

# **Quantitative Assessment Model for Social Risk of Arrest: Principles, Empirical Data, Challenges, and Optimization Paths**

MENG JUN<sup>1</sup>, ZHANG ZETIAN<sup>2</sup>

<sup>1</sup> School of Law, Beijing Normal University,

<sup>2</sup> School of Criminal Justice, China University of Political Science and Law

**Abstract:** The general requirements for arrest include evidentiary requirements, criminal elements, and the social danger factor, with the latter being pivotal in approving or issuing an arrest decision. Current legal provisions regarding the social danger factor for arrest suffer from issues such as abstract wording, inconsistent evaluation criteria, and ambiguous application conditions. To ensure more equitable, objective, scientific, and standardized application of arrest measures, judicial authorities have introduced a quantitative assessment mechanism for the social danger factor in judicial practice. Analysis of historical data, pilot programs, and international experiences reveals that enhancing the objectivity of assessment factors and weighting assignments, along with timely updates to sample data to balance quantitative evaluation with discretionary discretion, represent the primary directions for future optimization of the quantitative assessment model.

**Key words:** Arrest; Social danger; Quantitative assessment; Algorithmic regulation

## **Introduction**

With the evolution of China's crime structure and lenienter sentencing practices, the deepened implementation of the balancing leniency and severity criminal policy and the strengthened emphasis on judicial human rights protection have become pivotal guiding principles for reforms in the criminal justice system in the new era. The arrest measure, which restricts personal freedom and carries penalties comparable to other custodial sentences such as detention and fixed-term imprisonment, has seen its application curtailed amid the rising proportion of minor offenses---a development aligning with China's legal system advancement trends (Zhuang et al., 2021). This policy has yielded significant results, with China's arrest rate declining annually. However, compared to some international jurisdictions, China's arrest rate remains relatively high. The reasons include: first, law enforcement personnel generally prioritize crime control over human rights protection, viewing arrest as a primary tool for crime punishment; second, insufficient comprehensive understanding of arrest criteria---overemphasis on evidence proving criminal facts and crimes punishable by imprisonment or heavier penalties while neglecting social danger assessment, leading to arrest upon conviction; third, the abstract nature of social danger criteria requiring discretionary judgment based on case experience, resulting in inconsistent and

superficial assessments. To address this, legislators aim to refine social danger criteria through specific scenario listings. However, such enumerative approaches cannot exhaust all practical possibilities, suffering from persistent issues of imprecision and low operability, with subjective factors still decisively influencing arrests (Wan, 2021). Judicial authorities have introduced quantitative assessment models for social danger evaluation to enhance consistency, fairness, and accuracy in risk assessments, thereby reducing pre-trial detention rates. The quantitative assessment of social danger in arrest cases exhibits relative uniformity and standardization, reducing discrepancies in judicial personnel's interpretation of regulations and enhancing consistency in social danger evaluations. As early as 2005, the Shanghai People's Procuratorate established a feasibility assessment system for non-custodial measures for minors; however, initial scoring relied heavily on empirical judgment and subjectivity, failing to fully leverage the objectivity and scientific rigor inherent in quantitative assessment. Theoretical research has predominantly focused on qualitative approaches, with limited quantitative studies utilizing statistical methods, yet these efforts have guided continuous improvements to quantitative social danger assessment models (Wang, 2016). In 2021, the Supreme People's Procuratorate launched a nationwide pilot program for quantitative social danger assessment in arrests, pioneering innovative methods to evaluate suspects' social dangers more accurately, scientifically, and systematically.

## **I. Background and Principles of the Quantitative Assessment Model for Social Danger in Arrests**

### **1.1 Conditions for the Application of Arrest Measures**

The 2012 amendment to the Criminal Procedure Law classified arrests into three types: general arrest, immediate arrest, and converted arrest. General arrest requires three elements: evidentiary basis (existence of evidence proving criminal facts), criminal nature (the potential for a sentence of imprisonment or heavier), and social danger assessment (the inability of bail to effectively prevent social harm). Among these, the evidentiary and criminal elements serve as foundational requirements, while the social danger assessment constitutes the core criterion. The primary purpose of arrest is to ensure smooth litigation proceedings rather than conviction; thus, the evidentiary requirement should not be the central basis for arrest (Sun, 2017). The criminal element acts as a screening mechanism for assessing social danger. When the estimated sentencing range does not warrant imprisonment or heavier penalties, the law automatically presumes that alternative non-custodial measures can prevent social harm, thereby excluding the application of arrest. Consequently, the social danger assessment element is pivotal in limiting state power and implementing the principle of proportionality in criminal justice, warranting heightened attention

(Chen, 2019).

## 1.2 The Element of Social Danger in Arrest

According to the 1996 Criminal Procedure Law, three conditions must be met for an arrest: factual evidence, potential criminal liability, and the necessity of arrest, with social danger serving as a key criterion for determining arrest necessity. Due to the overly abstract legislative wording regarding the necessity of arrest, practical implementation faced challenges. The 2012 amendment refined the criteria for social danger into five specific scenarios, replacing the phrase necessity of arrest with an emphasis on social danger. The 2018 amendment further expanded the scope of social danger considerations to include the nature of the alleged crime, relevant circumstances, and the accused's plea and acceptance of punishment. Subsequently, the Supreme People's Procuratorate and the Ministry of Public Security issued the Provisions of the Supreme People's Procuratorate and the Ministry of Public Security on Several Issues Concerning the Conditions for Arresting Individuals with Social Danger (For Trial Implementation) (hereinafter referred to as the Provisions), along with related judicial interpretations, to provide detailed guidance.

The term social danger is mentioned in security measures, community correction programs, and criminal compulsory medical treatment procedures. Specifically, arrest-based social danger refers to the likelihood that the accused will commit new crimes, destroy or fabricate evidence, obstruct witness testimony, or engage in other socially harmful acts (Wang, 2018). Strictly speaking, there are no clear boundaries for arrest-based social danger; although the Criminal Procedure Law lists five specific scenarios, its practical application covers a broad range of situations (Chen, 2013). The criteria for arrest-based social danger require not only that the accused poses a social threat but also that non-detention measures such as bail are insufficient to prevent such threat. This concept emphasizes a dynamic state of danger that evolves over time, characterized by its potential for occurrence and ongoing nature (Chen & Tang, 2023). The elements of social danger encompass both substantive risks (e.g., recidivism or endangering public safety) and procedural risks (e.g., evidence tampering, escape, or suicide that hinder legal proceedings).

## 1.3 Principles of the Quantitative Assessment Model for Social Danger in Arrests

In the early stages of China's practice, the assessment of social danger during arrest primarily relied on the operational experience of case handlers. Although legislation regarding social danger has become increasingly refined, the evaluation factors, methods, and outcomes in practice remain highly subjective and arbitrary. To standardize the criteria for assessing social danger during arrest and enhance the scientific rigor and objectivity of the process, judicial practice has gradually

incorporated mathematical statistical methods into the assessment framework, integrating big data analytics and artificial intelligence technologies. The predominant quantitative model currently employed involves numerically quantifying the impact of various risk factors and then applying statistical methods to calculate the probability of a defendant exhibiting social danger behavior (Wang, 2016). Overall, social danger assessment systems can be categorized into comprehensive assessment systems and specialized assessment systems: the former provides an overarching evaluation of all relevant factors, while the latter focuses on specific aspects such as flight risk or recidivism risk. The comprehensive system offers the advantage of addressing all scenarios simultaneously, offering streamlined operational procedures and facilitating thorough evidence review, though it requires more sophisticated assessment scales. The specialized system delivers more precise predictions and facilitates clearer justification of conclusions. Chinese judicial practice predominantly adopts the comprehensive assessment approach. A complete social danger assessment system for arrest comprises four key components: evaluation criteria, weight assignment mechanisms, risk assessment algorithms, and the application of assessment results.

The evaluation criteria refer to the key factors guiding the implementation of arrest measures, with the most critical basis being legal provisions concerning social danger. Given that these provisions still contain terms such as possible and attempt, along with a catch-all clause for other potential scenarios, the practical experience of case handlers and international best practices serve as vital supplements to the evaluation framework. Weight assignment involves quantifying the specific impact of each evaluation criterion, typically manifested as risk coefficients in risk assessment sections. During the initial stages of quantitative social danger assessment models, empirical scoring was predominant---human judges assigned scores to each criterion before summing them to determine the individual's risk level. Subsequently, statistical methods emerged to calculate precise numerical values reflecting each criterion's influence on arrests. While statistical approaches are more scientific and objective than empirical rules, they cannot fully replace subjective judgment. Risk assessment algorithms utilize established criteria and weights to predict a case's social danger level. With advancements in big data and artificial intelligence, more sophisticated technologies may be integrated into future quantitative assessment models. Regarding the application of evaluation results, key considerations include how investigators effectively utilize these findings, establish appropriate arrest thresholds, and safeguard the rights of individuals assessed. Evaluation outcomes generally do not directly warrant arrest orders but serve as references---they reduce operational complexity for investigators without replacing their final decision-making authority. There are primarily two approaches to establishing the standard criteria for arrest: one uses a specific numerical threshold, where arrest is unnecessary if the risk

value falls below this threshold; the other involves categorizing risk levels, requiring arrest for individuals classified as high-risk. Regarding the remedy procedures for assessment results, Article 4 of the Regulations stipulates that public security authorities may verify relevant evidence through methods such as interrogating criminal suspects, questioning litigation participants including witnesses, and consulting defense attorneys.

## **II. Developing a Quantitative Assessment Model for Social Danger in Arrests Based on Existing Experience**

### **2.1 Historical Experience in Developing Quantitative Evaluation Models**

In judicial practice, early attention has been paid to the critical role of social danger assessment models in arrest decisions, with attempts to adopt mature evaluation frameworks from other jurisdictions. In 2005, the Minhang District People's Procuratorate in Shanghai referenced the UK's bail procedure risk assessment mechanism to establish a feasibility evaluation system for non-custodial measures for minors. The Taiyuan Municipal People's Procuratorate in Shanxi Province categorized arrest-related social danger assessment factors into individual and common factors. The Taizhou Municipal People's Procuratorate in Jiangsu Province required judicial personnel to provide reasoned justification for specific social danger scenarios, offering insights into whether categorical assessments should be incorporated into social danger evaluation systems. During the initial phase, China's arrest social danger assessment models primarily relied on empirical experience and legal provisions to determine evaluation factors, scoring criteria, and outcome classification standards, marking a preliminary exploration of such assessments. Statistical methods were employed to identify key assessment factors by screening among existing options, effectively avoiding the challenge of finding specific criteria within extensive datasets and reducing screening complexity. These efforts have accumulated valuable experience for developing a comprehensive arrest social danger assessment system in China.

### **2.2 Pilot Experience in Developing a Quantitative Evaluation Model**

To explore the practical application of social danger assessment in arrests, the Supreme People's Procuratorate launched pilot programs for a quantitative social danger assessment mechanism in 2021 across 11 provinces and municipalities---including Beijing, Hebei, and Shanxi---encompassing 38 pilot units. Through these pilots, several procuratorates have established relatively comprehensive assessment mechanisms with positive outcomes. Work reports from multiple regional procuratorates indicate declines in arrest reporting rates, non-arrest rates, and pre-trial detention rates, alongside an increase in the proportion of non-arrest decisions based on social danger assessments.

Pilot units primarily employ two approaches for social danger assessment. The first method follows the historical model described earlier: a Quantitative Social Danger Assessment Form is first developed, with relevant options selected based on actual circumstances, and the assessment score calculated by summing or subtracting corresponding values. Representative examples include the Measures for Quantitative Social Danger Assessment of Common Offenses issued by the Wuwei Municipal Procuratorate in Gansu Province and the Guidelines for Evaluating Suspects' Social Danger Before Arrest (Trial) implemented by the Suzhou Industrial Park Procuratorate in Jiangsu Province. The second approach utilizes statistical software or other information technology tools to analyze and calculate factors, determining their weights and risk threshold levels. For instance, the Nansha District Procuratorate in Guangzhou requires prosecutors to input case details into the system, which then automatically generates an arrest recommendation report using specialized algorithms (Zhong & Zhao, 2020). Similarly, the Tongzhou District Procuratorate in Beijing directly outputs suspects' risk scores and classification levels through its assessment system (Ma, 2023).

The classification and selection of evaluation factors vary across regions, with some areas encompassing as many as 207 factors while others include only 25. Despite these regional differences in factor selection, the social danger assessment criteria proposed by procuratorates within the same province or municipality are generally similar. Overall, categorizing evaluation factors remains the predominant approach, covering aspects such as personal danger, social harm, and litigation controllability, as well as personal danger factors, criminal factors, and factors hindering litigation. Additionally, some procuratorates classify factors based on legal provisions---such as guilty plea and acceptance of punishment or prior criminal records---or examine elements in chronological order, including criminal circumstances, post-crime conduct, and conditions for non-custodial compulsory measures.

### 2.3 Developing a Quantitative Evaluation Model Based on International Experience

Risk assessment tools can be categorized by evaluation content into violence risk assessment tools, sexual violence risk assessment tools, general offender risk assessment tools, and adolescent risk assessment tools; by evaluation methodology into clinical assessments, actuarial assessments, and structured expert assessments (Wen, 2014). Unlike domestic assessment systems, international risk assessment tools typically have clearly defined target populations and evaluation domains---for example, the HCR-20 and the Violence Risk Assessment Guidelines (VRAG) are designed for violence crime assessment, AssetPlus targets adolescents aged 10--17, and the most widely used COMPAS consists of specific risk prediction models, including general recidivism risk prediction and violence crime recidivism risk prediction (Northpointe, 2024). These tools

combine the efficiency and comprehensiveness of integrated assessment systems with the accuracy of categorical assessment systems. International risk assessment tools find extensive application, playing crucial roles in bail decisions, sentencing support, and parole determinations.

The development of pre-trial risk assessment outside the United States has undergone four distinct stages. The first generation of risk assessment tools emerged between the 1920s and 1970s, employing methods known as clinical assessment. Judges and relevant professionals---such as psychologists, psychiatrists, and social workers---made qualitative judgments that were significantly higher, close to the safety threshold, or acceptable, based on their expertise, adjudicative experience, clinical diagnoses, and unstructured interviews (Shi & Zhou, 2017). Clinical assessment typically lacked standardized evaluation criteria and relied heavily on the individual competence of assessors, raising questions about the accuracy and consistency of the results. Nevertheless, the risk assessment philosophy and methodologies developed during this period laid the foundation for modern risk assessment tools. The second generation of risk assessment tools appeared in the 1970s, emphasizing the use of statistical data and mathematical models, with greater reliance on actuarial and evidence-based approaches. These tools quantified the impact of risk factors into numerical values, which were then input into specific systems to generate risk outcomes. A representative example of this generation is PTRAs, which employ a Logistic regression model incorporating 26 risk factors---including violent crime, cannabis abuse, and unemployment---along with corresponding risk coefficients. Implementation requires simply determining whether any of these 26 scenarios occur in a case; a occurrence is assigned a value of 1, while absence is assigned 0. The model categorizes risks into five levels, each accompanied by corresponding intervention recommendations (Zhang & Liang, 2010). Compared to the first-generation risk assessment tools, the second generation demonstrated enhanced practicality and accuracy. The third-generation risk assessment tool, developed in the 1990s, incorporated dynamic evaluation factors based on the second-generation framework. These dynamic factors---such as the suspect's poor work attitude, relationships with family and friends, and unhealthy living environment---reflect evolving circumstances that guide criminal rehabilitation efforts. By modifying these factors, assessments can not only accurately evaluate pretrial risks but also enable targeted interventions for effective rehabilitation outcomes. Some scholars refer to these factors as criminogenic need factors (Wen, 2014). The third-generation tool employs a structured expert assessment approach that emphasizes individual characteristics over group data, combining elements of clinical and actuarial evaluations. The fourth-generation risk assessment tools share similarities with or even overlap with the third generation; while the latter's dynamic factors provide guidance for rehabilitation, they lacked specific intervention methods. The fourth

generation addresses this gap by integrating risk management with risk assessment. The AssetPlus pre-release risk assessment tool used by the Youth Justice Board (YJB) in England and Wales exemplifies this fourth-generation approach, comprising three modules: risk factor collection and description, behavioral analysis and risk evaluation, and post-release intervention and supervision recommendations. Each module is independently updated, allowing continuous risk assessment based on real-time information (Shi & Zhou, 2017).

The practice of extraterritorial risk assessment has a long history and has accumulated extensive experience. Reviewing the evolution of risk assessment tools, the methodologies have progressed from clinical evaluations and actuarial assessments to structured expert assessments that leverage the complementary strengths of both approaches; the evaluation criteria have shifted from relying solely on empirical intuition and statistical actuaries to integrating both approaches; assessment factors have evolved from random, fixed static elements to a combination of static factors and diversified dynamic descriptors; and the objectives have expanded from merely obtaining risk assessment results to achieving dual goals of risk identification and management. Although risk assessment tools have developed in these directions, this does not imply that the latest methods represent the optimal choice---they reflect rather diverse national approaches. Existing tools each possess distinct advantages and limitations: clinical assessments focus on individual suspects and defendants, offering high flexibility but lacking uniformity and consistency; actuarial assessments are simple to implement and efficient, yet require sophisticated scale development and prioritize population-level data, often overlooking individual-specific circumstances; structured expert assessments balance common and unique characteristics with standardized criteria to ensure reliability and consistency, though they demand deep knowledge of suspects' backgrounds and are less suitable for widespread adoption (Wen, 2014). A comprehensive approach that flexibly combines various assessment methods, considering both judicial practices and methodological strengths, proves to be the most viable pathway for effective risk assessment.

### **III. Practical Challenges in China's Existing Quantitative Assessment Model for Social Danger of Arrests**

As the Supreme People's Procuratorate's Pilot Program for Quantitative Assessment of Social Danger continues to advance, various regions across the country have successively implemented quantitative assessment schemes for assessing social danger. While drawing on China's experience, these efforts also reveal numerous shortcomings.

### 3.1 Differences and non-standardization in evaluation factors and weight assignment

First, there are regional disparities in evaluation criteria and weight assignments. One of the primary objectives of social danger assessment models is to standardize arrest risk evaluation criteria and limit judicial discretion. However, procuratorial authorities across regions differ not only in the variety and quantity of evaluation factors but also in their weight assignments and result calculation methods. In terms of factor categories, Beijing categorizes them into three groups---personal danger, social harm, and litigation controllability---with a total of 25 items; Shanghai's 206 System identifies 32 arrest risk evaluation indicators; while Guangzhou's Nansha District divides them into three components---personal danger, social harm, and litigation controllability---with 43 indicators. For weight assignment and result calculation, Beijing and Guangzhou utilize big data and statistical software to derive coefficient values and risk thresholds for each factor, creating data-driven quantitative assessment models. Prosecutors simply input actual data to obtain results through specialized algorithms. Chengdu's Jinniu District, meanwhile, establishes a quantitative assessment matrix specifying scoring criteria for each dimension, enabling direct calculation of social danger scores (Ma, 2023; He, 2021). These regional variations directly lead to inconsistent social danger evaluations, contradicting the goal of unified risk assessment and undermining consistency across jurisdictions.

Second, the evaluation factors and their weighting assignments exhibit significant subjectivity. Regarding evaluation criteria, incorporating sentencing factors such as voluntary surrender, truthful confession, and meritorious service undermines the objectivity of social danger assessment during arrests (Wu, 2023). Taking voluntary surrender as an example: According to Article 67(1) of the Criminal Law, voluntary surrender requires simultaneous fulfillment of voluntary surrender after committing the crime and truthfully confessing one's offenses, which involves both objective actions and subjective determination by case handlers. Moreover, voluntary surrender outcomes are categorized as mitigated punishment or reduced punishment; the specific degree of mitigation under mitigated punishment introduces subjectivity into factor weighting within the quantitative social danger assessment system. Beyond these inherently subjective evaluation factors, certain regions base their selection and weighting of objective assessment factors on intuition and experience, lacking sufficient data to establish clear risk thresholds for arrests. Taking Beijing, Guangzhou, and Chengdu as examples: Beijing's evaluation standards and weighting were derived from over 3,000 cases in its prosecutorial system, with statistical software used to determine risk level thresholds. Guangzhou's Nansha District established evaluation factors and weights through statistical analysis and empirical research. Chengdu's Jinniu District, however, provided no detailed explanations regarding its selection

criteria, weighting standards, or minimum score thresholds for arrests---its primary intent being to compel public security authorities to conduct social danger assessments. The quantitative assessment process lacks data support, which reduces the referential value of the social risk assessment in the region and may even render it merely formalistic, thereby having little impact on enhancing the objectivity, accuracy, and fairness of social risk evaluations.

Third, certain evaluation criteria lack legal coherence. According to the Criminal Procedure Law's provisions on direct arrest, individuals meeting the criteria of evidence proving criminal facts, potential sentencing of imprisonment or heavier penalties, and prior intentional crimes or unidentified status should be arrested immediately. This regulation has sparked widespread controversy: On one hand, the irrefutable presumption of social danger derived from prior intentional crimes raises doubts regarding its rationality, as such cases are highly complex; uniform arrest would overlook individual circumstances, undermining substantive justice and violating the principle of proportionality. On the other hand, compared to the recidivism provisions in the Criminal Law and the five-year restriction on prior intentional crimes under the criminal reconciliation procedure in the Criminal Procedure Law, the interval between offenses and the nature of penalties should be specified---this provision might even be eliminated in the fourth amendment to the Criminal Procedure Law (Chen & Cui, 2023). However, no clear legal norms or judicial interpretations currently support this approach. Some regions use terms like recidivism or prior convictions---which overlap with prior intentional crimes---as social danger assessment criteria for general arrest, creating inconsistencies with statutory requirements.

### 3.2 The representativeness and timeliness of the model sample data need to be enhanced

First, the sample data volume and representativeness are insufficient. To maintain the objectivity of the quantitative assessment system for arrest-related social danger, it is essential to ensure that the selection of evaluation criteria and weight assignment are supported by adequate sample data, thereby ensuring statistically significant results. Insufficient or non-representative sample data significantly compromise the accuracy and reliability of the outcomes. For less common offenses such as crimes endangering national security or crimes endangering national defense interests, an inadequate sample size prevents the assessment model from fully revealing underlying data patterns, thereby undermining its effectiveness. Insufficient sample representativeness means the sample fails to accurately reflect the true distribution or key characteristics of the overall population, leading to oversight of distinctive patterns in minority cases and resulting in biases in their assessment and prediction. Some research and development efforts for quantitative assessment systems for arrest-related social danger rely on extensive electronic case files, yet critical risk assessment information---such as personal details like

gambling, alcohol abuse, or mental health issues, as well as litigation obstruction history or court absence records---may not be contained within these documents. Additionally, judicial personnel's descriptive summaries and the system's keyword identification processes can result in the loss of certain information (Gao, 2021).

Second, the information lacks dynamic adjustment mechanisms. The assessment of social danger during arrests is time-sensitive; since it relies on population data from specific periods, it only reflects the behavioral characteristics of individuals under investigation during those periods. Therefore, sample data must be updated in real-time to ensure reliable assessment outcomes (Kenschall, 2009). Since 2019, Guangzhou's Nansha District has pioneered the development of a quantitative social danger assessment model, serving as a national benchmark. As documented in the 2022 Typical Cases for Advancing the Judicial Accountability System (Part V) released by the Supreme People's Procuratorate, the Nansha District Procuratorate conducted a comprehensive review of arrest cases since 2019, optimizing the weighting of 15 evaluation criteria. Some existing quantitative assessment models for arrest-related social danger failed to account for this update, and relevant authorities did not promptly validate their results, leaving their actual effectiveness uncertain. However, recent years have witnessed significant shifts in crime patterns, particularly a sharp rise in telecom fraud cases. Continued reliance on outdated data would be akin to marking a boat to seek a sword---resulting in inadequate adaptation to current realities and diminished assessment accuracy. Thus, integrating these updates into future arrest-related social danger assessment frameworks is imperative.

### 3.3 Insufficient accuracy of the evaluation results

The quantitative social risk assessment system inherently possesses limitations: in both criminal justice and other risk prediction domains, nearly all risk assessment outcomes fail to achieve 100% accuracy. Taking the international COMPAS system as an example, ProPublica compared COMPAS's predictions for recidivism risk and violent recidivism risk with actual recidivism rates two years later, revealing a 62% accuracy rate for recidivism prediction versus only 20% for violent recidivism (Larson et al., 2024). In China, social risk assessments typically consider factors such as education level, gender, and place of residence when deciding on arrest warrants. However, the actual impact of these factors on social risk has not been scientifically validated. Directly incorporating such variables into assessment criteria perpetuates biases, leading to disproportionately high risk estimates for individuals with specific educational backgrounds, genders, or residences outside the case investigation area. Three primary reasons explain this phenomenon: first, the quantitative assessment system mirrors the decision-making logic and psychological processes of law enforcement officers, relying on existing judicial precedents and

cognitive frameworks regarding criminal behavior---thereby limiting its quantitative capabilities beyond practical and cognitive constraints (Gao, 2021). Secondly, quantitative assessment is based on extensive sample data, reflecting the behavioral characteristics of a population. Ignoring individual traits and special circumstances in specific cases is an inevitable consequence of identifying universal patterns; therefore, predictions for individual defendants may not be entirely accurate. Thirdly, quantitative assessment systems inherently carry value orientations---investigators may prioritize political correctness and social justice by omitting certain assessment factors that could lead to discrimination or bias, even when these factors exhibit significant statistical significance. From another perspective, statistically significant factors do not necessarily reflect validity; they may represent accumulated historical biases and discrimination, such as educational background, income level, or gender. These objective limitations on assessment accuracy cannot be altered, meaning that quantitative social risk assessment models can only improve accuracy through measures like enhancing sample representativeness, timeliness, and algorithmic refinements to improve data fit. Accepting these constraints and designing error-tolerant mechanisms are crucial for practical application. Additionally, existing arrest-related social risk assessment systems have not undergone accuracy validation, leaving investigators uncertain about the reliability of such assessments.

### 3.4 Lack of supporting programs for the evaluation system

One primary reason why risk assessment tools face skepticism outside their intended domains is the lack of transparency in their algorithmic decision-making processes, which deprives defendants of the right to challenge unfavorable evidence. Firstly, the arrest social danger quantification system is primarily developed by public authorities, creating an inherent disparity between defense attorneys' capabilities and those of these authorities from the outset, thereby diminishing lawyers' ability to advocate for coercive measures (Gao, 2021). Secondly, due to the complexity and opacity of the algorithmic decision-making process, parties involved---including defendants, their defense attorneys, and case handlers---have no means to comprehend or challenge the assessment methodology. Defense teams cannot effectively raise objections regarding potential biases in the system's decisions, while case handlers remain passively reliant on the quantitative social danger evaluations without adequate opportunity for explanation or modification (Li, 2018). This algorithmic assessment system, which offers neither transparency, accountability, nor remedies, has been dubbed the algorithmic tyranny (Jiang, 2020). These issues cannot be resolved through model optimization alone; it is imperative to grant defense teams the right to access assessment processes and raise objections externally.

## **IV. Optimization Pathways for China's Quantitative Assessment Model of Social**

## **Danger in Arrests**

### **4.1 Scientifically and rationally establish evaluation factors and their weight assignments**

#### **4.1.1 Unified evaluation by specialized agencies**

Establishing unified national standards for social risk assessment enhances the fairness of compulsory measures implementation, yet this does not imply disregarding local considerations shaped by economic and cultural differences. Variations in assessment criteria and methodologies across China stem from the fact that data for the quantitative arrest risk evaluation system are derived from local case files, reflecting distinctive regional factors in arrest decision-making and aligning assessment outcomes more closely with local judicial practices. The optimal approach balances preservation of regional characteristics with standardization of assessment procedures. To prevent biased regional standards, delegating this task to specialized institutions proves advisable. Such institutions can aggregate comprehensive data from nationwide agencies---including public security authorities, courts, and prisons---over specified years, incorporating factors like gambling, alcohol abuse, and mental health issues not typically documented in case files, thereby ensuring holistic risk evaluations. Moreover, technical experts within these institutions possess specialized expertise and can effectively leverage advanced technologies such as blockchain and cloud computing. The U.S.-based COMPASS risk assessment algorithm, developed by Northpointe, has been widely adopted by state courts, while Shanghai High People's Court's 206 system incorporates technologies from iFlytek. Consequently, cross-disciplinary collaboration between legal and technological professionals---and the outsourcing of certain tasks---has become an inevitable trend (Zheng, 2020).

#### **4.1.2 Reduce the subjectivity of evaluation factors**

As previously discussed, incorporating sentencing factors such as confession and voluntary surrender into the quantitative assessment of social danger carries significant subjectivity in the evaluation system. Therefore, it is advisable to isolate the objective behavioral components of sentencing factors and treat them independently as quantitative indicators of social danger. Taking voluntary surrender as an example, its purpose lies in encouraging the accused's remorse and reducing judicial costs. Thus, assessing the subjective intent of voluntary participation constitutes a crucial element in establishing voluntary surrender, though this intent must also be verified through objective evidence. For general voluntary surrender, two conditions must be met: voluntary surrender after committing the crime and full and truthful confession of the offense. The objective elements of voluntary surrender include the timing, recipient, and method of surrender, while the subjective intent can be demonstrated through objective evidence---such as whether the

suspect was bound, resisted arrest, or attempted escape. The purpose or motive behind the surrender need not be considered; only its voluntariness and initiative matter. The requirement for full and truthful confession comprises three objective criteria: completeness of statements (covering key criminal details relevant to the charges), timeliness (before the first-instance judgment), and consistency of confession (absence of subsequent retraction). The term truthful pertains to the objectivity of the confession, i.e., its consistency with subsequent investigative evidence. Although subjective intent can be assessed through objective evidence, case handlers must exercise discretion in special circumstances. For instance, if a suspect is apprehended by relatives or friends without resistance, voluntary surrender does not apply; however, if the suspect remains at the scene despite knowing someone has reported the crime and does not resist arrest, voluntary surrender is valid. Similarly, the requirement of meritorious service can be replaced by conditions such as exposing others' criminal acts with verified facts or providing crucial leads that facilitate the resolution of other cases; whereas the requirement of legitimate self-defense can be substituted by elements including being subjected to unlawful aggression, acting to protect one's own or others' personal or property safety, and the aggressor being the perpetrator of the unlawful act (Gao, 2021).

#### 4.1.3 The established evaluation criteria conform to legal logic.

Many quantitative assessment systems include recidivism, prior convictions, and prior administrative penalties as evaluation factors, yet this approach does not align with legal logic. Arrests are categorized into general arrest, conversion arrest, and direct arrest, with prior intentional crime being one of the conditions for direct arrest. According to Article 65 of the Criminal Law, which defines recidivism, it constitutes a subordinate concept to the direct arrest requirement of prior intentional crime. Therefore, individuals with recidivism who meet both evidence proving criminal facts and the potential for a sentence of imprisonment or more severe punishment should be subject to immediate arrest. Prior convictions, generally understood as previous criminal convictions---including both negligent and intentional offenses---are broader than prior intentional crime and require further refinement (Yu, 2002).

#### 4.2 Enhancing the Representativeness and Timeliness of Model Sample Data

To enhance the quality of information extraction, two key approaches are essential. First, improve the representativeness of sample data by selecting cases that accurately reflect judicial practice for model development. Second, focus on refining extraction precision through advanced information extraction techniques that improve the accuracy of converting case file information into machine-readable formats. Additionally, investing human resources to assist machine

recognition can strengthen both comprehensiveness and accuracy of information extraction. Statistical analyses of sample data over a specific period reveal general patterns of social risk behaviors among individuals under prosecution. By continuously monitoring data drift and implementing periodic incremental learning to update sample datasets while discarding outdated information, we ensure timely assessment results and enhance predictive model accuracy.

#### 4.3 Rational Application of Assessment Results

Given the inherent limitations of quantitative assessments---such as their inability to account for individual circumstances and achieve absolute accuracy---they require supplementation with qualitative evaluations, which explains the necessity of expert assessments conducted by structural specialists outside the legal framework. Moreover, judicial personnel cannot rely solely on conclusions derived from quantitative social risk assessment systems; while these results serve as crucial reference indicators, they do not directly determine arrest decisions. First, in practice, no two cases are identical; each involves unique circumstances beyond statistical analysis, requiring judicial personnel to exercise subjective judgment regarding specific case details and the individual characteristics of the accused (Wang, 2016). Second, computer data can only represent risks using binary 0 or 1 values, whereas real-world situations often defy such simplistic categorization; judicial personnel must employ more nuanced descriptive approaches to enhance qualitative social risk assessments (Shi & Zhou, 2017). Finally, when a particular assessment factor's sample data significantly underestimates its impact compared to others, the validity of its coefficient becomes questionable, necessitating final judgments based on judicial expertise and analytical reasoning.

#### 4.4 Safeguarding the Defense Counsel's Right to Information and Right to Express Opinions

Given the inherent limitations of quantitative social risk assessment models that cannot be fundamentally resolved, granting the accused a certain degree of right to information regarding the assessment system and the right to challenge unfavorable results can help mitigate issues such as the algorithmic black box, algorithmic tyranny, and implicit algorithmic discrimination from external perspectives. Since arrest entails severe consequences involving restrictions on personal freedom, the procedural fairness of arrest decisions must be fully ensured to prevent the indirect deprivation of the accused's right to defense due to opaque assessment processes, lack of transparency, absence of explanations, or inability to challenge the results. In practice, case handlers may conduct detention hearings in accordance with the Measures for Detention Hearings by the People's Procuratorates, safeguarding the defense party's right to present arguments and thoroughly reviewing the defense arguments regarding the social risk assessment outcomes.

### Author Profile

Meng Jun, Associate Professor at the School of Law, Beijing Normal University; Zhang Zetian, Master's Student at the School of Criminal Justice, China University of Political Science and Law.

### REFERENCE

- Zhuang, Y. L., et al. (2021). Connotation, function and implementation of the criminal justice policy of less arrest, less prosecution and less detention. *People's Procuratorate*, 15, 37--40. (In Chinese)
- Wan, Y. (2021). Interpreting three key words of arrest system: "social risk," "necessity of arrest," and "necessity of detention". *Chinese Journal of Criminal Law*, 4, 70. (In Chinese)
- Wang, Z. H. (2016). Principle and construction of quantitative assessment model for social risk in examination and approval of arrest. *Tribune of Political Science and Law*, 2, 73--74. (In Chinese)
- Sun, Q. (2017). Studies on several issues of arrest under the background of judicial reform. *China Legal Science*, 3, 31. (In Chinese)
- Chen, W. D. (2019). Theory and practice of custody necessity review system. China Legal Publishing House. (In Chinese)
- Wang, A. L. (2018). Interpretation of revised provisions of the Criminal Procedure Law of the People's Republic of China. China Legal Publishing House. (In Chinese)
- Chen, Y. S. (2013). China's problems of arrest and institutional responses: Centering on the revision of the arrest system in the 2012 Criminal Procedure Law. *Tribune of Political Science and Law*, 4, 23. (In Chinese)
- Chen, P., & Tang, Y. F. (2023). Reshaping understanding of arrest under the background of new era justice. In *Shanghai Law Research* (Vol. 10). Shanghai People's Publishing House. (In Chinese)
- Zhong, Y. Y., & Zhao, Y. (2020). Exploration on establishing a quantitative assessment mechanism for social risk: Nansha District, Guangzhou, Guangdong. Retrieved from [https://www.spp.gov.cn/spp/zd gz/202010/t20201010\\_481455.shtml](https://www.spp.gov.cn/spp/zd gz/202010/t20201010_481455.shtml)
- Ma, L. J. (2023, March 22). Social risk assessment speaks with data. *Procuratorate Daily*, 001. (In Chinese)
- Wen, J. (2014). Research on personal risk assessment methods. China University of Political Science and Law Press. (In Chinese)
- Northpointe. (2024). Practitioner's guide to COMPAS core. Retrieved from <https://assets.documentcloud.org/documents/2840784/Practitioner-s-Guide-to-COMPAS-Core>

- Shi, L. M., & Zhou, Y. (2017). Risk investigation and assessment in criminal justice: From the perspective of pretrial release. In *Criminal Law Review* (Vol. 4). Law Press. (In Chinese)
- Zhang, J. X., & Liang, X. H. (2010). US Department of Justice pretrial risk assessment model and its enlightenment to China. *Chinese Journal of Criminal Law*, 7, 106--107. (In Chinese)
- He, Y. (2021). Significant year-on-year decrease in pretrial detention rate: What tricks does this pilot procuratorate have? Retrieved from <https://sichuan.scol.com.cn/ggxw/202107/58221925.html>
- Wu, Q. B. (2023). Research on optimization of quantitative assessment model for social risk examination of arrest. *Journal of Southwest University of Political Science and Law*, 5, 93--94. (In Chinese)
- Chen, W. D., & Cui, K. P. (2023). Reflection and construction of applicable conditions of arrest: From "general arrest" and "direct arrest". *Journal of Shanghai University of Political Science and Law*, 5, 31. (In Chinese)
- Gao, T. (2021). Research on quantitative assessment of social risk of arrest: From the perspective of automated decision-making and algorithm regulation. *Northern Legal Science*, 6, 136--137. (In Chinese)
- Kenschall, H. (2009). *Interpreting risk in criminal justice* (W. Wang, Trans.). China People's Public Security University Press.
- Larson, J., et al. (2024). How we analyzed the COMPAS recidivism algorithm. Retrieved from <https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm>
- Li, B. (2018). Artificial intelligence in American judicial practice: Problems and challenges. *China Law Review*, 2, 56. (In Chinese)
- Jiang, S. (2020). Automated decision-making, criminal justice and algorithm regulation: Reflections triggered by the Loomis case. *Oriental Law*, 3, 77. (In Chinese)
- Zheng, X. (2020). Application and regulation of artificial intelligence technology in judicial adjudication. *Chinese Journal of Comparative Law*, 3, 683. (In Chinese)
- Yu, Z. G. (2002). On the definition and connotation of criminal record. *Zhejiang Social Sciences*, 2, 112. (In Chinese)
- Min, F. J. (2018). Problems and solutions to the litigation-based examination and approval mechanism for arrest. *Western Legal Science Review*, 4, 55--57. (In Chinese)
- Cheng, G. (2023). Review and reconstruction of China's criminal compulsory measure system. *Journal of Hubei Police University*, 5, 145. (In Chinese)

which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited.

