

Attribution of Legal Effects and Risk Allocation in Unintended Contracting by AI Agents: Rethinking the Scope of the Material Mistake Doctrine under the PRC Civil Code

Xinan Zhu

School of Law, Panzhihua University, Panzhihua, Sichuan, China

Abstract: After the transition of artificial intelligence agents from "information assistance" to "autonomous action", users no longer need to manually determine the counterparty, subject and other terms of the transaction. When the output of the artificial intelligence exceeds the user's prior specific expectations, should this behavior still be attributed to the user? What kind of independent decision-making risks should be borne by which users voluntarily choose automation, and what unforeseen circumstances should be considered? Under what circumstances will it constitute a major misunderstanding? In the past, most studies have often shifted directly from "effects attributable to the influence of users" to error cancellation, which is easy to confuse actual unexpected consequences with formal expression errors. This article argues that attribution of conduct, defects affecting legal validity, and ultimate risk allocation must be treated as three distinct normative levels. Only where an output crosses the objective decision boundary and concerns a matter sufficiently important to the decision to contract, such as the nature of the transaction or the counterparty, should the inquiry proceed to material mistake. On this basis, the article develops a four-stage adjudicative framework—*attribution review, interpretive review, mistake review, and risk review*—which coordinates technological innovation, private autonomy, and transactional security without creating a special body of AI contract law.

Keywords: AI Agents; Unintended Contracting; Declaration of Intent; Material Mistake; Risk Allocation

1. Introduction

From “Machine Execution Error” to “Autonomous Deviation by AI Agents”

Changes in technological architecture have transformed the nature of 'error' in electronic transactions. In settings such as vending machines, automatic renewals, and fixed-algorithm trading, machines primarily execute predetermined rules mechanically, so human decisions and system outputs remain closely aligned. AI agents, by contrast, can decompose tasks, retrieve information, invoke tools, compare alternatives, and execute actions directly on the basis of abstract goals. Xu Ke classifies user involvement as assisted action, limited delegation, and full delegation. Under full delegation, the user may do little more than set an initial objective and approve key decisions, while the system completes complex tasks in an open environment. A user may therefore never make an immediate decision concerning a particular hotel, price, or counterparty, yet a transaction may still be concluded because the user deployed the agent for that purpose [1]. Disputes of this kind are often reduced to a simple question: what happens when the AI gets it wrong?

Yet 'getting it wrong' covers at least three distinct factual patterns. First, the system may select, within the user's clearly authorized range, an option with which the user is dissatisfied *ex post*. Second, the system may interpret an ambiguous instruction in a manner that differs from the user's subjective expectation but remains defensible on an objective interpretation. Third, the system may cross a key boundary expressly set by the user or generate an extreme output because of an external attack or anomalous data. Treating all three situations as merely 'beyond expectations' and moving directly to avoidance for material mistake would transform the doctrine from a mechanism for correcting accidental discrepancies between intention and declaration into a general technological right of regret. It would also turn the commercial value of AI agents—the ability to spare users from deciding every contractual detail—into a reason for undermining transactional stability.

Recent Chinese scholarship has already laid important foundations. One line of research, drawing on an objective theory of declarations of intent and risk attribution, argues that contractual effects generated by AI agents should in principle be attributed to the user, while allowing avoidance for mistake where the agent operates beyond the user's expectations. Another approach examines the tripartite structure of user, provider, and third party and uses concepts of digital delegation and agency to explain the agent's internal and external behavioral boundaries. You Wenting, in turn, develops the idea of the temporal advancement of declarations of intent to bring continuously autonomous AI conduct within the framework of civil juristic acts [2]. These studies demonstrate that denying AI agents independent legal subjectivity does not prevent existing private law from attributing and evaluating their conduct. The further question addressed here is why, once attribution has been established, an output should not enter the doctrine of material mistake merely because it exceeded the user's expectations.

The central claim of this article is that unexpectedness is a technological fact, whereas mistake is a legal characterization. In contracting through AI agents, the two should be connected by an intermediate concept: the 'objective decision boundary.' Where a user deliberately leaves particular contractual matters to the autonomous determination of an AI agent, the uncertainty of the resulting choice becomes part of the user's chosen mechanism for forming intent. The judicial judgment should be carried out according to the following four steps: first, determine whether the behavior of the intelligent body should be attributed to the user; second, according to the provisions of Article 142 of the Civil Code of the People's Republic of China (hereinafter referred to as the Civil Code), the objective boundaries entrusted by the user to the intelligent body for independent decision-making should be explained; again confirm Whether the boundary breakthrough has reached the significant level stipulated in Articles 147 and 19 of the Interpretation of the Supreme People's Court on the Application of the General Provisions of the Civil Code of the People's Republic of China; finally, the final risk distribution is through the transaction habit, the recognizable nature of the counterparty of the transaction and the trust interest after revocation. The compensation is realized. This can not only

avoid the formulation of a separate "AI contract law", but also make the existing civil law rules truly adapt to the automatic conclusion of contracts. Interpretation on Several Issues Concerning the Application of the General Provisions of the Civil Code (the 'Judicial Interpretation on the General Provisions'). finally, the final risk distribution is through transaction habits, the identifiable nature of the counterparty of the transaction and the letter after revocation It is achieved by the compensation of interests. This can not only avoid the formulation of a separate "AI contract law", but also make the existing civil law rules truly adapt to the automatic conclusion of contracts.

2. Normative Characterization and Attribution of Unintended Contracting

2.1 Delimiting the Inquiry: Autonomous Action Does Not Make an AI Agent an Independent Bearer of Intent

This article does not agree with the premise of "AI personality" or "AI qualification". The autonomy of the intelligent body only means that the system can independently formulate and execute specific plans according to the goal and environment, without the need for users to confirm each step immediately. Its legal significance lies in the transformation of human expression, that is, from "making immediate decisions on specific transactions" to "preparing decision-making mechanisms and decision-making space in advance". This leads to the fact that the formation of the expression of meaning is predetermined in time and content, but this does not change the fact that the ultimate rights and obligations still fall on users or other entities.

2.2 Distinguishing Attribution, Validity, and Risk Allocation

The most common theoretical controversy in the theoretical controversy about artificial intelligence agents is to combine the three issues for analysis. The first aspect is the attribution of the behavior: whose behavior is legally regarded as the contractual behavior executed by the system? The second problem is the validity defect: can the statement made by the user be revoked due to error, fraud or other reasons? The third risk is the final risk: after the contract is still valid or revoked, will the final loss be borne by the user, counterparty, service provider, developer or third party? These three are interrelated, but they have

different normative functions.

The UNCITRAL Model Law on Automated Contracting provides a clear reference for this distinction. Article 7 stipulates that when the parties have not reached an agreement on the attribution procedure, the act of the automated system execution shall be attributed to the person who uses the system for the relevant purpose. The third paragraph of Article 7 clearly stipulates that the attribution cannot be denied simply because the result is "unexpected". The fourth paragraph of Article 7 further clearly reserves room for adjustment of other laws on the consequences of ownership and liability [3]. The drafting guidelines of the UNCITRAL Model Law on Automated Contracting further emphasize that attribution and liability should be distinguished. Attribution is usually only a prerequisite for the application of other liability rules [4]. An unexpected result is therefore not, in the first instance, a question of whose act it is, still less a question of who must ultimately bear the loss; those issues arise only after attribution has been established.

To keep the inquiry within manageable bounds, 'unintended contracting by AI agents' in this article refers only to situations in which a user deploys or invokes an AI agent for the purpose of concluding a contract, the agent autonomously selects, generates, or confirms contractual elements in interaction with a third party, and the resulting contract was not specifically foreseen *ex ante* or is later alleged by the user to be inconsistent with the user's expectations. The article does not address torts committed by AI agents, personality rights, intellectual property, autonomous-driving product liability, or transactions wholly detached from the user's account and authorization environment through hacker impersonation. The latter mainly involves identity fraud, lack of intentional or security infringement, and cannot be confused with unexpected contracts of "accidental contracts caused by user deployment, but output deviations".

2.3 Attribution and Its Limits: Deployment, Purpose of Use, and Sphere of Control

In online transactions, the two parties to the transaction usually cannot and do not need to determine whether the other party is a natural person, a fixed program or an automatic process operation based on an uncertain model. As long as the instruction is generated from the user-

controlled account, interface or transaction environment and aims to form a contract, in principle, it should first be classified as the user's transaction behavior. If the validity of the statement is denied simply because the system is not certain or the user does not review each project one by one, the legal stability of the automatic contract will depend on the invisible technical details. Article 5 of the UNCITRAL Model Law on Automated Contracting adopts a technically neutral position and shall not deny the legal effect of the contract or related acts simply because there is no participation of a natural person [5].

Users bear the initial effect of the AI agency contract, which can be explained by three mutually reinforced factors. The first aspect is the deployment behavior. When the user activates the intelligent body and entrusts the contract target to it for processing, it marks that the transaction risk begins to enter the autonomous domain of AI. The second aspect is the specific purpose. The system does not accidentally generate transaction-related information; Instead, users will use this information to search, compare, quote, place orders or confirm, all of which are part of the contract signing process. The third aspect is the control field. Account permissions, payment methods, budget parameters, transaction categories, and whether the final execution is allowed are usually within the scope of the technical environment that users can set or choose. As long as the system expresses its intention to the outside world in this application and control field, in principle, it should belong to the user. If the model independently selects the counterparty or price that the user does not clearly expect in the transaction task authorized by the user, it does not mean that the behavior has exceeded the risk area. As long as the user is still the user of the system, the system will continue to run according to their transaction goals. Whether it "exceeds expectations" should be evaluated from the perspective of validity, not used to refute and deny the attribution of the behavior. We can get the answer to the first judicial question: unexpected circumstances will not terminate ownership, unless the system behavior deviates from the user-initiated transaction risk area due to unintended circumstances.

3. The Scope of Material Mistake: Interpretive Priority and the Objective Decision Boundary

3.1 Unexpectedness Is not Material Mistake

The term "unexpected" simply means that the actual output does not match the specific expectations envisaged by the user. "Dissatisfaction" refers to the evaluation of the value of the transaction result after the fact. The major misunderstanding is the defect in the validity of the civil law. It requires the person who expresses the meaning to have misunderstandings about the nature of the civil lawsuit, the identity of the other party, or the type, quality, specification, price, quantity, etc. of the subject matter when conducting a civil lawsuit. This kind of misunderstanding is a common understanding. Without this kind of misunderstanding, there would be no corresponding meaning [6]. The thresholds of these three aspects are completely different.

The core of the main misunderstanding is not whether the user wants to get the result, but whether the result is still within the scope of decision-making covered by the intention expressed by the user after explanation. This is also the fundamental difference between this article and the view of "correcting wrong transactions by exceeding expectations". The technical autonomy of the intelligent body itself does not naturally produce a new type of error. Judicial organs should still return to the traditional structure of civil law and examine whether there is an unintentional inconsistency between intention and expression. Courts should remain within the traditional private-law structure of asking whether intention and declaration are unintentionally inconsistent [7]. In Chinese law, material mistake substantially corresponds to the traditional doctrine of mistake in declarations of intent; materiality should be assessed by an objective reasonable-person standard, and mistake as to motive ordinarily does not affect validity. This doctrinal foundation requires the court first to determine what the declaration means and only then to ask whether a mistake occurred.

3.2 Determining the Objective Decision Boundary: Interpretation before Mistake

This article uses the term 'objective decision boundary' to identify the range of matters that the user objectively leaves to the AI agent's autonomous determination in a particular transaction. It is neither the scope of authority in the law of agency nor the maximum technical permissions available at the code level. Rather, it

is the range of contractual decisions that can be established objectively through interpretation of the user's declaration of intent. Instructions such as 'within a budget of RMB 5,000,' 'prefer high-speed rail and choose air travel only if no rail ticket is available,' 'purchase only from official flagship stores,' or 'select the fastest-delivering supplier among these three' all define elements of that boundary. Purely internal preferences that the user has not externalized to the counterparty ordinarily cannot define the external boundary of decision-making.

Article 142 of the Civil Code should govern the determination of this boundary. For a declaration of intent made to another person, the provision requires meaning to be determined in light of the words used, relevant terms, the nature and purpose of the act, usages, and the principle of good faith. Use methods such as meaning, system, purpose and habit to explain, and emphasize that integrity is more suitable as an explanation principle. Applying this interpretation framework to artificial intelligence contracts can transform the previously vague concept of "user expectations" into verifiable external evidence without questioning whether the model really "understands" the user [8].

Specifically, at least five types of materials should be checked. The first is the user's current natural language instructions and clear numerical parameters, which constitute the most direct semantic boundaries. Secondly, there are the preset technical parameters in the account, such as payment limit, merchant whitelist and category restrictions, which can be used as a structured representation of the user's intention. The third aspect is the purpose of the task. For example, for "arranging a business trip" and "arranging a luxury vacation", the interpretation of "suitable hotel" is obviously different. The fourth aspect is the previous transactions and continuous settings of the two parties, which are used to determine the stable meaning of vague terms in long-term human-computer collaboration. The fifth aspect is the relevant trading practices and platform rules, which are used to eliminate interpretations that are obviously inconsistent with the usual understanding of transactions.

Traditional civil law has embodied the concept of "interpretation before revocation". If the two parties only use a common but non-literal word to express, then the real common intention should be explained through Article 142, rather than directly applying the factual error principle [9].

AI scenarios should follow this order more strictly, because its essential function is to concretize abstract goals. If the court does not first clarify what matters the user will leave to the system for processing, but directly compares the "user's statement" with the "system output", then almost all independent choices may be interpreted as incorrect expressions.

3.3 Materiality Review after a Boundary Crossing

Even if the output of the intelligent body exceeds the boundary of objective decision-making, it cannot be immediately determined that a major misunderstanding has occurred. The judicial interpretation of Article 19 of the General Provisions stipulates the double standard of "listing specific matters + importance": the wrong object usually involves the nature of the act, the corresponding object and the type, quality, specification, price, quantity, etc. of the object [10]. According to common sense, if there is no such wrong perception, there will be no corresponding intention expression. Therefore, whether the violation of technical parameters has legal significance also depends on whether it involves important matters that determine the basis of the transaction.

4. Typological Treatment of Unintended Contracting and Allocation of Risk

4.1 Choices within the Boundary and Ambiguous Boundaries: No Mistake in Principle

The first type of situation is that AI makes choices within the boundary of objective decision-making, but the results are different from the preferences of the user. Its typical structure is that "the user only needs to determine the scope or goal, and the system determines the specific value". If the user sets the rule that "the total price of the ticket does not exceed 3,000 yuan, and priority is given to the ticket with the earliest arrival time", AI will buy the ticket at a price of 2,850 yuan. Later, the user found that there was another flight with a price of 2,500 yuan, so he asked to cancel the original reservation. Major misunderstandings should not be discussed at this time, because 2,850 yuan is the specific contract terms established by the system allowed by the user. The term "uncost-effective" only refers to problems related to the quality of the algorithm or the service contract, not to any defects outside the

expression of the intention of the contract.

The second type involves users' use of open expressions, such as words such as "reasonable price", "value for money", "as cheap as possible" and "upgrade when necessary", which causes the system selection to not match the user's subjective understanding. This kind of controversy is not a typical system failure, but a discussion about the interpretation of the intention. The court should first consider the literal meaning, the purpose of the task, the previous settings, the default rules of the platform and the transaction habits, and then match the decision-making scope that rational users and rational transaction counterparties can understand.

4.2 Crossing Key Boundaries and Abnormal Inputs: Distinguishing Contract Validity from Recourse

The third type of situation is that the AI agent obviously exceeds the key conditions that have been objectively defined by users. For example, the user requires a single transaction amount of no more than 5,000 yuan, but the final purchase total is 50,000 yuan; the user requires the purchase of brand-new equipment, but finally chooses the refurbished equipment; the user requires to purchase only from suppliers A and B, but finally signs a contract with an unfamiliar third party. At this time, the user's choice of the final price, the nature of the subject matter or the counterparty is obviously inconsistent with the description of the user's instruction. This situation gives users the legitimacy of inciting Article 147.

The fourth category is accidental transactions caused by external factors such as abnormal input or use by third parties. For example, data source errors, injection prompts, malicious merchant inducements or abnormal platform interfaces. From an external perspective, if the user is still a system user and the transaction instructions still come from his account and contract environment, the ownership and material misunderstandings should be determined according to the above rules. The contract cannot be invalidated just because of the existence of third-party technical reasons. Internally, where the system provider has failed to satisfy reasonable duties of security, fitness, or care, the user may separately seek recourse under the service contract, tort law, or product liability rules.

4.3 Protection of Reliance and Risk Allocation after Avoidance

Article 8 of UNCITRAL Model Law on Automated Contracting offers a comparative-law approach that deserves attention. Where an action by an automated system has already been attributed to a party, that party could not reasonably have been expected to anticipate the action, and the other party knew or ought to have known that the action was not expected, the other party may not rely on it [11]. Some scholarship similarly argues that, in AI-agent contracting, the counterparty should bear the mistake risk when it knew or should have known of the user's error [12]. Although this approach has evident appeal from the standpoint of transactional fairness, it should not be converted in Chinese law into an additional constitutive element of material mistake.

The reason is textual and systemic. Neither Article 147 of the Civil Code nor Articles 19 and 20 of the Judicial Interpretation on the General Provisions requires the counterparty to have participated in, known of, or been in a position to know of the mistake. Adding such a requirement would lack a textual basis and would narrow the doctrine's function of protecting private autonomy. The counterparty's interests can instead be balanced through post-avoidance damages [13]. Creating a special 'counterparty knowledge' requirement solely for AI transactions in the name of transactional security would make the same Article 147 operate with different constituent elements in conventional and AI-mediated transactions.

The counterparty's ability to recognize the abnormality nevertheless remains relevant. Refusing to make 'knowledge or constructive knowledge' an element of material mistake does not make the counterparty's state of knowledge irrelevant. Article 19(2) of the Judicial Interpretation on the General Provisions expressly provides that, even where the actor proves a material mistake, a court may deny avoidance where trade usages or other circumstances indicate that the actor has no right to rescind [14]. The open-ended reference to 'trade usages and other circumstances' leaves room for allocating risk between the parties in automated transactions.

Risk allocation after avoidance should then return to Article 157 of the Civil Code. That provision requires a party at fault to compensate the other party for losses resulting from an invalid, avoided, or ineffective civil juristic act and apportions responsibility where both parties are at fault. In

AI contracting, the inquiry should focus on whether the user exercised care appropriate to the transaction in setting the agent's permissions, parameters, and obvious risk controls, rather than treating the use of AI itself as fault. If the user has set the key boundary correctly and the abnormal situation is entirely caused by a system defect, the scope of compensation of the affected party can be appropriately limited. If the user allows the high-risk behavior to be automatically executed at will without setting an upper limit on the costs that may be incurred, the user shall bear the corresponding substantial compensation after canceling the contract, which shall be determined to have a legal and factual basis.

E-commerce price errors can only be used as an analogy, and price errors and AI unexpected contracts cannot be simply evaded. The former is usually the input and setting errors of operators or employees first, and the automatic system is only mechanically transmitted; the latter may not be wrong at the time of input, but the system forms specific content in the independent decision-making space left by the user. The revelation of the wrong price case is that the more electronic transactions rely on automatic systems, the more the court must distinguish between internal technical reasons and external transaction trust. The first step in the scenario of AI intelligence contracting should still be to explain the decision-making boundaries, rather than copying the rules of duty to pay attention in the bidding error of traditional transactions.

5. Judicial Application: A Four-Stage Adjudicative Framework

The first level is the examination of the affiliation. The court should first determine whether the user deploys or calls AI agents for contractual purposes, including account binding, task instructions, payment authorization and tool call permissions, etc. As long as the system operates according to the transaction purpose set by the user, in principle, the legal effect of its external behavior should be attributed to the user. Just because the output results are unexpected, without manual review or the model is uncertain, the corresponding attribution cannot be denied. Only when the system behavior deviates from the user's expected effect due to identity fraud, complete hacker takeover and other reasons, it should be controlled and adjusted at the ownership level.

The second level is explanatory comments. After

taking over the case, the court should first explain the standards and then make a judgment of attribution. The attribution judgment should be based on natural language instructions, numerical parameters, account permissions, task purposes, previous settings, platform rules and transaction habits, and clarify that AI agents can objectively determine the scope of matters entrusted by users to make independent decisions. To make a scope that meets the psychological expectations of the user at that time, we can't just take the expectations expressed by the user after the fact as a boundary.

The third level is error examination. Only when the system output exceeds the boundary of objective decision-making will it enter the stage of major misunderstanding review. The court should then ask whether the crossed condition concerns an important matter such as the nature of the transaction, the counterparty, or the type, quality, specification, price, or quantity of the subject matter, and whether, under an ordinary understanding, the actor would not have made the corresponding declaration absent the erroneous understanding. A choice within the boundary should in principle exclude material mistake; a crossing that is immaterial should likewise not justify avoidance.

Fourth, risk review. Even after material mistake is established, the court should examine under Article 19(2) of the Judicial Interpretation on the General Provisions whether 'trade usages and other circumstances' limit the right of avoidance and should evaluate, in light of the counterparty's ability to recognize the abnormality, whether the counterparty's reliance deserves protection. If the contract is avoided, reliance losses should then be allocated under Article 157. If the abnormality is caused by the breach of contract, infringement or security defect of the AI intelligence provider, developer or third party, the user can recover the ultimate responsible person in the internal relationship.

6. Conclusion

The non-expected conclusion of AI intelligence should adhere to three basic judgments: non-expected does not naturally deny the attribution of the behavior; non-expected does not naturally constitute a major misunderstanding; the establishment of a major misunderstanding does not mean that all risks are borne by the relative. "Objective decision-making boundary" should be taken as the intermediate concept, and carried out

according to the four steps of "attribution review - interpretation review - error review - risk review", and finally achieve a systematic linkage between Articles 142, 147 and 157 of the Civil Code and Article 19 of the judicial interpretation of the General Code, and the technical facts should be transferred. It is a civil law adjudication requirement that can be adjudicated. In this way, in the face of rapidly developing AI, the efficiency and benefits generated by the intelligent economy will not be at the expense of autonomy.

References

- [1] Xu Ke, 'The Private-Law Structure and Behavioral Boundaries of AI Agents,' *Oriental Law*, 2026, no. 1: 15–28.
- [2] You Wenting, 'Legal Characterization and Allocation of Liability for Automated AI Conduct: The "Lobster Farming" Model as an Example', 2026, no; 6-7.
- [3] UNCITRAL, *Model Law on Automated Contracting with Guide to Enactment* (United Nations, 2025, art. 7(1)–(4). Article 7(2).
- [4] UNCITRAL, *Model Law on Automated Contracting with Guide to Enactment* 2025, United Nations, Guide to Enactment, para. 71.
- [5] UNCITRAL, *Model Law on Automated Contracting with Guide to Enactment* 2025, United Nations, art. 5.
- [6] Civil Code of the People's Republic of China, art. 147; Supreme People's Court Interpretation on Several Issues Concerning the Application of the General Provisions of the Civil Code of the People's Republic of China, art. 19.
- [7] Zhai Yuanjian, 'The Doctrinal System and Normative Application of Material Mistake,' *Comparative Law Review*, 2022, no. 4: 156–170.
- [8] Yin Xiangrui, 'On the Interpretation of Declarations of Intent: With an Interpretation of Article 142 of the Civil Code,' *Journal of East China University of Technology (Social Science)*, 2026, no. 2: 56–64.
- [9] Zhai Yuanjian, 'The Doctrinal System and Normative Application of Material Mistake,' *Comparative Law Review*, 2022, no. 4. 3-6.
- [10] Supreme People's Court Interpretation on Several Issues Concerning the Application of the General Provisions of the Civil Code of the People's Republic of China, art. 19(1).

- [11] UNCITRAL, Model Law on Automated Contracting with Guide to Enactment 2025, United Nations, art. 8(1).
- [12] Zhou Xuefeng, 'The Meaning, Legal Status, and Attribution of Effects of Artificial Intelligence Agents,' Chinese and Foreign Law Review, 2026, no. 1:15-18.
- [13] Zhai Yuanjian, 'The Doctrinal System and Normative Application of Material Mistake,' Comparative Law Review, 2022, no. 4: 157–158.
- [14] Supreme People's Court Interpretation on Several Issues Concerning the Application of the General Provisions of the Civil Code of the People's Republic of China, art. 19(2).