

China's Anti-Unfair Competition Law on Data Crawling

Xinpeng Liu^{1,*}

¹Faculty of law, Nanjing Agriculture University, Nanjing, China

*Corresponding author. Email: 9182210108@njau.edu.cn

ABSTRACT

Data economy is an important development feature of the 21st century. Data has become an important competitive resource among enterprises. However, China's legal system does not properly manage data, and emerging competitive resources. Data capture competition law regulation is one of them. In judicial practice, courts generally rely on Article 2 of the Anti-Unfair Competition Law and other competition law provisions that meet the characteristics of individual cases to regulate data grabbing. However, China still lacks criteria for judging the legitimacy of data scraping. To better regulate, the judgment principle of "protecting the long-term interests of consumers" should be established. Through analysis, the final criterion for judging the unfairness of the judgment has two components. First, to distinguish whether the captured data is open or not and whether the data that is crawled is disclosed, it is generally judged to be legitimate. In the case of crawling non-public data, distinguish between search engine bot crawling and non-search engine bot crawling. Search engine robots crawl non-public data, and non-search engine robots disclose the data they have crawled, which is generally considered to be improper. For non-search engine robots, if their subsequent use constitutes a substantial substitute for the original operator's services, it should be judged that the crawling behavior is improper. In the case of non-search engine robots that do not disclose the data they have crawled, other crawling behaviors other than "substantial substitution" are generally considered legitimate. Second, if the captured party invests a lot of money in the process of collecting and processing data, the grabbing party should give the captured party economic compensation. Finally, the three-tier classification system of fair use, statutory licensing, and prohibited use should be drawn on intellectual property law, and the judge's rules should be incorporated into the Internet provisions of competition law.

Keywords: data collection, legal regulation, judgment standard.

1. INTRODUCTION

The data economy is an important development feature of the 21st century. Software of all kinds, based on big data, provides a wealth of services to the public. Data has become an important competitive resource among enterprises.

However, in contrast to traditional competitive resources such as raw materials, the legal system does not adequately regulate the new competitive resource of data. Many legal regulations, such as data ownership and data security, need improvement. This includes the study of competition law regulation of data scraping.

Data scraping is the process of using automated algorithms to traverse web content in a predetermined path to capture and save the subject information to a local

database. Data scraping is a frequent occurrence in market competition. In practice, cases of data scraping have taken shape on a large scale. The legal regulation of data scraping involves many theories of law and economics. To promote the sound development of the data economy, it is of undeniable significance to improve the competition law regulation of data scraping.

Research is helpful to improve China's anti-unfair competition legislation. Data capture behavior involves the interests of many subjects and involves a large number of legal relations, so it is difficult to judge the wrongness of the behavior with a certain standard. China has yet to produce a widely recognized, systematic, and operable standard. However, this study can establish a three-level classification rule that evaluates data capture behavior as fair use, legally permitted use, and improper use. The problem of improper identification of data

fetching behavior may be solved easily and at the same time. This could help develop Competition law in China.

2. CURRENT STATUS AND DILEMMAS OF CHINESE ANTI-UNFAIR COMPETITION LAW ON DATA SCRAPING

2.1. Current Status of Regulation

Article 2 of the Anti-Unfair Competition Law is a miscellaneous provision and it stipulates that in their production and operation activities, business operators shall observe the principles of voluntariness, equality, fairness, and good faith and abide by laws and commercial principles.

In judicial practice, data scraping is regulated by the courts mainly by Article 2 of the Anti-Unfair Competition Law, supplemented by other competition law provisions with individual characteristics. Other competition law provisions with case-specific characteristics should be judged in the context of individual cases to determine whether they meet the prerequisites for application.

All in all, the current state of the law on unfair practices can be summarized by the phrase "Article 2 of the Anti-Unfair Competition Law + other competition law provisions that fit the characteristics of each case".

2.2. Regulatory Dilemma

The dilemma in the regulation of data scraping in our competition law is the result of the lack of an unfairness standard.

Article 2 of the Anti-Unfair Competition Law stipulates that operators should follow the principles of voluntariness, equality, fairness, and honesty in their production and business activities, and comply with the law and business ethics. In judicial practice, the court often argues through reasoning why the act of data scraping violates the principle of good faith and business ethics. The Supreme People's Court in the Kelp Quota case clarified the three elements of the application of the general provisions: firstly, the conduct is not specifically defined by law; secondly, the legitimate rights and interests of other operators are harmed by the competitive conduct; thirdly, the competitive conduct is wrongful because it violates the principle of honesty and credit and accepted business ethics. In addition, the *Weibo co. v. Pulse co.*, the court established the rule of "user authorization → platform authorization → user re-authorization" (hereinafter referred to as the "triple authorization rule") to determine the legitimacy of the crawl. The court determined the legitimacy of the crawl. The standard for the legitimacy of data capture seems to have been established. However, this is not the case.

Firstly, the general rules of application set out in the Kelp Quota case do not provide a clear standard for determining the legitimacy of data capture. It does not crystallize the principle of good faith and accepted business ethics, but merely establishes a prerequisite for the application of the General Clauses.

Secondly, the triple authorization rule does not address the problems with the criteria for determining the impropriety of data capture. First, while the rule adequately protects information security, it severely hinders the flow of information. The rules require that operators collect data held by other operators, in addition to obtaining the consent of the other operators, also need to obtain the consent of the user. Obtaining the user's consent again undoubtedly makes the flow of data more difficult, which in turn hinders the production of market value. The practical value of the rules is subject to further consideration. Secondly, data has multiple attributes such as sharability and public access. The rules, however, provide near-exclusive protection for data. Unlike objects in traditional civil law, data is hardly a private right object due to its public nature. This private law approach to protection cannot exclude the impact of the public nature of data. Therefore, the rules lack sufficient doctrinal support to achieve logical self-consistency. Thirdly, there is a risk of inconsistency between the rules and other laws. For example, Article 4(2) and Article 13(6) of the Personal Information Protection Law provide that publicly available personal information may be handled within a reasonable range. However, in the context of the decision in *Weibo co. v. Pulse co.*, the rules do not distinguish whether data consisting of personal information is publicly available. In short, the triple authorization rule does not serve as a standard for determining the impropriety of data crawling.

Finally, the miscellaneous provision of the Anti-Unfair Competition Law does not provide a clear standard for determining the impropriety of data scraping. The provisions are essentially based on "business ethics" to determine whether or not unfair competition is constituted. However, the content of business ethics is uncertain. Different courts have held differently on whether conduct is consistent with business ethics. In *Weibo co. v. Pulse co.*, the court held that whether consent was obtained and whether the user's free choice was guaranteed was a recognized business ethic. In *Headline v. Baidu* and *Fairview co. v. Taobao co.*, the court judged whether the data captured was used in a homogenous manner. In the case of *Qihoo co. v. Baidu co.*, the court used the Internet industry's self-regulatory convention as the standard of judgment. This shows that China has not formed a set of clear and specific criteria for interpreting the content of business ethics in data capturing cases.

Given the above, it is necessary to establish a rule for determining the unfairness of data scraping.

3. COUNTERMEASURES AGAINST UNFAIR COMPETITION IN DATA SCRAPING

3.1. Fundamental Principles for Judging the Unfairness of Data Scraping

The so-called principle is the legal practitioner based on long-term judicial practice experience, the use of rational analysis, the variety of cases to be categorized, organized, so that they are gathered under a general standard or guideline and will be used to justify the reasonableness and legality of future decisions, to respond to the rapidly changing social situation [1]. Just as legal principles guide the application of the law, the principles that determine whether data capture is justified or not guide the interpretation and application of the rules of justification.

The standard of data capture regulation is, in essence, to strike a balance between the legal interests of the data capture party, the ISP (the party whose data is captured), and the user. The judgment of the legitimacy of competitive behavior has the inherent property of weighing interests [2]. The crawler has the right to search for commercially valuable data for its development. Internet service providers, who have worked on the data they have collected and processed, have a legal need to protect it. Users, on the other hand, have a legal right to their information.

The principles governing the legitimacy of crawling directly guide the rules on how to strike a balance between these three parties. The legal basis for the protectability of data capture is that in a market economy, data in circulation can often create greater economic value than static data. Internet service providers, in the process of collecting raw data, have made significant "inputs", at least in terms of basic labor. If their rights and interests are not recognized by law, they are likely to encourage "free-riding" and "cannibalization" and undermine the incentives of operators. Users have the right to their personal information. No operator should disregard the law and infringe on the legitimate rights of users.

Although data scraping cases are not uniform, this does not mean that we cannot extract the underlying principles for determining the impropriety of data scraping from complex cases. The primary legal interest objective of economic law is the public interest of society. The public interest in economic law (which in practice means the public interest of society) refers to the sound development of national economic democracy with the keynote of ensuring the interests of consumers [3]. Indeed, the protection of the long-term interests of consumers is the fundamental principle underlying that

judgment of data capture, justified or not. The fundamental principle of "protecting the long-term interests of consumers" should guide the application of the rules in cases where there is a problem with interpretation and where there is a lack of understanding.

The values of the Anti-Unfair Competition Law include encouraging every market actor to legitimately obtain the maximum benefit, providing a fair opportunity to compete, and making the outcome of competition for market operators fair [4]. The ultimate goal of protecting the competitive interests of operators is, ultimately, to protect the interests of consumers. Protecting the interests of competitors can maintain the stability of the market economic order. With a stable order, the market economy can prosper. The fundamental value of a prosperous market economy is to enable the general public, to enjoy quality services. It can be said that the protection of the interests of competitors is a means to an end, while the protection of the interests of consumers is the ultimate goal. The interests of consumers, however, are more abstract and difficult to concretize. It is difficult to say whether the interests of consumers are represented by lower prices in exchange for better quality services, or by the sustainability of service provision in the market, or other market characteristics. The interests of consumers, however, are easily crystallized in individual cases. The judge can judge how the decision can be made in such a way that the interests of the consumer are protected. In addition to this, the interests of the consumer are divided into short-term and long-term interests. Very often, short-term and long-term interests conflict. For example, a vicious competition between businesses for low prices will result in a great short-term benefit to consumers, but a loss of long-term benefits due to monopolies. The pursuit of short-term benefits, therefore, places the long-term interests of consumers in the hierarchy of legal benefits ahead of short-term benefits. The long-term interests of the consumer are the highest priority for the protection of the legal interests of competition law. The determination of whether data capture is justified or not should be in line with the basic values of economic law and competition law. Therefore, the rules on the determination of illegality should be designed around the "protection of the long-term interests of consumers".

In summary, "protecting the long-term interests of consumers" should be used as the fundamental principle to guide the design and application of the rules for determining the legitimacy of data scraping.

3.2. Construction of Rules for Judging the Illegitimacy of Data Scraping

The thinking that guides the establishment of the judgment criteria

The prerequisite for determining the criteria for judging the legitimacy of data capture is the

establishment of two types of thinking, namely "data sharing thinking" and "data contribution thinking". These two types of thinking are essential in guiding the development of the judging criteria. The sharing mindset is used to clarify that data is generally shared by the operator, while the contribution mindset is used to determine when data is not shared by the operator.

3.2.1. Data Sharing Thinking

The core of data sharing thinking is that "data is shared, and data is controlled only in certain circumstances". The reasons for this approach are as follows.

Firstly, traditional private law objects are scarce and it is necessary for the law to 'empower' them to stop disputes. Data is public, and it meets the core characteristics of a 'public good' in economics, namely non-rivalry and non-exclusivity. The privatization of data will undoubtedly have a substantial impact on the sharing of information. The traditional system of privatization in private law was initially based on the scarcity of objects (mainly tangible objects), which led to the need for a legal necessity to stop disputes. Information, which is plentiful and shared, belongs to a completely different theoretical category than the typical object in private law (the object) [5]. The greater the volume of data in circulation, the greater the economic value it creates, often. The use of empowerment to uniformly grant public data on internet platforms to a specific subject is detrimental to the sharing and mobility of data [6]. In addition, data is disseminated in large quantities on the Internet without creating physical damage such that it reduces economic value. In this way, the widespread use of data completely avoids the economic "tragedy of the commons".

Therefore, the traditional civil law system is unable to deal with a range of data law issues, including data capture issues. The construction of rules for judging the legitimacy of data capture is very necessary to break away from the traditional perspective of protecting private interests, facing the public interest, and promoting the development of data public order, with new thinking as the premise of construction.

Second, competition law is inherently modest. Competition in a market economy is, in principle, a "private matter" between operators. Free competition is the soul of a market economy, and the private law concept of "freedom without prohibition" must be firmly embedded in the system of competition law. As a typical intervention mechanism, the prohibition of subject-specific competition by competition law is essentially aimed at achieving more general freedom of competition, and the restriction of subject-specific freedom is limited to what is necessary to protect the order of competition [7]. So, in principle, operators can use data that is close to a public good as they wish. The law can only restrict it

in certain circumstances.

Thirdly, the data sharing mindset strikes a delicate balance between the full use of data and the protection of the interests of data collectors. It allows the economic value of data to be fully utilized in general terms, while at the same time restricting the flow of data where legal protection is necessary to protect the rights of data collectors.

In short, it is essential to use data-sharing thinking to address the issue of data capture.

3.2.2. Data Contribution Thinking

The core of data contribution thinking is that the extent to which an operator contributes to data determines the extent to which the law protects it. Only when an operator makes a "contribution" to the use of data does he or her behavior warrant legal protection. No contribution, no protection. It determines the circumstances under which the law can prohibit the sharing of data. The "homogenous use of data" is the reason why the law prohibits data sharing. This is important for the establishment of subsequent rules of judgment, which are derived as follows.

Firstly, once the premise of data sharing has been established, a "negative list" approach can be introduced to establish the standard of illegitimacy. The negative list is an enumeration of scenarios where data capture constitutes unfair competition. Once these scenarios have been excluded, the remaining data capture practices should generally be considered legitimate. In this way, the boundary between the intervention of public authority and the freedom of conduct of private law subjects can be delineated, avoiding the imputation of data capture parties and excluding excessive intervention of public authority, in line with the spirit of competition law.

Second, in judicial practice, the number of data crawling judged to be improper is high. With the development of Internet technology, the number of relevant cases will increase, and it is difficult to exhaustively list them in the negative list. Therefore, it is important to summarise the essential characteristics of improper data capture, extract its "public causality", and include the type of behavior in the negative list. This axiom is the "homogenization of data". Homogenization" means that the operator does not contribute to the growth of the value of the data. Naturally, there is no need for the law to protect their behavior.

In particular, operators often need to collect and process large amounts of data before they can provide their services, and more or less invest in them. However, competitors can capture the data they collect and replicate the services they offer at a fraction of the cost. This homogenization can distort the competitive relationship between operators of similar services. As a result, there

is a lack of incentive to improve services between competitors offering the same services. Such capture provides little new value to the market, but rather undermines the incentive for operators to innovate, leading to a less efficient market and damaging the long-term interests of consumers. Therefore, even if the data is premised on sharing, there are good reasons for the law to prohibit the homogenized use of data after capture.

In short, it is the operator's contribution to the growth of the value of the data that makes it possible for the law to protect the operator's interests. The homogenous use of data is a direct cause of the unjustified nature of data capture.

4. THE PROCESS OF JUDGING THE CONSTRUCTION OF RULES

4.1. The Cut-off Point for Rule Construction

After establishing the mindset of data sharing and contribution, we can gradually construct judgment rules in conjunction with the specific scenario of data capture. However, the specific scenarios of data capture are too complicated. It is necessary to choose an appropriate perspective to start the analysis. A possible approach is to use the question of whether the data is publicly available as a starting point, and gradually analyze it about the various factors that affect the judgment of legitimacy. The reasons for this are as follows.

As mentioned earlier, sharing thinking and contribution thinking are the basic thinking behind the issue of data capture. Data capture is, in general, carried out in a shared context and there is no case where the capture is not justified. It is generally the contribution of the operator to the data that blocks the contribution of the data and enables the law to find the data capture unjustified. Therefore, we can say that the rules for determining the legitimacy of a seizure are based on the premise of "data sharing" and the contribution of the operator to the data.

One of the states in which data is most closely associated with 'sharing' is the state of the public. Apart from open data, the rest of the data is closed data. These two types of data are not shared to the same extent. To study 'contribution' in the context of 'sharing', it is necessary to first distinguish between these two states of data. Then, in each case, we examine how the operator's contribution affects the determination of the legitimacy of the crawl.

It is, therefore, feasible to distinguish between the scenarios of whether the data is publicly available or not, and then to discuss separately how crawling should be determined in each case.

4.2. Analytical Process of Rule Construction

Having determined whether the data is public or not as an entry point, we can then proceed with the analysis.

If the data is already publicly available, then the crawl should generally be justified. There are two reasons for this. Firstly, the data is in principle shared by all operators. There are no barriers to accessing data that is publicly available to the operator. In principle, any operator is entitled to direct access to publicly available data. Secondly, while people should enjoy the convenience brought about by internet technology, they should moderately bear the risks associated with the disclosure of information. In the US case of *hiQ Inc. v. LinkedIn Corporation*, the court found that *hiQ* was justified in capturing LinkedIn's publicly available user information. The judge likened internet space, to physical space, holding that unauthorized entry into a private home is usually illegal, but for shops that are open for business, access is available to the general public. Like an open shop, the openness of public data across the internet dictates that anyone can access it. Some academics have criticized this decision, arguing that the judge ignored the psychological needs of LinkedIn users. *hiQ* crawled LinkedIn users' job search information and analyzed and processed it for the sake of the users' real-life employers, to be able to better assess the users. However, according to general social perceptions, few people would want to let employers know about their job search trends. While this criticism is certainly justified, it does not negate the reasonableness of the judge's decision, as LinkedIn users benefit from the disclosure of information, they should also bear the risk of disclosure. The employer's knowledge of the user's job search was through legitimate channels, and the negative impact on the user was not unacceptable but was a proportionate risk that the user should take. Therefore, the judge's finding was not inappropriate.

All in all, the crawling of publicly available data by an operator is generally found to be justified.

If the data is closed, a distinction should be made between search engine bots crawling and non-search engine bots crawling. The crawling of non-public data by search engine robots and the making public of the data crawled by non-search engine robots is generally considered improper. In the case of non-search engine robots, a crawl should be judged to be improper if its subsequent exploitation constitutes a substantial substitution for the original operator's services. Where a non-search engine robot does not disclose the data it has crawled, crawling other than where it constitutes a "substantial substitution" is generally considered to be justified. The reasons for this rule are as follows.

Closure often means that the operator is unwilling to share the data with the public. According to Locke's doctrine of the natural right to work, the world is common

to mankind, men are rational, and men make use of it to live. At the same time, every man has a right of ownership over his person, and the labor performed by his body and the work done by his hands belongs to him. Thus, he adds his labor to the object so that it gains a corresponding increase in value, and he also has the right to acquire it [8]. The operator, naturally, has a property interest in the data he has collected and processed. Thus, it would seem that if the seizure of data in a closed state infringes the property rights of the operator, it should always be considered improper.

In reality, however, this is not the case. The legal consequences of infringement of property rights must be determined by taking into account several factors, including the consequences of the infringement. It would be too arbitrary to make a blanket determination of impropriety without considering the impact of the crawl. In practice, the law may restrict property rights to a certain extent and legitimize certain minor infringements of property interests in the public interest, among other considerations. The most typical examples are the "fair use" and "statutory license" regimes in intellectual property law. In addition, data is, to some extent, a public good. No one can subjectively take public goods for himself or herself. Therefore, the determination of whether a crawl is justified or not can be based solely on the subjective will of the operator. The act of grasping alone has little to do with the determination of impropriety. Moreover, whether or not the act of crawling alone is consistent with business ethics and whether or not it is justified is of little significance in regulating unfair competition. What matters is how the data is used after it has been captured. For example, if the operator does not use the data after crawling, but keeps it in storage, this will not generally lead to unfair competition disputes. If, for example, the crawling of non-public data is done to make it known to others, the security of the Internet data storage is compromised and the action is naturally difficult to be found justified. Therefore, the impropriety of crawling closed data must be judged on how the data is used after crawling.

There is a significant difference between search engine bots and non-search engine bots in terms of the "manner of exploitation", and a distinction should be made. Search engine bots serve search engines, and their crawling of information facilitates the sharing of information. However, the prerequisite for this is that users can find relevant information without the need for a search engine bot. Robots are there to facilitate access to information, not to help users gather the information that others do not want to make public. If search engines were free to crawl through the information that others do not disclose, there would be no secrets in the online world, which would be terrifying. Similarly, if non-search engine bots crawl, will crawl the non-public data, the public naturally should also be judged as improper. In other cases, the search engine bot has accelerated the

sharing of information and the crawl should be considered legitimate.

For non-search engine robots, which are more common among operators, their crawling data is generally used for their operational development, rather than making the data public, so the crawling behavior should not be regarded as improper. It should be judged based on whether the services provided by the operator after the data capture constitute a homogeneous substitution. The reason is, as mentioned above, homogenous use of data, which "maliciously" diverts competitors' traffic through free-riding and does not make a new contribution to the market, is the public cause of improper data capture and is the reason why data capture is prohibited by law.

4.3. Financial Compensation after Judgment

As mentioned above, crawling is justified if it does not constitute homogenous exploitation. However, the data collectors have invested a great deal of effort and their interests should be worth protecting. After all, they have contributed to the collection of the data and, according to the 'contribution mindset', that contribution is the source of the legal interest.

Combining the labor doctrine of property rights with the theory of intellectual property. Locke argues that when labor is added to the object of labor, it separates it from the commonwealth, thus transforming it into a private property right [10]. In addition, many operators, who have collected data, have processed it in-depth, transforming the raw data into derivative data. These derivative data often have a commercial value far beyond the original data, and to a certain extent, even have originality and meet the standards of intellectual property rights. For example, in the case of the dispute between Vision and Taobao, the business reference product.

A viable approach is to financially compensate the crawler of the data. Data-sharing thinking is essentially a trade-off between the benefits of the market and the interests of the data collector. With data contribution thinking, the loss to the data collector can be offset using financial compensation for data sharing.

In short, if the data collector contributes to the data, it is necessary to compensate them financially.

4.4. Summary of the Rules for Judging Data Crawling

Combined with the above argumentation and analysis, we can finally arrive at the impropriety judgment standard, which consists of two points, specifically as follows.

First, distinguish between the crawling of data public or not, crawling of public data, is generally judged as

justified. In the case of crawling non-public data, a distinction is made between search engine bot crawling, and non-search engine bot crawling. A search engine bot crawling non-public data, and a non-search engine bot crawling "closed data" that is public, are generally considered improper. In the case of non-search engine robots, crawling should be considered improper if the subsequent exploitation constitutes a substantial substitution of the original operator's services. In the case of non-search engine robots, where the "closed data" from the crawl is not made public, crawling other than in the case of "substantial substitution" is generally considered to be justified.

Secondly, if the crawled party has made a significant investment in the collection and processing of the data, the crawler should provide financial compensation to the crawled party.

However, it is worth noting that as data technology develops, data capture cases will become increasingly complex. The above rules are only applicable in general. In the case of a particular data capture case, the application of the rules would result in a clear injustice to the case. Judges should take into account the circumstances of the case and the underlying principle of protecting the long-term interests of consumers to reach a decision that best reflects the fairness of the individual case.

4.5. Judgment Rules Legalization

4.5.1. Three-tier Classification System

After establishing the criteria for judging the impropriety of data scraping, it is necessary to establish the corresponding legal regulation to absorb the criteria. The three-tier classification system of prohibited use, statutory permission, and fair use, which is based on intellectual property law, is a good way to incorporate the standard and establish a logical and strict legal rule. This three-tier classification system balances the conflicting values of data flows and the labor input of data collectors. It ensures that, in most cases, the data is fully utilized and that the data collectors are rewarded for their work. By the aforementioned rules of judgment, the system of legal regulation of the impropriety of data capture should be constructed as follows.

Firstly, publicly available data can be crawled at will and considered fair use of data. This will ensure the circulation of data and promote information sharing, and also remind the public, to a certain extent, that they should bear the information risks brought about by the Internet era.

Secondly, data crawlers, using non-search engine bots, can crawl data at will without constituting a homogenous use of the data and without making the "closed data" they have crawled public, which is

considered a legally permitted use of the data. The data collector is financially compensated. The financial compensation is paid by the data collector. The reason for this is that the data capture party is the beneficiary of the data collected by the data collector to earn a financial benefit. A right entails a corresponding obligation. As the beneficiary of the data capture, the capturer is naturally obliged to pay financial compensation. Such a provision reflects the value of fairness in law. The amount of financial compensation is proportional to the labor input of the collecting party. Such a system is designed to maintain the incentive for operators to collect data and, to a certain extent, to balance the financial losses caused by data crawling.

Finally, the use of non-public data by search engine robots, the disclosure of "closed data" by non-search engine robots, and the homogenization of data collected by non-search engine robots, should be directly deemed improper and prohibited.

4.5.2. Incorporation of Legal Provisions

There is a wide range of opinions on how to incorporate the legal regime governing data crawling into existing legal provisions. However, incorporating a three-tier classification system into the Internet provisions is the most appropriate and logical approach to competition law.

Firstly, data scraping is already inextricably linked to the Internet industry. Including the rules governing it on the internet, the clause is in line with the spirit of the internet clause and is easy for lawmakers to find. Secondly, the Internet provisions are supposed to organize the types of unfair competition related to the Internet. The inclusion of a three-tier classification system is in line with the logic of the categorization of competition law provisions. The inclusion of data scraping in addition to the Internet provisions is not a surprise.

In summary, it would be a good idea to include a three-tier classification system for data capture in the Internet provisions of the competition law.

5. CONCLUSION

Internet technology has changed the outlook of China's economic development and promoted the development of the market economy. The governance of unfair competition brought about by Internet technology is conducive to the healthy development of China's market economy. This study puts forward a solution to the unfair competition of data capture, which is of unquestionable importance to maintain the order of the Chinese market economy and protect the rights and interests of operators and consumers.

To regulate data capture more effectively, the

fundamental principle of "protecting the long-term interests of consumers" must be established. At the same time, it is necessary to define a specific and clear regulation of unfairness, taking into account the nature of data, market efficiency, and business ethics. Firstly, a distinction should be made between data that is publicly available and data that is not, with data that is publicly available generally judged to be legitimate. In the case of crawling non-public data, a distinction is made between search engine bots and non-search engine bots. A search engine bot crawling non-public data, and a non-search engine bot crawling "closed data" that is public, are generally considered improper. In the case of non-search engine robots, a crawl should be considered improper if its subsequent use constitutes a substantial substitution for the original operator's service. In the case of non-search engine robots, where the "closed data" from the crawl is not made public, crawling other than in the case of "substantial substitution" is generally considered to be justified. Secondly, if the crawled party has made a significant investment in the collection and processing of the data, the crawler should provide financial compensation to the crawled party.

After establishing the criteria for judging the impropriety of data capture, the three-tier classification system of fair use, statutory license, and prohibited use should be borrowed from intellectual property law, and the judging criteria should be incorporated into the Internet provisions of competition law. In this way, we have a solution to unfair competition in data capture.

In this era of information explosion, the legal issues concerning data have been growing and declining. The legal issues of data, however, are often very complex. Fortunately, however, we live in a time of great talent. Scholars are coming and going to tackle data law challenges. We believe that one day we will be able to establish a comprehensive data law system.

REFERENCES

- [1] Li Ke, Some Issues of Principles and Rules, *Journal of Legal Studies*,2001.
- [2] Kong Xiangjun, On the Basic Paradigm of Anti-Unfair Competition, *Jurist*,2018.
- [3] Wang Baoshu, On the Legal Benefit Objectives of Economic Law, *Journal of Tsinghua University (Philosophy and Social Sciences Edition)*,2001.
- [4] Cao Xinming, On the Value Orientation of Anti-Unfair Competition Law, *Journal of Central South University of Political Science and Law*, 1996.
- [5] Mei Xiaying, Between sharing and control, the limitations of private law and the construction of public order in data protection, *Journal of Chinese and Foreign Law*,2019.
- [6] Li Xiaoyu, Reflection on the Empowerment and Protection of Internet Platforms in the Era of Big Data and the Approach to Legal Remedies, *Intellectual Property*,2021.
- [7] Zhang Zhanjiang, On the Humility of Anti-Unfair Competition Law, *Jurisprudence*,2019.
- [8] Li Zhaoyang, Reflection and Amendment of Data Capture Behavior Regulation from the Perspective of Anti-Unfair Competition Law, *Jinan Journal (Philosophy and Social Sciences Edition)*,2021.
- [9] Wang Ran, On the Rules of Data Benefit Distribution of Network Open Platform, *Electronic Intellectual Property*,2020.
- [10] Zhu Li, Property Rights Labor Doctrine and Intellectual Property – Can Labor Doctrine Justify Intellectual Property? , *Science and Technology and Law*,2006.