An Empirical Study on the Judicial Regulation of Data Crawling Unfair Competition

Jinjin Gu*
School of law, Anhui University of Finance and Economics, Bengbu, 233000, China
* Corresponding author: 2221074420@qq.com.

Abstract: In recent years, data crawling, as a new type of unfair competition on the Internet, has become a hot spot of concern in the theoretical and practical circles. This paper composes cases related to digital capture unfair behavior, and through the empirical analysis of typical cases, it is found that such cases present realistic problems in judicial practice such as blurred data ownership boundary, alienation of criteria for determining competitive relationship, non-uniform adjudication basis, and difficulty in determining the amount of damages; in this regard, this paper tries to put forward effective references from specifying the path of data regulation, clarifying the structure for determining competitive relationship, unifying the adjudication of unfair behavior, and clarifying the reference factors for damages. In this regard, this paper tries to put forward effective references from the aspects of specifying the path of data regulation, clarifying the structure of the determination of competition relationship, unifying the basis of improper conduct adjudication, and clarifying the reference factors of damages.

Keywords: Data scraping behavior, Unfair competition, Judicial regulation.

1. Theoretical Foundation: Analysis of The Connotation of Data Scraping Unfair Competition

Data crawling unfair competition goes hand in hand with legitimate data scraping. With the development of digital industry, security risks arising from data come along with it, and the competition among operators about data becomes more intense. However, in judicial practice, there is no clear concept of data crawling unfair competition, and the definition of data crawling unfair competition in this paper refers to the act of using data crawling to obtain data of data owners and then using the data to disrupt the market competition order. Essentially, it is to obtain others' data through data crawling to increase one's competitive advantage, or to weaken others' competitive advantage through data crawling.

To clarify the connotation of data crawling unfair competition, it should first be clear that the data crawled must be commercially valuable data, with commercial value of the main factors to be considered have data operators in the generation of data invested costs, is limited to the data provided by the data operators. After moderate processing such as desensitization, filtering, format adjustment, encryption and screening of the collected data, the data is controlled by technical means and provided to specific people only for commercialization, which includes the cost of human and material resources, and such crawling behavior causes the reduction of the competitive advantage of data operators, and such data does not belong to the "works" in the Copyright Law. Once the data information is property, the competitive advantage gained by the data operator after paying the cost, which belongs to the operator's capital, and the capital can bring profits to the data operator, once the data is captured and the captured data is used to provide products or services instead of competitors, it will fall into the category of unfair competition and infringe the legitimate rights and interests enjoyed by the data operator itself to the data.

Data crawling has a strong concealment. Its concealment is mainly reflected in the implementation of data capture and the implementation of the capture behavior are concealed, due to the data capture behavior from time to time and in the process of continuous change, the implementation of the subject of the act in the dark, resulting in the behavior is more concealed. Secondly, the data capture has a strong technical nature. For the implementation of data capture behavior, must be based on advanced Internet technology. Finally, the damage caused by the consequences are serious. The unfair competition of data capture is different from the unfair competition in the traditional sense. Due to the characteristics of virtualization of the Internet, the unfair competition of data capture transcends the limitation of time and territory, which makes the damage caused by the unfair competition of data capture extend rapidly and cause more serious damage than other general tort damage.

2. Case Cohort: Characterization of Data Crawling Unfair Competition Regulations

This paper relies on the "case group" induction method. The key words "data capture, unfair competition, new type of unfair competition on the Internet" were used in the search of Peking University's Fabao, Judgment Document Network, Law Letter, and the official website of the Supreme People's Court, etc. Since the connotation of data capture unfair competition defined in this paper is narrow, the time condition was relaxed to 2011-2021. After eliminating the cases of contract disputes that do not meet the research theme, only 15 cases of unfair competition were obtained, namely, "Mass Review v. Love Help case" "Mass Review v. Baidu case" "Baidu v. Qihu 360 Case" "GuMi v. YuanMi Case" "FenHan.com v. WuBa Case" "Taobao v. Mei Jing Case" "Alibaba v. Nanjing Code Note Case" "Weibo v. Ant Square Case" "Weibo v. FanYou Case" "Tencent v. Sis Case" "TikTok v. Taobao Case" "Weibo v. Momo Case" "WeChat v. Juketong Group Control Software Case" The cases of "Byte
The competitive relationship is not consistent. Competition is not uniform, and the understanding of relationship in the determination of data capture unfair in the current judicial practice, the positioning of competitive the new law in judicial practice is very limited. The application of unfair competition behavior, judges still habitually follow the illustration that in the judicial determination of data capture 2017 added the Internet exclusive provisions, the judicial provisions of Article 2 of the Anti-Unfair Competition Law, than 10 were determined in accordance with the relevant application, in the determination of data crawling unfair competition, although academics have continued to pay high attention to the hot issues of data-related research and have also developed a series of theories of data rightsization, there is currently no consensus among different academic theories, including: "data property rights theory" "data claims theory " "New Data Property Rights" "Data Intellectual Property Theory" and "Data Intangible Property Rights Theory".

2.2. Identify inconsistencies: tend to be predicated on competitive relationships

From the perspective of the positioning of competitive relationship, in the 15 cases of data capture unfair competition in this paper shows that although it is not directly clear in judicial trials that data developers and operators enjoy property ownership of data, they all agree in essence that they have corresponding property rights and interests in data and should be protected by law. Although academics have continued to pay high attention to the hot issues of data-related research and have also developed a series of theories of data rightsization, there is currently no consensus among different academic theories, including: "data property rights theory" "data claims theory " "New Data Property Rights" "Data Intellectual Property Theory" and "Data Intangible Property Rights Theory".

2.3. Tendency to generalize: reliance on principle clauses

From the perspective of the basis of the final adjudication application, in the determination of data crawling unfair competition, among the 15 cases collated in the text, more than 10 were determined in accordance with the relevant provisions of Article 2 of the Anti-Unfair Competition Law, even after the newly revised Anti-Unfair Competition Law in 2017 added the Internet exclusive provisions, the judicial practice for data crawling unfair competition behavior The determination of the use of Article 12 alone is only 2 , in addition to this, there are a few is the application of Article 2 and Article 12 of the relevant provisions of the joint determination, after 2017, the separate application of Article 2 of the data capture behavior still exists, which fully illustrates that in the judicial determination of data capture unfair competition behavior, judges still habitually follow the trial mode and trial thinking before 2017. The application of the new law in judicial practice is very limited.

2.4. Relative scarcity: the application of pre-litigation ban is relatively blank

From the cases, it can be seen that only one case of pre-litigation injunction has been applied in the process of determining the data crawling behavior. The application of pre-litigation injunction is more common in the determination of new types of unfair competition behavior on the Internet as a whole. Especially since 2017, the application of pre-litigation injunction has become the norm in the process of regulating new types of unfair competition on the Internet. However, the application of pre-litigation injunction in data crawling unfair competition is far from sufficient. The application of pre-litigation injunction can protect the realization of data rights and interests to a certain extent, but from the current situation, the role of pre-litigation injunction in the process of identifying data scraping behavior is far from enough. At present, there is only one case of pre-litigation preservation, accounting for only 6.7% of the number of data capture cases, and only in the case of "TikTok v. Brush Treasure" the court applied a pre-litigation injunction, which is the first case of data capture injunction, so it can be seen that the cases of applying preservation measures in data capture unfair competition are very limited.

2.5. The difference is obvious: it is difficult to determine the amount of compensation awarded by the court

From the perspective of the basis of the final adjudication application, for the final compensation of data crawling unfair competition, although the judicial interpretation issued in 2022 has made it clear that the determination of compensation for new types of network unfair competition can be calculated in accordance with the provisions of Article 17 of this Law, it cannot specifically identify the hidden nature of the occurrence of infringement and the irreversibility of damages, nor can it make up for the infringement caused by other means The damage caused by the infringement. This provision indicates that the punitive measures for the act are relatively simple and the relevant remedy can only be carried out by way of damages. Although in practice, the court will stop the infringement and apologize, the damage already caused cannot be repaired. In the court's decision we can see that in the relevant compensation, the final amount of compensation and the data rights and interests of the requirements of a large gap, and the plaintiff for their own damage to the facts and the amount of the relevant proof, which is undoubtedly also difficult for the plaintiff, which is also one of the reasons for the large gap in compensation.

3. Problem Tracing: The Real Dilemma of Judicial Regulation of Data Scraping Unfair Competition

3.1. Cognitive Chaos: Blurring Data Ownership Boundaries

Taobao v. Mei Jing case, the Hangzhou Intermediate People's Court denied Taobao's property ownership of the data product based on the principle of "Statutory property rights"; although the data product invested a lot of labor and also brought Taobao some economic benefits and substantial exchange value, the court only affirmed Taobao's interest in the data product. Although the data product had invested a lot of labor and brought Taobao certain economic benefits, and had substantial exchange value, the court only affirmed Taobao's interest in the data product. Thus, it can be seen that the court did not directly respond to the issue of data ownership, but recognized the property interest of the network operator in the data collection in its possession from the perspective of property interest protection. However,
different courts have not formed a uniform understanding and practice on the nature and content of property rights and interests. With the continuous emergence of such disputes, this issue must be addressed and resolved. This is because it is directly related to the limit of the protection of data property rights and interests in the field of anti-unfair competition law.

In conclusion, the application of the general provisions of the Anti-Unfair Competition Law to data protection has the following characteristics: First, there is no absolute right over data. In the above-mentioned cases, the courts did not find that the operators enjoy property ownership whether it is public data, user data, or data products. Compared with absolute rights, the anti-unfair competition law protects data less strongly. The court affirmed that the operator has legal rights and interests in the input of data, and at the same time, the law allows others to make use of and imitate the data of data controllers to a certain extent. Secondly, the law against unfair competition is a behavior regulation law, which regulates the behavior of data capture or use. Data cannot be directly protected data in all aspects until it gets the civil code. In judicial practice, the relativity of the competitive relationship in the Internet economy is disappearing, and the competitive relationship is no longer a prerequisite for determining unfair competition.

3.3. Ambiguity in application: diversity of adjudication bases

3.3.1. The application of the legal basis of data crawling unfair competition is not unique

In judicial practice, there is a tendency to diversify the basis for judging the legitimacy of data crawling behavior. Before 2017, there was no special legislation on new types of improper acts on the Internet, then we could only base on the general provisions for determination. However, after the addition of special provisions in 2017, there are cases in judicial practice that cite the relevant provisions of Article 12, but their proportion is small. The tendency to diversify the applicable provisions has put judges in a confusing situation in their reasoning, either avoiding or being vague about the application of the law. For example, in the cases of Sina Weibo v. Momo and Weibo v. Fanyou, the defendants argued that the applicable law was wrong, but the court did not explain the application of the law. The court did not interpret the law when applying it. In a small number of cases, both Article 2 and Article 12 were cited, but the reasoning was mainly based on the general provisions, so we do not know the starting point of Article 12.

3.3.2. Lack of clear basis for adjudication of data crawling unfair competition

China’s legislation on unfair competition has set up a bottom clause, and the same is true for the legislation on new types of unfair competition on the Internet. In the whole process of adjudication of new Internet unfair competition cases, not only the application of the tuck-in clause shall be made with the help of the principle provisions of Article 2, but also the identification of new unfair competition behaviors clearly listed in Article 12 cannot be separated from the invocation of the provisions of Article 2, i.e. the general provisions become the primary path to solve the dispute of unfair competition of data capture. The application of the general provisions of Article 2 has accumulated rich experience in long-term legal practice, the most representative one being the ruling of the Supreme People’s Court in the "Kelp quota case" in 2009. The ruling determined the three elements of the application of the general provisions, which played a pivotal role in future judicial practice. The application of the general provisions can be divided into two aspects: the formal elements and the substantive elements. The formal element refers to the behavior involved in the case belongs to the "Anti-unfair competition law” within the scope of regulation of a certain behavior, but the law does not specify the behavior, or the behavior cannot be included in the law of the specific type of behavior. In this case, the general provisions assume the function of underwriting the application of the improper behavior. Its substantive elements involve the analysis and judgment of the legality of the act, which is also the core issue of anti-unfair competition law.
In terms of formal elements, the principle of "avoiding evasion to general terms" should be adhered to. In terms of substantive elements, most of the cases basically follow the judgment standard of "operator damage + violation of the principle of good faith and recognized business ethics" of improper behavior, which is also the view of the three elements proposed by the Supreme People's Court in the "seaweed quota case". Among them, business ethics, as a more abstract concept in the second general clause, refers to the moral ethics that market operators should follow in the process of business activities. From the existing judicial precedents, the judgment of business ethics in practice is entirely based on the subjective understanding of judges, and the judges' understanding of business ethics tends to be rigid. And due to the lack of unified standards and understanding of abstract business ethics in judicial practice, we certainly cannot use the traditional evaluation of business ethics in the real economy to judge the unfair competition behavior caused by data capture. The understanding of business ethics in the field of data should be more open and inclusive, taking into account multiple factors. For example, the final judgment in the case of Popular Dianping v. Baidu states that the determination of whether the conduct in question violates business ethics should not only take into account the characteristics of information sharing and interconnection in the Internet industry, but also weigh the interests of the parties who obtained the information, users of the information and third parties.

At the same time, the "Anti-Unfair Competition Law" Article 12 of the Internet-specific provisions in the legislation also has too general a problem, so judges in judicial practice, in the identification of such unfair competition, judges often do not cite the bottom clause of Article 12 as the relevant basis, the bottom clause provides for other acts to prevent and damage the order of market competition, but because of the market competition is inherently harmful. However, since market competition is naturally harmful, it is not possible to find that the act is unjustified simply by obstructing or disrupting the order of market competition, and thus it is not possible to find that the data capture act is justified or not simply by the conditions stipulated in the bottom clause.

3.4. Uneven: difficulty in determining the amount of compensation

In the case of Sina Weibo v. Momo, the plaintiff's claim was $10 million, and the court ultimately awarded 2 million in damages; in the case of Weibo v. Super Star Rice Group, the court ultimately awarded 10.23 million in reasonable damages. "In the case of Sina Weibo v. Byte Jump, the court awarded Weibo 20 million yuan and reasonable expenses of 1.157 million yuan. It can be seen that the court has a large discretion in determining the amount of compensation, and the amount of compensation varies from court to court. The main reason is that there is no basis for the standard of compensation for data crawling unfair competition. As the data capture unfair competition is based on the network platform, it is often manifested as the diversion of the users of the network operator and the replacement of the market advantage, with strong uncertainty and concealment. The causal relationship between the conduct and the damage is very complex, and the boundary of the infringement loss is very vague, which directly leads to the unclear scope of compensation and the lack of basis for the measurement of anticipatory benefit loss. Under such circumstances, it is difficult for the court to rely on a single general principle of statutory compensation to determine the specific amount of compensation, and different courts consider different factors, resulting in a lack of predictability and accuracy of the decision.

4. Logical Correction: Exploring the Path of Judicial Regulation of Data Crawling Unfair Competition

4.1. Rational Response: The Path to Concrete Data Regulation

4.1.1. Do not create a right of possession for data.

The purpose of regulating competition is to maintain the order of competition, not to protect specific rights. Beyond the existing legal protection system, it is not necessary to create a new exclusive right to data, and it is also contrary to the legislative purpose and objective of the Anti-Unfair Competition Law. In the absence of sufficient knowledge and consensus on the properties and development prospects of data, granting some new legal rights to data, especially exclusive rights, may lead to the "Oolong effect" of rights.

4.1.2. The path of regulation is specified from data types.

The classification and grading of data is the key to establishing data rules and clarifying the boundary of data protection, and is the prerequisite for developing a differentiated protection system. In data capture cases, the data types involved can be roughly divided into four categories: information data of users, data released by users, data collected by network operators, and derived data information. These include data property rights and interests and data personality rights. In the civil rights system, personality rights are higher than property rights in application, and citizens' personality rights are also higher than property rights in constitutional terms. Therefore, the author believes that personality rights and interests in data rights and interests should take precedence over property rights and interests. Specifically, user data as primary data and underlying data should take precedence over derived data to obtain protection. Specifically in the data capture dispute, the network operator in acquiring the user's native data bears the obligation to protect the user's personal information, and the user's right to personal information and the right to independent choice enjoyed on the native data are limited to obtain protection, and in the process of authorization to third parties, the user's authorization should also be obtained as a priority. Before the application of the anti-unfair competition law, if there is a situation where specific rights such as personal information and copyright are infringed, the specific legal provisions shall be preferred for regulation. Only when there are obvious difficulties in direct remedy, such as the situation where the actual damage is unknown and difficult to prove, then the role of dual regulation of administrative and judicial anti-unfair competition law is played to solve the difficulties in civil litigation.

4.2. Step by step refinement: clarify the structure of competition determination

4.2.1. Competitive relationship in a broad sense of understanding

Data crawling is carried out by the Internet, and the process of data crawling cannot be presented in a visual carrier; therefore, it is difficult to visually determine whether the data
crawling is justified. Therefore, it is necessary to adopt a broad competitive relationship in identifying unfair competition in data crawling cases. A broad competitive relationship is a necessary requirement for the regulation and development of the digital economy. In the digital economy, once the data operator's data is captured by others, regardless of whether the other party belongs to the same industry operator, it may cause the loss of the data operator's competitive advantage and may infringe upon the data operator's right to fair competition. In this case, it is necessary to break through the understanding of competitive relationship in the traditional economy in order to better maintain the order of market competition in the digital economy.

4.2.2. Clarify the criteria of broad competition relationship
Excluding the previous judicial practice to the narrow "scope of business" as the standard for the identification of competitive relations, the current stage of identification, the need to consider the following aspects of the competitive relationship: one is the actual business conduct of the operator as the standard, "actual business conduct" is refers to the data operator actually implemented business behavior, rather than the scope of business license contained in the business. The second is the competitive interest, which refers to whether or not the act increases its competitive advantage by harming the data rights of others. The third refers to the consumer population for which the product is intended, i.e. whether the user group that the operator ultimately competes for is the same user group. For these three factors that need to be considered, there will be overlap between the determination criteria, and only if the judgment is made in a comprehensive manner will there be a stronger persuasive force in the judgment.

4.3. Unification of rules: unified data crawling unfair competition adjudication basis
4.3.1. Clarify the conditions of application of Article 2 and Article 12
In the determination of data crawling unfair competition, before the formulation of the Internet article, the determination is mostly in line with the conditions of application of Article 2, but with the formulation of the Internet-specific article, many cases are determined in accordance with Article 12, then there is no clear standard on whether the data crawling behavior still meets the conditions of application of Article 2, so clarify the circumstances of application of these two articles, for data crawling unfair. Therefore, it is important to clarify the applicable circumstances of these two articles for data crawling improper competition. From the aforementioned cases, we can see that the invocation of the principle of Article 2 of the data capture unfair competition will lead to a more general judgment, and the principle of the provisions of the specific behavior is difficult to produce guidance, it is difficult to produce direct regulation of behavior, making it difficult for market participants to have a reasonable expectation of their behavior. Therefore, in the judicial determination process, only the application of the above two conditions to make specific provisions, in order to make the process of data capture to determine the applicable legal basis tends to be consistent.

4.3.2. Adjustment of the application of Internet-specific underwriting provisions
From the cases of data crawling unfair competition in the past two years, it can be seen that the application of the relevant provisions of the fourth paragraph of the Internet article has become a trend, and there are also scholars calling for the introduction of judicial interpretation on the underwriting clause, which shows that the application of Article 12 has become a consensus in the theoretical community and judicial practice. In judicial practice, a pragmatic interpretation of the touting clause has been adopted in order to give full play to the role of the Internet clause in regulating new Internet competition behaviors.

4.3.3. Adequate justification when applying general provisions
The German jurist Hefe 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 capture, he or she must explain why and 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 these issues. When judges use the standard of "good faith" to determine whether data scraping violates the so-called business ethics in the data market, they must discuss the meaning of "good faith" in the framework of the actual Internet data trade, and must not confuse it with the traditional general ethics. If judges cite industry rules as the standard for evaluating business ethics, they should review industry standards to ensure that they do not contradict laws and regulations. In conducting the review, the focus should be on whether the industry rules are consistent with the purpose of the laws and regulations, and the citation of inappropriate industry rules or practices to disrupt market competition should be strictly prevented.

4.4. Progressive results: clarifying the reference range of damages
As a new type of unfair competition, the illegal data capturing behavior is still a tort in essence, and the resulting damages should be aimed at restoring the original competitive advantage. Article 17(3) of the Anti-Unfair Competition Law actually determines the priority order of damages, with the first order being the actual loss and infringement proceeds, the second order being the license royalty, and the third order being the statutory compensation. Therefore, the damages suffered by the data operator should be the core, taking into account the degree of subjective fault of the data capturing party, the nature of the capturing behavior, the environment in which it is located, the difficulty of data collection by the data operator, the reasonable cost of the plaintiff, the market position of both parties, etc. Also, due to the civil remedy attribute of the Anti-Unfair Competition Law, the court is required to take into full consideration the following factors when reviewing the evidence in court in conjunction with the opinions of the prosecution and the defense. In civil litigation, the court should take into full consideration the negative consequences of failure to adduce evidence and flexibly apply the rules of disclosure, exclusion of evidence and preponderance of evidence. Thus, the evidence submitted by the data operator has a more direct reference value compared to the defendant's claim, and should be used as the basis for calculating
In determining the damages, the principle of prudence and flexibility should be established, the interests of all parties should be integrated, and the amount of damages finally determined should match the actual loss suffered by the network operator to make up for the damage suffered by the network operator as a guideline to effectively stop the unfair competition of data scraping.

5. Conclusion

The flow of data is an inevitable requirement for the free development of the Internet market, and healthy data competition is a guarantee for the flow of data, which determines that the protection model of the Anti-Unfair Competition Law will continue to play an important role in handling data competition disputes. At the same time, due to the multiple attributes of data and the diversity and complexity of data acquisition and use, the Anti-Unfair Competition Law should maintain the necessary modesty when regulating, be careful to uphold the idea of "presumption of innocence" when using state intervention, and balance the interests of data controllers with the value pursuit of promoting data flow.

Acknowledgment

This article was supported by a grant from “Postgraduate Research Innovation Fund Project of Anhui University of Finance and Economics (Project No.: ACYC2022304)”

References