

The Burden of Proof Dilemma and institutional Improvement of AI Infringement of Personality Rights under the Black Box of Algorithms

Qian Xu

The National Police University for Criminal Justice, Baoding, Hebei, China

Abstract: With the rapid development of deep intelligent production, the infringement of personality rights such as portrait rights, name rights, reputation rights and personal information rights by artificial intelligence products has gradually increased. The existence of algorithmic black boxes leads to an asymmetry in the evidential capacity between the parties and the algorithmic developers, and the "he who asserts must prove" burden of proof allocation rule is difficult to adapt to the requirements of the trial practice of online personality rights cases. To sum up, the current academic research on issues such as the regulation of algorithmic black boxes, AI compensation subjects and distribution principles, and the burden of evidence mainly focuses on the above three aspects, but it lacks multi-perspective integrated research, does not explore the root causes comprehensively and deeply, and is too general in the design of specific rules, and does not closely combine theory with judicial practice. The author aims to focus on the core proposition of how to balance the unfair distribution of evidence due to technological superiority, avoiding the two extreme solutions of "full responsibility of the plaintiff" and "reversal of burden of proof", and adhering to "matching risk with control and matching proof with responsibility", Create a stepwise dynamic burden of proof allocation rule and system supporting measures and specific situation guidelines of "first stage of proof - second stage of explanation - third stage of discretion" to reduce the difficulty of proof for rights holders while overcoming the predicament of "difficulty in proof", and coordinate the relationship between algorithmic rights protection and algorithmic innovation through gradient algorithmic explanation obligations. Ultimately achieve a balance

between the comprehensive protection of personality rights and the sound development of the artificial intelligence industry.

Keywords: Algorithm Black Box; Artificial Intelligence Infringement of Personality Rights; Burden of Proof Dilemma; Dynamic Allocation of Burden of Proof

1. Introduction

With the rapid development of intelligent generation technology, artificial intelligence technology services have been widely applied in various fields of social life, and as a result, disputes over the infringement of the rights and interests of natural persons by artificial intelligence products have been increasing. From AI deepfake videos (audio) infringing on portrait rights and voice rights, to AI fabricating rumors damaging reputation rights, to intelligent chatbots leaking and abusing personal information infringing on privacy rights, AI has not only changed people's content production patterns, but also challenged the existing personality rights protection system.

In cases where AI infringes on personal rights, the unexplainability of the algorithmic system is a major obstacle for rights holders to assert their rights. The "technical black box" nature of machine learning algorithms, combined with the application of the principle of "non-disclosure of information" by algorithm operators, leaves the right holders with no knowledge of the internal operation of the algorithms and unable to fulfill the burden of proof for the two elements of subjective fault and causality in the composition of infringement. First of all, the traditional burden of proof allocation method of "he who asserts must prove" is difficult to apply to such cases. Although the academic community has recognized the challenge posed by the algorithmic black box to the traditional rule of evidence of "who asserts, who proves", the

related discussions have mostly been limited to proposing the idea that the scope of the burden of proof should be expanded, and have not attempted to provide a solution that is both theoretically justified and operable. Nor does it respond to the contradiction between "revealing algorithms to protect the rights of the parties" and "not disclosing algorithms to encourage technological innovation".

The research idea of this paper is: on the basis of a comprehensive summary of previous research results, to identify the deficiencies in existing research, and in combination with relevant laws and regulations and judicial practice, to explore the real predicament and multiple causes of the burden of proof in the case of infringement of personality rights in the AI black box era; Break through the traditional two-choice cognition and construct a threelevel dynamic burden of proof transfer model; Ultimately, formulate relevant rules and guidelines to resolve disputes over AI infringement of personality rights and promote the coordinated development of personality rights protection and intelligent technology regulation in the online environment.

2. Current Status of Academic Research and Problems to be Solved

2.1 Main Research Approaches and Schools of Thought

Based on the core issues and theoretical perspectives of the literature, this study divides existing academic achievements into four key areas to construct a complete analytical framework of academic development context.

First, basic theoretical research and governance model exploration on algorithmic opacity.

The algorithmic black box problem is a fundamental proposition in the legal regulation of artificial intelligence. On this issue, there are mainly two viewpoints in the academic circle regarding the governance path. The first is the critical governance dilemma theory (anti-mandatory disclosure theory), represented by Yuan Zeng[1], Mei Xiaying[2], etc., which holds that mandatory disclosure algorithms are neither advisable nor necessary. The reasons are that forced disclosure is difficult to implement in practice, violates the principle of technology neutrality, and has adverse effects on the relevant subjects, etc.[3] It is advocated to abandon the concept of "technology openness" and build a legal mechanism of "overall

controllability"; The second is the hierarchical control theory (conditional openness theory), led by Ling Qiushi[4], Liu Jian and Wu Licai[5], among others, which holds that algorithmic black boxes are an important reason for the governance difficulties of algorithmic discrimination. However, instead of simply demanding full openness, precise control should be achieved through the establishment of an algorithmic information reporting system, the improvement of regulatory communication mechanisms, etc., to ensure the supervisability of algorithmic operation.

Secondly, research on the legal nature and liability determination of personality rights infringement in the context of artificial intelligence.

As a major subject in the study of personality rights law in the Internet age, there are currently three major schools of thought: First, constitutive elements strictly speaking, represented by scholars such as Zhu Xiaofeng and

Li Yongjun[6], who adhere to traditional tort law theory, advocate the constitutive elements theory as the basis for liability, and strictly distinguish the differences between the right to claim protection of personality rights and the right to claim damages for infringement. The second is the innovation of the dissemination mechanism, represented by Yang Liu[7] and others. Without changing the traditional distribution pattern of liability centered on content expression, considering the characteristics of diversified tortfeasors, diverse dissemination methods and fast dissemination speed in the

Internet age, they propose that the tort liability system of personality rights should be constructed from the perspective of information dissemination; Third, contextualized liability determination, represented by Liu Wenjie[8], the research group of Beijing Internet Court[9], etc., based on the technical characteristics and application environment of artificial intelligence services, it rejects the extensive principle of imputation and advocates the establishment of multi-level and specific imputation principles for courts to refer to in individual cases.

Third, theoretical debate and institutional analysis on the allocation of burden of proof in cases of infringement of personality rights. This issue is an interdisciplinary topic between civil procedure law and personality rights law, and there are three major theoretical points of

contention: First, the dual attribution and stratification theory. Representative viewpoints such as Yang Xianbin[10] and Lyu Ziyi[11] advocate that a differentiated multi-level allocation model of burden of proof should be constructed based on different types of personality rights infringements and different attribution bases. The second is the debate between the "expansion and prudence" of the burden of proof inversion. The former holds that the burden of proof inversion should be extended to special tort areas where there is a significant imbalance in the burden of proof of the parties, while the latter, considering the risks that such expansion may lead to, holds that the relationship between procedural justice and substantive justice should be taken into account.[12] The fourth is the dynamic adjustment theory. Zhu Xiaofeng[13] et al. believe that in a lawsuit involving immaterial compensation for mental distress, the court can make appropriate adjustments to the allocation of burden of proof and the degree of proof based on the principle of interest measurement to suit the burden of proof among different parties.

Fourth, a study on the compatibility of the algorithmic black box phenomenon with the burden of proof for infringement of personality rights.

Algorithmic responsibility, as a core topic running through the entire text, has been put forward by existing research results in three categories of propositions on how to solve the proof problem: The first is the inverted burden of proof scheme proposed by the mainstream theory, that is, based on the asymmetry of the burden of proof ability of both parties caused by the algorithm black box, it is proposed to allocate the burden of proof of core facts to the algorithm operator; The second is the gradient of platform burden of proof theory, which denies the absolute inversion rule and advocates establishing a hierarchical burden of proof system based on the different functional positioning of AI platforms to balance the protection of algorithmic intellectual property rights and the relief needs of rights holders; Finally, there is the comprehensive governance supporting theory, which holds that merely adjusting the burden of proof rules is difficult to fundamentally solve the problem and requires the establishment of a system that combines pre-emptive prevention and post-emptive relief covering the entire process governance of

artificial intelligence.

2.2 The Limitations of Existing Research and the Innovative Perspective of This Paper

To sum up, academic research on algorithmic black box governance, artificial intelligence tort liability and burden of proof is abundant; However, there is a lack of specialized research on "the burden of proof for the infringement of personality rights by artificial intelligence under the algorithmic black box", and there are four shortcomings: this is also the entry point of this study.

The first is that the cross-research is not systematic enough, and the integration among the three core issues is not high. Most studies only focus on the governance of algorithmic black boxes or the determination of liability for artificial intelligence infringement or the construction of a burden of proof system in a single dimension. Only a few articles pay attention to the combination of algorithmic black boxes and burden of proof, lacking a comprehensive understanding of the relationship among the three, "the technical characteristics of algorithmic black boxes - the unique types of artificial intelligence infringement - the fit of the burden of proof system" There is no in-depth exploration of the impact of algorithmic black boxes on the applicable basis of traditional personality rights infringement burden of proof, and there is a lack of sufficient pertinence.

Second, there are obvious deficiencies in the analysis of the root causes of the burden of proof dilemma, and the theoretical applicability needs to be improved.

Existing research, when discussing the proof dilemma caused by algorithmic black boxes, merely attributes the cause to the complexity of the algorithm itself without a comprehensive analysis of the four aspects of technical attributes, normative requirements, subject characteristics, and institutional environment: In terms of technical attributes, it confuses the essence of "the algorithm itself is difficult to explain" and the difference of "artificially creating information asymmetry", and the regulation is too broad. From a legal perspective, the impact of the algorithmic black box on the proof procedures of various elements in infringement of personality rights is not considered;

From the perspective of the subjects, the differences in the ability to provide evidence

among the parties are not taken into account, and there is a lack of fairness.

In addition, the "mandatory disclosure + liability inversion" scheme, which is currently widely advocated in the academic circle, is essentially contradictory to the critical theory of algorithmic governance and has not yet formed a logically consistent theoretical system.

Thirdly, there is a lack of detailed operational guidance, which deviates to some extent from the scenario setting. Many research findings remain at the level of slogans such as "the scope of liability inversion should be further expanded" and "the burden of proof for the platform should be enhanced", without proposing different ideas for the allocation of burden of proof based on different types of liability disputes (such as different types of AI involving portrait rights, reputation rights, and even privacy rights); Nor did they provide clear and definite answers on how to determine whether an inversion of liability is true, how to reduce the difficulty for plaintiffs to prove, and the corresponding information disclosure obligations. Some institutional design proposals fail to strike a balance between protecting algorithmic intellectual property rights, encouraging technological innovation and providing relief to victims, which may impose an excessive liability burden on algorithmic developers and dampen their enthusiasm for technological innovation. [14]

Forth, there is a lack of guidance at the practical operation level and a lack of refinement of empirical rules. Most of the existing literature has adopted the method of normative analysis for discussion, and very few scholars have analyzed and discussed cases as samples. It has not systematically summarized the specific rules for the allocation of the burden of proof regarding the infringement of AI personality rights in China's judicial practice, nor has it further responded to the dilemma of "he who asserts must prove", the scope of "free judgment" authority, and the difficulty of presenting evidence. This results in an asymmetry between theoretical research and judicial practice, and the institutional recommendations are not very operational.

3. Difficult Evidence Presentation and Cause Analysis in Cases of AI Infringement of Personality Rights in the Context of Algorithmic Black Box

3.1 An Examination of the Current Norms and Judicial Practice Status

In terms of substantive law, China does not have special rules of proof for cases where AI infringes personality rights at present. Both the personality rights section and the tort liability section of the Civil Code apply the principle of fault liability; In terms of procedural law, the Civil Procedure Law and relevant judicial interpretations merely follow the general rule of "he who asserts must prove" and only stipulate the liability inversion system in cases of personal injury compensation disputes such as environmental torts and medical malpractice torts.[15] Secondly, the Interim Measures for the Administration of Generative Artificial Intelligence Services only stipulates the general compliance obligations of service providers and does not stipulate the relevant rules for reversing the burden of proof.

The existing legal framework fails to effectively address the structural imbalance in the burden of proof caused by algorithmic black boxes, resulting in a significant legal regulatory void.

In fact, there have been many cases in real life involving the protection of personality rights related to AI, such as the AI voice copyright case, the AI portrait rights case, and the AI reputation rights case, etc. In most of these cases, courts insist that the victims collect the evidence themselves and bear the responsibility for failing to provide evidence, and do not introduce a third-party technical appraisal mechanism to assist judges in ascertaining the relevant facts. In the vast majority of cases, the burden of proof was reduced at the discretion of the plaintiffs based on the principle of free evaluation of evidence, or the burden of proof was shifted to the network service providers to some extent. Overall, under the current laws of our country, there is no set of fixed rules for the allocation of liability, and the phenomenon of different judgments in the same case is widespread, resulting in the burden of proof for rights holders. Many cases have been lost due to the failure to fully prove fault and causality.

3.2 Core Problems in the Allocation of Burden of Proof

The algorithmic black box phenomenon has led to three major predicaments for traditional personality rights infringement proof rules in the field of artificial intelligence.

The first is the predicament in proving the elements of infringement: it is difficult to objectively prove the elements of fault and causal relationship. In traditional disputes over personality rights based on claims, the plaintiff can prove the defendant's fault and causal relationship through direct evidence such as the content of the infringement and the way of conduct. In the case of AI, because the operating mechanism of the algorithm is in a black box state for the right holder, it is difficult for the right holder to prove the fault of the algorithm controller in model training and content generation. It is also difficult to effectively prove the causal relationship of legal interests between the consequences of the damage and the operation of the algorithm. The objective problem of "knowing that there is infringement but being unable to provide evidence" cannot be solved.

Secondly, the problem of the uniformity of rule application: the mechanical application of proof standards in different tort situations. Artificial intelligence personality rights infringement encompasses a variety of specific scenarios, including infringement involving specific personality elements such as portraits and voices, as well as cases of defamation of the right to reputation and honor, and cases where intelligent services infringe upon the right to privacy and personal information rights. There are significant differences in the extent of the influence of algorithmic black boxes, the ability of the subjects to provide evidence, and the importance of rights in various scenarios, but the current laws and judicial practices have not formulated differentiated proof rules for different scenarios and still adopt uniform standards, making it difficult to meet the proof requirements of various scenarios.

Third, the standards of discretion are inconsistent: the extent to which discretionary power is exercised is ambiguous. Faced with the problem of proof caused by algorithmic black boxes, some judges use discretionary power to transfer the burden of proof or relax the requirements of proof standards, but due to the lack of a unified standard to guide, the application of discretionary power lacks specific boundaries and factors to consider.

Different courts and judges have different opinions on whether to transfer the burden of proof and to what extent. This results in extremely inconsistent standards of judgment,

which not only undermines the credibility of the judiciary but also fails to form behavioral expectations for the parties involved.

3.3 A Multi-Dimensional and Deep Analysis of the Causes of the Predicament

To sum up, in the algorithmic black box context, it is difficult to prove the damage to individual rights and interests caused by the application of AI technology, which is the result of multiple reasons such as the characteristics of AI technology, the construction of legal norms, the interaction between subjects, and institutional Settings.

Firstly, technically speaking, the algorithmic black box has dual attributes, which make it difficult to determine the facts of the case. On the one hand, deep learning algorithms themselves have the characteristic of being unexplainable. Due to their complex neural networks and nonlinear computing mechanisms within them, the results of their operation are often difficult to be generally expressed; On the other hand, algorithmic controllers, considering factors such as confidentiality rights and patent rights, usually do not disclose algorithmic programs, algorithmic samples, and operation models, etc. On the other hand, there is a lack of corresponding technical means to verify them in judicial practice, which makes it difficult to prove.

In terms of the burden of proof, the traditional principle of allocation of burden of proof is difficult to adapt to the emerging types of artificial intelligence infringement. Based on the traditional personality rights infringement lawsuit, the principle of no-fault liability is usually applied and follows the distribution model of "he who asserts, he who proves", which implies the precondition that all parties involved in the lawsuit have equal ability to prove. However, artificial intelligence infringement often presents the characteristics of high technological content, opaque process of causing harm, and complex path of damage resulting, with numerous participants and unclear division of liability, which goes beyond the application preconditions of traditional rules. Due to the fact that legislation has not promptly improved the rules of proof for this new type of infringement, the gap between the lag of legal provisions and reality is too large.

First, there is an imbalance of evidence resources between the two parties to the lawsuit.

AI infringement disputes typically involve multiple subjects such as rights holders, developers, network service providers and users, while developers and network service providers, as controllers of AI, hold key evidence materials such as core information and technical parameters of the algorithm. At the same time, the alleged infringer, as a highly technical right holder, does not have the corresponding technical background and available data support, and its ability to provide evidence is relatively weak compared to that of the plaintiff, which is also one of the factors contributing to the difficulty in providing evidence.

Second, the existing rules of evidence are not well matched with the algorithmic governance system. To solve the burden of proof problem caused by the black box of algorithms, it is not enough to merely adjust the burden of proof rules. A series of supporting institutional arrangements such as algorithm filing, compliance review, evidence disclosure, and expert assistance should also be established. As far as China's current legal system is concerned, the preventive system for algorithm governance is not effectively connected with the rules of proof in litigation, and the evidence disclosure mechanism for algorithm black boxes, the technical investigation system, and the third-party audit system are not yet complete, making it difficult to guarantee the institutional support needed for the application of the rules of proof.

4. Paradigm Reconfiguration: The Core construction of the Three-stage Dynamic Burden of Proof Allocation Mechanism

Based on the problems and causes mentioned above, this paper breaks away from the bipolar situation of "whoever asserts all requirements must prove" and "inversion of the burden of proof for all requirements" in previous theories, and under the principles of "proportionate risk management, equivalent burden of proof, contextualized distinction, and moderate balance", Under the value orientation of giving priority to the full protection of personality rights, taking into account the incentives for technological innovation, and coordinating procedural justice with substantive justice, a new allocation scheme of burden of proof is constructed. At the same time, a dynamic distribution model of "proof first - gradient clarification - discretionary distribution" is proposed for the first time, which, while

achieving a reasonable allocation of the burden of proof, forms a responsibility distribution pattern with clear responsibilities, clear procedures and clear boundaries.

4.1 The First Stage: the Basic Burden of Proof of the Right Holder

The purpose of establishing the basic burden of proof for right holders is to facilitate rights protection, solve the relief predicament of "difficulty in presenting evidence", prevent abuse of rights, and safeguard the due process interests of all parties in the dispute.

First, define the circumstances under which the basic burden of proof applies and the key elements of proof.

Right holders should provide preliminary evidence on three aspects: First, they have legal personality rights, and the AI work can be identified with their personality rights, that is, the public can associate the AI-created work with a certain subject; Second, there are objectively existing acts and consequences of infringement, that is, the content created by AI itself has an illegal nature of infringing upon personal rights and has been publicly disseminated and caused damage to personal interests; Third, prove that the algorithm operator is an AI content service provider and the actual controller of the algorithm, and there is a formal correlation of infringement.

For the two key elements of subjective fault and causality, the right holder does not need to bear the basic burden of proof, thereby substantially reducing the difficulty of presenting evidence.

Second, the standard of proof for the basic burden of proof of "reasonable likelihood". "Reasonable likelihood" is much lower than the "high likelihood" standard of proof in traditional civil litigation. The burden of proof is fulfilled simply by initially proving that there is a reasonable likelihood of the occurrence of the tortious fact, which takes into account the level of proof ability of the right holder in cases where the algorithm cannot be effectively controlled, effectively avoiding the problem of inability to prove due to an excessively high standard of proof.

Third, it is determined that the legal effect resulting from the completion of the basic burden of proof is the statutory transfer of the burden of proof. After the right holder has completed the basic burden of proof, the burden of proof is transferred to the algorithm operator,

who is responsible for fulfilling the burden of proof that he is not at fault and that there is no causal relationship, and fulfilling the corresponding obligation to explain the algorithm. If the algorithm operator fails to provide evidence, it shall bear tort liability.

4.2 Phase 2: Classification of Algorithm Operators and Burden of Proof

Hierarchical explanation and burden of proof is a core part of the entire mechanism. Its innovation lies in discarding the inertia thinking of absolute liability inversion and "mandatory disclosure", and adopting contextualized and differentiated liability design, which ensures that rights holders obtain effective relief while also taking into account the protection of algorithmic trade secrets and intellectual property rights. It does not impose too much compliance burden on algorithmic operators and stifle technological progress.

First, establish a gradient evaluation mechanism for algorithmic interpretation obligations. This paper attempts to cascade algorithmic interpretation obligations from four aspects: "the nature of the black box, the type of infringement, the role of behavior, and the status of rights and interests": In terms of the attributes of the black box, under the

"artificially created invisible" type of black box, algorithm providers should bear a stronger algorithmic interpretation obligation; Under the "unexplainable" type of black box, the principle of strict literal interpretation should be adopted, and only a general statement of legitimacy is required. In terms of the specific protection of personality rights, in areas such as the right to life and the right to the body, strict literal interpretation rules should be adopted; Moderate literal interpretation rules should be adopted in areas of typical personality rights such as portrait rights and voice rights; In the realm of personification rights such as the right to reputation and the right to privacy, they should be interpreted differently based on the nature of the infringement. From the perspective of subject type, higher obligations of proof and explanation should be imposed on the parties who actually produce AI content, while for service providers who merely provide technical support, the standard of "knowing or should have known" can be adopted, and the obligation of proof they bear is commensurate with the duty of care they have.

Third, determine the key points of the burden of proof in each type of liability. The burden of proof for algorithm operators mainly consists of two parts: regarding fault, they should be able to prove that they have fulfilled the compliance and care obligations stipulated by law, including the prevention of infringement during training, the review of content during the generation of content, the establishment of infringement early warning and handling mechanisms in the system, etc., to prove that they have no subjective fault for the occurrence of infringement results; In terms of causality, it should be proven that there is no legal causal connection between the loss of the right holder and the operation of the alleged AI system or that there are legal grounds for exemption. In terms of the obligation to explain, it is only necessary to explain the operation of the algorithm related to the infringement, rather than requiring the disclosure of trade secrets such as the source code and core parameters of the algorithm, in order to achieve the governance purpose of "effective supervision without disclosing the internal mechanism of the algorithm".

Finally, define the scope of application of the obligation to interpret algorithms and the exceptions. The performance of algorithmic interpretation obligations does not require the provision of critical technical information involving trade secrets and intellectual property rights, such as source code, training data sets, key algorithmic programs and technical indicators. Where an algorithm operator can provide evidence that fulfilling the algorithm interpretation obligation would result in the leakage of its critical trade secrets or the loss of its technical advantage, the corresponding interpretation obligation may be waived, but the interpretation act may be replaced by an independent algorithm audit by a third party. An audit opinion from a third-party institution is used to strike a balance between rights relief and trade secrets.

4.3 Phase 3: Dynamic Application of Judicial Discretion and a Mechanism for Adjusting Proof Standards

The exercise of judicial discretion is the final stage of the three-stage dynamic burden of proof system, mainly to achieve fairness in the presentation of individual cases through the legitimate use of judicial discretion in this stage, and to exercise judicial control over the

allocation and application of the burden of proof to avoid abuse of discretionary power.

First, establish the dynamic consideration elements in judicial discretion.

Courts should consider the following factors when determining the allocation of burden of proof and the extent of proof: First, the level of interest. In cases involving personal rights and interests such as personal safety and reputation, the difficulty of proof for the plaintiff can be appropriately reduced, while that for the platform can be increased; The second is the difference in the difficulty of obtaining evidence for both parties. According to factors such as the professionalism of both parties, the amount of information available, and financial resources, the first is the disparity in ability between the two parties. In cases where one party is significantly stronger than the other, the burden of proof for the algorithm operator should be increased even more. Third, the type of opacity of the algorithm and the extent of the infringement result, for artificially set technical obstacles or infringement cases that cause significant social impact, the standard of proof for the algorithm operator should be raised; Fourth, the compliance of algorithm operators. For algorithm operators who have established a sound compliance system, it may be considered to appropriately reduce their burden of proof to encourage algorithm operators to attach importance to pre-compliance work.

Second, set differentiated standards for the degree of proof. On the basis of "high likelihood", the requirements for evidence presented to the algorithm operator are differentiated based on the above considerations: when it is an important or significant violation of personal interests, a higher standard of "excluding reasonable doubt" is adopted; When there is technical unpredictability and the necessary duty of care has been fulfilled, it can be appropriately relaxed to "greater likelihood". At the same time, it is required that judges clearly state in their judgments the factual basis and legal basis for the change in the standard of proof, so that the trial activities can be known and supervised by the parties and the general public.

Secondly, limit the scope and boundaries of the adjustment of the burden of proof discretion. Judges are required to adjust the scope and extent of proof of the algorithm operator only on the basis that the right holder has fulfilled his

initial burden of proof, and not beyond its boundaries. It is not allowed to reduce the right holder's initial burden of proof, nor to transfer the burden of proof entirely to the algorithm operator. At the same time, by means of systems such as similar case search, adjudication guidelines, and review by the court of first instance, the discretionary space should be appropriately limited to ensure uniformity and stability of judicial decisions and avoid the phenomenon of different judgments in the same case.

5. Judicial Implementation: Systematic Supporting Measures for the Three-stage Dynamic Burden of Proof Mechanism

The effective implementation of the three-stage dynamic burden of proof mechanism cannot rely solely on unilateral adjustments to the burden of proof rules. It also requires the establishment of a complete supporting system covering pre-event, during-event and post-event, achieving an organic connection between the burden of proof rules and the algorithmic governance system, and formulating specific adjudication guidelines for different types of infringement situations to ensure the enforceability of the mechanism.

5.1 Preparation in advance: A Full-Cycle Compliance and Risk Control System for Algorithms

The core function of the pre-compliance mechanism is to reduce the likelihood of infringement from the source and to lay a factual basis for evidence in litigation.

On the one hand, establish an algorithm filing and compliance review mechanism, clearly stipulating that generative AI service providers need to file algorithm filing materials with regulatory authorities, and establish a regular algorithm compliance inspection mechanism, entrusting third parties to assess the effectiveness of algorithm infringement prevention. Such appraisal reports can serve as one of the effective pieces of evidence that the algorithm provider does not bear liability during the trial stage; On the other hand, establish a pre-emptive duty of care for AI platforms, that is, make it clear that AI platforms should adopt generally recognized technical means in the industry to prevent copyright infringement, and make comprehensive arrangements for whether the training materials are legal, whether the generated works are suspected of infringement,

and whether infringement is dealt with in a timely manner, The fulfillment of these obligations will be an important factor in determining liability for fault in litigation.

5.2 In-Process Support: Evidence Support and Professional Intervention System in Litigation

The key to the litigation support system lies in addressing the difficulty in fact-finding caused by algorithmic opacity and providing procedural guarantees for the implementation of the burden of proof system.

The specific measures include: First, establish evidence disclosure rules for the black box characteristics of algorithms, which stipulate that after the right holder has completed the initial presentation of evidence, upon the right holder's application, judicial authorities have the right to require the algorithm operator to submit evidence materials related to the infringement, such as algorithm operation data, content generation logs, and review system records, if the algorithm operator refuses to submit without reasonable grounds, the right holder's claim can be presumed to be valid; Second, improve the participation mechanism of algorithmic technical experts, clarify that in cases of infringement of artificial intelligence personality rights, both litigants have the right to hire algorithmic field expert assistants to express opinions on professional technical issues, and fully implement the technical investigation officer system in professional courts, where technical investigation officers assist judges in clarifying the technical facts involved in cases; Third, introduce an independent third-party algorithm evaluation mechanism. When the case involves complex algorithm technical issues or the algorithm operator refuses to explain on the grounds of trade secrets, the court may appoint a qualified third-party institution to conduct a neutral evaluation of the algorithm involved in the case, and the evaluation report issued by it can serve as an important reference for judgment.

5.3 Post-Event Support: Unified Adjudication and Diversified Dispute Resolution Mechanisms

The main purpose of the post-event system is to ensure the consistency of judicial standards and to provide rights holders with diverse and convenient means of rights relief. This includes: First, the highest judicial authority formulates

judicial interpretations on infringement cases of artificial intelligence personality rights and issues guiding cases, clarifies the application rules of the dynamic burden of proof system, unifies the standards of adjudication in similar cases, and regulates the adjudication behavior of courts at all levels through the compulsory search system for similar cases; Second, build a diversified dispute resolution system that suits the characteristics of infringement of artificial intelligence personality rights, promote the formation of a dispute resolution mechanism in which industry mediation, administrative mediation and judicial mediation work in coordination, guide parties to resolve disputes efficiently through non-litigation means, and stipulate that compliance commitments and compensation agreements reached during the mediation process can be used as evidence in subsequent litigation.

5.4 Scenario-based Guidance for Adjudication

For specific forms of infringement of personality rights in the application of artificial intelligence technology, the rules for the transfer of burden of proof at different stages should be determined respectively, "tailored to the case".

The first category is AI-created works that contain content with personal identification attributes (facial information, voice information, names, etc.). The core of such disputes is the issue of distinctiveness. The claimant only needs to provide evidence that the personal identification content produced by AI can form a clear connection with themselves and has been used for commercial purposes or improper uses without permission; Therefore, the platform is obligated to provide evidence that it has obtained the relevant authorization and that the content it produces is not identifiable, as well as the effectiveness of the content review mechanism.

Secondly, in cases where AI works infringe upon the right to reputation and honor. In such infringement disputes, the plaintiff must provide evidence that the content of the AI work contains insulting, defamatory or other false words, and be able to associate such content with it, thereby reducing its social reputation; The defendant, as a supporter of the technology, should prove the authenticity of the AI work and that reasonable measures have been taken to verify it. If the infringing information was generated by user instructions, the defendant

should further prove that he had no subjective fault with the infringing information and had exercised reasonable care, and that necessary measures had been taken to stop it.

Secondly, in cases where AI technology service providers infringe upon privacy rights and personal information rights. The victim only needs to prove that their personal information or privacy data has been illegally collected, processed and disclosed; AI service providers only need to prove that they have obtained the explicit consent of the victim and met the legal collection standards, and prove that the information disclosure is not related to the algorithm model.

6. Research Summary and Future Direction

The structural imbalance of the ability to provide evidence due to algorithmic black boxes in AI infringement is the core problem of AI infringement at this stage and an important issue that urgently needs to be regulated. Neither the principle of "he who asserts must prove" nor the extreme approach of "inversion of all elements of proof" can deal with the new situation of AI infringement, nor can it balance the protection of personality rights and the promotion of technological progress.

This paper breaks away from dualism and constructs a three-tier dynamic liability allocation model of "preliminary proof - layered explanation - dynamic discretion", which has dual value: on the one hand, it resolves the differences between the "forced transparency" theory and the "anti-forced transparency" theory in algorithmic governance, and harmonizes the contradictions and conflicts among algorithmic governance theory, tort liability law and the burden of proof system. In practice, reasonable initial evidence standards have been set, significantly lowering the threshold for rights protection and resolving the relief predicament of "difficulty in presenting evidence". At the same time, through the hierarchical design of algorithmic explanation obligations, it avoids excessive compliance pressure on technology operators and achieves a value balance between full protection of personality rights and the healthy development of the artificial intelligence industry.

It should be noted that the development of artificial intelligence technology is very rapid, and the ways and means of its infringement of personality rights are constantly changing.

Therefore, the design of the burden of proof rules should also be updated and iterated in accordance with the development of The Times and the progress of technology. Future research could focus on issues related to the protection of personality rights in frontier technologies such as general artificial intelligence and embodied intelligence, and finally, summarize judicial practice experience and improve the specific application standards of the three-stage dynamic burden of proof allocation rules to provide a more complete theoretical basis and institutional foundation for the protection of personality rights in the digital age.

References

- [1] Yuan Zeng. Governance Myths and Solutions of Algorithmic Black boxes [J]. Chinese Maritime Law Studies, 25,36(04):22-31.
- [2] Mei Xiaying. Fundamental Issues in AI Legislation: Computational Law Training and Human-Machine Relationships in Its Scenarios [J]. Jurist,2025,(06):16-30+191-192.
- [3] Wang Yanfang. Flaws and Reconstruction of the Burden of Proof System for Infringement of Trade Secrets [J]. China Legal Review,2023,(03):46-56.
- [4] Ling Qiushi, Ma Wanli, Zhang Hairu. Research on the Legalization Path of Algorithmic Discrimination in the Context of Artificial Intelligence: Typical Scenarios, Regulatory Dilemmas and countermeasures [J]. Finance and Economics Problem Studies,2025,(10):39-52.
- [5] Liu Jian, Wu Licai. Black Box and Regulation of Algorithmic Governance: A Perspective Based on the Governance Interface [J]. Learning and Practice,2025,(11):34-45.
- [6] Li Yongjun. On the Basic Norms of Claims in the Personality Rights Section of the Civil Code - Can and How to Distinguish It from the Norms of tort Liability? [J] Contemporary Law Chinese Studies,2022,36(02):19-30.
- [7] Yang Liu. Logical Transformation of Liability Allocation for Personality Rights Infringement in the Digital Age [J]. Jiangnan Forum,2025,(12):135-140.
- [8] Liu Wenjie. Platform Attribution path for Agent Infringement of Personality Rights [J].

- Social Sciences Journal,2026,(01):146-156+207+209.
- [9] Research Group of Beijing Internet Court. Legal determination of the Infringement of Voice Rights by AI-generated Sounds - in the case of Yin Moumou v. a certain intelligent technology company in Beijing et al. Personal Rights Infringement Case of Rights Infringement [J]. Application of Law,2024,(09):123-133.
- [10] Yang Xianbin. On Compensation for Mental Damages in Infringement of Personality Rights [J]. Comparative Law Studies,2025,(04):104-121.
- [11] Lu Ziyi. The allocation and implementation path of burden of proof in Criminal compliance cases [J]. Law of Finance and Economics,2022,(01):149-166.
- [12] Xu Weigong, Zhang Yajun. The development of human rights protection in cross-border personality rights infringement conflict law [J]. Human Rights,2023,(06):138-160.
- [13] Zhu Xiaofeng. On Constitutive Elements and Trade-off of Interests in the Determination of Liability for Infringement of Personality Rights [J]. Legal Review, 24,42(01):98-111.
- [14] MAO Muran, Dong Xiaomei. Institutional norms for Ethical risks of infringement of personal and property rights in information privacy [J]. Science & Technology Review, 25,43(04):75-81.
- [15] She Liyong, Liu Wei, Hu Bin, Yang Siqi, Zhu Chao. Analysis of the Application of the Inverted Burden of Proof System in the Determination of collusive bidding [J]. Construction Economy,2021,42(09):33-35.