Blacklists and Redlists in the Chinese Social Credit System: Diversity, Flexibility, and Comprehensiveness

Severin Engelmann, Mo Chen, Lorenz Dang, Jens Grossklags
Professorship of Cyber Trust
Department of Informatics
Technical University of Munich
Munich, Germany

ABSTRACT

The Chinese Social Credit System (SCS, 社会信用体系) is a novel digital socio-technical credit system. The SCS aims to regulate societal behavior by reputational and material devices. Scholarship on the SCS has offered a variety of legal and theoretical perspectives. However, little is known about its actual implementation. Here, we provide the first comprehensive empirical study of digital blacklists (listing "bad" behavior) and redlists (listing "good" behavior) in the Chinese SCS. Based on a unique data set of reputational blacklists and redlists in 30 Chinese provincial-level administrative divisions (ADs), we show the diversity, flexibility, and comprehensiveness of the SCS listing infrastructure. First, our results demonstrate that the Chinese SCS unfolds in a highly diversified manner: we find differences in accessibility, interface design and credit information across provincial-level SCS blacklists and redlists. Second, SCS listings are flexible. During the COVID-19 outbreak, we observe a swift addition of blacklists and redlists that helps strengthen the compliance with coronavirus-related norms and regulations. Third, the SCS listing infrastructure is comprehensive. Overall, we identify 273 blacklists and 154 redlists across provincial-level ADs. Our blacklist and redlist taxonomy highlights that the SCS listing infrastructure prioritizes law enforcement and industry regulations. We also identify redlists that reward political and moral behavior. Our study substantiates the enormous scale and diversity of the Chinese SCS and puts the debate on its reach and societal impact on firmer ground. Finally, we initiate a discussion on the ethical dimensions of data-driven research on the SCS.

CCS CONCEPTS

• **Social and professional topics** → **Government technology policy**; • **Security and privacy** → **Social aspects of security and privacy**.

KEYWORDS

China's Social Credit Systems; Reputation Systems; Digital Socio-Technical Systems; China.

ACM Reference Format:
Severin Engelmann, Mo Chen, Lorenz Dang, Jens Grossklags. 2021. Blacklists and Redlists in the Chinese Social Credit System: Diversity, Flexibility, and Comprehensiveness. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society (AIES '21), May 19–21, 2021, Virtual Event, USA.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3461702.3462535

1 INTRODUCTION

In 2014, the Chinese government published the *Planning Outline for the Construction of a Social Credit System (2014-2020)* as part of its 12th five-year plan [32]. Following its release, media and research have offered various perspectives on the Chinese Social Credit System (SCS). Some Western media have characterized the SCS as a mass surveillance apparatus, with the purpose of calculating a digital "sincerity score" for each Chinese citizen based on a wide range of personal data [3, 23, 28]. Below a certain point level, citizens would face multiple restrictions, such as exclusion from air travel and high-speed trains. A positive score, on the other hand, would lead to discounts and preferential treatment for a variety of products and services. This "dystopian perspective" sees the unification of an authoritarian regime's policies and artificial intelligence (AI) to enforce social order by means of a sincerity score. Some media outlets have since revised their original viewpoints regarding such comprehensive sincerity scoring [16, 26].

Academic scholarship on the SCS has largely been theory-driven, which has led to the independent development and discussion of different conceptualizations. The SCS has been defined as a novel administrative policy program with the main goal of strengthening compliance of citizen and organizations with laws and regulations [1, 7]. The novelty consists in the public (at least temporary) disclosure of already existing citizen and organizational records on so-called digital blacklists and redlists. Blacklists publicly showcase non-complying individuals and organizations, while redlists, as their normative counterpart, show complying entities. In this perspective, the SCS deploys reputational tools with some similarity to company rankings or background checks on individuals in Western economies.

Other authors have called the SCS a big data empowered system that collects, processes, and evaluates vast amounts of personal data [6]. These data are ultimately aggregated and published as public credit information (PCI) on digital platforms. This line of research argues that PCI creates transparent citizens, not least due to the lack of a sufficient legal framework that protects personal data in China [22]. Some scholars have noted an all-encompassing application of credit to society's political, economic, and social activities. Thereby, the SCS marks the emergence of a so-called reputation state [9, 24]. As a governance tool, the SCS seeks to harness reputational information for purposes that go beyond neoliberal notions of regulating market failure. Still other perspectives frame the SCS as a social

AIES '21, May 19–21, 2021, Virtual Event, USA
© 2021 Copyright held by the owner/author(s). Publication rights licensed to ACM. This is the author's version of the work. It is posted here for your personal use. Not for redistribution. The definitive Version of Record was published in *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society (AIES '21), May 19–21, 2021, Virtual Event, USA*, https://doi.org/10.1145/3461702.3462535.

management program [36]. Drawing on concepts from systems engineering, a social management program considers society to be a complex system that can be optimized using digital technologies.

While these accounts disagree in many important regards, three points of agreement can be identified: first, multiple independent initiatives have been labelled as "SCS" [35]. One SCS is driven by the apps and services of big data companies (e.g., Sesame Credit) that distribute scores to consumers in voluntary promotion programs [5, 18]. Here, "voluntary" denotes consenting to the terms and conditions of the service. Second, *local* governments have tested SCSs that integrate different scoring systems in "prototype cities" (社会信用体系建设示范城市), such as Rongcheng and Suzhou. Participation in these local "credit scoring experiments" is mandatory for residents in these areas. Such policy experiments [14] can serve as models for other local SCS implementations but they are not necessarily a model for national implementation. Third, government-led SCS measures have been realized nationally. There are various types of blacklists (黑名单) and redlists (红名单) run by government agencies at different levels of administrative divisions (ADs) including municipalities and provinces, but also government departments at the national level. These platforms publicly display information to "shame"[1] or "praise" natural and legal persons (e.g., companies) for non-compliance or compliance with a variety of legal and social norms [10, 15, 19, 22, 30]. No entity can opt out from being listed. Depending on the type of list, entities are subjected to different types of reward or punishment over a wide range of areas, a process that has been termed "joint reward and punishment mechanism" (JRP) by the Chinese government [32]. Both natural and legal persons on specific blacklists or redlists will be punished or rewarded under the rules defined in Memoranda of Understandings (MoUs). Different government agencies have jointly signed and started enforcing these MoUs [8].

To summarize, the government-run SCS operates blacklists and redlists throughout the entire country. It enforces regulations with reputational and material means and requires mandatory participation. *This* SCS has regulatory "teeth". However, no research has conducted an empirical analysis of this nationwide SCS blacklist and redlist infrastructure.

This lack of knowledge is troubling, as the SCS will likely shape the behavior of about 1.4 billion Chinese citizens and all companies doing business in China. Further, important international long-term technology policy challenges are dependent on the success of systems such as the SCS, as highlighted by Antony Blinken in his confirmation hearings, when he argued that "*whether techno democracies or techno autocracies are the ones who get to define how tech is used (...) will go a long way toward shaping the next decades*" (2021 U.S. Secretary of State confirmation hearings [11]).

This study investigates the design and technical implementations as well as the number and types of blacklists and redlists across 30 Chinese provincial-level ADs. Our exploratory study shows the diversity of SCS lists in granular detail and outlines the informational consistency between social credit records of the same type of list on different SCS platforms. We find that SCS listings focus on economic activities but also capture reputational rewards for moral and political behavior. Moreover, we show that the SCS listing infrastructure is flexible, as observed in a second round of data collection during the COVID-19 outbreak: when necessary, new types of lists can regulate novel forms of transgression and thereby help accomplish new policy goals.

[1]The authors use quotation marks to communicate a neutral standpoint towards SCS-specific normative concepts (e.g., "positive", "negative", "reward", "sanction/punishment"). For the remainder of the article, quotation marks will be omitted for the sake of reader-friendliness.

2 STUDY PROCEDURE

2.1 Policy-making in China: Provinces implement blacklists and redlists

SCS implementation is largely left to regional rather than central government, a common trait of China's policy-making process that tends to follow a principle of "centralized planning, decentralized implementation" [12, 13]. As a planning polity, central policy-makers outline policy goals in top-level policy documents valid for a specific policy-making cycle. Commonly, a first policy document (called *jianyi*/建议) includes general guidelines for a new cycle of policy-making. A second, more refined, but still broad, policy outline (called *gangyao*/纲要) sets more specific policy goals [14].[2] Importantly, the *implementation* of the policy goals outlined in top-level policy documents is left to provincial, county, and city governments. This also applies to the SCS: provincial-level administrative authorities (i.e., those in charge of provinces, municipalities under the direct administration of central government, and autonomous regions) are, to some extent, free to determine *how* they implement nationwide policy goals for their AD [27, 31].

The SCS's *gangyao* includes vague instructions regarding social credit record applications for broadly defined commercial and social sectors (e.g., [6, 8, 22]). SCS implementation rests on the commitment of provincial-level ADs[3] to realize general instructions laid out in top-level policy documents. As such, understanding the nationwide SCS listing infrastructure requires an empirical assessment of all SCS platforms at the provincial level. As each province is responsible for the implementation of its own SCS blacklist and redlist, we expected to find differences in the technological setup, interface design, and list types (i.e., differences in types of rewards and sanctions) between the provincial-level SCS platforms.

We conducted two rounds of data collection. First, between June 2019 and December 2019, we collected data on blacklists and redlists from 30 Chinese provincial-level ADs comprised of 22 provinces, 5 autonomous regions and 4 municipalities under the direct administration of central government. Second, in February 2020, we started collecting data on blacklists and redlists related to the coronavirus outbreak.

As we describe in more detail in the methodology section, our study approach is fundamentally *exploratory*. Data collection and analyses were intended to understand SCS implementation with regard to three high-level research questions, as follows.

- `RQ1`: Are there technological and design differences in credit lists and records between the provincial SCS platforms?

[2]Generally, policy-making in China is accompanied by a multitude of other policy documents. Engaging in a comprehensive description of Chinese policy-making would go beyond the scope of this study.
[3]In China, provincial-level ADs comprise provinces (e.g., Sichuan), municipalities under the direct administration of central government (e.g., Beijing, Shanghai) and autonomous regions (e.g., Inner Mongolia, Tibet).

Figure 1: Screenshot of an overview of the SCS information platforms of the different ADs listed on the national SCS platform "creditchina.gov.cn". Taiwan, Hong Kong and Macao were previously listed together with other ADs on the landing page of the "Credit China" website, but without a valid link. The listings were then removed in July 2019. Data collection was conducted via the SCS platform of each AD. Color-coding: orange represents municipality under the direct administration of central government; blue represents provinces; purple represents autonomous administrative regions; green represents the Xinjiang production and construction corps (Bingtuan), an economic and paramilitary organization in the Xinjiang Uyghur Autonomous Region, which is not included in our analysis due to an insignificant amount of credit data. Translations of AD names added by the authors.

- RQ2: How do provincial SCS platforms differ in the number and types of blacklists and redlists?
- RQ3: How do SCS blacklist and redlist *records* of the same type of list differ in terms of the information displayed across provincial SCS platforms?

2.2 Methodological approach

2.2.1 Data. Our analysis pertains to blacklists and redlists implemented at the AD level from June 2019 to December 2019. Data collection was aimed at provincial-level blacklists and redlists from 31 ADs (22 provinces, 5 autonomous regions, 4 municipalities under the direct administration of central government) listed on China's national SCS platform "creditchina.gov.cn" (Figure 1).[4] For the follow-up study of coronavirus-related lists, we inspected the same SCS platforms again between February 2020 and April 2020.

Data collection primarily refers to a) the types of lists implemented in each AD (RQ2) and b) retrieving individual credit records from the most commonly implemented blacklist and redlist across all 31 ADs (RQ3). Collecting list types and credit records enabled an analysis of the technical realization and interface designs of SCS platforms and credit records (RQ1).

Our data collection was organized to produce a *descriptive* study of SCS implementation. Our core analyses focused on the diversity of list types across ADs and the structural differences between list records, in particular, their interface designs and the information provided in individual credit records. For several reasons, we did not conduct a quantitative analyses on published records. First, during data collection, we observed that the number of published SCS records changed on a day-to-day basis for all SCS platforms. We refrained from drawing general inferences on SCS credit records based on a onetime quantitative analysis. Second, when we began to scrutinize different SCS platforms, we observed large differences in the amount of credit records uploaded. Some SCS platforms had not published any credit records, while some displayed multiple millions (note that only a few SCS platforms indicated the total number of credit records). Third, given the early stage of SCS development, a comprehensive quantitative analysis of the economic and societal impacts of credit records was not possible at the time of data collection. This impact may need several years to materialize as SCS measures begin to influence the economy, government administration, and social processes at large. Fourth, as we discuss in the next subsection, we encountered challenges in accessing and retrieving public credit information from SCS platforms.

2.2.2 Data collection obstacles. The first obstacle was obtaining access to the 31 AD SCS platforms. Access from our location was severely impeded, so we tested the accessibility of different SCS websites from various locations. To accomplish this, we sent web requests from 44 servers spread around the world to each AD's SCS website.[5] SCS server accessibility from outside China was generally possible but unstable.[6] To investigate SCS platforms, we used a virtual private network of servers located in China. Requests from China provided more stable access to SCS servers than from other locations. All SCS servers, apart from the SCS server of the municipality of Chongqing, responded to requests from a Chinese server. For the server of the municipality Chongqing, no data could be retrieved at any time, as the server did not respond to requests for the entire data collection period from any location. Thus, our final data collection represented 30 ADs. Overall, it took 6 months to access all SCS platforms and to document the different types

[4] This list also included the Xinjiang production and construction corps (Bingtuan). However, we did not include these data in our analysis for two reasons: first, Bingtuan is a unique state-owned economic and paramilitary organization in Xinjiang and, second, at the time of data collection, Bingtuan's SCS platform had published only a very small amount of credit information (9 blacklist and 7 redlists entries).

[5] The analysis was conducted with the Uptrends online monitoring service (www.uptrends.com). Data available from the authors.

[6] The most frequent return values were: HTTP connection failure, HTTP protocol error, HTTP timeout, and TCP connection failure.

of blacklists and redlists, verify them through revisits, and collect credit records for each AD.

While documenting the different types of lists for each province, we observed that each AD operated a different web server with different implementations of front-end, back-end and database design. Moreover, we did not find a public API on any of the AD SCS platforms. Taken together, this made data collection for *credit records* complicated, as each AD SCS platform required the programming of a unique web crawler and scraper.

The systematic sampling of public credit records from each blacklist and redlist on all SCS platforms was not possible for several reasons. First, the number and therefore types of lists implemented varied between the ADs. Some ADs had more than 10 types of lists, while others only displayed a single list (see Results). We saw that some ADs with only a single implemented blacklist or redlist used this list to present different types of sanctions or rewards. Second, some ADs had only one list but no records to show at all. Third, SCS platforms differed in how credit records were displayed. For example, some SCS platforms displayed a number of credit records on a single page and offered page tabs that opened the next page, displaying the next set of credit records. This interface style allowed page visitors to go through all available credit records. Other SCS platforms only showed a selection of credit records and instead of page tabs provided a search bar for specific queries. Here, visitors could not see all available credit records. Finally, some AD SCS platforms deployed captchas and bot blockers that sometimes led to time-out denials such as temporary or even permanent IP address suspension.

Given these restrictions on the collection of credit records, systematic and unbiased sampling of credit records across all SCS platforms was not possible. However, the goal of our study was not to measure effects between credit record samples to generalize to the SCS as a single system. Instead, for the credit record analysis, our research goal was to explore informational differences in credit records across the SCS platforms. For this purpose, homogeneous convenience sampling was sufficient to compare the information provided on credit records on the same list between SCS platforms. Homogeneous convenience sampling differs from conventional convenience sampling by constraining sampling by one factor (see e.g., [17]). We did not sample any credit record on any type of list (i.e., we did not conduct conventional convenience sampling). We directed the analysis of credit records toward the most frequently implemented type of blacklist and redlist across all SCS platforms. Consequently, different crawling and data extraction (scraping) robots were programmed to extract pre-specified information on credit records from the most common type of blacklist and redlist.[7] The two main frameworks and tools used for the crawling and scraping process were ThoughtWorks Limited open source headless browser Selenium and Scrapinghub Limited open source framework called Scrapy. The extracted data were eventually pushed into a noSQL database (MongoDB) as a horizontally scaling non-relational database was the better solution given the different SCS platform implementations.

Finally, the obstacles described above naturally led to credit record samples of varying size. On some SCS platforms, we managed

[7]We provide a code example of a crawler and a spider in the Auxiliary Material.

Figure 2: Shanghai's "Dishonest legal persons subjected to enforcement" (Lao Lai) blacklist of companies only displayed 10 record entries, requiring visitors to make a targeted search query. Translations by the authors.

to retrieve thousands of public credit records. On other platforms we obtained less than a hundred; some platforms did not have *any* credit records at all during the entire data collection period (for an overview of sampling results, see Table 2 in the Auxiliary Material). The differences in sample size were not due to any systematic sampling error committed by us but reflected the arbitrariness of the credit record display across the SCS platforms during the data collection period.

3 RESULTS

3.1 Technical implementation and design of blacklists and redlists

Each SCS platform operated a different web server with its own front-end, back-end and database design. We observed that the designs of the blacklists and redlists differed between ADs but was, overall, simple and plain.

All SCS platforms implemented either a Hypertext Markup Language (HTML) document with classic Cascading Style Sheet (CSS) structure or advanced dynamic scripting technology (JavaScript) for lists and individual records.

The majority of ADs (21) displayed only a selection of records but enabled targeted queries via a search bar. The remaining ADs showed all available social credit records with the help of a page tab. For example, on Guangxi's SCS platform, blacklist records could

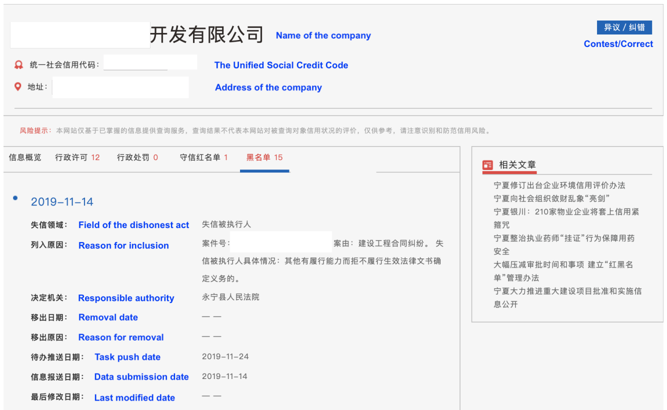


Figure 3: A two-column example credit record of the "Lao Lai" blacklist published on Ningxia's SCS platform. Translations by the authors.

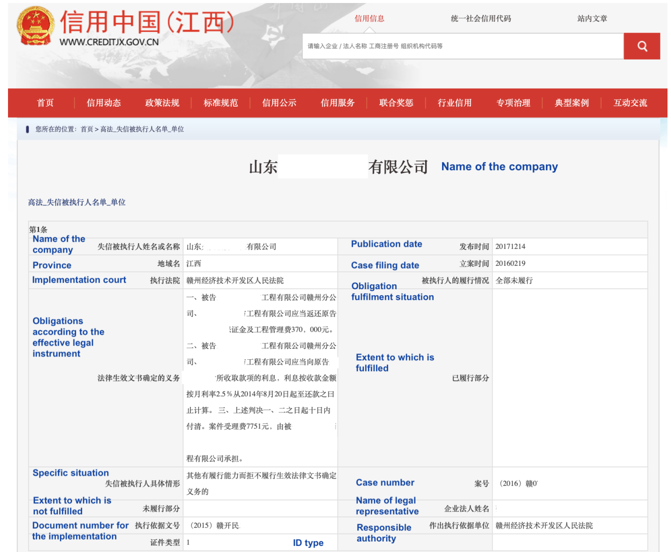


Figure 4: A multi-column example record of Jiangxi's "Lao Lai" blacklist (失信被执行人名单). Translations by the authors.

be accessed via 6852 tabs, each displaying 10 records. By contrast, Shanghai's blacklists showed ten blacklist records with no option to access more entries other than with a targeted query (Figure 2).

The design differences extended to individual credit records. Blacklist and redlist records were either structured as two column tables (Figure 3), multiple column tables (Figure 4) or continuous text documents.

Inner Mongolia and Shandong enabled sharing of blacklist and redlist records through Chinese social media platforms (e.g., Wechat, Sina Weibo, and Baidu Tieba). We found that eight SCS platforms offered citizens and organizations the possibility to contest published social credit records via a standardized interface option (e.g., Figure 3 top right corner). Our data indicate that there are technological and design differences in credit lists and records between provincial SCS platforms (RQ1). The current design and implementation of SCS platforms prioritize the display of social credit records rather than any aspect of their reputational effects. All SCS platforms had a binary rating system for good and bad behaviors – redlists and blacklists. Other than this binary classification, however, ADs did not apply other rating measures, such as numerical or continuous scoring. Indeed, we did not observe any social credit score at all communicated on any provincial-level SCS platform across China. Different types of lists were not put into relation with each other by means of a sorting or ranking. For example, no system of reputational ordering was found between individual records that highlighted severe transgressions more prominently than less severe cases. Five ADs showed numerical aggregation when a citizen or company had multiple social credit records. Entities with additional record entries were not displayed more prominently than entities that had a single credit record entry. Currently, the design of the SCS lists serves as a digitally accessible repository for citizen and company records and does not use any advanced features characteristic of other digital reputation systems [25].

3.2 Diversity and comprehensiveness: Number and types of blacklists and redlists

In response to RQ2, our data provide evidence for substantial differences in the number and types of lists between ADs (compare Figures 5 & 6). This confirms that regional governments determine the number and types of blacklists and redlists for their administrative region. For example, Beijing, Tianjin, Tibet, Guangdong, Hunan, Shanxi and Qinghai each operated more than ten different types of blacklists and redlists. In contrast, Inner Mongolia, Ningxia, Gansu, Guizhou, and Hebei each had implemented only one blacklist *and* one redlist. At present, it is impossible to say why some ADs run multiple lists and some only a single list. The number of lists did not correlate with economic, demographic, or geographic factors (data not shown).

In total, more blacklists (273) were published than redlists (154). We first grouped the 273 blacklists into 41 categories and the 154 redlists into 45 categories. We then created a taxonomy consisting of eight types of blacklists and eight types of redlists that currently make up the entire SCS AD listing infrastructure (Table 1). Note that different types of lists emphasize compliance with the legal and social norms that an AD wants to improve on. Thereby, the SCS influences behavior through two common reputation strategies [2]. With a minimum threshold strategy, blacklisting stresses the need for conformism. This technique tries to bring all entities to the same level of compliance. Redlisting, on the other hand, highlights praiseworthy performers that are intended to serve as behavioral role models.

The majority of blacklists displayed companies and citizens that have not fulfilled a court order, have committed commercial or transactional fraud, or have not complied with specific industry regulations. *All* ADs had implemented a "List of Dishonest Persons subject to Enforcement" also called the "Lao Lai" blacklist. This

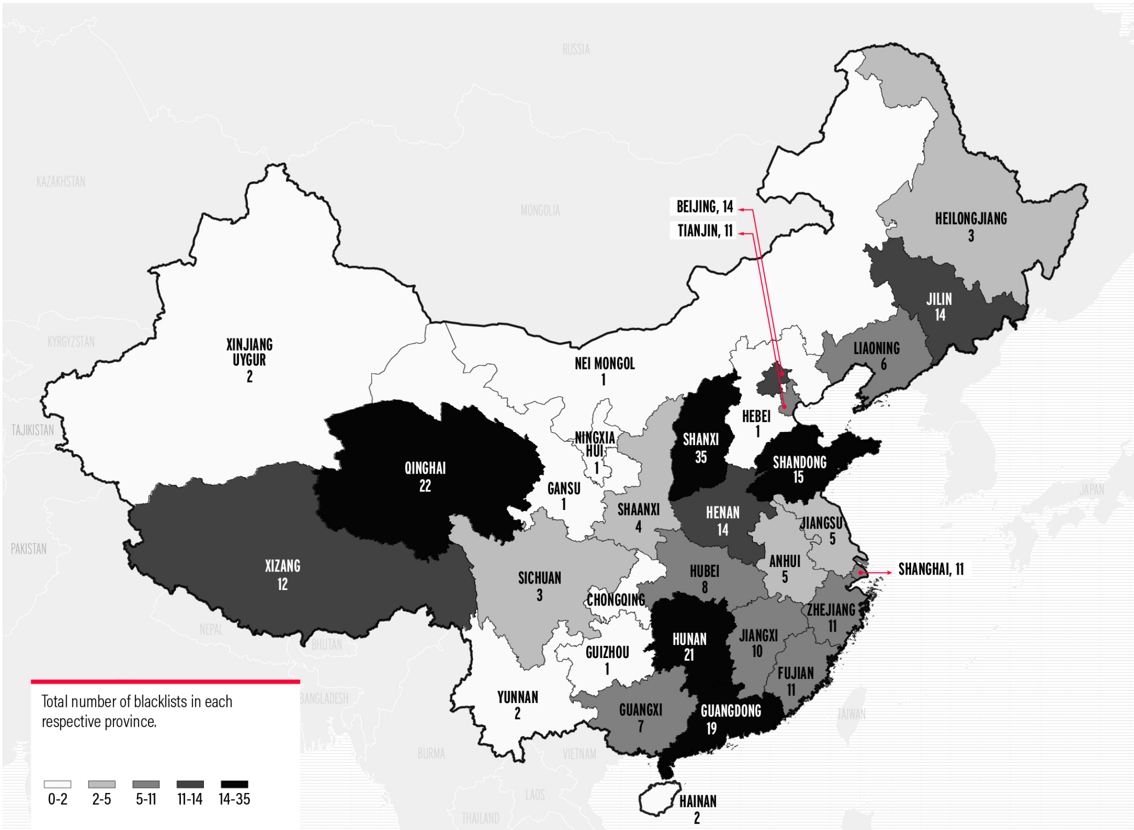


Figure 5: The number of blacklists implemented across 30 ADs. Shanxi had implemented most blacklists (35), followed by Qinghai (22), Hunan (21), Guangdong (19) and Shandong (15).

blacklist published information on citizens and companies that have failed to fulfill a court order. The "Lao Lai" blacklist aims to tackle China's court order enforcement problem [8, 9]. It forms a critical part of the JRP by which listed citizens face multiple restrictions, such as being banned from taking flights and high speed trains. Restrictions for "Lao Lai" companies include denial of licenses, reduced possibility to win bids for public contracts, or being subject to additional requirements for mandatory government approval for investments in sectors where market access is usually not regulated. Beyond the "Lao Lai" blacklist, we did not find any other type of blacklist implemented on all SCS platforms. The other types of blacklist most commonly found targeted non-compliance in tax payment (12 out of 30 ADs), untrustworthy behavior in financial activities (9/30), illegal import or export of products (8/30), delay or failure to compensate migrant[8] workers (8/30, companies only), or failure to protect the environment (7/30, companies only). We found blacklists that sanctioned fraud in marriage registrations or charity donations (social fraud), companies that had failed to comply with product quality standards (especially in food and drug production), or companies that had bad employment relationships.

The most frequently implemented redlists displayed entities that complied with tax law (18 out of 30 ADs) and import and export regulations (10/30). Usually, redlists serve to reward particularly "praiseworthy" behaviors. We made the surprising observation that many types of redlists highlighted regular compliance with laws and regulations. Some redlists, however, showcased individuals and companies that distinguished themselves politically or morally. For example, Beijing's SCS platform published a list called "4th Beijing Excellent Builders of Socialism with Chinese Characteristics", and Jiangxi and Tianjin listed citizens that had been rewarded the "May Fourth Medal". Tianjin had implemented two lists titled "Tianjin Good Man" and "Tianjin Ideological and Moral Model". Tibet had a similar redlist called "Moral Models & Good Political Ideology" (Figure 7). Other redlists were dedicated to citizens that had volunteered, given to charity or won awards in education, science or technology. Overall, the redlist infrastructure was less elaborate than its blacklist counterpart: not a single type of redlist existed in all ADs. Three ADs had published a single redlist with no data (Xinjiang, Gansu, and Jilin).

3.3 Informational consistency on credit records of the most common blacklist and redlist

To address RQ3, we explored the informational differences among the credit records of the most frequently implemented types of lists: the "Lao Lai" list (blacklist) and the "Class A Taxpayer" list

[8]"Migrant" here refers to rural citizens moving into urban centers for employment.

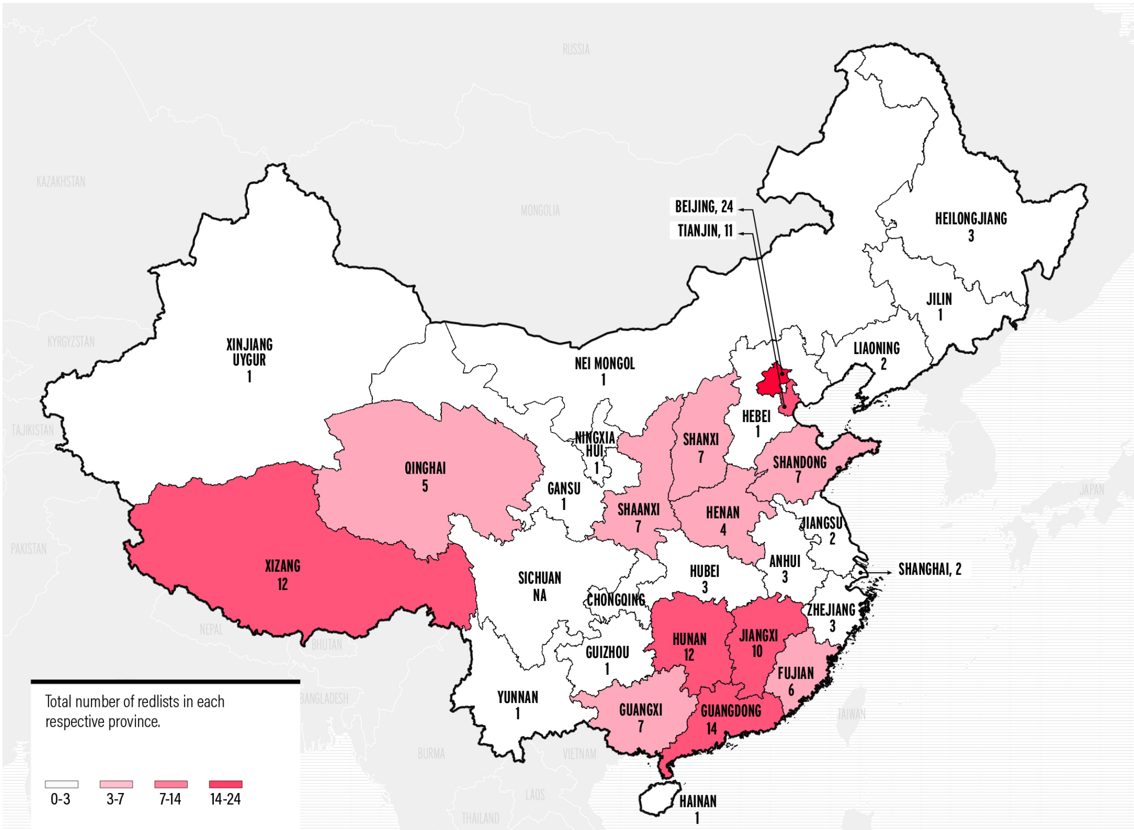


Figure 6: The number of redlists implemented across 30 ADs. Beijing had implemented the most redlists (24), followed by Guangdong (14), Xinjiang (12), Hunan (12), Tianjing (11), and Jiangxi (10).

(redlist). With the exception of Jilin and Tibet, the remaining 28 ADs had published credit records in their "Lao Lai" lists. We compared ADs based on the provision of five types of information in "Lao Lai" credit records: 1) the unified social credit code (companies) or identification number (natural persons), 2) specification of a data source or responsible authority, 3) reasons for listing (i.e., a justification), 4) information on the fulfillment of the requirements, and 5) information on a future removal date of the record (see Figure 8).

3.3.1 Information on "Lao Lai" blacklist credit records. Based on the samples of credit records obtained, out of the 28 different ADs, only 14 ADs had provided either the unified social credit code (8/28) or the natural person's identification number (6/28). The remaining ADs either listed an organization code (3/28) for companies or simply the name of the natural person listed (3/28). 23 ADs specified the data source of the record (i.e., where the data had been generated), the name of the executive court (12/28) or a responsible agency. In all, 24 ADs provided at least some explanation for why an entity had been listed. In the majority of cases, the credit records referred to a specific law that was to be enforced. Finally, 12 ADs indicated whether the requirement had already been fulfilled or not, and only 6 ADs displayed the removal date of the record.

3.3.2 Information on "Class A Taxpayer" redlist credit records (including unspecified redlists). For ADs without a "Class A Taxpayer" list, we inspected records from the only list available. 25 ADs had provided redlist records on their SCS platforms. 17 ADs had explicitly used the term "unified social credit code" in their records, and 7 listed a "taxpayer identification number". The remaining ADs simply presented the name of the listed entity. All ADs that published redlist records provided some form of identifying information. Of these, 21 ADs indicated the responsible authority for the case in question, and 16 ADs included a justification for being listed (commonly termed "reason for inclusion" or "honor content"). 6 ADs indicated the record's expiration date. An example record of a Class A Taxpayer List is shown in Figure 9.

3.4 Flexibility: Blacklists and redlists regulate behavior during the COVID-19 epidemic

Finally, we found that novel types of norm transgression can be quickly subjected to blacklisting and redlisting. Between February 27 and March 30, 2020, we collected data from the same SCS platforms to understand whether blacklisting and redlisting were used to regulate social behavior in an exceptional state of emergency. During this second round of data collection, we had access to 25 of

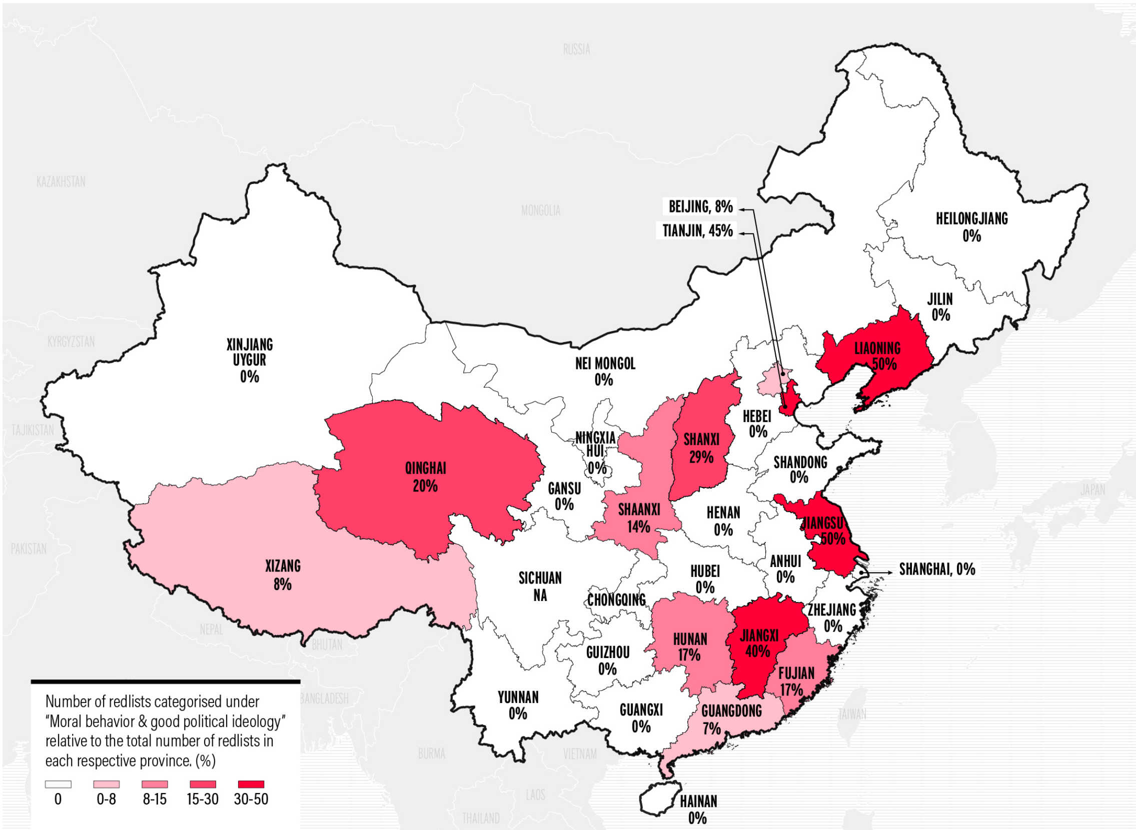


Figure 7: Ratios of redlists for moral behavior and good political ideology to total redlists across the 30 listed Chinese ADs.

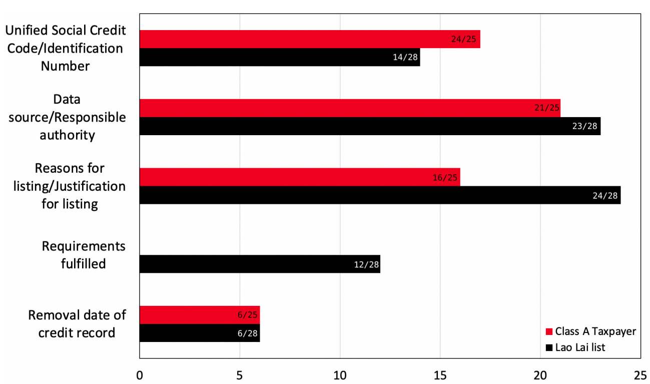


Figure 8: A comparison of the information provided on credit records collected from the most frequently implemented type of blacklist and redlist across all ADs.

the 31 ADs.[9] We identified coronavirus-related blacklists in 15 ADs and redlists in 10 ADs. Pursuant to our first analyses, blacklist and redlist records targeted natural persons and companies. We found that coronavirus blacklists included entities for selling fake preventive health products, violating quarantine regulations, organizing or participating in gatherings during lockdown, or illegally operating transport vehicles as ambulances. Blacklists were presented in different formats across the 15 ADs: they were either given in a row-and-column format (5) or in narrative-like news reports (10) (see Figure 10). Coronavirus redlists reported on devoted professionals such as doctors, nurses, volunteers, and border control officials, as well as on companies and individuals that had donated health products. All coronavirus redlist records were presented as narrative news reports.

[9]We did not have access to the SCS platforms of Jilin, Beijing, Fujian, Qinghai, Chongqing, and Hainan.

4 SUMMARY AND CONCLUDING ANALYSIS

We conducted an empirical investigation on the diversity, flexibility, and comprehensiveness of provincial-level SCS blacklists and redlists in China.

Overall, we highlighted that SCS listing designs facilitate public access to social credit records. The majority of SCS platforms display a selection of credit records and enable targeted queries. SCS platforms serve as digital reputation systems because redlists and blacklists digitally showcase entities' good and bad behaviors. However, with the exception of a few ADs that aggregated credit records for a single entity or allowed sharing of credit records to

Types of blacklists	Blacklists of untrustworthy entities (general)	Blacklists of commercial and transactional activities	Other blacklists	Employment relationship blacklists	Product quality blacklists	Blacklists of financial fraud	Industry blacklists	Blacklists of social fraud	Coronavirus blacklists
Beijing	2	4	3	0	1	4	0	0	NA
Shanghai	4	2	2	0	0	2	1	0	1
Tianjin	6	1	0	1	0	3	0	0	1
Hebei	1	0	0	0	0	0	0	0	1
Shanxi	5	5	5	1	3	2	12	2	1
Liaoning	2	1	3	0	0	0	0	0	0
Jilin	1	6	3	1	2	1	0	0	NA
Heilongjiang	3	0	0	0	0	0	0	0	1
Shandong	2	6	2	1	1	2	1	0	1
Jiangsu	2	0	2	0	0	1	0	0	1
Zhejiang	3	4	2	0	1	1	0	0	1
Anhui	4	0	1	0	0	0	0	0	1
Fujian	3	2	2	2	0	1	1	0	NA
Jiangxi	2	2	1	1	0	2	2	0	1
Henan	5	4	1	0	1	1	2	0	1
Hubei	2	5	0	0	1	0	0	0	0
Hunan	5	5	3	1	4	0	3	0	0
Sichuan	2	1	0	0	0	0	0	0	0
Guangdong	1	5	5	4	2	1	1	0	0
Gansu	1	0	0	0	0	0	0	0	0
Hainan	2	0	0	0	0	0	0	0	NA
Qinghai	3	4	2	1	3	2	7	0	NA
Guizhou	1	0	0	0	0	0	0	0	1
Yunnan	2	0	0	0	0	0	0	0	1
Shaanxi	1	1	1	1	0	0	0	0	1
Tibet	1	4	0	6	0	0	1	0	0
Inner Mongolia	1	0	0	0	0	0	0	0	1
Guangxi	3	1	1	2	0	0	0	0	0
Ningxia	1	0	0	0	0	0	0	0	0
Xinjiang	2	0	0	0	0	0	0	0	0

Types of redlists	Redlists of trustworthy entities (general)	Redlists of commercial and transactional activities	Redlists of moral behavior & good political ideology	Employment & customer relationship redlists	Product quality redlists	Industry redlists & redlists of professions	Redlists of Intellectual property & awards	Other redlists	Coronavirus redlists
Beijing	1	5	2	2	2	4	8	0	NA
Shanghai	1	1	0	0	0	0	0	0	0
Tianjin	0	3	5	0	0	0	3	0	0
Hebei	1	0	0	0	0	0	0	0	1
Shanxi	0	3	2	0	0	1	0	1	0
Liaoning	0	1	1	0	0	0	0	0	0
Jilin	1	0	0	0	0	0	0	0	NA
Heilongjiang	1	2	0	0	0	0	0	0	1
Shandong	1	3	0	0	1	1	0	1	1
Jiangsu	1	0	1	0	0	0	0	0	0
Zhejiang	2	1	0	0	0	0	0	0	1
Anhui	0	3	0	0	0	0	0	0	1
Fujian	1	3	1	1	0	0	0	0	NA
Jiangxi	0	2	4	0	0	3	1	0	0
Henan	0	3	0	0	0	0	1	0	1
Hubei	1	2	0	0	0	0	0	0	0
Hunan	1	2	2	0	0	3	4	0	0
Sichuan	0	0	0	0	0	0	0	0	0
Guangdong	0	7	1	0	0	1	5	0	0
Gansu	1	0	0	0	0	0	0	0	1
Hainan	1	0	0	0	0	0	0	0	NA
Qinghai	0	2	1	0	0	1	1	0	NA
Guizhou	1	0	0	0	0	0	0	0	1
Yunnan	1	0	0	0	0	0	0	0	1
Shaanxi	1	3	1	0	0	2	0	0	1
Tibet	0	4	1	0	2	2	1	2	0
Inner Mongolia	1	0	0	0	0	0	0	0	0
Guangxi	0	1	0	0	0	0	6	0	0
Ningxia	1	0	0	0	0	0	0	0	0
Xinjiang	1	0	0	0	0	0	0	0	0

Table 1: The different types of blacklists and redlists implemented by ADs in China. Shading indicates the number of blacklists or redlists for a given type. N/A denotes no access to the SCS platform.

Figure 9: A screenshot of a redlist record from the "Class A Taxpayer List" published on the Fujian SCS platform. Translations by the authors.

social media platforms, we did not observe any automated classification, ranking or scoring on any of the current SCS listings.

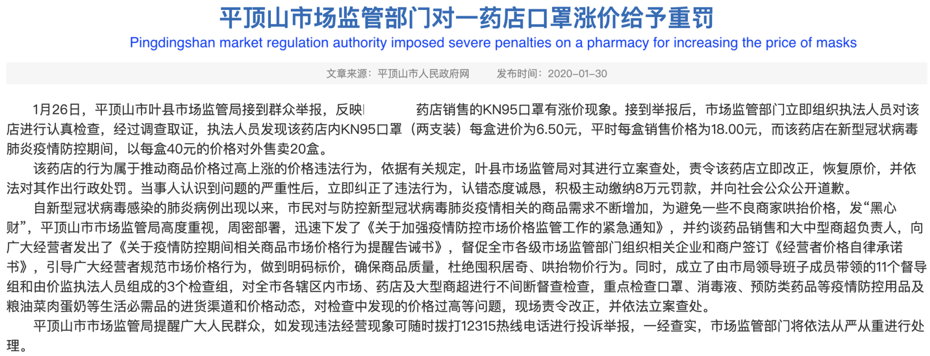
The SCS comprises hundreds of blacklists and redlists across provincial-level ADs. Currently, the majority of these types of lists target compliance with a wide range of laws and regulations. Thereby, SCS blacklists focus on "Lao Lai" entities, which are citizens and companies that have not fulfilled a court order. The SCS first displays "Lao Lai" on its digital listings and hence excludes them from future cooperative opportunities through its JRP. Based on these two mechanisms, the SCS seeks to turn "Lao Lai" into cooperators by attaching an exceptionally high cost to defection. We also observed redlists that highlight praiseworthy political and moral behaviors. Further development of lists that go well beyond legal or regulatory norms could substantially increase the social control characteristics of the SCS.

We have exemplified the flexibility of SCS listings by a case study on the COVID-19 outbreak. Digital blacklists and redlists might be a particularly powerful regulatory measure because they can be adapted to help accomplish novel policy goals quickly and at relatively low costs.

There are several outstanding questions for future research. For example, will SCS platform design incorporate more reputational affordances? Will the governmental and commercial branches (i.e., big data apps) of the SCS cooperate to share and analyze different data streams? Will SCS mechanisms really produce their intended regulatory effects? We believe that asking such questions is crucial and we hope to have laid a useful foundation for future empirical and conceptual studies on the SCS.

5 ETHICAL DIMENSIONS OF THE STUDY

We now turn to initial ethical considerations of data-driven research on SCS implementation. First, our analysis was based on publicly available data found on key platforms of China's SCS. These data are posted to enable public scrutiny. Our paper includes screenshots from the currently available implementations (see Figures 1, 2, 3, 4, 9, 10). Our data collection and analyses are privacy-preserving: we blurred any personally identifiable data to protect the privacy of



平顶山市场监管部门对一药店口罩涨价给予重罚

Pingdingshan market regulation authority imposed severe penalties on a pharmacy for increasing the price of masks

文章来源：平顶山市人民政府网　　发布时间：2020-01-30

1月26日，平顶山市叶县市场监管局接到群众举报，反映　药店销售的KN95口罩有涨价现象。接到举报后，市场监管部门立即组织执法人员对该店进行认真检查，经过调查取证，执法人员发现该药店内KN95口罩（两支装）每盒进价为6.50元，平时每盒销售价格为18.00元，而该药店在新型冠状病毒肺炎疫情防控期间，以每盒40元的价格对外售卖20盒。

该药店的行为属于推动商品价格过高上涨的价格违法行为，依据有关规定，叶县市场监管局对其进行立案查处，责令该药店立即改正，恢复原价，并依法对其作出行政处罚。当事人认识到问题的严重性后，立即纠正了违法行为，认错态度诚恳，积极主动缴纳8万元罚款，并向社会公众公开道歉。

自新型冠状病毒感染的肺炎病例出现以来，市民对与防控新型冠状病毒肺炎疫情相关的商品需求不断增加，为避免一些不良商家哄抬价格，发“黑心财”，平顶山市市场监管局高度重视，周密部署，迅速下发了《关于加强疫情防控市场价格监管工作的紧急通知》，并约谈药品销售和大中型商超负责人，向广大经营者发出了《关于疫情防控期间相关商品市场价格行为提醒告诫书》，督促全市各级市场监管部门组织相关企业和商户签订《经营者价格自律承诺书》，引导广大经营者规范市场价格行为，做到明码标价，确保商品质量，杜绝囤积居奇、哄抬物价行为。同时，成立了由市局领导班子成员带领的11个督导组和由价监执法人员组成的3个检查组，对全市各辖区内市场、药店及大型商超进行不间断督查检查，重点检查口罩、消毒液、预防类药品等疫情防控用品及粮油菜肉蛋奶等生活必需品的进货渠道和价格动态，对检查中发现的价格过高等问题，现场责令改正，并依法立案查处。

平顶山市市场监管局提醒广大人民群众，如发现违法经营现象可随时拨打12315热线电话进行投诉举报，一经查实，市场监管部门将依法从严从重进行处理。

Figure 10: Screenshot of the coronavirus blacklist from the SCS platform for Henan province. Translation: ***On January 26, the Market Supervisory Authority of Ye County Pingdingshan City received reports from the public reporting that ** Pharmacy increased the price of KN95 masks. After receiving the report, the authority immediately sent out law enforcement officers to conduct a serious inspection of the store and found that the purchase price of the KN95 masks (2 pieces in one package) was 6.5 RMB for the store and the sale price was usually 18 RMB. However, the pharmacy sold 20 packages of the masks at the price of 40 RMB during the epidemic period. The pharmacy was thus in violation of the price regulation. Following relevant regulations, the Market Supervisory Authority filed a case for the investigation and ordered the pharmacy to restore the price to its original level. The authority also imposed administrative penalties on the pharmacy according to law. The pharmacy realized the seriousness of the problem and immediately halted the illegal behavior, admitted its misconduct, proactively paid a fine of 80,000 RMB, and apologized to the public.*** **Translations by the authors.**

listed companies and citizens. Our methodological approach does not result in any unfavorable consequences or costs for any of the data subjects. We are transparent in our methodology and provide a representative code example of a web crawler and spider we used in this study (see Auxiliary Material).

Second, our account adheres to the principles of ethical web crawling and scraping [20, 29, 33, 34]. For each SCS platform, we checked for a specified *robots.txt* file. At no point during our data collection did we find a robots.txt file that specified rules for web crawlers. Accordingly, when platforms make data publicly available, do not specify a robots.txt file, and do not provide a data collection interface (e.g., API), then robots are free to gather data (see, e.g., [29, 33]).

Third, the purpose of our study is ethically justifiable on its own. In the absence of systematic empirical accounts, uncertainty will inevitably help foster misconceptions about the SCS (whether overly positive or negative). Given China's geopolitical prominence, governments of other countries may be inspired to copy China's SCS [24]. This is particularly likely for neighboring countries [37]. Data-driven research on SCS implementation can help prevent hasty SCS adaptations by other governments based on false assumptions. Empirical and conceptual analyses on the SCS allow for a more informed public debate about the development of digital socio-technical systems. As our data indicate, *currently*, there is little evidence that blacklists and redlists operate as AI-driven reputation systems. Apart from two SCS platforms that enable sharing of credit records to social media platforms, at the moment, there is no evidence that credit records are subjected to other means of digital reputation mechanisms such as classification, ranking, or profiling based on AI. It is possible that future developments might implement AI-based reputation mechanisms. As we have argued, additional empirical work on the SCS is necessary given that Chinese policy-making rests on often vaguely formulated policy goals. We show a considerable diversity of SCS blacklist and redlist implementation that cannot be concluded from policy analysis alone. Our study raises important questions that also matter for non-Chinese citizens and organizations. For example, is stable access to blacklists and redlists from outside China justifiable when non-Chinese citizens and companies are listed [4, 10]? Should China distribute licenses or special APIs to allow non-Chinese entities to ascertain whether they are listed? Or will Chinese authorities directly notify non-Chinese entities when they are listed?

The Chinese SCS is already one of the most comprehensive reputation systems in the world. Given that the government generates the reputation signals, we believe that SCS blacklisting and redlisting could have a strong influence on societal behavior at large.

Finally, this research extends growing calls for more open data in computational social science [21] with a case for more data availability *in China*. As this body of research has shown, open government data can significantly improve our understanding of societies' most important challenges in the context of equality, health, or employment. Even if data collection obstacles are likely to persist, we hope that our study underlines the importance of future data-driven research on the Chinese SCS.

ACKNOWLEDGMENTS

We thank Rogier Creemers, Bilge Kobas, and Marianne von Blomberg for their helpful input. We also thank the anonymous reviewers for their constructive comments. We gratefully acknowledge funding support from the Bavarian Research Institute for Digital Transformation (bidt). Mo Chen further thanks for the support through a postdoc research stipend of the Fritz Thyssen Foundation. Responsibility for the content of this publication rests with the authors.

REFERENCES

[1] Shazeda Ahmed. 2019. The messy truth about social credit. *Logic* 7 (2019).

[2] Geoffrey Brennan and Philip Pettit. 1993. Hands invisible and intangible. *Synthese* 94, 2 (1993), 191–225.

[3] Charlie Campbell. 2019. How China is using "Social Credit Scores" to reward and punish its citizens. *Time* (2019). Available at: https://time.com/collection/davos-2019/5502592/china-social-credit-score/.

[4] Mo Chen, Kristina Bogner, Joana Becheva, and Jens Grossklags. 2021. The transparency of the Chinese Social Credit System from the perspective of German organizations. In *Proceedings of the 29th European Conference on Information Systems (ECIS)*. Completed research paper.

[5] Mo Chen and Jens Grossklags. 2020. An empirical analysis of the commercial arm of the Chinese Social Credit System. *Proceedings on Privacy Enhancing Technologies* 4 (2020), 89–110.

[6] Yongxi Chen and Anne Cheung. 2017. The transparent self under big data profiling: Privacy and Chinese legislation on the Social Credit System. *The Journal of Comparative Law* 12, 2 (2017), 356–378.

[7] Martin Chorzempa, Paul Triolo, and Samm Sacks. 2018. *China's Social Credit System: A mark of progress or a threat to privacy?* Policy Briefs, Peterson Institute for International Economics No. PB18-14.

[8] Rogier Creemers. 2018. *China's Social Credit System: An evolving practice of control.* SSRN Working Paper Nr. 3175792.

[9] Xin Dai. 2018. *Toward a reputation state: The Social Credit System project of China.* SSRN Working Paper No. 3193577.

[10] Severin Engelmann, Mo Chen, Felix Fischer, Ching-Yu Kao, and Jens Grossklags. 2019. Clear sanctions, vague rewards: How China's Social Credit System currently defines "good" and "bad" behavior. In *ACM Conference on Fairness, Accountability, and Transparency (FAT*)*. 69–78.

[11] Karen Hao. 2021. The Biden administration's AI plans: what we might expect. *MIT Technology Review (22 January 2021)* (2021).

[12] Sebastian Heilmann. 2008. From local experiments to national policy: The origins of China's distinctive policy process. *The China Journal* 59 (2008), 1–30.

[13] Sebastian Heilmann. 2009. Maximum tinkering under uncertainty: Unorthodox lessons from China. *Modern China* 35, 4 (2009), 450–462.

[14] Sebastian Heilmann. 2018. *Red Swan: How Unorthodox Policy-Making Facilitated China's Rise.* Chinese University Press.

[15] Samantha R. Hoffman. 2017. *Programming China: The Communist Party's autonomic approach to managing state security.* Ph.D. Dissertation. University of Nottingham.

[16] Jamie Horsley. 2018. China's Orwellian social credit score isn't real. *Foreign Policy* 16 (2018).

[17] Justin Jager, Diane Putnick, and Marc Bornstein. 2017. II. More than just convenient: The scientific merits of homogeneous convenience samples. *Monographs of the Society for Research in Child Development* 82, 2 (2017), 13–30.

[18] Genia Kostka. 2019. China's Social Credit Systems and public opinion: Explaining high levels of approval. *New Media & Society* 21, 7 (2019), 1565–1593.

[19] Theresa Krause and Doris Fischer. 2020. An economic approach to China's Social Credit System. In *Social Credit Rating*. Springer, 437–453.

[20] Vlad Krotov and Leiser Silva. 2018. Legality and ethics of web scraping. In *Proceedings of the Twenty-fourth Americas Conference on Information Systems (AMCIS)*.

[21] David Lazer, Alex Pentland, Duncan J. Watts, Sinan Aral, Susan Athey, Noshir Contractor, Deen Freelon, Sandra Gonzalez-Bailon, Gary King, Helen Margetts, et al. 2020. Computational social science: Obstacles and opportunities. *Science* 369, 6507 (2020), 1060–1062.

[22] Fan Liang, Vishnupriya Das, Nadiya Kostyuk, and Muzammil M. Hussain. 2018. Constructing a data-driven society: China's Social Credit System as a state surveillance infrastructure. *Policy & Internet* 10, 4 (2018), 415–453.

[23] Alexandra Ma. 2018. China has started ranking citizens with a creepy 'Social Credit' system – Here's what you can do wrong, and the embarrassing, demeaning ways they can punish you. *Business Insider* 29 (2018).

[24] Daithí Mac Síthigh and Mathias Siems. 2019. The Chinese Social Credit System: A model for other countries? *The Modern Law Review* 82, 6 (2019), 1034–1071.

[25] Sergio Marti and Hector Garcia-Molina. 2006. Taxonomy of trust: Categorizing P2P reputation systems. *Computer Networks* 50, 4 (2006), 472–484.

[26] Louise Matsakis. 2019. How the West got China's Social Credit System wrong. *Wired (29 July 2019)* (2019).

[27] Andrew Mertha. 2009. "Fragmented authoritarianism 2.0": Political pluralization in the Chinese policy process. *The China Quarterly* 200 (2009), 995–1012.

[28] Simina Mistreanu. 2018. Life inside China's Social Credit Laboratory: The party's massive experiment in ranking and monitoring Chinese citizens has already started. *Foreign Policy* (2018). Available at: https://foreignpolicy.com/2018/04/03/life-inside-chinas-social-credit-laboratory/.

[29] Zeina Mneimneh, Josh Pasek, Lisa Singh, Rachel Best, Leticia Bode, Elizabeth Bruch, Ceren Budak, Pamela Davis-Kean, Katharine Donato, Nicole Ellison, et al. 2021. *Data Acquisition, Sampling, and Data Preparation Considerations for Quantitative Social Science Research Using Social Media Data.* Working Paper.

[30] Mareike Ohlberg, Shazeda Ahmed, and Bertram Lang. 2018. Central planning, local experiments: The complex implementation of China's Social Credit System. *Mercator Institute for China Studies* 43 (April 2018), 1–15.

[31] Gunter Schubert and Björn Alpermann. 2019. Studying the Chinese policy process in the era of 'top-level design': The contribution of 'political steering' theory. *Journal of Chinese Political Science* 24, 2 (2019), 199–224.

[32] State Council. 2014. Notice of the State Council on issuing the outline of the plan for building a Social Credit System (2014-2020); (in Chinese).

[33] Yang Sun, Isaac Councill, and Lee Giles. 2010. The ethicality of web crawlers. In *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*. 668–675.

[34] Mike Thelwall and David Stuart. 2006. Web crawling ethics revisited: Cost, privacy, and denial of service. *Journal of the American Society for Information Science and Technology* 57, 13 (2006), 1771–1779.

[35] Alexander Trauth-Goik. 2019. "Constructing a culture of honesty and integrity": The evolution of China's Han-centric surveillance system. *IEEE Technology and Society Magazine* 38, 4 (2019), 75–81.

[36] Marianne von Blomberg. 2018. The Social Credit System and China's rule of law. In *Social Credit Rating*, Oliver Everling (Ed.). 111–137.

[37] Yau Tsz Yan. 2020. *Exporting China's Social Credit System to Central Asia.* Retrieved Apr 28, 2021 from https://thediplomat.com/2020/01/exporting-chinas-social-credit-system-to-central-asia/

A AUXILIARY MATERIAL FOR “BLACKLISTS AND REDLISTS IN THE CHINESE SOCIAL CREDIT SYSTEM: DIVERSITY, FLEXIBILITY, AND COMPREHENSIVENESS”

A.1 Documentation: Example crawler and spider for Guangdong province

The following code sections are an excerpt of the crawling and scraping methodology to systematically collect data from public blacklists and redlists of the Chinese Social Credit System. The crawler for collecting relevant data and the spider for extracting specific information from the data are demonstrated for the example of the Guangdong province below. Please note that the collection methodology may have to be adjusted, if the collection site is undergoing changes. You also may want to revisit the discussion on the ethics of data crawling in our paper (see Section 5).

Crawler example Guangdong province:
This section shows how the link lists are created, in particular, the methodology to collect the deep links that lead to the entry records of blacklists and redlists. A headless browser (like Selenium) is used, which is basically a normal web browser remotely controlled by a programmed robot.

In the following, an example of a web crawler is given:

```
class GuangdongSelenium():
    def crawl_red(self):
        link = 'https://credit.gd.gov.cn/opencreditAction!getOpencreditList_new.[...]&tbType=1'
        print_start("Guangdong␣Redlist")
        linkliste = []
        file = open("linklist_guangdong_red.txt", "a")

        driver.get(link)
        driver.find_element_by_css_selector('#newtype␣>␣option:nth-child(8)').click()
        driver.find_element_by_css_selector('label.search_button').click()

        while '下一页' in driver.page_source:
            try:
                categorylist = driver.find_elements_by_css_selector('tbody␣>␣tr:nth-child(1)␣>␣td␣>␣div␣>␣a')
                for i in categorylist:
                    print(i.get_attribute('href'))
                    s = i.get_attribute('href')
                    linkliste.append(s)
                driver.find_element_by_css_selector('a.next').click()
                time.sleep(10)
            except():
                print ("Error,␣no␣next␣page␣available!")
                break

        print("Length␣of␣final␣linklist:␣", len(linkliste))
        linkliste = list(dict.fromkeys(linkliste))
        print("This␣is␣the␣lenght␣of␣the␣list␣after␣removing␣all␣duplicates:␣", len(linkliste))
        for e in linkliste:
            file.write(e + "\n")

        print("Crawled␣links␣are␣written␣into␣the␣final␣file.")
        print("File␣created")
        file.close()
        driver.close()
        sys.exit()

    def crawl_black(self):
        link = 'https://credit.gd.gov.cn/opencreditAction!getOpencreditList_new.[...]&tbType=2'
        print_start("Guangdong␣Blacklist")
```

```
        linkliste = []
        file = open("linklist_guangdong_black.txt", "a")
        driver.get(link)
        driver.find_element_by_css_selector('#newtype > option:nth-child(2)').click()
        driver.find_element_by_css_selector('label.search_button').click()
        try:
            while '下一页' in driver.page_source:
                wait = WebDriverWait(driver, 10)
                wait.until(ec.visibility_of_element_located((By.CSS_SELECTOR, 'a.next')))
                time.sleep(10)
                categorylist = driver.find_elements_by_css_selector('tbody > tr:nth-child(1) > td > div > a')
                for i in categorylist:
                    print(i.get_attribute('href'))
                    s = i.get_attribute('href')
                    file.write(s + "\n")
                    linkliste.append(s)
                driver.find_element_by_css_selector('a.next').click()
                time.sleep(5)
        except:
            pass
            print("Error, no next page available!")
        print("File created")
        file.close()
        driver.close()
        sys.exit()
```

The desired output should be a collection of links stored in corresponding files ‘linklist_guangdong_black.txt’ or ‘linklist_guangdong_red.txt’.

```
https://credit.gd.gov.cn/infoTypeAction!getAwardAndGruel.[...]id=FF89EED12BC14E21BF36360E9044FC45
https://credit.gd.gov.cn/infoTypeAction!getAwardAndGruel.[...]id=FF89EED12BC14E21BF36360E9044FC45
[...]
https://credit.gd.gov.cn/infoTypeAction!getAwardAndGruel.[...]id=FF89EED12BC14E21BF36360E9044FC45
https://credit.gd.gov.cn/infoTypeAction!getAwardAndGruel.[...]id=FF89EED12BC14E21BF36360E9044FC45
```

Spider example Guangdong province:
This section shows a web scraping spider, a methodology that follows the web crawling process. A web scraper’s task is to sequentially work through the web crawler’s link list and extract specific data.

In the following, an example of a web scraper is given:

```
import scrapy, re

class GuangdongSpider(scrapy.Spider):
    name = "guangdong"
    file = open("linklist_guangdong_black.txt", "r")
    start_urls = [i.replace("\n", "") for i in file]

    def parse(self, response):
        table = response.css('table > tr > td')
        yield{
            'case_number' : table[1].css('::text').extract_first(),
            'lost_trustee_name' : table[3].css('::text').extract_first(),
            'gender' : table[5].css('::text').extract_first(),
            'age' : table[7].css('::text').extract_first(),
```

```
        'ID_number_desensitization__organization_code' : table[9].css('::text').extract_first(),
        'corporate_legal_person_name' : table[11].css('::text').extract_first(),
        'executive_court' : table[13].css('::text').extract_first(),
        'execution_basis_number' : table[15].css('::text').extract_first(),
        'basis_for_execution' : table[17].css('::text').extract_first(),
        'obligation_established_by_the_law' : table[19].css('::text').extract_first(),
        'implementation_of_the_person_being_executed' : table[21].css('::text').extract_first(),
        'untrustworthy_enforcer' : table[23].css('::text').extract_first(),
        'release_time' : table[25].css('::text').extract_first(),
        'filing_time' : table[27].css('::text').extract_first(),
        'fulfilled_part' : table[29].css('::text').extract_first(),
        'unfulfilled_part' : table[31].css('::text').extract_first(),
        'hyperlink' : response.url
    }
```

A.2 Table: Summary of credit record collection for blacklists and redlists

AD	No. of blacklist records	Avg. size blacklist record	No. of variables	No. of redlist records	Avg. size redlist record	No. of variables
Municipalities						
Beijing	100	1700 B	35	50	776.9 B	27
Shanghai	10	156.5 B	3	10	157.8 B	3
Tianjin	1501	1100 B	5	2000	306.6 B	5
AR						
Guangxi	30281	265.7 B	8	27692	547.5 B	15
Inner Mongolia	10	795.9 B	15	10	319.5 B	5
Ningxia	20	853.3 B	12	19	714.5 B	12
Xinjiang	3	1100 B	12	no data	-	-
Tibet	no data	-	-	no data	-	-
Provinces						
Anhui	190	926.5 B	15	190	315.8 B	6
Fujian	99	477.6 B	9	78	380.5 B	7
Gansu	20	1200 B	21	no data	-	-
Guangdong	160	1900 B	17	90	476.1 B	6
Guizhou	38	1600 B	6	39	2900 B	6
Hainan	40	817.3 B	17	40	654.6 B	13
Hebei	311	663.9 B	11	652	515.2 B	11
Heilongjiang	24	804.2 B	6	7	939.7 B	14
Henan	180	218.0 B	2	180	218.0 B	2
Hubei	50	588.4 B	11	50	465.5 B	8
Hunan	20	174.1 B	4	79	129.9 B	3
Jiangsu	50	1700 B	26	50	440 B	8
Jiangxi	2413	1600 B	16	482	1300 B	13
Jilin	no data	-	-	no data	-	-
Liaoning	4	1100 B	14	8	356.1 B	8
Qinghai	19	1000 B	15	18	928.6 B	15
Shaanxi	49	1100 B	15	47	748.6 B	15
Shandong	100	672.3 B	14	100	361.5 B	7
Shanxi	53	2100 B	21	73	1100 B	21
Sichuan	320	226.4 B	10	10	650.9 B	10
Yunnan	50	752.0 B	9	42	516.8 B	9
Zhejiang	1950	163.0 B	4	5580	217.0B	5
$\sum$	38065			37596		

Table 2: The "No. of blacklist records" and "No. of redlist records" indicate the number of credit records retrieved from each AD SCS platform for the most commonly implemented type of blacklist and redlist, respectively. Numbers show varying sample sizes due to several data collection obstacles (see Section 2.2). "Avg. size blacklist record" denotes the average byte size of a blacklist record for each sample. "No. of variables" indicates the number of informational variables on each credit record in the sample.