Proceedings of the 4th International Conference on Educational Innovation and Philosophical Inquiries
DOI: 10.54254/2753-7048/14/20230933

Conflict of Interest in the Value Judgement of Law of Artificial Intelligence from the Perspective of Game Theory

Yaxin Deng[1,a,*]

[1]*Law school, Shenzhen University, Xueyuan Dadao, Shenzhen, 518055, China*
a. 2021030210@email.szu.edu.cn
**corresponding author*

Abstract: Artificial intelligence's intervention in judicial adjudication will lead to the alienation of adjudication process, which will lead to conflicts among different interest groups. This problem is especially prominent in the process of value judgment. The same case obtains the same judgment result, which is the core element of judicial justice. At the same time, the new legal system needs to emphasize the proper application of law and the just result of judgment. This requires the system to have flexible conflict rules to effectively balance the contradictory conflict. In the field of civil and commercial law, the introduction of game theory to solve the conflict of interest has a long precedent. Based on the Game Theory, according to certain legal premises, considering AI's involvement in value judgment, an effective judgment result is finally obtained. This provides theoretical consideration for the structure of the subsequent specific legal system.

Keywords: Jurisprudence, game theory, algorithmic justice, judging, new technology and law

1. Introduction

The quality of legal reasoning determines the quality of the rule of law. The value judgment of law refers to the value judgment of legal reasoning. According to the legal cultures, whatever based on common law system or civil law system, the judge is usually the subject of the value judgment of law. The theory of "free evaluation of evidence" depends on the judge making value comment on individual cases. If court rulings are impacted partly the judge's value judgments, the multi-party interests will be reflected in different strata, what mainly include the parties to the case, the court (the judges or legislators) and the public.

With the development of science and technology, artificial intelligence is associated with legal reasoning. There is a possibility that the original multi-party coordination of interests will be broken, resulting in the value judgment of the conflict of interests. The conflict doubts whether artificial intelligence can make value judgment and meet the interests of different groups. First of all, the academic theory of artificial intelligence debated about the value judgment of jurisprudence, though developing slowly, is a trend of development. The CJTS system of Singapore courts provides consultation and mediation for the parties. Indeed, perhaps in recognition of these innovations, the European Union recently enacted ethical guidelines for companies working to develop and implement AI [1]. Secondly, there is a debate about the legitimacy and rationality of the existence of robot judges

DOI: 10.54254/2753-7048/14/20230933

in jurisprudence. The basis of the debate is whether artificial intelligence can make value judgments. Proponents argue that the essentially social nature of law can be reproduced by machines, no matter how sophisticated [2]. In contrast, the opponents argue that "calls for robot judges and juries are typically met with derision" because machines are unable to take into account "softer" goals of the criminal justice system such as dignity, equity, and mercy [3].

From a deep level of analysis. In essence, the conflict is the reconstruction of the pattern of distribution of benefits between different subjects of rights. This is also why manual can only be corrected for the breakthrough of past value judgment standards and more about the legal orientation of AI and the limits of legal intervention. What position does AI occupy in the law?

2. Jurisprudential Reflections on the Theory of Original Value Judgment

The objectivity of moral judgment is a fundamental issue in the debate between cognitivism and non-cognitivism. Cognitivism generally affirms that morality is objective, and asserts that moral judgments are descriptions of natural or non-natural realities, and therefore are cognitive. The central leitmotif of cognitivist theories is the idea that moral judgments are truth-apt; there are moral facts that our moral judgments could be true (or false) in relation to [4]. Cognitivism in psychology and philosophy is roughly the position that intelligent behavior can be explained (only) by appeal to internal 'cognitive process's-that is, rational thought in a broad sense [5]. Non-cognitivism generally denies that morality is objective, and holds that moral judgments are expressions of emotions or attitudes, or prescriptions of volition, which explain it is non-cognitive. Kantian constructivism, developed in response to Rawls and Christine M. Korsgaard, attempts to go beyond the debate between cognitivism and non-cognitivism, asserting that moral judgments are neither descriptions of the facts of the world, nor expressions of emotions or attitudes, but solutions to practical problems.

According to non-cognitivism, AI obviously cannot make value judgments. The representative of non-cognitivism is Hume's Theory of Human Nature. Hume holds that the original cause of morality is emotion, and the distinction of morality is not derived from reason. He denies the existence of truth of value, and rejects the objectivity of ethics. Ethical Theory - Part I The Status of Morality - Introduction summarizes two of Hume's arguments. The first one：

All claims that can be known by reason are either empirical matters of fact, or conceptual truths.

Moral claims do not represent empirical matters of fact.

Moral claims do not represent conceptual truths.

Therefore reason cannot give us moral knowledge.

Here is another argument taken from Hume's classic work:

Moral judgments are intrinsically motivating.

Beliefs are not intrinsically motivating – they need desires to generate motivation.

Therefore moral judgments are not beliefs.

There seems to be a tension between Hart's non-cognitivism and his rejection of sanction-based and predictive theories of law [6]. According to the present technical development, the machine can not have the "belief", "emotion" and "intuition" mentioned by Hume. We take the view that most, if not indeed all, approaches that seek to bring Al to the activity of judging mistake the nature of law. It generally is seen there simplistically, as a traditional Austin-style "command backed by sanction" [7].

For non-cognitivists, "whether AI can make value judgments" is still a controversial proposition. The problem for non-cognitivism, Joyce suggests, is that "if moral judgments were nothing more than expressions of the speaker's conative attitudes, then they too would be equally irrelevant to others' deliberations (unless those others happened to care about the speaker's inner states) [8]". Bruner's learning theory shows that knowledge learning is to form a certain knowledge structure in students'

DOI: 10.54254/2753-7048/14/20230933

minds. This kind of knowledge structure is composed of the basic concepts, basic ideas or principles in the subject knowledge. The structural form of knowledge structure is formed by the encoding of human encoding system. Encoding has two forms: one is formal encoding, which takes the form of a certain logical principle, or subsumes it into a certain logical principle. The other is informal encoding, whose basic form is generalization, either inductively or intuitively. And it can be represented by three modes of representation. The value of a knowledge structure depends on its ability to simplify data, generate new propositions and enhance the ability to use knowledge. These abilities can already be achieved by AI's information-processing computation and huge database.

In Gagne's model of information processing, executive control and anticipation are two important structures, which can stimulate or change the processing of information flow. The former is the influence of previous experience on the current learning process, which plays a regulating role, and the latter is the influence of motivational system on learning, which plays a directing role, which can regulate and supervise the whole information processing process. The former can be done by a lot of computation. But the theory of "motivational system" itself is subjective, and AI does not seem to be able to generate the "motivation" in this model.

Even if AI can meet the requirements of cognitivism for factual judgment, it does not mean that non-cognitivism can provide a theoretical basis for AI to make value judgment. That is, Kant questioned: Can factual judgments be derived from value judgments?

3. The Application of Game Theory

Based on the theory of value judgment, it is necessary to reconcile the conflict between different stakeholders in the process of value judgment of artificial intelligence. Game theory can be used as a good academic tool.

Game theory and law summary the game theory represents a mathematical theory and methodology which is used for solving conflicting and partly conflicting situations in which individuals have conflicting interests [9]. Considering situations in which two or more subjects make decisions in the conditions of interest conflict has been named [9]. Game theory is essentially the study of strategic communication of information through language in a rigorous and stylized way [10]. People try to use the theory of game theory to find the inherent mechanism and logic of law, and design the exquisite legal system to adapt to the legal order of the age of artificial intelligence technology. From the type, game theory can be divided into cooperation game and non-cooperation game.

First, the concept of game theory has been used more and more in the field of legal system design. Second, AI's value judgments may be trapped in the prisoner's dilemma and the battle between the sexes. The prisoner's dilemma is an individual choice, not the best choice for both sides. If the prisoner's dilemma is repeated by multiple people, it may further develop into the tragedy of the commons. But this paper only discusses it in the context of the prisoner's dilemma.

Also, Prisoner's dilemma as one of the most popular and most used models of game theory and its application in selected branches of law [9]. The battle between the sexes is the imbalance between the two sides. When AI makes a value judgment, it can either use cognitivism to reject its rationality, or use non-cognitivism to convert the value judgment into data. Both may fall into a crisis of excessive randomness or lack of humanity. The residual decision-making rights of each party cannot be realized, and in some cases it is impossible to make an ethical decision that meets the public's expectations.

Thirdly, cooperative game and non-cooperative game make coordination possible. Cooperative game is a binding agreement of rules between cooperators, and each party makes a rational choice based on obeying the agreement of rules. In addition, non-cooperative games are united on the "Nash Program", each player knows what he can do, the results of different members' joint actions, and the preferences and utility of the results. Rational individuals respond to the needs of real life, to a certain

DOI: 10.54254/2753-7048/14/20230933

extent, break the rigidity of legal regulations. When constructing the legal system, When constructing the legal system, the designer should not set up a set of abstract rules, but a flexible and dynamic system that can respond to the real situation. In theory, cognitivism and non-cognitivism achieve a static coordination. According to the characteristics of the case and the actual situation of the AI robot's value judgments to achieve dynamic coordination. In other words, using the characteristics of AI's computational logic rationality and judgment of stochastic universality and existing legal theory, the interests of the game.

4. Theoretical Model

Any case may not only have a unique disposal result, whether administrative adjudication, civil adjudication or criminal adjudication, a judgment only means that it may be the final disposal, the final decision, but may not be the only result, or may there be a correct decision. There is no right to be found, it is right as long as it is final. " What does this mean? That is, all cases are right only because it is final, not because a result can be deduced as in the natural sciences through theorems or laws. Unlike the natural sciences, which are always involved in a case, under the influence of discretion, the only right is that you are final. For the judge, since complete justice cannot be achieved directly, eliminating the unfavorable option is obviously the best strategy. Under this premise, the judge's game behavior in different situations can be analyzed.

Suppose: A had been subjected to B's violence for a long time, and finally could not take it anymore and killed him. The public sympathized with A's experience. The panel now consists of three judges, C, D, and E. If A is now required to serve a specified term of imprisonment of 10 to 14 years by the panel judge. Assume that in ideal circumstances, a 12-year prison sentence is the most just sentence. The more the judgment deviates from this value, the less "just" it is. And the more likely it is to produce such a "just" judgment, the more reasonable the value judgment. That is to say, every possible alternative to the "fair result", the negative benefit to the judicial fair is 1 value unit. This is a reasonable basis on which the same judgment should be entered in similar circumstances so as to uphold justice.

Scenario one: C, D, and E all base their decisions on non-cognitivism (emotion, intuition). Each judge makes each choice with equal probability, one in five. The probability that the judgment will last for 12 years is also one in five. For A and B, the value outcome may be different from time to time. Both A and B try to arouse the judges' "emotions" to achieve greater interests, but the cost and outcome are still incalculable.

In scenario 1, each judge's choice is to choose an outcome at random. Although traditional theory claims that this is based on "intuition," the behavior is still random. The generation mechanism of the final outcome is no different from the random probability of a dice throw. To be outcome-oriented, AI can also make the same choice by lottery. At this time, AI's involvement in value judgment can replace human beings.

Given these assumptions, the introduction of AI would reduce the randomness of value judgments. The AI could combine the underlying logic programmed into a large number of cases in the database to choose a solution. In other words, the AI would be limited to eliminating solutions that were significantly different from the same cases in most of the databases, and the randomness would be one in four or three. In other words, the intervention of AI will make it more likely that the final judgment will be "a just result" and less arbitrary. The negative benefit is reduced from 4 to 2.

Scenario two: C, D, and E all base their decisions on cognitivism. Set the following items as the criteria for increasing length of imprisonment. The value of this case can be decomposed into one or more natural attributes: 1) the violence used by B on A, 2) the plight faced by A when hurting B, 3) the duration for which A was subjected to B's violence, 4) the means used by A to kill B, 5) the mental state of A. Legally valid evidence is also required. For each attribute, the university also has

DOI: 10.54254/2753-7048/14/20230933

objective standard regarding how much influence each of these attributes has on the final determination. Compared with Assumption 1, both A and B will claim their own interest by submitting relevant evidence as much as possible. The result of such interest should be estimable and measurable. C, D and E all use the same "reasonable person" test in their discretion, and the result should ideally be consistent to guarantee judicial justice. That is, the judgment result may not necessarily be 12, but shall be infinitely close to 12. The numbers closest to 12 shall be 11 and 13. At this time, the probability of the judgment to be "fair" is one third.

AI data processing quantifies these innumerable natural attributes into computable data. The proportion of each attribute is fixed numerically. Under the cognitivist theory, the standardization system has filtered the database of solutions that are very similar. The number of remaining solutions is related to the set of cognitive criteria. The more specific and detailed the criteria, the more the solution passes through the standard. Ideally, as long as the standard is sufficiently detailed, the result can be infinitely close to 12. Setting the standard requires the accumulation of a large number of cases, which is the same way that people set the standard. So under the same standard adopted by the AI and the previous judge, the probability of AI to obtain a "just" result should be the same. It is concluded that in this case the negative benefit does not become 2.

Scenario three: C, D base their decisions on non-cognitivism, and E base their decisions on cognitivism. The negative effectiveness of C and D at this time are both 4, and E is 2. The average of them is more than 3 but less than 4. After intervention by AI, C and D become 2, but E remains 2. The average value is reduced to 2.

Scenario four: C, D base their decisions on cognitivism, and E base their decisions on non-cognitivism. The negative efficiency of C and D at this time are both 2, and E is 4. The average of them is more than 2 but less than 3. After intervention by AI, the values of C and D shall remain unchanged, but E becomes 2. The average value is reduced to 2.

The comparison between Scenario 1 and Scenario 2 shows that the justification of judgment is supported by cognitivism. Scenarios 3 and 4 are the situations in a judicial trial. Before the intervention of AI, the negative benefit of Scenario 4 is smaller than that of Scenario 3, so the coordination between them should also be cognitivist. In either case, the intervention of AI in value judgment reduces the negative effects of the coordination between non-cognitivism and cognitivism. Compared with Scenario Three, Scenario Four has less changes, and the judicial system can achieve the same effect at a lower cost.

5. Conclusion

The intervention of AI in the rule of law has a great impact on the original legal order. Does the trial of law still belong to the rule of law after the intervention of AI? According to Fuller, the rule of law is the avoidance of evil rules. AI avoids the generation of evil rules by a lot of computation, but part of the process is not countable. The subjective attribute of value judgment belongs to this part. It is difficult for AI to make the general explanation that accords with human's intuition. This phenomenon makes the appearance of AI may cause the dissimilation of judgment process and result in different judgment results. At present, the tightness of AI and people is getting higher and higher, but AI cannot fully realize the rule of law. So, the limit of the combination of AI and people, that is, what position should AI occupy in value judgment?

Both cognitivism and non-cognitivism have theoretical defects. Both viewpoints cannot solve the conflict problem brought by AI. Therefore, the viewpoint of combination and balance of cognitivism and non-cognitivism is put forward. Non-cognitivism is man's "value faculty". The value judgments made by human beings are not absolute, and are dominated by will. AI cannot make value judgments in the traditional non-cognitivist context. Even without the aid of cognitivism, the importance of human beings should not be weakened. But the value source of non-cognitivism needs the support of

DOI: 10.54254/2753-7048/14/20230933

cognitivist theory. Apparent primitives (language, culture, etc.) can be perceived by humans and can be digitized. This feature simultaneously realizes the justifiability of the judgment elaborated above.

According to the theoretical model of this paper, under the game theory, the model with cognitivism as the main and non-cognitivism as the auxiliary is better. Facing the same event, the non-cognitivism as the main and the cognitivism as the auxiliary are not the same. Apparently, the former is to use human subjectivity as an important factor to assist AI to reach the final value judgment conclusion, human "preferences" greatly affect the final result but do not determine the result. The latter is the person regards AI as the thinking tool, the final result still makes the explanation according to the person's preference. The reason why the latter is preferred in practice is that there are algorithmic "black box" problems behind the algorithm. Even the latter, which relies less on AI algorithms, is still challenged by "black box" risks. But this does not mean that the latter is a better model. The real-time, optimal, and general advantages of AI's powerful computing power remain significant. AI makes value judgments more deterministic and reduces the cost of errors in post-event analysis. In theory, the introduction of AI is a better model for making fairer and more efficient judgments.

References

[1] *Epps, Willie J. Jr. ,Warren, Jonathan M. (2020) Artificial Intelligence: Now Being Deployed in the Field of Law. Judges' Journal, 59(1),16-19.*

[2] *Morison, John, Harkens, Adam. (2019) Re-Engineering Justice: Robot Judges, Computerised Courts and (Semi) Automated Legal Decision-Making in Legal Studies. The Journal of the Society of Legal Scholars, 39(4),618-635.*

[3] *Simmons, Ric. (2018) Big Data, Machine Judges, and the Legitimacy of the Criminal Justice System in U.C. Davis Law Review, 52(2),1067-1118.*

[4] *Swaminathan, Shivprasad (2016) Projectivism and the Metaethical Foundations of the Normativity of Law in Jurisprudence, 7(2), 231-266.*

[5] *Patterson, Dennis (2003) Fashionable Nonsense in Texas Law Review, 81(3), 841-894.*

[6] *Rodriguez-Blanco, Veronica (2012) Social and Justified Legal Normativity: Unlocking the Mystery of the Relationship in Ratio Juris, 25(3), 409-434.*

[7] *Morison, John, Harkens, Adam. (2019) Re-engineering justice? Robot judges, computerised courts and (semi) automated legal decision-making in Legal Studies ,39, 618-635.*

[8] *Morison, John, Harkens, Adam. (2019) Re-engineering justice? Robot judges, computerised courts and (semi) automated legal decision-making in Legal Studies ,39, 618-635.*

[9] *Bojanic, Ivana Barkovic, Eres, Maja. (2013) Game Theory and Law in Pravni Vjesnik,29(1), 59-76.*

[10] *Arfi, Badredine (2006) Linguistic Fuzzy-Logic Game Theory in Journal of Conflict Resolution, 50(1), 28-57.*