital context. Following this approach, by considering the responsibilities, authorities, and cognitive contents of each entity, it is sufficient to attribute the harmful consequences to the corresponding entities, thereby avoiding the improper expansion of the scope of punishment.

4.2. Intentional crimes are determined based on the will to act and are attributed to the principal offender or accomplice

According to the aforementioned, for both active and passive types, bribery is established when there is an intent to accept bribes, but a systematic analysis of the allocation of responsibilities among multiple participants is required. In terms of the determination of the principal offender, the degree of the actor's dominance over the causal process formed by the algorithm bias is considered to determine the subject of the algorithm bias content as the principal offender, given that they have a dominant position. For example, the operator who sets inappropriate weights in the parameters, or the manager who has the authority to affirm or deny the output results but allows it to persist, are principal offenders. Regarding accomplices, the boundary for the establishment of the aiding offense must be defined for neutral business behaviors. The algorithms delivered by developers usually have broad and legitimate uses, and their accomplice liability cannot be pursued merely because they were later abused. Only when the developer specifically develops a certain algorithm, reserves backdoors, or modifies the algorithm logic knowing that a specific solicitation of benefits exists, and their technical behavior forms an agreement with the harmful consequences, can it be considered an aiding offense. It should be noted that the responsibility of technicians should be analyzed on a case-by-case basis. Technicians do not constitute a separate principal offender for bribery crimes because they do not have an official position. If they provide key technical assistance knowing that a bribe transaction exists, it constitutes an aiding offense; if they provide technical support only according to normal business operations without knowing about the bribe transaction, it is not included in the criminal evaluation.

4.3. Negligent Crimes Based on the Allocation of Supervisory Responsibility for Obligations

Unlike intentional crimes, in negligent crimes, no one actively pursues the outcome. Therefore, it is impossible to attribute responsibility based on the "who controls the outcome" logic. For the criminal liability of the risks of criminal tools used by artificial intelligence entities, the criminal responsibility for intentional or negligent crimes should be pursued separately by the algorithm designers, product manufacturers, and management users, depending on the objective form of the criminal act committed by the offender and the form of their subjective fault[15]. Specifically, before the algorithm is delivered, the developers bear the obligation of conducting security tests and detecting deviations; during the deployment and acceptance process when the algorithm is launched, the deployers and acceptors have the obligation of conducting compliance reviews and tests; during the operation of the decision-making process, the users have the obligation of monitoring and responding to anomalies. In addition, when determining whether the reviewers have violated the duty of care, it is necessary to specifically distinguish between two types of duty of care. One is supervisory negligence, where the person in the handling position fails to supervise and inspect the specific outputs and the operation personnel, resulting in the occurrence of harmful consequences; the other is management negligence, where the person with leadership responsibilities fails to establish deviation detection, manual review, and abnormal handling mechanisms at the organizational level,

ISSN: 2688-8653

resulting in the occurrence of harmful consequences. The former pertains to specific business and specific individuals, while the latter pertains to the security system itself. In supervisory negligence, the behavior of the actor and the result generally involve the behavior of the person being supervised, presenting a multi-causation-one-effect characteristic[16]. That is to say, when multiple faults accumulate, a simple joint liability determination should be avoided. For instance, in cases where the algorithm developer has flaws, the reviewer fails to detect them, and the user does not verify them either, each party should be questioned separately. Assuming that each party fulfilled their due diligence, would the same outcome still occur?

4.4. The reasonable control of the scope of criminal regulation by the principle of reliance

In automated decision-making, the actor often defends by claiming that the algorithm is trustworthy. The principle of reliance has been fully discussed in the research on autonomous driving. Some scholars believe that applying the principle of reliance to trust between humans and machines is an extension of the organizational model, and it will not dilute the responsibility[17]. The allocation of attentional obligations in the human-machine co-driving mode is based on the principle of reliance[18]. The principles of risk and reliance should be allowed to be incorporated into the substantive judgment of criminal liability[19]. It can be seen that the core of the reliance principle lies in the fact that when the system operates reliably, the actor has no reason to anticipate its failure, and thus the possibility of anticipation is excluded. However, the allocation of administrative resources by automated systems is public resources, and the consequences of errors are borne directly by the unspecified public. The reliance principle can be applied only on the premise that the algorithm has been legally approved and operates reliably; the applicable scope only covers regular outputs, and if the actor fails to review abnormal signals but does review, the reliance is from the very beginning non-existent; at the same time, the actor's review obligation should be commensurate with the impact of the decision on the rights and interests of the counterpart. Therefore, once an automated decision-making system has been legally approved, tested, and accepted, and has no known loopholes or abnormal records, the actor can perform their duties based on the reliance principle. If the biased results are caused by the hidden defects of the algorithm, no liability for negligence is incurred; however, if the actor discovers abnormalities or the decision involves significant public interests, the review obligation cannot be waived, and the actor still needs to bear the liability for negligence.

5. Conclusion

The automation of decision-making has shifted the mode of exercising public power from the traditional human-centered approach to a new human-machine integrated form. In fact, it has changed the manner and timing of power-for-money transactions, creating a structural tension in the traditional interpretation of bribery crimes in algorithmic scenarios. In response to this issue, criminal law cannot simply adhere to the traditional indulgence of possible crimes or become overly radical and fall into the quagmire of "science fiction law". The criminal evaluation of algorithmic-driven acts of benefiting others in automation decisions should be conducted from two dimensions: behavior characterization and liability attribution. At the qualitative level, as a type of actor, by embedding and deploying biased parameters in the algorithm, they essentially pre-emptively carry out official acts using the algorithm as a tool. The time of the bribery implementation act should be advanced to the time when the algorithm is launched. In the non-action type, tolerating algorithmic bias has the same homogeneity in terms of harm to legal interests as actively setting biased algorithms. "Benefiting others" can be achieved through an inactive manner, but it must be limited by the behaviorist having actual control over the algorithm and the existence of a generalized bribery intent. In the negligent

ISSN: 2688-8653

type of behavior, since there is no awareness of the exchange of power and money, bribery does not constitute a crime and may only bear the responsibility of dereliction of duty when the violation of the duty of care and significant losses coexist. At the liability attribution level, the attribution basis should shift from the control of the behavior to the control of the algorithm. In intentional crimes, the responsibility of the principal and accomplice is determined based on the will control, while in negligent crimes, the supervision and management responsibility is distributed based on the violation of the obligation. The embedding of algorithmic technology in public decision-making is an inevitable trend in the development of digital governance. The technology itself is a neutral tool, but the power rent-seeking and official illegal acts implemented through algorithms still require criminal law to respond promptly to solve the problems of criminal application brought about by the development of digital technology.

References

[1] Wang, Z., & Yan, J. (2021). The publicness of algorithmic administration in the era of artificial intelligence: From the perspective of human-machine relationship. Exploration, 4.

[2] Zhang, L. (2019). The rise, alienation and legal regulation of algorithmic power. Journal of Law and Commerce, 4.

[3] Lu, C. (2024). Institutional advantages and risk responses of automated administrative licensing. Chinese Public Administration, 6.

[4] Huang, Z., & Wang, X. (2025). Legal pathways for building trust in algorithmic administration. Shanghai Jiaotong University Law Review, 5.

[5] Liu, X. (2025). Reanalysis of theoretical research on the evolution of criminal responsibility subjects in the era of artificial intelligence. Law Science, 9.

[6] Han, C. (2024). Public law governance of algorithmic purpose anomie in automated administration. Journal of Comparative Law, 2.

[7] Chen, Y. (2024). On the “human-in-the-loop” governance model for algorithmic decision risks in automated administration. Administrative Law Review, 4.

[8] Ma, H. (2024). Connotation characteristics, basic principles, and construction paths of a digital rule-of-law government. Journal of East China University of Political Science and Law, 3.

[9] Bai, Y. (2025). Review framework for the status of digital administrative organizations under law. Oriental Jurisprudence, 4.

[10] Wang, Y. (2024). Digital reform of administrative approval: Operational mechanism and rule-of-law guarantee. Chinese Public Administration, 6.

[11] Lin, Z. (1989). Chinese criminal law course. People’s Publishing House, 640–641.

[12] Wang, Z., & Chen, X. (1991). A new exploration of the composition of the crime of acceptance of bribes. Political Science and Law Tribune, 1.

[13] Zhang, M. (2004). On “seeking benefits for others” in the crime of acceptance of bribes. Political Science and Law Tribune, 5.

[14] Bai, L. (2025). The basic conduct for result imputation is negligent conduct: Advocating a hierarchical approach to imputation. Tsinghua Law Journal, 2.

[15] Liu, R., & Cao, B. (2021). Criminal risks and imputation of artificial intelligence agents. Jiangxi Social Sciences, 8.

[16] Xie, X. (2016). “Two-step judgment” of causation in supervisory negligence. Politics and Law, 5.

[17] Yang, N. (2025). The negligence liability of safety officers when human-machine shared driving fails: From the perspective of the principle of reliance. Jurist, 2.

[18] Zhou, M. (2026). On the allocation of duty of care in human-machine shared driving mode for autonomous vehicles. Politics and Law, 7.

[19] Chen, Q., & Sun, D. (2025). Negligence imputation for traffic accidents involving autonomous vehicles: Difficulties, pathways, and limitations. Jiangxi Social Sciences, 11.