Singapore Journal of Legal Studies
[March 2026 Online] 1–22

AGAINST TRADE SECRETS PROTECTION FOR "SEMI-PUBLIC" DATABASES

YANG CHEN*

This article examines whether trade secrets law should be applied to protect "semi-public" databases with frontend data access that is open to the public. It argues that the incentive-based justification, whether grounded in the traditional rationale or reframed through an investment lens, does not provide a compelling basis for extending protection. The business efficiency rationale may arguably support trade secrets protection only where frontend access is restricted to a clearly defined and limited group of users. By contrast, when access is open to an indefinite public, ambiguous legal standards fail to mitigate inefficiencies and may even intensify the technological arms race between data holders and scrapers. Although deterrence could theoretically yield efficiency benefits, such an effect rests on a flawed assumption and risks suppressing activities that serve the public interest. Moreover, given the powerful and arguably overprotective alternatives already available to database holders, introducing trade secrets protection in this context risks further distorting the balance between private and public interests.

I. INTRODUCTION

There is a rich literature mapping, discussing, and critiquing the legal actions that could potentially be invoked by businesses to protect their databases or data compilations from misappropriation by third parties. The most widely discussed legal avenues for protection are copyright law,[1] a *sui generis* protection regime,[2] anti-intrusion provisions (such as the United States Computer Fraud and Abuse Act ("CFAA")),[3] standalone property protections (such as Article 13 of the People's

* Assistant Professor, City University of Hong Kong School of Law. I would like to thank David Tan, Guobin Cui, Gideon Parchomovsky, Hui Jing, Tianxiang He, Jyh-An Lee and other participants at the Intellectual Property and Technology In the 21st Century: Challenges in the Next Decade Conference held in August 2025 at the National University of Singapore for their valuable comments and suggestions on this paper.

1 See, *eg*, Lothar Determann, "No One Owns Data" (2019) 70(1) Hastings LJ 1 at 18–20.

2 See, *eg*, Peter K Yu, "Data Producer's Right and the Protection of Machine-Generated Data" (2019) 93(4) Tul L Rev 859 at 867–879; Matthias Leistner, "The Existing European IP Rights System and the Data Economy — An Overview with Particular Focus on Data Access and Portability" in Sebastian Lohsse, Reiner Schulze & Dirk Staudenmayer, eds. *Data Access, Consumer Interests and Public Welfare* (Baden-Baden: Nomos, 2020) 209 at 223–232.

3 Computer Fraud and Abuse Act, 18 USC (US) § 1030 (2018) [CFAA 2018 (US)]. See, *eg*, Han-Wei Liu, "Two Decades of Laws and Practice Around Screen Scraping in the Common Law World and Its Open Banking Watershed Moment" (2020) 30(1) Wash Intl LJ 28 [Liu, "Screen Scraping & Open Banking"] at 32–39.

Republic of China Anti-Unfair Competition Law ("AUCL")),[4] and intellectual property ("IP")-like rights.[5]

The application of trade secrets law to protect enterprise databases has only recently been recognised by courts and discussed by scholars.[6] There are no salient objections to using trade secrets law to protect databases that are genuinely private and confidential to the corporation and that limit or deny public access.[7] However, this does not capture some of the common forms of enterprise databases that are not entirely confidential but are increasingly open, particularly in the data economy.

Some business models require that companies allow customers frontend access to individual data points within their databases, while maintaining secrecy at the backend.[8] This type of database has a unique "semi-public" feature because it remains secret at the backend, but the public availability of most of its data points at the frontend undermines this secrecy, distinguishing it from traditional private databases and rendering the application of trade secrets law questionable. This is because one of the critical requirements for applying the law is that the information being protected is secret, either because it is "not publicly available" or "not readily ascertainable" ("NRA").[9]

On its face, the semi-public nature of the databases considered here conflicts with the secrecy requirement. However, there has recently been judicial recognition that some semi-public databases may qualify for trade secrets protection.[10] It is important to note that presently only a few cases recognise such databases as trade secrets. For example, so far, in the United States ("US"), there is only one federal court of appeals decision that directly applies trade secrets law in this context, and Chinese courts have only recently started to realise the potential role of trade secrets protection, particularly in the past two years.[11] Despite this, it is still valuable to examine whether such protection should be extended to semi-public databases. These early cases, albeit few, may nonetheless send a strong signal to data holders,

---

4 Anti-Unfair Competition Law of the People's Republic of China (2025 Revision) [AUCL 2025 (PRC)], Art 13.

5 See, *eg*, Cui Guobin, "The Eligibility Requirements for Legal Protection of Publicly Accessible Datasets" [2022] 4 Intellectual Property 18 [Cui, "Legal Protection of Publicly Accessible Datasets"].

6 See, *eg*, Tanya Aplin *et al*., "The Role of EU Trade Secrets Law in the Data Economy: An Empirical Analysis" (2023) 54(7) International Review of Intellectual Property and Competition Law 826 [Aplin *et al*., "Role of EU Trade Secrets Law"]; Geoffrey Xiao, "Data Misappropriation: A Trade Secret Cause of Action for Data Scraping and a New Paradigm for Database Protection" (2022) 24(2) Colum Sci & Tech L Rev 125 [Xiao, "Data Misappropriation"].

7 See Cui, "Legal Protection of Publicly Accessible Datasets", *supra* note 5; Yang Chen, "The Promise and Perils of Enterprise Data as Trade Secrets" (2026) 29(1) Stan Tech L Rev 1 [Chen, "Enterprise Data as Trade Secrets"] at 20–34.

8 See Part II below.

9 Uniform Law Commission, *Uniform Trade Secrets Act with 1985 Amendments* (1985) [UTSA (US)] at § 1(4)(i); Directive on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use and disclosure, EU Directive 2016/943, [2016] OJ L 147/1 [TSD (EU)]; Provisions of the Supreme People's Court on Several Issues Concerning the Application of Law in the Trial of Civil Cases Involving Infringements upon Trade Secrets (Interpretation No 7, 2020) (PRC) [2020 SPC Interpretation (PRC)] at Art 3.

10 See Part II below.

11 See Part III below.

encouraging them to resort to trade secrets protection, which in turn may prompt more courts to follow suit in the future.

This article argues that trade secrets law should not protect any semi-public database if it offers the public virtually unlimited frontend access to data. The argument is advanced according to the following analytical structure. First, applying trade secrets law to semi-public databases does not serve the law's purported critical aims—preserving incentives to innovate or invest and enhancing business efficiency—to justify the accompanying protection costs. Second, another major benefit of trade secrets law, its deterrent effect on third-party data scraping, may be counterbalanced by its potential chilling effect on publicly beneficial data scraping. Third, introducing trade secrets protection in this context risks further distorting the balance between private and public interests, given the already powerful and arguably overprotective alternatives available to database holders.

This article adopts a comparative law approach, focusing on examples from the US and China.[12] These two jurisdictions are selected for comparison because they are at the forefront of regulating the digital economy, and their legal positions on related issues exert significant influence globally. Furthermore, the judicial and scholarly discourse on the protection of semi-public databases is comparatively more developed in these two jurisdictions than elsewhere.[13] Part II of this article elaborates on the nature of semi-public databases and documents the current judicial recognition of these as trade secrets. Part III addresses the tricky question of whether semi-public databases can satisfy the secrecy requirements, notably whether the information is NRA. Part IV presents the normative arguments against trade secrets protection for semi-public databases. Part V outlines the existing protection alternatives that further diminish the necessity and appropriateness of extending trade secrets protection. Part VI concludes that trade secrets law should not be applied to protect any semi-public databases that allow unlimited frontend access to the public.

II. Semi-Public Databases: A Common Form of Enterprise Data

This Part offers a precise definition of "semi-public" databases and presents representative examples. It emphasises the distinctiveness of these databases and why it matters for potential legal protection in general, and trade secrets protection in particular. The discussion then turns to current judicial practices concerning their protection and the extent to which trade secrets law has been invoked in both the US and China.

Semi-public databases are a type of data compilation that is kept wholly secret at the backend. However, due to the nature of their business models, companies must allow public access at the frontend to some or most of the data points. Normally,

---

[12] For the purposes of this article, "China" refers to the People's Republic of China, excluding the Hong Kong Special Administrative Region, the Macau Special Administrative Region and the Taiwan Region.

[13] The discussions in the European Union ("EU"), another digital empire, mostly centre on the protection of private and confidential databases rather than "semi-public" databases: see Anu Bradford, *Digital Empires: The Global Battle to Regulate Technology* (Oxford: Oxford University Press, 2023); Aplin *et al*., "Role of EU Trade Secrets Law", *supra* note 6 at 839–846.

these companies maintain a website or an application where consumers or users can interact with to request and retrieve the individual data points at the frontend. Technically speaking, the database remains secret at the backend, but the public availability of most of its data points at the frontend undermines this secrecy. For this reason, in this article, such databases are referred to as "semi-public". Figure 1 below depicts a simple version of such a semi-public database.

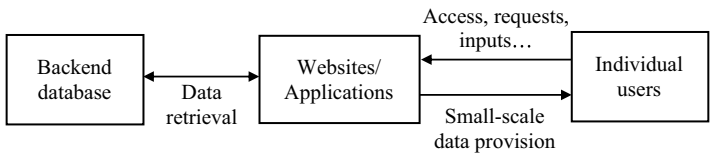

Figure 1: A Simple Version of Semi-Public Databases

Databases of a semi-public nature are prevalent primarily because there are many business models in the digital economy that rely on large-scale databases while providing services and, in turn, data to their users.[14] Most users of airline websites or applications interact or engage with this type of database. Airlines normally maintain a backend database compilation of flight-related data, such as flight times, routes, and customer loyalty points. Their websites or applications allow users to interact through an interface, send search requests, and retrieve specific flight data at the frontend, with or without the need to log into an account.[15] When consumers purchase insurance online, they enter their information on a website. Corresponding data on suitable insurance is then retrieved from confidential backend databases and made accessible to users to view and select.[16] Accordingly, the databases that compile numerous individual insurance quotes also have a semi-public nature. Another notable example is the Chinese company, AMap, which created and maintained a backend confidential database comprising traffic-congestion prediction data for multiple Chinese cities and allowed application users to access and view such data for particular cities at specific times at the frontend.[17]

These semi-public databases are unique by virtue of the fact that companies allow the public to access and retrieve individual data points through the frontend user interface, leaving the entire backend databases vulnerable to data scraping. For instance, relying on current data-scraping technology, third parties may deploy digital bots to send millions of requests to the user interface of these websites or applications, retrieving majority of the available data points and recompiling these into datasets that substantially resemble the backend databases kept secret by businesses.[18] This

---

[14] Typical business models include those represented by social media platforms, online shopping platforms, airline portals, and insurance websites.

[15] See, *eg*, *Air Canada & Aeroplan Inc v Localhost LLC*, No 23-1177-GBW, 2024 WL 1251286 (14 March 2024) (D Del, US) [*Air Canada v Localhost*].

[16] See, *eg*, *Compulife Software Inc v Newman*, 959 F 3d 1288 (2020) (11th Cir, US) [*Compulife* (2020)]; *Compulife Software Inc v Newman*, 111 F 4th 1147 (2024) (11th Cir, US) [*Compulife* (2024)].

[17] *Wan Ltd v G Yuntu Ltd & G Software Ltd* (2024) Beijing Intellectual Property Court, Civil Appeal Judgment 1761 (PRC) [*Wan v G&G*].

[18] See Melany Amarikwa, "Internet Openness at Risk: Generative AI's Impact on Data Scraping" (2024) 30 Rich JL & Tech 533 at 538–540.

tactic has resulted in numerous disputes in the US and China between database holders and third-party scrapers over whether the latter should be liable for appropriating data from the former.[19]

Indeed, the plaintiffs in most of these disputes did not view trade secrets law as a suitable cause of action,[20] likely because their data was publicly accessible at the frontend, which, as noted, appears to fail the secrecy requirement. In 2020, however, the US Court of Appeals for the Eleventh Circuit, the only federal appellate court in the US to have ruled in this way, held in *Compulife Software Inc v Newman* ("*Compulife*") that the large-scale scraping of insurance quotes, accomplished by sending millions of automated requests to the plaintiff's website, constituted the misappropriation of trade secrets in the form of the backend database.[21] In other words, even though individual quotes were accessible to the public, the plaintiff, an insurance company, could invoke the law on trade secrets to protect its backend database by preventing third parties from using scraping technology to massively crawl frontend insurance quotes. In 2024, the same court reaffirmed the application of the law on trade secrets in this scenario.[22]

In China, there has been growing recognition of the role of the law on trade secrets in protecting large-scale data compilations and data products, prompting scholarly debate on the potential role of trade secrets law in database protection.[23] This shift in judicial attitude in China is evidenced by court decisions issued in different years concerning the protection of the same type of database products.[24] These cases concern Business Advisor, a database developed by Taobao, a leading online shopping platform in China. The database comprised various forms of derivative data related to goods and stores on the platform that were valuable for

---

19 For US cases, see, *eg*, *Ryanair DAC v Booking Holdings Inc*, No 20-1191-WCB, 2024 WL 3732498 (17 June 2024) (D Del, US); *Southwest Airlines Co v Kiwi.com Inc et al*., No 3:21-CV-00098-E, 2021 WL 4476799 (30 September 2021) (ND Tex, US); *Air Canada v Localhost*, *supra* note 15; *Maplebear Inc dba Instacart v Cornershop Techs Inc et al*., No 2:20-CV-00240 (16 July 2020) (ED Tex, US); for Chinese cases, see, *eg*, *Shenzhen City X Technology Ltd v Wuhan X Technology Ltd & ors* (2017) Shenzhen City Intermediate People's Court, Civil Trial Judgment 822 (backend data compilation of massive real-time bus data); *Beijing Lianjia Property Management Ltd & anor v Beijing Shenyingcheng Communication Technology Ltd & anor* (2021) Haidian Primary People's Court, Civil Trial Judgment 9148 (PRC) (data compilation of massive housing information accessible through user searches); *Wan v G&G, supra* note 17 (derivative data on daily traffic congestion predictions for multiple cities).

20 In the US, the most common causes of action are breach of contract or CFAA claims: see Xiao, "Data Misappropriation", *supra* note 6 at 129–141. In China, courts normally invoke the general provision of the AUCL: see Cui Guobin, "Towards a Theory of Limited Exclusive Right to Big Data" (2019) 41(5) Chinese Journal of Law 3 [Cui, "Limited Exclusive Right to Big Data"] at 9.

21 *Compulife* (2020), *supra* note 16.

22 In the first appeal, the 11th Circuit reversed the District Court's decision and remanded the case for reconsideration: *Compulife* (2020), *supra* note 16. The District Court then issued a new decision, leading to this second appeal: *Compulife* (2024), *supra* note 16.

23 See, *eg*, Cui, "Legal Protection of Publicly Accessible Datasets", *supra* note 5; Lu Chunxin, "Constructing a Trade Secret-Like Path for Data Protection" [2024] 3 Intellectual Property 88.

24 *Taobao (China) Software Ltd v Anhui Meijing Information Technology Ltd* (2018) Hangzhou Intermediate People's Court, Civil Appeal Judgment 7312 [*Taobao v Anhui*] (decided under general provision of the AUCL). *Cf Zhejiang Taobao Internet Ltd v Taoshu Ltd & ors* (2023) Nanjing Intermediate People's Court, Civil Trial Judgment 4082 [*Zhejiang Taobao v Taoshu*]; *Miao v Yuhang District Market Regulation Agency & ors* (2024) Zhejiang High People's Court, Administrative Appeal Judgment 862 [*Miao v Yuhang MRA*] (decided under trade secrets law).

analysis by store operators. The entire database is confidential and not available to the public. Members of the public can nevertheless become subscribers and gain access to different types of data within the database, depending on the particular subscription plan purchased. This database is not as public as that in the *Compulife* case, since it restricts frontend access to a limited pool of paid users, whereas the Compulife database imposed virtually no limitations on access. However, given the frontend data accessibility, the Business Advisor database, while more limited, is still semi-public.[25] In 2018, the Hangzhou Intermediate People's Court invoked the general provision of the AUCL to sanction the defendant for using technical means to acquire a substantial portion of the database for third-party use.[26] However, in 2024 and 2025, in cases involving strikingly similar fact patterns, the Zhejiang High People's Court and the Nanjing Intermediate People's Court held the database as trade secrets and granted protection under trade secrets law, rather than merely sanctioning the scraping activities under the general provision of the AUCL.[27]

Critical questions thus persist regarding the application of trade secrets law to semi-public databases. First, how can the semi-public nature of these databases be reconciled with the law's secrecy requirement? Second, to what extent should trade secrets law be deployed to protect such databases?

III. The Focus of Trade Secrets Protection: "Not Readily Ascertainable"

This Part primarily addresses the first question outlined above. It identifies the key doctrinal parameter determining whether a semi-public database can be protected under trade secrets law: the requirement that it be NRA. The positions taken in current cases before the courts in the US and China are then considered to determine whether such databases satisfy the NRA condition. It argues that, from a doctrinal standpoint, semi-public databases may meet the requirement if sufficiently robust measures are implemented to restrict frontend access or prevent data scraping, such that the databases are not readily replicable through public scraping activities.

For trade secrets protection, the backend databases should not be readily ascertainable by other parties through independent efforts. In the traditional context, this condition means that if the information can be replicated with relatively little effort or can be easily reverse-engineered, it does not qualify for trade secrets protection.[28] For instance, if certain information can be directly observed from a product sold in an open market or if all information within a compilation can be easily assembled from public sources, the NRA condition may not be satisfied.[29] When most individual data points within a semi-public database are publicly accessible and retrievable,

---

[25] The legal analysis of the applicability of trade secrets law is, however, different from that in *Compulife* scenario. See Part IV below.

[26] *Taobao v Anhui*, *supra* note 24.

[27] *Zhejiang Taobao v Taoshu*, *supra* note 24; *Miao v Yuhang MRA*, *supra* note 24.

[28] James Pooley, *Trade Secrets* (New York: Law Journal Press, 2019) [Pooley, *Trade Secrets*] at § 2.03[2][d].

[29] *Ibid.*

the determinative question becomes whether the backend database can still be considered NRA and replicable through the process of requesting and aggregating the frontend data points. Indeed, what distinguishes a semi-public database from the traditional case is that reverse-engineering could potentially allow direct access to its core components for recompilation; acquiring trade secrets in the traditional context typically requires more indirect means, such as analysing a product or gathering information from public sources. As such, the semi-public nature of these databases makes satisfying the NRA condition more counterintuitive, and a nuanced analysis is required.

The US Eleventh Circuit in the *Compulife* case adopted a lenient and expansive approach to the interpretation of the NRA requirement for semi-public databases. In that case, the plaintiff's website imposed no restrictions on accessing or scraping frontend data: there was no account login required, no terms of service prohibiting scraping, and no technical measures in place to prevent large-scale data extraction.[30] As such, it would appear that the defendant could legitimately rely on data-scraping technology and use a data-scraping bot to send numerous requests to the website during a four-day period to retrieve more than 43 million insurance quotes, which constituted a significant portion of the backend database.[31] However, the court entirely disregarded the absence of access restrictions and found the database to be a protectable trade secret:

> [E]ven if individual quotes that are publicly available lack trade secret status, the whole compilation of them (which would be nearly impossible for a human to obtain through the website without scraping) can still be a trade secret.[32]

According to the court's reasoning, the fact that a human cannot manually extract all the data from the website without the assistance of scraping technologies indicated the difficulty of replicating the database, thereby rendering it NRA. As I argue elsewhere, the *Compulife* NRA standard is not aligned with prior jurisprudence, in which the ease of ascertainment is typically evaluated based on what can be achieved through human capabilities aided by existing technology.[33] This standard imposes an unduly low threshold, effectively allowing any company with a large-scale backend database to claim trade secrets protection, regardless of the extent to which its frontend data is publicly accessible. The NRA standard adopted in prior jurisprudence permits broader public use of platform data through available data scraping technologies. Accordingly, it entails lower risks of intensifying concerns about the already-alarming platform data lockout problem and strikes a more appropriate balance between private interests and the public interest in accessing, collecting, compiling, and analysing data distributed across digital platforms.[34]

---

30 *Compulife* (2020), *supra* note 16 at 1296–1300.

31 *Ibid* at 1299–1300.

32 *Compulife* (2024), *supra* note 16 at 1161.

33 Chen, "Enterprise Data as Trade Secrets", *supra* note 7 at 40–41.

34 *Ibid*; see Niva Elkin-Koren, Maayan Perel (Filmar) & Ohad Somech, "Unlocking Platform Data for Research" (2025) 100(4) Ind LJ 1479 [Elkin-Koren *et al*., "Unlocking Platform Data"] at 1492–1494.

As a point of comparison, in the recent Business Advisor case in China, the court adopted an acceptable understanding of the NRA condition when applied to semi-public databases. In this case, the database holder, Taobao, implemented multi-layered frontend measures to prevent third-party appropriation, specifically data scraping. It provided tiered access to different categories of subscribed users and required each user to sign a non-disclosure agreement. In addition, it deployed a range of technical safeguards to deter scraping, including login authentication, Internet Protocol blocking, and plugin control mechanisms.[35] The Chinese court held that, taken together, these measures pointed to the database being secret, *ie*, not publicly available and readily ascertainable, and subject to reasonable secrecy measures ("RSM").[36] Interestingly, in the same case, when the data scraped by the defendant were publicly accessible without meaningful frontend restrictions—that is, akin to the fact pattern in *Compulife*—the court relied on the general provision of the AUCL rather than trade secrets law.[37]

A key insight is revealed by the comparison between the Chinese case of Business Advisor and the US *Compulife* decisions, along with critiques of the latter's reasoning: properly understood, the NRA condition implies that, for their backend databases to qualify as NRA, companies must impose access restrictions on their frontend data that raise the cost or difficulty of scraping using current technology. Conversely, if no sufficient measures are implemented at the frontend, the backend database may fail to satisfy both the NRA condition and the corresponding requirement for RSM; the two requirements are thus related and, in this context, effectively merged.[38] In essence, whether the NRA condition is satisfied depends on the sufficiency of frontend access restrictions, which requires courts to exercise their discretion based on the specific facts of each case.

However, uncertainty immediately arises regarding the following: what types of frontend restrictions are deemed sufficiently reasonable for the backend database to not be easily replicable by third parties using data-scraping technology? The US Eleventh Circuit completely ignored this question, and the Chinese court likewise gave no reasoning, simply listing the measures taken before reaching a conclusion. Some scholars also halt their analysis at this level, merely pointing out the doctrinal possibility of applying trade secrets law to the database so long as reasonable measures are in place to maintain secrecy; they offer neither clues as to how the reasonableness of these measures can be evaluated, nor do they articulate the threshold for satisfying the NRA condition.[39] This article acknowledges the inherent uncertainty of the NRA condition. However, it argues in the next Part that, as a result of such uncertainty, it is inefficient and inappropriate to apply trade secrets law to a database where frontend access is not restricted to a limited number of users.

---

35 *Zhejiang Taobao v Taoshu*, *supra* note 24.

36 *Ibid*.

37 *Ibid*.

38 As such, throughout this article, the discussion of the NRA condition is understood to inherently encompass the analysis of the RSM requirement. For a detailed introduction of the two requirements, see Pooley, *Trade Secrets*, *supra* note 28 at § 2.03[2][d]–[e].

39 However, one should note that Professor Guobin Cui was the first scholar in China who had identified the doctrinal possibility. See Cui, "Legal Protection of Publicly Accessible Datasets", *supra* note 5 at 17–18.

IV. The Case Against the Application of Trade Secrets Law

The theoretical justifications for trade secrets protection generally fall into two main categories: the incentive rationale and the business efficiency rationale.[40] The incentive rationale suggests that granting legal protection over trade secrets encourages companies to innovate and invest in developing informational assets, while the business efficiency rationale posits that such protection helps firms avoid wasteful expenditures on excessive self-help measures to safeguard their information. However, when trade secrets law is applied to semi-public databases—those whose frontend data is accessible to a broad or indefinite group of users—these justifications become contested, particularly given the significant costs of restricting access to data that could serve the public interest.

This Part critically assesses whether either rationale, or a combination of both, convincingly supports extending trade secrets protection to semi-public databases. It proceeds in three steps. First, it evaluates whether the incentive rationale—both in its traditional form and its investment-focused variant—provides persuasive grounds for such protection. Second, it examines the business efficiency rationale, considering whether trade secrets law truly enhances productive efficiency or instead exacerbates inefficiencies and public costs in this context. Third, it summarises the findings and concludes that trade secrets law should not extend to protect semi-public databases that are effectively open to the public.

A. *Trade Secrets Law's Incentive Rationale?*

In the US, trade secrets law is sometimes justified on the grounds that it incentivises continued innovation by granting information holders a form of exclusivity, mitigating the problem of free riding.[41] However, this traditional understanding of the incentive rationale carries little weight in the context considered here. Unlike Coca-Cola's fizzy drink formula or KFC's fried chicken seasoning recipe which are clearly closely guarded trade secrets, companies collect, produce, and maintain semi-public databases to serve business needs, often not in anticipation of needing protection under the law on trade secrets. Otherwise, when this application of the law is currently limited, as evidenced by the small number of cases recognising such

---

[40] See, *eg*, Robert G Bone, "A New Look at Trade Secret Law: Doctrine in Search of Justification" (1998) 86(2) Cal L Rev 241; Robert G Bone, "The (Still) Shaky Foundations of Trade Secret Law" (2014) 92(8) Tex L Rev 1803; Mark A Lemley, "The Surprising Virtues of Treating Trade Secrets as IP Rights" (2008) 61(2) Stan L Rev 311 [Lemley, "Treating Trade Secrets as IP Rights"]; Yang Chen, "Development of China's Trade Secrets Law in the US' Shadow: Negative Consequences for China and Suggestions" (2022) 17(1) U Pa Asian L Rev 138 [Chen, "Development of China's Trade Secrets Law"] at 169–171.

[41] See, *eg*, Lemley, "Treating Trade Secrets as IP Rights", *supra* note 40 at 329–332; Chen, "Development of China's Trade Secrets Law", *supra* note 40 at 169–171. This justification is akin to what justifies other IP laws or rights on intangible information: see, *eg*, Yang Chen, "Reviving 'Computer-Generated Works': Should Hong Kong Copyright Law Adapt the Rule to Harness AI Opportunities?" J Intell Prop L & Practice (forthcoming, 2025) [Chen, "Reviving 'Computer Generated Works'"] at 8–9; Yang Chen, "Is Chinese Law Well-Prepared for AI Songs?: A Note of Caution on the Over-Expansion of Personality Rights" (2024) 42(2) Cardozo Arts & Ent LJ 261 at 275–277.

protection,[42] one would expect it would result in a decline in the prevalence of these types of databases within the digital economy. Yet, this is evidently not the case. On the contrary, data collection, compilation, and creation remain vibrant, and platform data continues to proliferate.[43] It is unlikely that there is such an undersupply of these databases that would warrant trade secrets protection as an incentive.

There is, however, another interpretation of the incentive rationale—one that shifts the focus from incentivising the creation of informational goods to encouraging investment across the business chain.[44] One argument centres on the role of the law of trade secrets in preserving companies' continuous investments in their overall business operations. The idea is that, by allowing firms to protect key informational assets that support their business, the law indirectly safeguards the substantial resources invested in building, maintaining, and expanding the business. As such, companies' incentives to continue their investment in the business can be preserved and even promoted. This article contends that even this type of incentive may be somewhat limited in the context of semi-public databases. This is because there are several legal—albeit indirect—avenues already available to companies to sue data scrapers who may threaten their business operations, notably through contract law and anti-intrusion provisions. These do not directly protect the underlying databases being appropriated and have their own limits.[45] Nevertheless, companies are actively utilising these legal remedies against data scrapers,[46] with some achieving desirable results.[47] These tools offer companies some assurance that they can maintain their competitiveness and that continued investment in the business is worthwhile. It is doubtful that trade secrets law can provide any meaningful additional incentive. Needless to say, many investment decisions are not driven solely by the availability of legal protection. There are numerous non-legal factors, most notably market and strategic considerations such as lead time (*ie*, wanting to be first to market), market conditions, competitive pressure, and managerial strategy that may sufficiently support profitability and thus sustain investment incentives.[48]

As such, neither the traditional incentive nor the investment-based rationale is a convincing justification for extending trade secrets protection to semi-public databases by further sanctioning data scrapers.

---

42 See Part III above.

43 See Elkin-Koren *et al*., "Unlocking Platform Data", *supra* note 34 at 1487–1489.

44 Reto M Hilty, Jörg Hoffmann & Stefan Scheuerer, "Intellectual Property Justification for Artificial Intelligence" in Jyh-An Lee, Reto M Hilty & Kung-Chung Liu, eds. *Artificial Intelligence and Intellectual Property* (Oxford: Oxford University Press, 2021) 50 at 61–62; Chen, "Reviving 'Computer Generated Works'" *supra* note 41 at 8.

45 See Part V below.

46 Elkin-Koren *et al*., "Unlocking Platform Data", *supra* note 34 at 1495–1496.

47 See *ibid* at 1494–1499; Chen, "Enterprise Data as Trade Secrets", *supra* note 7 at 35–41.

48 See, *eg*, Bronwyn H Hall, Christian Helmers, Mark Rogers & Vania Sena, "The Choice between Formal and Informal Intellectual Property: A Review" (2014) 52(2) J Econ Lit 375.

B. *Trade Secrets Law's Business Efficiency Rationale?*

With the incentive rationale offering no persuasive justification for costs, the task falls to the business efficiency rationale. Trade secrets law can function to reduce production costs and, in turn, enhance productive efficiency. According to this rationale, trade secrets law offers information holders assurance that they do not need to waste their resources and invest in excessive self-help measures to deter every possible information-appropriation behaviour.[49] Absent legal protection, information holders may invest heavily in self-help precautions against all possible ways of obtaining their secrets, even though many of these ways are neither predictable nor expected.[50] A wasteful "arms race" may emerge between those holding trade secrets and those who might appropriate those secrets, with the excessive protective measures of the former being constantly met with new methods to circumvent them. Trade secrets law mitigates the inefficiencies of this arms race by providing assurances that owners need only implement reasonable measures to preserve secrecy. The law then supplements these efforts by protecting against unforeseeable acts, such as espionage or unauthorised appropriation.[51]

In the classical trade secrets scenarios, the business efficiency offered by trade secrets protection is legitimate and substantial. In a traditional context, holders of trade secrets do not need to allow the public access to any components of their trade secrets directly; they refrain from doing so to avoid their information losing its secret status.[52] They only share their trade secrets in highly restrictive situations, for example, with a very small number and select group of internal staff and outside business partners. Under these circumstances, even though the legal standards for RSM and NRA are uncertain and fact-specific, those holding trade secrets generally understand the measures needed to maintain that secrecy under the current law. So long as companies are not required to proactively disclose any components of their trade secrets to the wider public, the current trade secrets law merely requires that they implement measures that prevent unauthorised disclosure. These could include internal technical measures and non-disclosure agreements being concluded with a clearly defined group of information recipients (*eg*, employees or prospective business partners).[53] Trade secrets law eliminates the need for additional measures to guard against appropriation by unforeseeable third parties. As a result, wasteful

---

[49] Lemley, "Treating Trade Secrets as IP rights", *supra* note 40 at 333–334.

[50] *Ibid* at 334.

[51] Douglas Lichtman, "Property Rights on the Frontier: How the Law Responds to Self-Help" (2005) 1(2) J L Econ & Pol'y 215 at 232; Michael Risch, "Why Do We Have Trade Secrets?" (2007) 11(1) Marq Intell Prop L Rev 1 at 42–43.

[52] See, *eg*, *Rockwell Graphic Systems, Inc v DEV Industries*, *Inc*, 925 F 2d 174 (1991) (7th Cir, US).

[53] In fact, both US and Chinese law provide detailed guidance on the typical measures trade secrets holders must take to preserve secrecy, most of which focus on preventing disclosures by parties who receive the information from the holders. See David S Almeling, Darin W Snyder & Michael Sapoznikow, "A Statistical Analysis of Trade Secret Litigation in Federal Courts" (2009) 45(2) Gonz L Rev 291 at 322; David S Almeling, Darin W Snyder, Michael Sapoznikow & Whitney E McCollum, "A Statistical Analysis of Trade Secret Litigation in State Courts" (2010) 46(1) Gonz L Rev 57 at 81; Yang Chen, "Demystifying China's Trade Secrets Law in Action: A Statistical Analysis" (2023) 13(2) Queen Mary J Intell Prop 198 at 237. Chinese law even directly stipulates a list of typical measures in its judicial interpretation: 2020 SPC Interpretation (PRC), *supra* note 9 at Art 6.

investment in excessive self-help measures can be avoided, yielding business efficiency to justify the accompanying costs.

These business efficiency gains are not present in the case of semi-public databases. The key distinction between these databases and traditional trade secrets is that the former has an inherently public nature—the components are proactively disclosed to any users who interact through the interface. The distinction matters because a different type of arms race may arise between database holders and data scrapers, one that trade secrets law is ill-suited to alleviate. As mentioned, satisfying the RSM and NRA requirements requires that companies adopt certain frontend measures to limit user access to data, such that frontend data scraping is neither too easy nor too inexpensive to carry out.[54] *Compulife* is an outlier case in that the plaintiff adopted no such measures, a point raised by some scholars to question whether the application of the trade secrets law was appropriate.[55] In practice, many companies already implement technical measures to restrict and slow down data scraping activities and also to increase the difficulty and cost of any such attempts.[56] Standard measures include Internet Protocol address blocking, a requirement for login credentials, application programming interface ("API") restrictions, rate and data limits, technical checks to verify human behaviours, and detection and cancellation of suspicious accounts.[57] The measures currently available slow down data scraping and make it more challenging. Furthermore, ongoing technological advancement means that more sophisticated tools will undoubtedly become available to database holders.[58] However, evolving technology also benefits data scrapers. In fact, scraping technologies continue to advance and are becoming increasingly accessible to the general public. Notably, artificial intelligence ("AI") appears to be a powerful enabler of data-scraping activities.[59] As such, there is already an arms race at the level of frontend data access/scraping that has escalated with the evolution of technology.

Adding trade secrets law as a possible remedy for database holders does little to alleviate this challenge. Unlike the general understanding of the reasonable measures needed to preserve secrecy for traditional trade secrets, holders here are given no meaningful guidance by the law; it is unclear what frontend restrictions would be reasonably sufficient to deter data scraping and satisfy the NRA requirement for the backend of their semi-public databases. In other words, the uncertainty

---

[54] See Part III above.

[55] See, *eg*, Peter J Toren, "A Dubious Decision: Eleventh Circuit Finds Scraping of Data from a Public Website Can Constitute Theft of Trade Secrets (Part I)", *IPWatchdog*, 2 July 2020 <https://ipwatchdog.com/2020/07/02/dubious-decision-eleventh-circuit-finds-scraping-data-public-website-can-constitute-theft-trade-secrets-part/id=123029/>; Cui, "Legal Protection of Publicly Accessible Datasets", *supra* note 5 at 17–18; Elkin-Koren *et al*., "Unlocking Platform Data", *supra* note 34 at 1511–1512.

[56] Orin S Kerr, "Norms of Computer Trespass" (2016) 116(5) Colum L Rev 1143 at 1166–1176; Melany Amarikwa, "Internet Openness at Risk: Generative AI's Impact on Data Scraping" (2024) 30 Rich JL & Tech 533 at 555.

[57] Chen, "Enterprise Data as Trade Secrets", *supra* note 7 at 43.

[58] Jay Peters, "Reddit will block the Internet Archive", *The Verge*, 12 August 2025 <https://www.theverge.com/news/757538/reddit-internet-archive-wayback-machine-block-limit>.

[59] See, *eg*, ScrapingAPI, "The Rise of AI in Web Scraping: 2024 Stats That Will Surprise You", 4 December 2024 <https://scrapingapi.ai/blog/the-rise-of-ai-in-web-scraping>; Eddy Hage-Youssef & Maxime C Cohen, "Generative AI for Data Scraping" (16 July 2025), available in SSRN, No 5353923.

surrounding the NRA condition and, correspondingly, the RSM requirement, offers database holders no clear assurance as to the point at which they can stop imposing restrictions at the frontend to obtain trade secrets protection of the remainder. The inefficient arms race at the frontend is not alleviated by trade secrets protection. As a result, the business efficiency benefits are, at best, rather limited.

The availability of trade secrets protection may even exacerbate the above inefficiencies. When companies perceive the possibility of relying on this powerful cause of action against data scrapers if they adopt measures deemed reasonable under the law, some may be incentivised to implement even more technical restrictions. Then, the application of trade secrets law may impose more costs by reducing business efficiency.

However, there is one exception: when frontend access to a semi-public database is also limited to a definite number of users rather than the public. This was the case in the Business Advisor case, where only an ascertainable number of users who subscribed and paid a fee may access the frontend data.[60] Meanwhile, these identifiable paying users are required under user agreements not to share or resell any non-public data they access through the frontend,[61] obligations that closely resemble those imposed by standard confidentiality agreements in traditional trade secrets contexts. Therefore, this type of database arguably mirrors the traditional trade secrets scenarios as sharing data with these paid users is akin to disclosing trade secrets internally to companies' employees and externally to business collaborators. Enjoying trade secrets protection requires that secret holders adopt measures that are sufficiently reasonable to limit the further disclosure of shared information to the public, a situation similar to the traditional context discussed above. As such, the reasonable measures required by the law are relatively unambiguous as well. As illustrated in the Business Advisor case, trade secrets law may apply to relieve the plaintiff of the need to adopt additional measures against other forms of appropriation, such as access restrictions to acquire the data or the ongoing monitoring of each paid user's follow-on data use and sharing activities. Accordingly, applying trade secrets law to this type of semi-public database serves the law's theoretical aim of enhancing business efficiency to justify the costs.

Arguably, applying trade secrets law as a sanction against data-scraping behaviours can bring significant benefits as a deterrent. In recent years, the general level of trade secrets protection has significantly increased in both the US and China. In the US, many scholars have questioned the application of trade secrets protection in a way that overly favours plaintiffs by relaxing requirements and abandoning limiting doctrines.[62] In recent years, trade secrets law in China has undergone remarkable developments, notably through the increase in the statutory

---

60 See Part III above.

61 See, *eg*, *Zhejiang Taobao v Taoshu*, *supra* note 24.

62 See, *eg*, Camilla Alexandra Hrdy & Mark A Lemley, "Abandoning Trade Secrets" (2021) 73 Stan L Rev 1; Deepa Varadarajan, "The Trade Secret-Contract Interface" (2018) 103(4) Iowa L Rev 1543; Camilla A Hrdy, "The Value in Secrecy" (2022) 91 Fordham L Rev 557; Deepa Varadarajan, "Trade Secret Fair Use" (2014) 83(3) Fordham L Rev 1401; Joseph P Fishman & Deepa Varadarajan, "Similar Secrets" (2019) 167 U Pa L Rev 1051 [Fishman & Varadarajan, "Similar Secrets"]; David S Levine & Sharon K Sandeen, "Here Come the Trade Secret Trolls" (2015) 71 Wash & Lee L Rev Online 230; Sonia Katyal & Charles Tait Graves, "From Trade Secrecy to Seclusion" (2021) 109(6) Geo LJ 1337.

damage amount caps, the introduction of punitive damages, burden-shifting provisions, a preliminary injunction system, and enhanced enforcement mechanisms.[63] Misappropriating trade secrets may even incur criminal liability,[64] further amplifying the powerful impact of trade secrets law.

Given this powerful impact, the potential application of trade secrets law to semi-public databases may effectively deter many potential data scrapers from scraping frontend data. This may, in turn, leave database holders less anxious about losing their core informational assets, and they may stop adopting excessive self-help measures. Following this line of reasoning, despite its uncertain standards, the applicability of trade secrets law could reduce the arms race through deterrence, enhancing business efficiency.

The perceived deterrent effect rests on one important assumption: data scrapers will be deterred. This does not necessarily hold true in practice. Precisely because of the uncertainty regarding the RSM and NRA requirements in the context of semi-public databases, there is very little guarantee that every database appropriated by scrapers qualifies for protection. Ensuring the law's deterrent effect requires that database holders send strong signals to any potential data scrapers that their databases are likely protected by trade secrets law. Unfortunately, it is exactly the frontend self-help measures that create this signal. The more access restrictions a database imposes, the higher the perceived likelihood of protection, enhancing the deterrence effect with respect to data scraping. Consequently, to maximise deterrence, database holders are incentivised to implement as many frontend restrictions as possible. Any reduction in the arms race would thus be modest at best.

Even if we assume that data scrapers can be deterred from accessing and collecting frontend data, irrespective of whether the semi-public database genuinely qualifies for trade secrets protection, the purported business efficiency gains are very likely to be offset by the accompanying chilling effect on publicly beneficial data collection and use. The reason is straightforward: such deterrence, if it occurs, indiscriminately targets all data scrapers, regardless of the purpose of their activities. After all, the misappropriation of trade secrets, another important prong of trade secrets liability, focuses on whether appropriators *acquire* the information via

---

[63] Notably, in 2024, the Supreme People's Court ordered a damage award (combining compensatory and punitive damage awards) of RMB 630 million in a trade secrets case, highlighting China's determination in sanctioning the misappropriation of trade secrets: *Zhejiang Ji X Holdings Ltd & ors v Wei X Automobile Technology Group Ltd* (2023) Supreme People's Court, Civil IP Appeal Judgment 1590. For details on the development of China's trade secrets law and its specific improvements, see, *eg*, Chen, "Development of China's Trade Secrets Law", *supra* note 40; Peicheng Wu & Charlie Xiao-chuan Weng, "Implications of the China–US Trade Agreement on the Civil Protection of Trade Secrets in China: Is It a Game Changer?" (2021) 28(2) Asia Pacific L Rev 1; Yang Chen, "Under Double Shadows: How U.S.–China Trade Relations and Path Dependence Shape China's IP Preliminary Injunction System" (2025) 33(1) Asia Pacific L Rev 68; Yang Chen, "Rebalancing the Burden of Proof for Trade Secrets Cases in China: A Detailed Scrutiny and Comparative Analysis of Article 32" (2023) 84 U Pitt L Rev 827.

[64] Pooley, *Trade Secrets*, *supra* note 28 at § 13.01; Hebe Chau, Jennifer Che & Sallly Yu, "Stealing Trade Secrets: How the Chinese Court Criminally Sentences Employees that Steal Trade Secrets — 2019 China's Top 50 Representative IP Cases", *China Patent Strategy*, 11 June 2020 <https://chinapatentstrategy.com/how-the-chinese-court-criminally-sentences-employees-that-steal-trade-secrets-2019-chinas-top-50-representative-ip-cases/>.

improper means or disclose it in breach of a duty of confidence.[65] Limited attention has been given to the purposes underlying such acquisition or disclosure.[66] Thus, if the deterrent effect is presumed, it applies to all scrapers.

Not all data scrapers engage in data extraction for competitive or commercial gain. Many utilise platform data for scientific and academic purposes, for instance, to study "mental health, substance (ab)use, diagnosis, and weight and physical activity".[67] Others employ scraped data to research diverse phenomena in the humanities and social sciences.[68] Faced with the legal risks of incurring liability from trade secrets law, these researchers may choose to steer clear of any research using data accessed and scraped from database holders, regardless of whether the research might benefit society.[69] Aided by the mere potential for trade secrets protection, companies are further empowered to exercise exclusive control over their data and are positioned to cherry-pick which parties can access and use their data based on whether the research serves their interests.[70] Other data-driven research that may benefit society but not the company, such as investigations of potential discriminatory practices or other illegitimate activities of the company, can be deterred and suppressed.[71]

Some data scraping activities undertaken for competitive purposes may nonetheless yield public benefits. This occurs when scrapers engage in cumulative innovation—developing a competing but superior product based on the scraped data. In such cases, the resulting innovation may bring significant social benefits, compensating for any harm suffered by the original database holder and generating positive net social gains. This can happen under the "efficient breach" scenario in contract law, but is hard to envisage in trade secrets law.[72] Trade secrets law generally allows the remedies of disgorgement and injunctions, effectively eliminating any surplus gains to scrapers and thereby deterring their cumulative innovation, irrespective of the potential social benefits.[73]

---

[65] It is admitted, however, that whether scraping frontend data amounts to improper means to acquire is another legal uncertainty harming the possibility for trade secrets protection and its related deterrent effect. Because of the space limits, this article cannot address this question. For relevant analysis, see, *eg*, Camilla A Hrdy, "Keeping ChatGPT a Trade Secret While Selling It Too" (2025) 40 BTLJ 75; Cui Guobin, "Legal Nature of Anti-Crawler Measures in Cyberspace" [2023] 6 China Law Review 157; Chen, "Enterprise Data as Trade Secrets", *supra* note 7 at 48–54.

[66] Pooley, *Trade Secrets*, *supra* note 28 at § 6.01. However, when it comes to unauthorised use of trade secrets by someone who obtains them through legitimate channels, the purposes of the use may matter: see Fishman & Varadarajan, "Similar Secrets", *supra* note 62.

[67] Elkin-Koren *et al*., "Unlocking Platform Data", *supra* note 34 at 1485–1487.

[68] *Ibid*.

[69] *Ibid* at 1483.

[70] They have already been aided by contract law and anti-intrusion provisions. The data lockout problem is already salient; *ibid* at 1494–1499.

[71] *Ibid* at 1493.

[72] See Part V below; see Xingguang Zou & Yang Chen, "Unveiling the Mysterious Role of Contractual Disgorgement: A Comparative and Functional Approach" (2025) 27(2) U Pa J Bus L 377 [Zou & Chen, "Unveiling the Mysterious Role of Contractual Disgorgement"] at 283–284.

[73] See, *eg*, Mark A Lemley & Philip J Weiser, "Should Property or Liability Rules Govern Information?" (2007) 85 Tex L Rev 783; Deepa Varadarajan, "Trade Secrecy Injunctions, Disclosure Risks, and eBay's Influence" (2019) 56 Am Bus LJ 879.

The indiscriminate deterrence thus exerts a salient chilling effect on publicly beneficial activities, including competitive, commercial, or non-commercial activities, based on data scraped. These unignorable public costs may well cancel out any efficiency gains to the business.

C. *Summary*

To summarise, Part IV has critically examined whether the costs of extending trade secrets protection to semi-public databases can be justified. It concludes that no compelling justification exists for protecting databases where their frontend access is open to an indefinite number of users.

First, the incentive rationale—whether focused on spurring innovation or ongoing business investment—does not justify extending trade secrets law to semi-public databases. The digital economy already thrives on extensive data collection and maintenance, indicating that trade secrets protection is not a necessary motivator. Companies also utilise alternative legal mechanisms, such as contract law and anti-intrusion statutes, to combat data scraping, and investment decisions are primarily influenced by market and strategic factors rather than legal protections. Thus, expanding trade secrets law would not meaningfully enhance innovation or investment incentives.

Second, the business efficiency rationale is similarly unconvincing. Unlike traditional trade secrets, semi-public databases require ongoing, costly technical measures to deter data scraping due to their public accessibility. This perpetuates an inefficient arms race between database holders and scrapers, further aggravated by legal uncertainty over secrecy standards. Instead of reducing inefficiency, extending trade secrets law may worsen it by encouraging ever more restrictive practices. Even the potential deterrent effect of trade secrets law is problematic. It risks chilling a wide range of publicly beneficial activities, including scientific research and socially valuable innovation, by empowering companies to exercise greater control over data access. Ultimately, the costs to public interest and business efficiency outweigh any purported benefits. Accordingly, trade secrets law should not extend to protect semi-public databases that are effectively open to the public.

V. Alternatives as Comparison

Having argued that trade secrets law is an unsuitable remedy for protecting semi-public databases, this Part turns to the question of whether other legal mechanisms—either existing or proposed—can more appropriately balance the interests of database holders and the public. Part V proceeds in three steps. First, it examines the current indirect alternatives, specifically anti-intrusion provisions and contract law, which regulate data-scraping behaviors rather than protect databases as intangible assets. Second, it considers more direct alternatives, including the recognition of competitive interests under China's AUCL and the prospects for property or IP-like rights in databases. Third, it assesses whether these alternatives genuinely address

the shortcomings of trade secrets law or risk replicating similar types of inefficiencies and imbalances.

A. *Indirect Alternatives: Anti-Intrusion Provisions and Contract Law*

Instead of direct protection of semi-public databases as intangible assets, most current alternatives, notably anti-intrusion provisions and contract law, offer indirect protection by regulating the data-scraping behaviours. Regarding anti-intrusion provisions, the CFAA in the US provides for criminal and civil liability if someone "intentionally accesses a computer without authorization or exceeds authorized access, and thereby obtains […] information from any protected computer".[74] Courts there have had no difficulty in interpreting the term "computer" to include a website or a platform, and there is thus potential liability for web data-scraping under the CFAA.[75] The key area of uncertainty for courts' interpretation is defining "unauthorized access".[76] CFAA claims have undergone an over-expansion trend in their use against data scraping as courts previously adopted a broad definition on "unauthorized access", extending it to include not only access without authorisation, but also authorised access coupled with unauthorised use, or access in violation of terms of service without any technical intrusions.[77] This trend was halted by the Court of Appeals for the Ninth Circuit, one of the most influential circuit courts in the US, in *hiQ Labs, Inc. v. LinkedIn Corp.*, holding that accessing a publicly available website ("the gate is down") does not violate the CFAA unless access is contingent upon authorisation, such as password authentication ("the gate is up").[78]

It appears then that post-*hiQ*, the protection offered by the CFAA to database holders may not be that strong if they open frontend data access to the public. However, the *hiQ* case does not weaken the availability of CFAA claims to data holders as significantly as assumed. As long as the website or platform requires that a user register an account with a username and password, the gate will be deemed to be up; in this case, the CFAA will apply if someone circumvents the password authentication to scrape data.[79]

---

[74] CFAA 2018 (US) § 1030(a)(2)(C); Benjamin L W Sobel, "A New Common Law of Web Scraping" (2021) 25 Lewis & Clark L Rev 147 at 156.

[75] Andrew Sellars, "Twenty Years of Web Scraping and the Computer Fraud and Abuse Act" (2018) 24 BUJ Sci & Tech L 372 [Sellars, "Twenty Years of Web Scraping"] at 388; Liu, "Screen Scraping & Open Banking", *supra* note 3 at 32.

[76] See, *eg*, Sellars, "Twenty Years of Web Scraping", *supra* note 75 at 388–412; Liu, "Screen Scraping & Open Banking", *supra* note 3 at 32–39; Andrew M Parks, "Unfair Collection: Reclaiming Control of Publicly Available Personal Information from Data Scrapers" (2022) 120 Mich L Rev 913 at 913–921; A G Fontana, "Web Scraping: Jurisprudence and Legal Doctrines" (2025) 28(1) J World Intell Prop 197 at 201–203.

[77] W C McRory, "Let the Bots Be Bots: Why the CFAA Must Be Clarified to Prevent the Selective Banning of Data Collection Facilitating Private Social Media Information Monopolization" (2021) 16 Brook J Corp Fin & Com L 279 at 288–290; Dalton Sjong, "Access Denied? Unauthorized Access After hiQ Labs v LinkedIn" (2021) 25(2) Marq Intell Prop & Innov L Rev 133 at 134–135.

[78] *hiQ Labs, Inc v LinkedIn Corp*, 31 F 4th 1180 (2022) (9th Cir, US) at 1197–1198.

[79] *Ibid*; Elkin-Koren *et al*., "Unlocking Platform Data", *supra* note 34 at 1498–1499.

Those who scrape data through automated registered accounts will incur contractual liability, as many websites' terms of service include anti-scraping provisions.[80] These provisions are included within boilerplate contracts so that any users would have to agree on a take-it-or-leave-it basis.[81] Although the doctrine of adhesion contracts might help data scrapers escape contractual liability,[82] this is not often the case because courts have generally found clickwrap (where the user must affirmatively click 'I agree' before accessing the website) and scrollwrap (where the user must scroll through the contract and click 'I agree') contracts enforceable.[83] It is only in limited circumstances, where "the user agrees to the contract merely by using the website" (browsewrap) without the actual or constructive knowledge of the terms, that the contracts might be unenforceable.[84] Accordingly, most users of websites or platforms are subject to valid terms of service and would be liable for contractual breach if they scraped data. This is particularly true when, as mentioned, the availability of CFAA claims prompts data holders to erect gates on their websites—these effectively require all users to access the site through authenticated accounts, to which the terms of service are invariably attached.[85]

Contractual claims may be seen as less robust than a claim under trade secrets law since contract law offers comparatively limited remedies, and thus cannot adequately protect the interests of data holders and sufficiently deter data scraping. There are two reasons this is not the case. First, it is acknowledged that contractual liability may not deter some competitors' data-scraping behaviours because disgorgement is rarely available, and injunctions and punitive damage awards are generally unavailable for contractual claims.[86] Some competitors would still scrap data and have the financial resources to defend against contractual claims.[87] It appears that database holders' private interests may be harmed in this scenario, but, overall, societal interests may be promoted. This is because, under contract law, competitors are normally only required to compensate database holders for losses, without the need to disgorge profits or face injunctions. As a result, under the efficient breach theory,[88] they are incentivised to efficiently breach terms of service when they believe the scraped data can be used for more innovative and socially beneficial purposes. In this scenario, a social surplus may arise from the breach, even after fully compensating database holders for their loss. Thus, what is perceived as the weakness of contractual claims may, in fact, represent a more efficient alternative to the trade secrets approach.

Second, contractual claims, even with relatively limited remedies, are already being weaponised by database holders to deter data scrapers who may put data toward more publicly-regarding uses. Again, researchers, unlike competitors,

---

80 Liu, "Screen Scraping & Open Banking", *supra* note 3 at 43.
81 Elkin-Koren *et al*., "Unlocking Platform Data", *supra* note 34 at 1494.
82 *Ibid*.
83 Xiao, "Data Misappropriation", *supra* note 6 at 132.
84 *Ibid*.
85 Elkin-Koren *et al*., "Unlocking Platform Data", *supra* note 34 at 1498.
86 See Zou & Chen, "Unveiling the Mysterious Role of Contractual Disgorgement" *supra* note 72.
87 Elkin-Koren *et al*., "Unlocking Platform Data", *supra* note 34 at 1498–1499.
88 See Gregory Klass, "Efficient Breach" in Gregory Klass, George Letsas & Prince Saprai, eds. *Philosophical Foundations of Contract Law* (Oxford: Oxford University Press, 2014) 362.

normally lack the financial capacity to fight contractual claims because they do not generally make profits from scraped data. The legal risks of breaching the terms of service can easily deter them from scraping and using data for public beneficial purposes, which is not socially desirable.[89] Indeed, there are already instances in which platforms have threatened litigation based on contractual claims against researchers they disfavour, effectively halting ongoing research projects.[90] These concerns are not hypothetical. When contractual claims have already contributed to data monopolisation and platform lockout, allowing companies to discriminate in granting access to their data, introducing a stronger form of legal protection may exacerbate, rather than correct, the imbalance.

The current CFAA and contractual claims, particularly when combined, are powerful tools for data holders to protect their interests. This raises further questions about the need for trade secrets protection.

B. *Direct Alternatives: Competitive Interests and Property or IP-like Rights*

Unlike the case of the US, where database holders primarily rely on anti-intrusion provisions and contract law, China stands out by providing protection through the AUCL, which essentially recognises data holders as having competitive interests in their large-scale databases.[91] Before the 2025 amendments to the AUCL, this was achieved by applying the general provision of the AUCL (*ie*, Art 2). In terms of this provision, the plaintiffs' large-scale databases were assessed to determine if they constituted protectable competitive business interests and whether the data-scraping activities amounted to acts of unfair competition.[92] Data holders in China have used this provision extensively to sue and protect their interests against data scrapers.[93] Much has been written about the problems with the overuse of this general provision, which is seen as excessively abstract, leaving courts with significant discretion, resulting in inconsistent judicial applications.[94] There nevertheless appears to be a general consensus among courts regarding the key factors that should be considered and balanced: (1) the investment or labour expended by database holders in constructing the databases (*eg*, whether they have a competitive edge and whether their investment was substantial); (2) the protection of consumer interests (*eg*, personal information); and (3) the technical or managerial measures adopted by data holders that have been circumvented.[95]

---

89 Elkin-Koren *et al*., "Unlocking Platform Data", *supra* note 34 at 1499.

90 *Ibid* at 1480–1481.

91 AUCL 2025 (PRC) at Art 13.

92 Cui, "Legal Protection of Publicly Accessible Datasets", *supra* note 5 at 18–20; L Fei (Lanfang), "A Comparative Study on Public Interest Considerations in Data Scraping Disputes" (2024) 20(4) Intl J Law in Context 568 at 574–576.

93 Wei Liu & Liyang Hou, "The Legal Boundary of Data Scraping" (2025) 15(2) Queen Mary J Intell Prop 219 [Liu & Hou, "The Legal Boundary of Data Scraping"] at 222–223. See also New Legal Professor, "Summary Catalogue of Data Law Cases in the Past Five Years", *WeChat*, 5 March 2025 <https://mp.weixin.qq.com/s/ImpIoSrMNHucjzzXMAjziw> [New Legal Professor, "Catalogue of Data Law Cases"].

94 See, *eg*, Cui, "Limited Exclusive Right to Big Data", *supra* note 20 at 9.

95 See, *eg*, Liu & Hou, "The Legal Boundary of Data Scraping", *supra* note 93 at 224.

This triple-interest framework shares some of the problems encountered in the application of trade secrets law. First, although some Chinese courts have recognised the competitive and innovative benefits of data scraping, such as its potential to foster cumulative innovation,[96] they have yet to consistently mandate this critical factor in their balancing analysis, let alone articulate how it should be weighed in the overall framework. Indeed, to date, there appears to be no case in which a defendant has escaped liability based on the innovative benefits of their data scraping.[97] Relying on the general provision, database holders have generally been able to prevail against data scrapers. As one study indicates, out of 36 cases analysed, the plaintiffs achieved a success rate of 80.56% in the court of first instance and 86.7% on appeal.[98] One acknowledged reason for these high percentages is that, in these cases, the defendants did not produce any innovative products but clearly were freeriding on the plaintiffs' databases to replicate or substitute their businesses. Scholars have also noted that some courts merely mention this factor without engaging in any meaningful balancing effort.[99] Given this state of affairs, without a meaningful legal pathway to escape liability when they genuinely create innovative products based on scraped data, competitive data scrapers are likely to refrain from such potentially socially beneficial activities. The result may mirror the effect of applying trade secrets law; given that the Chinese AUCL also provides for disgorgement and injunctive relief as remedies, "efficient breach" attempts would be less likely to happen.

Second, the current emphasis in the AUCL analysis on the technical or managerial measures adopted by data holders raises similar problems to the application of the RSM and NRA requirements in trade secrets law. By signalling to database holders that their frontend access restrictions are crucial for succeeding in a legal claim, the law does not curb the arms race between data holders and scrapers, and may intensify it. Rather than promoting efficiency, this incentive structure may lead holders to adopt increasingly stringent measures, resulting in outcomes that are, at the very least, inefficient.[100]

These problems are not addressed in the 2025 amendments to the AUCL. The amended Art 13 explicitly provides that any circumvention of technical and managerial measures adopted by database holders to acquire and use their legitimate data

---

[96] Jie (Jeanne) Huang, "The Rise of Data Property Rights in China: How Does It Compare with the EU Data Act and What Does It Mean for Digital Trade with China?" (2024) 27(2) J Intl Econ L 462 [Huang, "The Rise of Data Property Rights in China"] at 575–576; see, *eg*, *Beijing Weibo Technology Ltd v Beijing Chuangrui Media Ltd* (2021) Beijing IP Court, Civil Appeal Judgment 1011; *Beijing Weimeng Internet Technology Ltd v Guangzhou Jianyi Telecommunication Technology Ltd* (2022) Guangdong High People's Court, Civil Appeal Judgment 4541; *Shenzhen Tengxun Systems Ltd & ors v Zhejiang Soudao Internet Technology Ltd* (2019) Hangzhou Railway Court, Civil Trial Judgment 1987.

[97] For a list of data scraping cases apparently decided in China and their results, see New Legal Professor, "Catalogue of Data Law Cases", *supra* note 93.

[98] Liu & Hou, "The Legal Boundary of Data Scraping", *supra* note 93 at 223.

[99] Wu Peicheng & Tong Yujie, "Rule of Law Safeguards for the Market-Oriented Governance of Data-Driven Competitive Conduct: An Analysis of Art 13 of the Draft Revised Anti-Unfair Competition Law" [2025] 5 China Market Regulation Research 17 [Wu & Tong, "Governance of Data-Driven Competitive Conduct"] at 20–21.

[100] Cui, "Legal Protection of Publicly Accessible Datasets", *supra* note 5 at 50–52.

may constitute unfair competition.[101] The new law brings a degree of legal clarity by mandating that, rather than relying on other general legal provisions, all future courts apply this article in the case of data-scraping disputes.[102] Unfortunately, the analytical framework likely remains the same and, as such, the problems are unsolved.[103] Accordingly, when the current AUCL already offers a specific cause of action that strongly favours database holders, it is all the more questionable to extend the protection of trade secrets law, which raises similar problems when applied to semi-public databases.

Establishing a standalone system of property or IP rights to offer protection for databases in the future, as currently envisaged by China,[104] may draw on judicial practice under the AUCL. However, if this is the case, the current analytical framework should be replaced to better address existing problems and strike a balance between private and public interests. Notably, the new system could carve out an exception for innovative activities based on scraped data, even when such activities directly compete with the original data holder. Courts determining whether such an exception should apply could be authorised to weigh the innovative benefits against the substitution effects of the scraping conduct.[105] Also, the existence of sufficient technical or managerial measures to restrict data access should not be a relevant factor, as it risks exacerbating the arms race. As Guobin Cui argues, as long as the subject matter protected under the new system is strictly defined by the quantity of data, the substantiality of investment, and the public nature of the data points within the database, there is no need to require frontend restrictions.[106] Such requirements unduly burden the data holders while inconveniencing public access to data.[107] Meanwhile, an explicit exception could be introduced to permit data scraping and use for non-competitive purposes, such as personal study and research. This would help mitigate the potential chilling effect that a new rights system might have on publicly beneficial activities.

In sum, unless all these critical elements are incorporated, introducing a standalone rights system would be as inappropriate and unnecessary as extending trade secrets protection to most holders of semi-public databases.

VI. Conclusion

This article has argued that trade secrets law should not be applied to protect any semi-public databases that allow unlimited frontend access to the public. The incentive-based justification, whether grounded in the traditional rationale or reframed through an investment lens, does not provide a compelling basis for extending trade

---

[101] AUCL 2025 (PRC) at Art 13.

[102] Wu & Tong, "Governance of Data-Driven Competitive Conduct", *supra* note 99 at 19.

[103] *Ibid* at 19–21.

[104] Bingwan Xiong, Jiangqiu Ge & Li Chen, "Unpacking Data: China's 'Bundle of Rights' Approach to the Commercialization of Data" (2023) 13(1) Intl Data Priv L 93; Huang, "The Rise of Data Property Rights in China", *supra* note 96 at 473.

[105] For similar insights, see Liu & Hou, "The Legal Boundary of Data Scraping", *supra* note 93 at 235–236.

[106] Cui, "Legal Protection of Publicly Accesible Datasets", *supra* note 5 at 51–52.

[107] *Ibid*.

secrets protection to semi-public databases. The business efficiency rationale may support such protection only in cases where frontend access is restricted to a clearly defined and limited group of users. In contrast, when access is open to an indefinite group or the public at large, the ambiguous legal standards for protection fail to mitigate inefficiencies and may even intensify the arms race between data holders and scrapers. Although the deterrent effect of trade secrets law could theoretically have efficiency benefits, this effect is predicated on a flawed assumption, that scrapers would be deterred. Even if the assumption were to hold, the resulting deterrence could suppress activities that serve the public interest, nullifying any purported efficiency gains.

In fact, existing alternatives to trade secrets law already provide database holders with robust legal tools to pursue claims against data scrapers, tools that are actively and extensively utilised in practice. The anti-intrusion provisions, exemplified by the US CFAA, although recently narrowed by judicial interpretation, remain a potent remedy, particularly when database holders impose gatekeeping measures such as mandatory account registration. Where scrapers circumvent such authentication requirements, CFAA liability may still attach.

Those who scrape data using registered accounts are typically bound by anti-scraping clauses in the platform's terms of service and are thus almost certainly liable for breach of contract. While the remedies based on contractual claims may be comparatively limited, they are already being weaponised to deter a wide range of scrapers, including those who seek to use data for socially beneficial purposes. It is acknowledged, that the limited nature of contractual remedies may allow competitors to efficiently breach to pursue innovative and competitive uses of scraped data. Given the potential efficiency gains of such conduct, the limited remedies should be embraced rather than criticised.

China stands out for offering additional and distinctive avenues of protection. The AUCL, through either its general provisions or its newly added Article 13, offers strong protection to database holders. However, these provisions often favour private interests at the expense of public benefits and, like trade secrets law, fail to mitigate—and may even worsen—the inefficiencies resulting from an escalating technological arms race between data holders and scrapers.

Accordingly, extending trade secrets protection to cover semi-public databases is both theoretically unwarranted and practically unnecessary. Given the already powerful and arguably overprotective alternatives available to database holders, introducing trade secrets protection risks further distorting the balance between private control and public interest.