

An Empirical Study on the Judicial Regulation of Unfair Competition in Data Scraping Behavior

Jinjin Gu^{1,*}, Jinrui Gu²

¹School of Anhui University of Finance and Economics, Anhui, China;

²School of Bengbu Academy, Anhui, China.

*Corresponding Author

Abstract

Currently, China's data-related legislation is still in the initial stage of gradual exploration. Due to the unfair competition disputes involving data capture with a variety of data types, complex technical means, capture behavior hidden and other characteristics, resulting in the judge in the trial of data capture cases mostly choose to apply the general provisions of the "Anti-Unfair Competition Law", that is, the second article to be adjudicated, which makes the data capture behavior of the cases arising from the dispute is more and more, the data capture behavior whether it constitutes unfair competition, This has led to an increasing number of cases and disputes arising from data-scraping behavior. Therefore, this paper takes the data capture behavior as the research object, takes some typical cases as examples, discusses the adjudication results and thoughts of the cases, and in the process, finds that in the judicial practice of such dispute cases, there exists the difference of the applicable law, the judgment standard of the legitimacy of the data capture behavior is not clear, and the understanding of the damages caused by the unfair competition behavior of the data capture is not the same, and other problems. In view of this, in judicial practice, Article 2 and Article 12 should be clearly applied, and importance should be attached to the criteria for judging the legitimacy of data capture; finally, through the theory of balance of interests and the principle of proportionality, the interests of data platforms, consumer interests and public interests of the society involved in data capture should be taken into consideration, thus realizing the balance of interests of multiple parties, and achieving the balance between data protection and data flow through dynamic justice.

Keywords

Data Grabbing Behavior, Illicit competition, Division of Law and Regulation.

1. Introduction

The continuous emergence of new industries, new business forms and new modes cannot be separated from the driving role of data, which has become an important resource for market players in the digital economy. Against the background of the ever-eminent value of data, the protection of data has also received constant attention. 2022 On December 2, the Central Committee of the Communist Party of China (CPC) and the State Council announced the Opinions on Building a Data Base System to Better Play the Role of Data Elements, which further clarifies that a data base system is being built with a focus on data property rights, circulation and trading, revenue distribution, and security and governance. At the same time, the State and various localities are actively exploring data-related legislation and practices. In recent years, there has been a gradual increase in the number of cases involving unfair competition in data capture heard by courts across the country, with the cases often involving emerging network technologies and business models, and unfair competition against platform data presenting

features such as complexity and covertness. However, at present, China's data-related legislation is absent, what should be taken to regulate the improper use of platform data capture, how to reasonably balance the interests of all parties, and how to protect the rights and interests of the data while promoting the circulation of data and other issues, it is urgent to continue to explore in theory and practice. This paper starts from the empirical research of judicial decisions to explore the justification rules of judicial determination of data capture behavior, with a view to further releasing the effectiveness of data elements and promoting the development of the digital economy while providing references for judicial practice.

2. A Clarification of the Connotation of Unfair Competitive Behavior in Data Scraping

2.1. The Concept of Unfair Competitive Behavior in Data Scraping

Data scraping unfair competition behavior goes hand in hand with legitimate data scraping. With the development of the digital industry, the security risks generated by the data come along with it, and the competition between operators about the data also becomes more intense. However, in judicial practice, there is no clear concept of unfair competition behavior of data scraping. The definition of unfair competition behavior of data scraping in this paper refers to the behavior of disrupting the order of market competition triggered by the use of data scraping to obtain the data of the data owner, and then to make use of the data. Essentially, it is to obtain others' data through data scraping to increase one's own competitive advantage, or to weaken others' competitive advantage through data scraping.

Clarify the connotation of unfair competition behavior of data capture, first of all, it should be clear that the data captured must be commercially valuable data, with commercial value of the main considerations are the data operator in the generation of data invested in the cost of the data is data operators limited to the provision of data. After proper processing of the collected information such as desensitization, screening, format optimization, encryption and screening, it is managed using technological methods to provide commercial data only to a specific group of people. Including the cost of human and material resources, this kind of capture results in a reduction of the competitive advantage of the data operator, and this kind of data does not fall into the category of "work" in the Copyright Law. Once the data are captured, and the captured data are used to provide products or services in place of competitors' products or services, the data will fall into the category of unfair competition, infringing on the legitimate rights and interests of the data operators themselves in respect of the data, as the data information is property, and the data operators have gained a competitive advantage by paying for the costs, which belongs to the operator's capital, and which can bring profits to the data operators.

2.2. Characteristics of Unfair Competitive Behavior in Data Scraping

2.2.1. Diversification of data types

Given the multiple attributes of platform data, different types of data may present completely different characteristics. The identification of data types in judicial practice presents diversified characteristics, such as "Tencent v. Taozhuo" case, which divided data into authorized data and unauthorized data; "Weimeng v. Jane Yixun" case, the court of first instance classified data into user-generated, service-generated, and data products, and the court of second instance classified data into user data and platform data. The court of second instance categorized data into user data and platform data. In addition, there are original data and derivative data, public data and non-public data, single data and data collection. Scientific data classification can provide a reference for clarifying the basis of the platform operator's rights and interests and judging the legitimacy of data capture behavior.

2.2.2. Diversification of data types

In disputes involving unfair competition involving data capture, complex technical measures are often involved. While the defendant adopts technical means to capture platform data, the plaintiff also invests in human and technical costs to fend off the data capture behavior of other subjects. The technical protection measures taken by the plaintiff mainly include setting up Robots agreement for public data, setting up access rights for unpublished data, setting up database key, picture verification mechanism, etc.. The technology adopted by the defendant, in addition to crawling through the network crawler, also contains the means of calling the platform API port to realize server interaction, bumping the library, associating the external network account, Xposed plug-in, simulating the manual clicking operation of the RPA technology, etc. In the case involving unfair competition in data crawling, the identification of the technical means is indispensable in the judicial practice. On the one hand, the defendants in some cases will use technology neutrality and technological innovation as a defense. For example, in *Taobao v. Meijing*, Meijing claimed that the accused behavior should be exempted from liability according to the principle of technology neutrality. On the other hand, the technical measures taken by the defendant may also be one of the factors to be considered in determining whether it constitutes unfair competition. For example, in the case of *Palm v. China Services*, the defendant adopted the technology of "database crashing" to try to log in to the plaintiff's website with its own unencrypted member database account and password, and after verifying that some of the accounts and passwords were the same or basically the same, the court finally completed the logging in to the plaintiff's website, and the court thus found that the defendant was using the technology-neutral principle to avoid liability. Login, the court concluded that the defendant is through improper means to obtain data.

2.2.3. Data scraping behavior is covert

The lack of identification of technical means in judicial practice is mainly due to the covert nature of the act of data capture. The actors, technical means and effects of data capture are all hidden. First, it is difficult to investigate the behavioral subject of data capture. Compared with large-scale Internet platforms that control data, most of the data capture actors are unknown small and medium-sized enterprises, or even natural persons, and there is no business cooperation between the two sides, so it is difficult for the platform to investigate the main body of information of the data capture actors. Second, it is difficult to identify the technical means of data capture. The data grabbing party may grab data by dispersing multiple IP addresses, simulating user clicks, crashing the library, etc. The results presented by the above behavioral patterns are consistent, making it difficult to trace the technical means of data grabbing. Third, the effect of the behavior is more hidden. For example, the front-end data capture behavior is shown as repeated filtering and screening of computer information, etc., but in the process, the competitive purpose it wants to achieve is to achieve such capture effects as tampering with the user's browser homepage, differentiating between browsers of different vendors, algorithmic discrimination, and data 'hitchhiking' through data capture behavior. "

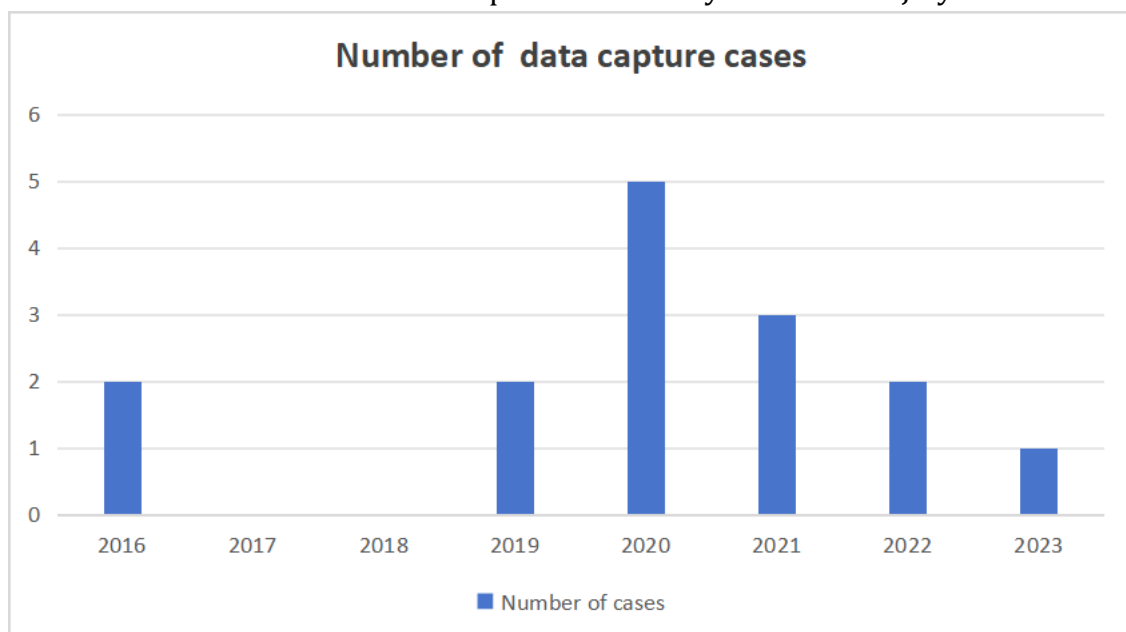
3. Sorting out and Analyzing Cases of Unfair Competitive Behavior in Data Scraping

In order to ensure the completeness and scientificity of the samples, this paper retrieved 15 adjudication documents on China Adjudication Documents Online with the search keywords of data crawling and unfair competition, and used them as the basis for analyzing and researching the number and distribution of this type of cases, as well as the adjudication results and bases of the cases.

3.1. Number and distribution of cases

From the beginning of 2014 to the end of July 2023, among the cases published on the China Judgement and Documentation Network, the people's courts at all levels closed a total of 14 cases in the form of judgments with the keywords of data crawling and unfair competition, and 1 ruling. Its number began to show an upward trend in 2019 with the development of the Internet economy and showed a relative decline after reaching a maximum of 5 in 2020 (see Table 1), which is related to the fact that China has gradually begun to pay attention to this kind of cases and has gradually regulated the behavior of unfair competition on the Internet. Moreover, among these 15 adjudication instruments, 9 were made by Beijing People's Courts at all levels, and the number of adjudication instruments made by the rest of the provinces and cities was 1 or 2. It can be seen that the number of cases of data crawling as well as its distribution is closely related to the economic development of different regions in China. Data capture is a competitive behavior that infringes on the rights and interests of the captured party through the medium of the Internet, and the captured party is usually a large Internet enterprise, such as Tencent, Baidu, etc. Beijing, as the capital of China and also an important region for the development of the Internet economy, can attract most Internet enterprises and provide more development opportunities for these enterprises, and thus becomes the headquarters of most Internet companies. As a result, the data capture cases accepted by the people's courts at all levels in Beijing accounted for 57.14% of the total number of cases accepted nationwide, which is in line with the positive correlation shown between the development of the Internet economy and judicial practice.

Table 1: Number of data capture cases early 2014-end of July 2023

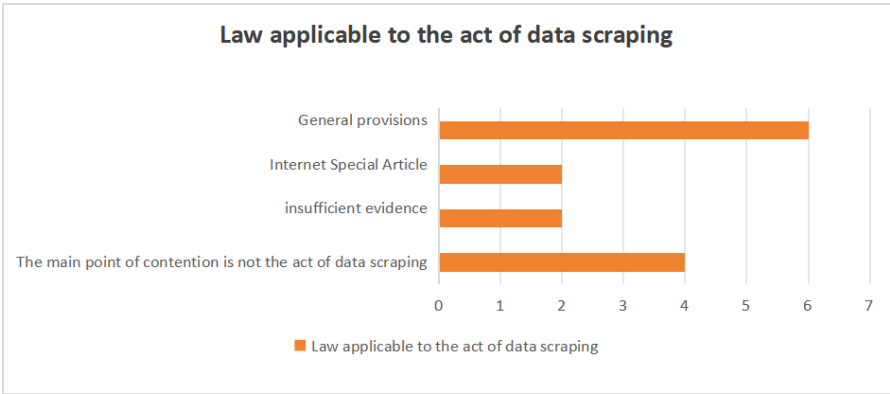


3.2. Findings of the Court and the basis thereof

At present, China's data capture behavior is generally regulated by the "Anti-Unfair Competition Law", and the "Anti-Unfair Competition Law" did not set up special provisions, resulting in judicial practice, the judge often cited the general provisions of the Internet or commercial secret provisions of the judgment[1]. This paper found that 8 of the 14 judgments used general provisions to determine the illegality of data scraping, 1 judgment chose to use the Internet article to determine the illegality, and the rest of the judgments rejected the plaintiff's litigation requests for lack of evidence or the main point of contention in the case was not the data

scraping behavior (see Table 2). It can be found that in judicial practice, the application rate of the general provisions for data crawling is much higher than that of the Internet special provisions and trade secret provisions. This is because, first, although the data capture behavior as a technical means, can not be separated from the Internet, but for the Internet special provisions of the three kinds of behavior, namely, traffic hijacking, interference with the operation of the malicious incompatibility can not be well encompassed by the data capture behavior[2]. Therefore, the application rate of the Internet special article in the data capture behavior is low. Secondly, the premise of the application of commercial secret provisions for the capture of data for commercial secrets, and the determination of commercial secrets in addition to the data itself has the secret, but also requires the captured enterprise to take confidentiality measures for the data, however, with the technological iteration, the cracking means more and more sophisticated, cracking the confidentiality of the low difficulty of confidentiality measures is difficult to be recognized as the confidentiality of the measures taken, resulting in the commercial secret provisions are difficult to be applied to the act of data capturing.

Table 2: Laws applicable to data scraping



4. The Realistic Dilemma of the Judicial Regulation of Unfair Competition in Data Scraping Behavior

4.1. Differences in applicable law

Different ways and means of data capture behavior may lead to differences in the application of the law. Of the 14 judgments analyzed above, six applied general provisions and two applied Internet-specific provisions. Even for the same conduct, judicial decisions still have different perceptions of which provisions of the Unfair Competition Law should be applied to regulate such disputes. The current unfair competition disputes involving data scraping usually cannot be included in the specific types of unfair competition in Chapter 2 of the Anti-Unfair Competition Law, but mainly rely on Internet-specific articles and general provisions. Among them, the fourth item of the second paragraph of the Internet Article is the main one, and the legal elements of the Article include: whether it utilizes network technical means, whether it affects the choice of users or other subjects, whether it targets the network products or services legally provided by others, and whether it hinders or destroys the normal operation of the said products or services of others. The behavior adjusted by the Internet Article has certain limitations and is not sufficient to regulate all unfair competition behaviors involving data capture, so the platform operator will advocate for relief through the general provisions. In most disputes involving unfair competition in data scraping, plaintiffs usually claim both the general terms and the Internet special provisions in order to maximize their own interests and

try to obtain double insurance. This tendency also reflects the lack of awareness of most parties in choosing the application of the law.

In judicial practice, different courts have ambiguous and intersecting understandings of the elements of application between specific provisions and between specific and general provisions. Two cases, the Jitterbug False Brush Volume Case and the Tencent Video False Brush Volume Case, involved the issue of increasing the number of clicks and broadcasts for videos, with the former case being regulated by the network provisions of Article 12 of the Anti-Unfair Competition Law, and the latter case applying both the second paragraph of Article 8 and the second paragraph of Article 12 of the Anti-Unfair Competition Law. In the case of Tencent v. Cloud Power for Unfair Competition, the court of first instance held that the manner of determining the constituent elements of unfair competition through the Internet-specific article was not complete, and that it should be combined with the constituent elements and judgmental paradigm of the general provisions of article 2.

4.2. Unclear standards for judging the legitimacy of data scraping behavior

According to Article 2 of the Anti-Unfair Competition Law, operators shall abide by the “principle of good faith” and “business ethics”. In the absence of specific and clear provisions in the Anti-Unfair Competition Law to regulate a certain competitive behavior, the court generally tends to apply the “principle of good faith” and “business ethics” to measure the legitimacy of data scraping behavior. However, “ethical norms are not engraved in marble, but exist in all aspects of social life, and their content is characterized by diversity and uncertainty. The lack of sufficient certainty makes it largely difficult for ethical norms in the context of market competition to rise to the level of a specific and stable rule of competition.[3]” The act of data capture mostly involves new fields, technologies and modes, and the plaintiffs and defendants have their own opinions on business ethics, which also leads to more difficult moral judgments in judicial practice.

In judicial practice, the court often points out in the reasoning part of the judgment that a certain data capturing act violates the recognized business ethics and the principle of honesty and credit at the same time, harms the lawful rights and interests of other operators, disrupts the social and economic order, and conforms to the general provisions of the Anti-Unfair Competition Law, so as to find that the data capturing act has the impropriety[3]. However, in the Internet era, all kinds of business rules are still in the groping stage, the right boundary of market subjects is not clear, some behaviors may bring benefits to the market and consumers while harming the interests of some market subjects. This will not only make it difficult to judge whether some new competitive behaviors violate business ethics, but also lead to the blurring of the boundaries between business ethics and the principle of honesty and trustworthiness.

In judicial practice, there are a variety of standards for determining the legitimacy of data capture, such as the “triple authorization” principle proposed by the court of second instance in the case of “Weimeng v. Taoyou”; the court in the case of “Weimeng v. Taoyou” explicitly proposed the “triple authorization principle” for the commercial use of user data and information in the Internet, i.e., open platform parties need to obtain user authorization when directly collecting and using user data; third-party developers need to obtain user authorization through open platforms. The court in the “Weimeng v. Taoyou” case explicitly proposed the “triple authorization principle” for the commercial use of user data and information in the Internet, i.e., the open platform party needs to obtain authorization from the user for the direct collection and use of the user's data; the third-party developer indirectly obtains the user's data through the OpenAPI interface of the open platform, which should obtain the authorization from the user and the authorization from the platform party at the same time. The lack of any authorization constitutes a violation of the law. The principle of triple authorization adopts a higher standard in judging the legitimacy of data capture, and the

separate authorization of any party cannot make the data capture behavior legitimate, which strengthens the protection of user information and platform data. However, reference to the principle of triple authorization to determine the legitimacy of data capture behavior should be more cautious, and should not ignore the differences between individual cases for general application.

"Hantao v. Baidu, etc." case put forward by the principle of substantive substitution; with the continuous development of society, some judicial decisions that, although the specific form of behavioral manifestations of different embodiments, but the act being sued usually includes two links, that is, data capture and data use, and therefore the use of "capture and the use of behavioral dichotomy. Therefore, the "dichotomy of capture and use" was adopted as the evaluation criterion, i.e., the data capture and use behaviors were discussed separately, and an overall judgment was made at the end.

In judicial practice, there are different evaluation standards for the legitimacy of data capture and use. In the data capture process, the court's considerations mainly include whether the means of data capture is justified, whether it is authorized by the platform operator, the type of data captured, whether the capture will cause leakage of personal information or affect the security of the data, and whether the capture will increase the operating costs of the platform. For data use, the court focused on whether the use of the data was a substantial substitute for the operator, whether it exceeded the scope of authorization, and whether the use of the data belonged to the copy-and-paste type of use, or whether the data was used for secondary development to create new products.

Judicial decisions, in their overall judgment of the legitimacy of the act, consider that if both the act of capturing and the act of using the data are per se improper, the act complained of can be evaluated negatively. However, if the acts of data capture and use are not both improper, a distinction should be made on a case-by-case basis. In cases where the act of seizure is justified and the act of use is not, judicial decisions usually consider that even if the means of capturing the data is justified, it does not mean that the subsequent use of the data is justified, and that if the act of use is not justified, it should be subject to the regulation of the law against unfair competition. Judicial decisions in the overall judgment of the legitimacy of the act that if the act of capturing and using the data itself is not legitimate, then the act being sued can be evaluated negatively. However, if the acts of data capture and use are not both improper, a distinction should be made on a case-by-case basis. In the case where the act of capture is legitimate and the act of use is not, judicial decisions usually consider that even if the means of capturing the data is legitimate, it does not mean that the subsequent use of the data is legitimate, and that if the act of use is not legitimate, it should be subject to the regulation of the anti-unfair competition law.

For example, in the case of Tencent v. DuoFlash, there were two different views on judicial decisions in cases where the act of capture was improper. One view holds that the act of capture itself is an independent act, and if the act acquires data in an improper manner, the subsequent use of the act can hardly be considered proper, regardless of the specific form. The second view is that the act of capture is not legitimate, and if the act of use does not harm the interests of the operator and is conducive to the interests of consumers and the interests of society, it does not constitute unfair competition. The adjudication standard of the "Dichotomy of Capture and Use" can provide clear guidelines for market players by breaking down and analyzing the stages of data circulation as the basis for judging the legitimacy of the act.

However, as mentioned above, some judicial decisions believe that the legitimacy evaluation results of data capture behavior directly affects the legitimacy of the use of behavior, and denies the legitimacy of the subsequent use of behavior due to the improper capture behavior, which seems to fall back into the path of evaluation of behavior as a whole, and that this kind of determination is only formally differentiating behavioral links, but ultimately returns to the

value measurement of the interests of the operator, the interests of the consumer and the interests of society. In the end, it is still necessary to return to the measurement of the value of the interests of the operator, consumers and society. As mentioned above, the standards and reasoning of judicial decisions on the legitimacy of data capture vary, resulting in case-by-case norms that cannot be universally applied and are not conducive to market players clarifying the boundaries of their behavior.

4.3. Differing perceptions of harm caused by unfair competitive practices of data scraping

4.3.1. Operators' rights and interests

Substitution of the operator's products or services, resulting in a loss of trading opportunities for the operator. In judicial practice, judges often consider whether an operator's products or services have been negatively affected by a data grabber as a factor in determining whether the defendant has acted improperly, and also as an important reason for awarding damages to the defendant. In *Weibo v. Super Star Rice Group*, the judge held that Super Star Rice Group's capture of Weibo's celebrities' microblogging content, likes, comments and other data actually diverted Weibo's potential user traffic, affected Weibo's commercial revenues that it could have gained from advertisements, tickets and other commercial revenues, and affected the market transaction opportunities and business development opportunities for Weibo's use of the data it possessed in the development of data derivatives or other businesses. It also affects Weibo's ability to utilize its data to develop data derivatives or other business opportunities and the corresponding revenue. Some of the versions capture Weibo's data and then display it directly and completely. Based on normal reading habits, the probability that a user will revisit Weibo after having browsed the relevant content using Super Star Rice Ball is very low. As a result, Super Star Rice Ball is a substantial substitute for Sina Weibo's services. In *Ali v. Code Note*, the court held that the Defendant Code Note's display of corporate information on the 1688 website on its own website to replace some of the functions of the 1688 website would lead to a loss of 1688 users, thereby harming the interests of 1688, affecting the operation and security of the system, and increasing the cost of maintenance.

In data-scraping cases, since the act of scraping itself can hardly affect the system functions or product services of the data controller, and the independent use of the act can hardly affect the normal operation of the plaintiff's network services, it is difficult to apply the special Internet regulations. Unless the plaintiff claims that the defendant's crawling behavior is an "improper" means that can affect the operating speed of the system, destroy the system firewall, or set up plug-ins on the plaintiff's system, thus changing the normal functions and services of the user to use the product. In the former case, such as *Weibo v. Super Star Rice Group*, the Court held that the Defendant CloudIntelligence Company's capture of some of the background non-public data that users could not view even if they were logged in could only be realized by destroying or bypassing the technical data protection measures, which would inevitably aggravate the pressure on Weibo's servers and even lead to security or other technical hazards, thus increasing the operating and maintenance costs to be invested by the company in this regard. This will inevitably increase the operating pressure on Weimeng's microblogging servers and even lead to security or other technical hazards, thus increasing the operation and maintenance costs that Weimeng should invest in. In the latter case, as in *Tencent v. Chivalry Group Control*, the judge held that the Respondent's use of group control technology to forcibly change the Claimant's WeChat functionality damaged the WeChat ecosystem, and that the Respondent was obtaining its own benefits in a way that harmed the interests of others. Thus, Article 12(2) of the Anti-Law was applied."

It can be seen that the current application of the Internet special article in the case of data capture is not for the data capture behavior itself, in this type of group control software, data

capture is only a small part of the action, the defendant through technical means can modify the function and interface of the software in question, capture the data therein and realize the purpose of automated operation on the software, and thus belongs to the situation stipulated in Article 12 of the Anti-Law, which impeded the normal operation of the software in question. The normal operation of the software in question, which also undermined the plaintiff's normal business interests and destroyed the business model of the operator's business.[5] "In the case of Tencent v. Luo Mou Software, the court held that the group control product operated on WeChat platform in bulk, relying on the WeChat platform's huge WeChat user resources, competing for WeChat user's attention, thus increasing its own market trading opportunities and competitive advantage in the market, in fact, it was the undue use of the WeChat ecosystem constructed by the Second Plaintiff to achieve market results.

4.3.2. Consumer interest

In the administration of justice, plaintiffs often argue that the defendant's actions are detrimental to consumer interests on the grounds that the defendant's actions violate the data rights of its users and disrupt the consumer's experience. Defendants argue that their actions were authorized by the users to capture their information, and that their innovative business models are beneficial to consumers' interests. In recent years, judges have also begun to consider the impact of the complained against behavior on consumers' rights and interests, but tend to make a negative evaluation of the complained against behavior. For example, in Tencent v. Chivalry Group Control, the judge pointed out that the defendant's group control software provided micro-businesses with functions such as batch number raising, automated one-click liking of the circle of friends, and batch access to friends' and users' information, pointing out that the defendant's product infringed on the WeChat user's data rights, and jeopardized the user's right to self-selection, the right to be informed, the right to privacy, and other consumer rights. It may also cause users to be deceived by false marketing of fake WeChat users manipulated by the group control system, damaging other legitimate rights and interests.

In the case of Weibo v. Six Worlds, the defendant Six Worlds used technical means to obtain and use data related to Jitterbug users' live streaming bounty records and anchors' bounty revenues, which involved sensitive personal consumption data such as users' interests, consumption preferences, and economic capabilities, and were often data that anchors and users were unwilling to disclose to the outside world, thus infringing the personal information rights of users and anchors. These data involve sensitive personal consumption data such as users' interests, consumption preferences and economic capacity, which are often data that anchors and users do not want to disclose to the public, violating the personal information rights of users and anchors. In the case of MileagePlus v. Billion Resumes, the judge affirmed the legitimacy of the action from the perspective of the user, holding that "the defendant's platform did not compel or deceive the user to use the Linked Account function, but only suggested the user to use the function, and there was no obvious impropriety in the manner of behavior. Automated processing of resumes collected from all websites, HR users do not have to log in separately to different platforms, which can obviously improve efficiency and facilitate market players."

4.3.3. Public interest

In recent years, although judges have made different final determinations on data scraping, more and more judges are recognizing the need to consider the impact of scraping on the public interest and innovation. For example, in the case of Baidu v. Flash Push, the judge pointed out that many of the so-called interfering behaviors in the Internet field are technologically innovative, and that although they may harm certain rights and interests of the operator, they may be beneficial to consumer welfare and the interests of society as a whole. The legal interests protected by the anti-law are diversified, and the whole market competition

environment and more interested parties should also be considered. "The judge in the Tencent v. Luo Software case held that the group control software's large-scale automated publishing and liking of friends' circles seriously interfered with the society's judgment of the Internet traffic economy, interfered with investors' judgment of the value of network products and market prospects, and also affected the real choices made by network users, which seriously disrupted the fair and orderly network business environment[6].

5. Exploring the Path of Judicial Regulation of Data Scraping Unfair Competition Behavior

5.1. Clarification of the conditions for the application of Articles 2 and 12

In the determination of unfair competition in data scraping, before the formulation of the Internet provisions, the determination is mostly in line with the conditions of application of Article 2, but with the formulation of the Internet special article, many cases are determined in accordance with Article 12, so there is no clear standard on whether the data scraping behavior still meets the conditions of application of Article 2, so clarifying the application of these two articles has a crucial role to play in the determination of unfair competition in data scraping. legitimate competition has a crucial role.

From the aforementioned cases, we can see that invoking the principle provisions of Article 2 to determine the unfair competition of data capture will lead to a more general reasoning in the judgment, and it is difficult for the principle provisions to guide the specific behavior, and it is difficult to produce a direct regulatory effect on the behavior, which makes it difficult for the market players to have a reasonable expectation of their own behavior[7]. Therefore, in the process of judicial determination of the above two conditions for the application of specific provisions, in order to make in the process of data capture determination of the applicable legal basis for convergence.

As can be seen from the cases of unfair competition behavior of data crawling in recent years, the application of the relevant provisions of the fourth paragraph of the Internet Article has become a trend, and some scholars have called for the introduction of judicial interpretation on the touting clause, so it is evident that the application of Article 12 has already become a consensus in the theoretical circles and judicial practice[8]. In judicial practice, in order to give full play to the Internet provisions of the regulatory role of new Internet competition behavior, the touting clause has adopted a pragmatic interpretation.

The German jurist Hefei once said that the most basic condition for the realization of justice is the prohibition of arbitrariness. When a judge invokes the General Clauses to determine data-scraping behavior, he or she must explain why the General Clauses should be applied, how the General Clauses should be applied, and what factors should be taken into account when applying the General Clauses, and must provide detailed arguments and reasoning on the above issues. When the judge uses the standard of "honesty and credit" to determine whether the data capture behavior violates the so-called business ethics in the data market, the meaning of "honesty" should be discussed within the framework of the actual Internet data trade, and should not be confused with the traditional general ethics. It should not be confused with traditional ethics in general[7]. If judges cite industry rules as the standard for evaluating business ethics, they should review the industry standards to ensure that they do not conflict with laws and regulations. In conducting the review process, emphasis should be placed on whether the industry rules are consistent with the purpose of the laws and regulations, and inappropriate industry rules or practices should be strictly prevented from being invoked to disrupt the order of market competition.

5.2. Clarify the criteria for judging the legitimacy of data scraping behavior

Data scraping is an important means for enterprises to obtain user information, and with the continuous development of Internet technology, data scraping behavior has also derived from a variety of forms, and as to whether it is unfair, the “Anti-Unfair Competition Law (Draft Revision for Solicitation of Opinions)” puts forward the corresponding judgment criteria (see Figure 1)

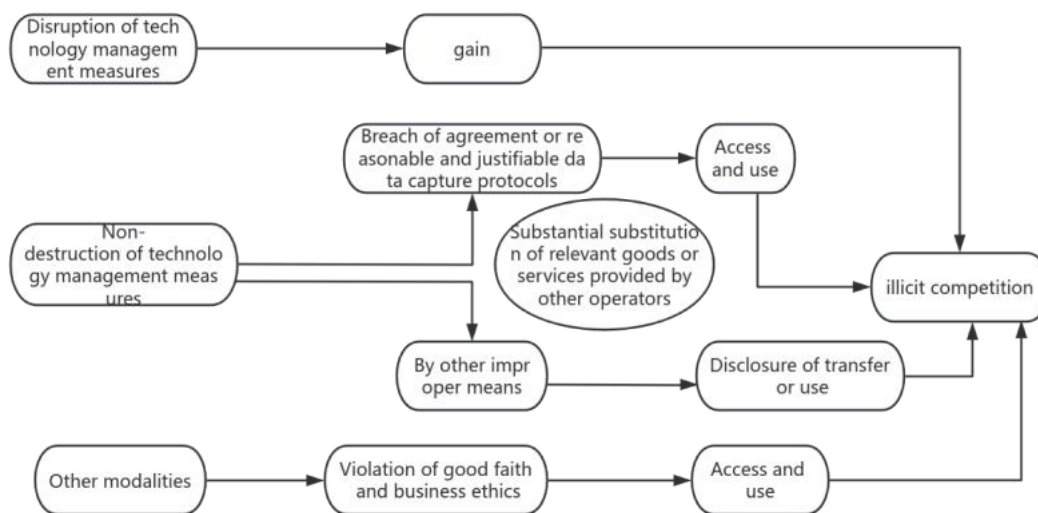


Figure 1: Criteria for determining the impropriety of data crawling

However, it still needs to be strengthened and improved in order to achieve effective regulation of improper data-scraping behavior. Article 18(4) of the Exposure Draft is a touting provision, which is not fundamentally different from the general provision and paragraph 2(4) of the Internet-specific provision, except for the different scope of application, and does not add special considerations for data-scraping behaviors. Therefore, the provision cannot generally stipulate that violation of honesty and business ethics is an unfair competitive practice, but rather, it should make corresponding provisions in a targeted manner according to the characteristics of data-scraping behavior, such as whether it is authorized, whether it is beneficial to society or consumers, and whether there is a competitive relationship, and so on[9]. In addition, the second and third subparagraphs of the Article both refer to substantial substitution as a prerequisite for the impropriety of data-scraping behavior, which will lead to the possibility that a scraper's scraping behavior will be found not to be impropriety when the scraper, after obtaining the data, transforms and uses the data without constituting a substantial substitution with the data of the person to be scrapped. For example, the court of first instance in the case of Microcast Vision v. Six Worlds held that the behavior of Six Worlds in capturing the reward and revenue data of the Jitterbug platform and providing it to the users of its website in the form of payment constituted unfair competition[8]. However, if according to this provision, the Zucchini product of Six Worlds was very different from the service provided by the Jitterbug platform, and the captured data had been transformed and used, it could not constitute a substantial substitution, and the capturing behavior could not constitute a substantial substitution. use, it does not constitute a substantial substitution, and its behavior is likely to be found to be legitimate competitive behavior. Therefore, on this basis, the concept of substantial substitution should be clarified and its scope should be limited, the standard of measurement for judging the legitimacy of data capture behavior should be implemented, the

scope of the fight against improper data capture behavior should be expanded, and the law should be enforced and complied with in the field of data capture.

5.3. Introduction of the principle of balance of interests and proportionality

The main reason for the difficulty and complexity of determining the legitimacy of data-scraping behavior is the complexity of the interests of the parties behind it. From a positive point of view, data capture is conducive to technological innovation, promoting data circulation and information sharing, enabling consumers to obtain better quality services, and enhancing user experience; from a negative point of view, there are some disadvantages of data capture, such as the risk of leakage of consumers' personal information, affecting data security, and the loss of platform operator's interests, affecting the competitive order and development of the industry. Therefore, it is important to establish reasonable rules for determining the legitimacy of data capture and to clarify the boundaries of data protection, so as to prevent data parasitism in the name of "data circulation" and to be wary of data blocking in the guise of "data protection". Article 2 of the Anti-Unfair Competition Law implies the protection of multiple interests, and Article 3(2) and (3) of the aforementioned Judicial Interpretation puts forward comprehensive considerations and industry norms for reference by way of non-exhaustive enumeration. This article believes that for the consideration of multiple interests, specific to individual cases can be integrated with the interests of operators, consumer rights and interests, and public interests, and can be judged by the principle of proportionality as an analytical framework.

Conflicts may exist between the interests involved behind data capture, and as the law does not provide for a hierarchy of values between the interests of the operator, the consumer and the public interest, which cannot be directly judged by a simple upward or downward hierarchy, it is necessary to introduce the principle of proportionality as an analytical framework for weighing the interests between the three. The principle of proportionality in the narrow sense requires that an appropriate proportional relationship be maintained between the means of causing nuisance and the purpose pursued, and this analytical framework not only has some economic analysis, but is not confined to stereotypical quantitative analysis, and is more in line with the actual needs of the application of the law[4]. In *Hantao v. Baidu*, the court of second instance also applied the principle of proportionality, holding that the defendant's act of capturing and displaying the review information in Baidu Maps provided convenience to the consumers, but the social benefits generated were not balanced with the damages inflicted on the plaintiffs, and exceeded the necessary limit. In summary, the interests involved in the data capture behavior are diversified and complex, and there is no order of stages between the interests of multiple parties[10]. The judicial determination of the legitimacy of the behavior needs to be made through the principle of proportionality and the theory of balance of interests, and the interests of the platform operator, the interests of consumers, and the interests of the public interest of society involved in the data capture behavior need to be considered, so as to achieve a balance of the interests of multiple parties, and achieve the balance of data protection and data flow through the dynamic judicial system. The dynamic balance between data protection and data circulation is realized through dynamic justice[11].

However, it still needs to be strengthened and improved in order to achieve effective regulation of improper data-scraping behavior. Article 18(4) of the Exposure Draft is a touting provision, which is not fundamentally different from the general provision and paragraph 2(4) of the Internet-specific provision, except for the different scope of application, and does not add special considerations for data-scraping behaviors. Therefore, the provision cannot generally stipulate that violation of honesty and business ethics is an unfair competitive practice, but rather, it should make corresponding provisions in a targeted manner according to the characteristics of data-scraping behavior, such as whether it is authorized, whether it is beneficial to society or consumers, and whether there is a competitive relationship, and so on.

In addition, the second and third subparagraphs of the Article both refer to substantial substitution as a prerequisite for the impropriety of data-scraping behavior, which will lead to the possibility that a scraper's scraping behavior will be found not to be impropriety when the scraper, after obtaining the data, transforms and uses the data without constituting a substantial substitution with the data of the person to be scrapped. For example, the court of first instance in the case of *Microcast Vision v. Six Worlds* held that the behavior of Six Worlds in capturing the reward and revenue data of the Jitterbug platform and providing it to the users of its website in the form of payment constituted unfair competition. However, if according to this provision, the Zucchini product of Six Worlds was very different from the service provided by the Jitterbug platform, and the captured data had been transformed and used, it could not constitute a substantial substitution, and the capturing behavior could not constitute a substantial substitution[12]. use, it does not constitute a substantial substitution, and its behavior is likely to be found to be legitimate competitive behavior. Therefore, on this basis, the concept of substantial substitution should be clarified and its scope should be limited, the standard of measurement for judging the legitimacy of data capture behavior should be implemented, the scope of the fight against improper data capture behavior should be expanded[13], and the law should be enforced and complied with in the field of data capture.

6. Conclusion

As an "oil" resource that plays a central role in the digital economy, data contains enormous commercial value and is of great importance to platform operators. In the absence of specialized laws and regulations to fully protect the rights and interests of platform data, prior relief through the general provisions of the Anti-Unfair Competition Law will play a major role in dealing with disputes over unfair competition in data scraping. As an emerging area of the application of the anti-unfair competition law, the path of determining the legitimacy of data scraping behavior is a major difficulty in judicial practice[14]. This requires the judicial authorities to play a dynamic role in seeking a balance between data protection and data flow, taking into account the interests of platform operators, consumers and the public, providing judicial guidance on the protection of platform data pursued by platform operators, and providing judicial support for the legislative work on the protection of data rights and interests.

Acknowledgements

This work was supported in part by a grant from "Postgraduate Research Innovation Fund Project of Anhui University of Finance and Economics(ACYC2022304)".

References

- [1] S.M. Deng, Y.L. Hou: The Realistic Dilemma of Enterprise Data Competition Law Protection and Its Way Out, *Technology and Law* (in Chinese and English)(2021) No. 05, p.1-10.
- [2] B. Chen, X.R. Ma: Challenges and Improvements in the Judicial System of New Unfair Competition Behavior on the Internet, *People's Justice*(2022) No. 02, p.18-22.
- [3] D.F. Lu,Y.Q. Zhang: From the Protection of Rights and Interests to the Measurement of Interests: Optimization of the Path of the Determination of the Justification of Data Scraping Behavior, *Journal of Henan University of Finance and Economics and Law*(2022) No. 06, p.59-70.
- [4] J. Sun,T. Feng: Review of Judicial Decisions on Unfair Competition Disputes over Data Scraping in the Digital Era, *Law Application*(2022) No. 6, p. 112-120.
- [5] X. Zhang, S.D. Li: The Path of Protection of Platform Data: Dilemmas, Controversies, and Choice: Dilemma, Controversy and Analysis Based on the Principle of Balance of Interest, *Electronic Intellectual Property Rights (EIPR)*(2021) No. 11, p. 79.

- [6] C. Zhong, Z.Y Wang: Judicial Practice and Reflection on the Dispute of Unfair Competition in Data Reflection, Journal of Beijing University of Aeronautics and Astronautics (Social Science Edition) (2022) No. 01, p. 24.
- [7] Z.S. Xie: The Infringement of Data Scraping under the Background of Personal Information Protection Law Remedies, Politics and Law Forum(2021) No.6, p. 34.
- [8] D.Zhao,C. Shen heng: Examination of Judicial Review Elements of Data Scraping Unfair Competition Disputes and Reflection ,Technology and Law (in Chinese and English)(2023) No. 02, 52-59.
- [9] D. Yang and Z.S. Li: On the Reconstruction of the Anti-Unfair Competition Law: Taking Data Regulation System as the Core, Law Application(2022), No. 11, p. 22.
- [10]Z.S. Xie: Data Scraping Infringement in the Context of the Personal Information Protection Law Remedies, Politics and Law Forum (2021), No. 6, p. 34.
- [11]N. Ni, T. Y. Wu: Application of Web Crawler Behavior to the Anti-Unfair Competition Law General Provisions . China Price Regulation and Antimonopoly(2023), No. 03, P.51-57.
- [12]B. Chen: Competition law attributes and regulatory implications of big data. Law, (2018), No. 08.
- [13] C.Z. Cai: Competition Law Regulating Data Scraping Behavior, Comparative Law Studies (2021), No. 04, p. 176.
- [14]Z.Y. Li: Reflection and Amendment on the Regulation of Data Scraping Behavior from the Perspective of the Anti-Unfair Competition Law, Jinan Journal (Philosophy and Social Science Edition)(2021), No. 06, p. 69.