

Advance Data Justice Research and Practice

China Report

Shanghai SODA Data Tech Co,Ltd.
(Open Data China)

March 2022

Table of Contents

Table of Contents	2
Chapter I: Introduction	1
Research Background	1
Report Structure	1
Quick Findings	2
About Us	2
Chapter II: Internal Assessment	4
Introduction and Methodology	4
Internal Review	4
Desk Research	6
China: Building a Legal and Ethical Data Innovation Environment	6
Data Justice: A new concept in China	8
The Popular Data Justice Cases	10
Big Data Discriminatory Pricing	10
Health Passport	10
Delivery Workers Trapped in the Algorithm	11
Chapter III: Interview	12
Selection and invitation of interviewees	12
Focus of Interview	14
Protocol of interview	14
Findings of Interview	15
Understandings of Data Justice	15
Overview of Keywords	15
Unpacking conversations on Keywords	16
Equity: human rights perspective versus consumer perspective	16
Privacy: the concern of surveillance and chaos on facial recognition	16
Distribution of data and its benefits: rise of interest in individual data rights	16
Inclusiveness: a conversation around digital divide	17
Reflection: Is it really about data?	17
Data (in)Justice Stories in China	18
Can Participation work in China?	19
Foundation: Legal and Culture Environment is being improved	19
Strong Legal Foundation	20
Shift Social Narrative	20
Scaffolding: Capacity building and Intermediaries	21
Ethics and Tech for good: A New Lesson inside Corporations	21

Building Data Literacy : Possible but also Difficult	22
A Demand for Intermediaries	22
What can China offer to the world on Data Justice?	23
Chapter IV: Workshop	25
Workshop Task Design	25
Workshop Protocol	26
Workshop Delivery and Participants	27
Workshop Results	28
Task1: Three Keywords	28
Task2: Case Studies	28
Task3: Relevance and Usability of Pillars in the context of China	31
Relevance	32
Usability	33
Task4: Relevance and Usability of Pillars in Selected Case Studies	33
Score Analysis	35
Reflection Questions	35
Chapter V: Discussion and Suggestions	37
Focus of Data Justice in China	37
Landing Six Pillars Framework in China	37
Suggestions	38
Annex A: Interview Question	40
Annex B: Workshop Activity Sheet	43
Annex C: Interview Transcripts	44
Interviewee #1	44
Interviewee #2	88
Interviewee #3	131
Interviewee #4	147
Interviewee #5	200
Interviewee #6	231
Interviewee #7	257
Interviewee #8	277
Interviewee #9	312
Interviewee #10	340
Interviewee #11	379
Interviewee #12	411

Chapter I: Introduction

Research Background

The "Advance research and practice of Data Justice"(ADJRP) project is about to examine the broader understanding of data justice beyond the traditional privacy-oriented or ethic-oriented data governance approach while making and regulating data innovation. The project is led by the Data Governance Working Group, Global Partnership for AI and currently Alan Turing Institute is commissioned by CEIMIA (International Centre of Expertise in Montreal on Artificial Intelligence) to lead the project research.

A preliminary guide has been created by Alan Turing targeting three stakeholder groups including policy maker, developer and impacted community to help understand data justice and tackle the issue based upon a list of reflection questions grouped into six pillars including power, equity, identity, knowledge, participation, and access.

To include more diverse perspectives, especially non-western perspectives and low-to-middle income countries' perspectives, 12 global partners referred to as "policy pilot partner" are selected through an open call and tasked to deliver local research findings.

We, Open Data China, are part of this global network to help conduct an internal assessment, a series of interviews and organize a workshop. We are expected to summarize local understanding of data justice and assess the usability of the preliminary guide in the context of China. Suggestions for how to improve the preliminary guidance and how to further research and practice data justice especially in China are also expected to be made during the research.

Report Structure

The report is structured into five chapters

- Chapter I Introduction: It gives readers the brief background of the research, who we are, the overview of the report structure and quick findings for reference.
- Chapter II Internal Assessment: In this chapter we assess the usability of the preliminary guide drafted by Alan Turing Institute based upon our teams' own understanding of data justice as well as our desk research about domestic academic research and social reporting on the topic of data justice. We will lay out key concerns we have about using the preliminary guide in the context of China.
- Chapter III Interviews: We will then report how we conduct the required interviews and the key findings or themes emerging from the series interviews
- Chapter IV Workshops: This chapter will report our workshop design and delivery. The outputs of the workshop and findings from it will be reported.

- Chapter V Discussions and Suggestions: The final chapter will summarize our findings through internal assessment, interviews and workshops then suggest how Alan Turing may incorporate our findings into their research to improve preliminary guide as well as point out future directions of data justice research.

Quick Findings

- In China, data justice is perceived by general citizens as conflicts between consumer and big tech platform or between labor (gig worker) and big tech platform. People in China tend to trust and rely on the government so they are less concerned about the government's use of data.
- Typical Data Justice stories in China involve tangible inequity in consumer rights (i.e. Discriminatory pricing or service) or labor rights (i.e. unfairly controlled by algorithms).
- "Identity" Pillar is the one Chinese people feel less connected to if it's not put in a specific story where people can put them into the shoes of the "weak" side.
- Participation as an approach to address data injustice is believed to have its room to work in China but it will take time to shape China's culture, build capacity of multi stakeholder and probably grow intermediaries who can professionally represent individuals to address data injustice issues.
- In addition to the growing civic voice, China has a strong government which is more capable of taking fast and powerful actions than western governments to effectively regulate corporations on collecting, processing and using data for justice or common good.
- One argument our interviewees make and western countries may learn from is: even though it may violate individual rights, especially those protected by GDPR, the ultimate utilization of data can deliver delicacy governance, which in turn helps protect society's freedom and welfare. This is exactly what we observed during COVID-19 where Shanghai government collects and uses personal data without explicitly asking for consent while the result is it can help narrow the scope of quarantine to as small as a single 20 meter square bubble tea shop and let most people stay in normal life.

About Us

Open Data China is the very first civic group and social enterprise based in Shanghai China working on building up an open digital future. Originally set up as a local group of Open Knowledge Foundation, we primarily worked on open data and incorporated ourselves as Shanghai SODA Data Tech Co,Ltd following the success of operating the SODA program, a challenge-based open data ecosystem program unlocking over 60 datasets and generating over 10 million RMB investment values.

We now focus on three working streams valuing openness: data governance, digital rights and corporate social responsibility. We try our best to research and advocate how to apply openness principles in the above three working streams as well as work with stakeholders directly in frontline to promote and practice 'open'. More information about who we are and what we are doing can be found blog.opendatachina.org

As contracted by CEIMIA to work on the ADJRP project, we set up an internal research team composed of Dr Feng Gao, our director and Liu Jinyi, our researcher to deliver the research activities and write up.

If you have any questions regarding the report or want to discuss further about data justice in the context of China, we can be reached at contact@opendatachina.org.

Chapter II: Internal Assessment

Introduction and Methodology

Internal assessment is conducted to identify possible issues and gaps in implementing the current version of the preliminary guide in the context of China. We conducted internal assessment by following three steps:

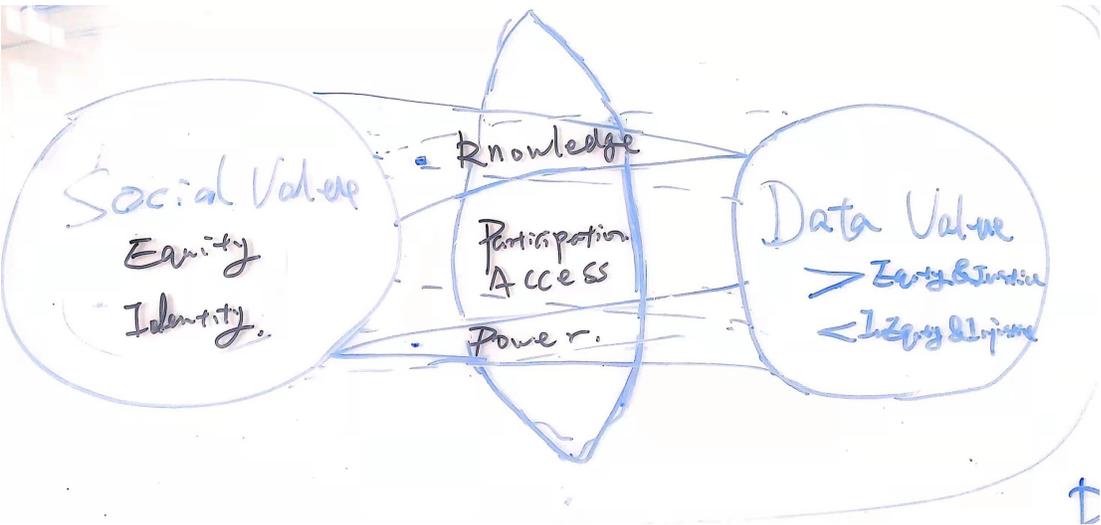
Step 1: Our team reviewed the preliminary guide and completed the online survey required by Alan Turing.

Step 2: Our team conducted desktop research. On the one hand, we tried to find out whether there is any relevant research on data justice in China from the academic field; on the other hand, we paid attention to social hot spots, thus reflecting the hot events of data justice currently focused in China.

Step 3: We synthesized our findings in Step 1 and 2 and decided on the focused list of research questions we need to explore in the interviews and workshops.

Internal Review

We spent a week reading, digesting and discussing the content of the preliminary guide. To help us better summarize the content of the preliminary guide, especially the pillars, we use visual drawing methods. One example below describes our understanding of relationship and interaction of six pillars:



So what we understand from reading the material is that the main idea of data justice is to investigate the process of projecting social values such as equity and identity into data values which are embedded into data innovations. In the middle, we imagine there is a scaling mirror mapping social values into data values. How citizens participate in this process and have access to information affects how this mirror works: either amplify or reduce the influence of existing power structure and dominate knowledge. If data innovations are more towards good and citizens can benefit more from data innovations, then it creates a positive feedback loop to also have impacts on real world society to improve social equity and identity recognition.

Based upon our understanding, we had an internal discussion about the relevance and usability of the six pillars model in China as well as how data justice as a new concept can be introduced into China. We in particular raised three interesting questions:

(1) Is Data Justice a completely new concept to China? How can we associate it with existing terms or ideology people are familiar with locally? And how can we translate it ?

The first question we have is about how we are going to translate "Data Justice" and introduce it into China. Justice in Chinese can be translated to "正义" (closed to the meaning of righteous) or "公平" (closed to the meaning of fair) or "公正"(closed to the meaning of impartial) which is the better option? Can we build it upon existing research or existing practice in social discussion? That's what we expect to further find out in desk research.

(2) When talking about data justice, whether Equity and Identity are the most concerned social issues or social values in China? Whether China has its own focus area for data justice?

Equity and identity in our view come from the context of human rights issues, and in China, such issues may usually be far away from daily life. Although they may become hot spots in the newspaper media for some reason at a certain stage, the long-term women rights, education issues for left-behind children, and the rights and interests of the disabled are not the daily concerns of ordinary people. The problems of gender discrimination and racial discrimination that are often seen in data and algorithmic ethics in Europe and the United States do occur less frequently in China. Therefore, we are more worried about the relevance of these two pillars in China, or how they will be interpreted and understood.

(3) Participation is the key of the data justice pillars model to address and tackle injustice. Can it work in China?

Finally we question the feasibility of Participation in China. On the one hand, we live in a society where people generally believe in and rely on the government, and usually do not directly

participate in any election or social governance, but through representative mechanisms, and are accustomed to a way of life arranged by the government; on the other hand, although the government has tried to increase the proportion and manner of participation in the government governance process under the promotion of more democratic social governance, the interest and awareness of the public may be low due to the long-term lack of participation. It may be even more difficult for participation to occur between citizens as consumers and big tech platforms.

Desk Research

In this section, we provide a short summary of desk research we conducted. The desk research serves three purposes: First, to give you as a reader sufficient background on China's digital environment: how digitized China is, what people's general attitudes towards data innovation and especially China's legal and ethical development on regulating data innovation. Secondly, we summarized what we found in literature review with a focus on investigating existing research on data justice in China and discussing the best way to translate data justice in Chinese. Finally, based upon a dig into social news and literature as well as our own observation, we report several typical stories on data justice or injustice in China.

China : Building a Legal and Ethical Data Innovation Environment

China's data protection regime is in a period of change and as of now The Personal Information Protection Law ('PIPL')¹, Cybersecurity Law ('CSL')² and the Data Security Law ('DSL')³ together constitute the three basic laws on cybersecurity and data protection in China. In addition, there are specific requirements in laws and regulations governing specific industry sectors, such as telecommunications, finance, healthcare, web services, consumer, e-commerce and transportation. In addition, China's Constitution⁴ and Civil Code⁵ address the right to privacy, and there are provisions in the Criminal Code⁶ regarding the unauthorized sale of personal information. In addition, China also has some restrictions on critical industries through the Regulations of

¹ Personal Information Protection Law of the People's Republic of China China § 2021.
<https://www.dataguidance.com/legal-research/personal-information-protection-law-peoples>

² China Cybersecurity Law China § 2017.
<https://d-russia.ru/wp-content/uploads/2017/04/China-Cybersecurity-Law.pdf>

³ Data Security Law, China § 2016.
<https://www.dataguidance.com/legal-research/cybersecurity-law-2016-unofficial-translation>

⁴ Constitution of the People's Republic of China, China § 2004.
<https://www.dataguidance.com/legal-research/constitution-peoples-republic-china>

⁵ The Civil Code of the People's Republic of China, China § 2021.
<https://www.dataguidance.com/legal-research/civil-code-peoples-republic-china>

⁶ Criminal Law of the People's Republic of China, China § 1997.
<https://www.ilo.org/dyn/natlex/docs/ELECTRONIC/5375/108071/F-78796243/CHN5375%20Eng3.pdf>

Security Protection of Critical Information Infrastructure⁷, as well as Draft Measures on Cross Border Data Transfers on cross-border data transfers⁸.

The name of the law	overview
The Personal Information Protection Law (PIPL)	The PIPL establishes the mechanism of personal information protection in China. It introduces several important concepts, such as personal information, sensitive personal information, and processing. It explicitly stipulates its extraterritorial jurisdiction, and provides the traditional elements for data protection, such as principles of personal information processing, consent and non-consent grounds for processing, cross-border transfer mechanisms and rights of data subjects.
Cybersecurity Law (CSL)	The CSL contains personal information protection requirements which are applicable to all enterprises that operate a computerized information network system. The CSL is the fundamental law regulating cyberspace, focusing on multi-level protection of cybersecurity, the protection of critical information infrastructure, cybersecurity reviews, and inspection as well as the certification of key network devices and special cybersecurity products.
Data Security Law (DSL)	The DSL is the fundamental law for data security, and it designs a series of policies – including those regarding data categorisation and classification, data risk controls, contingency responses for data security, data security reviews, export controls and anti-discrimination – to ensure data development and use, as well as industry development. The specific rules for implementing these policies are expected in the future, and may include supporting laws, regulations, and

⁷ Regulations of Security Protection of Critical Information Infrastructure, China § 2021. http://www.gov.cn/zhengce/content/2021-08/17/content_5631671.htm

⁸ Draft Measures on Cross Border Data Transfers, China § 2021. <http://www.djbh.net/webdev/web/HomeWebAction.do?p=getGzjb&id=8a818256641b29b90164409250320021>

	guidelines.
--	-------------

Chinese officials will also actively take action on data compliance and privacy protection through non-legal regulations, for example, in the first half of 2021 the Guangzhou Market Supervision Bureau revealed that in order to actively explore the problem of "Big Data Discriminatory Pricing" (大数据杀熟)⁹, the Supervision Committee jointly held a guidance meeting with the Commerce Bureau, after which, 10 Chinese leading Internet companies, including Jingdong and Meituan, signed the Platform Enterprises' Commitment to Maintain Fair Competition Market Order. Similar examples include China's National Computer Virus Emergency Response Center (NCERT) in late 2021 issuing news of privacy non-compliance for 17 apps and requiring these apps to be rectified¹⁰ and the Ministry of Science and Technology in 2021 issuing a non-mandatory guidance document recommending AI companies to set up technology ethics review committees¹¹.

On the enterprise side, some leading Chinese companies are gradually starting to focus on the social value of innovation and the ethical norms of data. Tencent, for example, held a multi-party research and action platform, Tech for Social Good¹², in Beijing in 2018. The platform aims to invite the government, the business community, academia, the general public and the media to stay aware of the changes brought about by new technologies and to guide technologies and products to amplify the goodness of humanity.

Data Justice: A new concept in China

There is a small amount of research on data justice in China, and is more focused on philosophical research and legal research. When searched using China's leading search engines for academic papers, Wanfang (万方) and Zhiwang (知网), the number of papers with the phrase "data justice" ("数据正义" and "数据公正") is relatively low, with 26 papers. Although some papers contain the keywords of data justice, they actually refer to the application of big data technology

⁹ Bytes develop independent e-commerce App, benchmarking Taobao; Meituan and other ten platforms promise not to kill big data. (2021, April 09). Houchang Morning News, A14.

¹⁰ 17 apps such as Hello Travel and Tesco have privacy non-compliance Violating users' privacy while precision marketing. (2021, December 21). Financial network.

¹¹ Ministry of Science and Technology: artificial intelligence companies may need to set up science and technology ethics review committee. (2021, July 31). Southern Metropolis Daily

¹² Tech for Social Good. (2018). Retrieved February 21, 2022, from <https://baike.baidu.com/item/科技向善/22493758?fr=aladdin>

to make social justice easier to achieve in the rule of law. In addition, few of the papers on data justice discuss specific cases of algorithmic discrimination or digital exclusion, but more often discuss data justice from an ethical or jurisprudential perspective.

Two noteworthy and major studies on data justice in China will be presented below, one of which explores the roots of data justice and the other proposes another term similar to data justice, namely computational justice (计算正义)¹³, and discusses it jurisprudentially.

Fu argues that data injustice is caused by the injustice that exists in society itself and argues that if data injustice is to be addressed, it should also be addressed from the root causes first in his paper¹⁴. Fu raises the following phenomena of data injustice: first, the inequitable appropriation and distribution of data resources, i.e., the vast majority of data resources are now monopolized by governments and corporations. Second is the digital divide, including the digital divide between developed and developing countries, and the digital divide between urban and rural areas. Third is algorithmic discrimination caused by different causes. Fourth is social exclusion, which is mainly manifested as structural unemployment caused by the development of AI technology, violation of human privacy, fracture of social structure and social stratification. Fu believes that these phenomena are rooted in the uneven development of the digital economy and thus have an impact on social justice issues, and believes that the unregulated development of smart technology is also a key factor. As to how to solve the problem of data injustice, Fu believes that data justice can be achieved only after distributive justice is realized, and thus solves the technical problems and institutional design that exist on this basis, and he also believes that promoting judicial justice will also help to promote data justice.

Zheng proposes the term "computational justice" to explore the relationship between law and algorithms. He argues that algorithms are not law, that their functions should not be exaggerated, and that he does not suggest that law should be used only as a regulatory tool. He argues that "computational justice" can be used as a principle to understand how law receives the impact of computational decision-making mechanisms and to explore how algorithms can be used for good under this principle. He defines computational justice in two dimensions, one is value-based, i.e., algorithms should not be denied because of the privacy and discrimination problems they generate, but rather a fair solution should be found. The second dimension is institution-based, he argues that algorithms should be used in specific application scenarios to achieve goodness.

¹³ Zheng Y. Computational justice: a jurisprudential construction of the relationship between algorithm and law. *Politics and Law*, 2021(11):91-104. DOI:10.15984/j.cnki.1005-9512.2021.11.008.

¹⁴ Fu, Z. Exploration of social justice issues in the age of intelligence. *Changbai Journal*, 2020(05):60-67. DOI:10.19649/j.cnki.cn22-1009/d.2020.05.009.

The Popular Data Justice Cases

Big Data Discriminatory Pricing

The public is very concerned about cases related to economic damage. A good example of this is the previously highly publicized lawsuit against Pinduoduo Price Chopper and the comparison of taxi-hailing software. The former is Liu Yuhang found that the bargain-hunting software Pinduoduo, no matter how many times the share reproduction bargain-hunting turned, still did not reduce the price as its software promised, so he was angry and sued Pinduoduo. This matter has aroused much attention in China and public opinion is overwhelmingly in favor of Liu Yuhang and believes that the behavior of Pinduoduo is fraud. The latter is a professor from Fudan University, Sun Jinyun, who conducted more than 800 experiments with taxi hailing software in different cities and found the existence of big data killing (大数据杀熟). For example, many taxi software will call more expensive cars for Apple phone users as well as give them a lower discount. Both of these cases have attracted a lot of attention, which shows that the Chinese public cares a lot about cases involving financial interests.

Health Passport

The Chinese public has shown a more accepting attitude toward possible privacy violations and the use of technology even though they are lived with a much larger and more granular scale of surveillance. Take the contact tracing code during the epidemic as an example. The Western media and the public are more skeptical about it. For example, some British media said that the digital surveillance during the epidemic was undoubtedly deprive some of the basic freedom enjoyed in 2019¹⁵. Whereas China, as a country that has applied surveillance including the all five areas of pandemic digital technology mentioned by Whitelaw et al, track disease activity in real time; screen individuals and populations; identify and track individuals who may have been in contact with infected individuals; identify and track infected individuals and implement quarantine; monitor clinical status the Chinese are also generally more accepting of the privacy invasion of COVID-19.¹⁶

¹⁵ Politics.co.uk staff (2020) The death of privacy: What happens to the covid-tracing apps after the virus? Available at: <https://www.politics.co.uk/comment-analysis/2020/06/01/the-death-of-privacy-what-happens-to-the-covid-tracing-apps-after-the-virus/> (Accessed: 17 March 2021).

¹⁶ Alsan M. et al. (2020) Civil Liberties in Times of Crisis. Cambridge, MA: National Bureau of Economic Research

Delivery Workers Trapped in the Algorithm

The issue of labor and algorithms is significant. The most famous example is the "Delivery workers trapped in the algorithm". Primary food delivery software platforms such as Meituan (美团) and Eerme (饿儿么) use a lot of objective data and rider behavior data to make predictions and improve rider efficiency, and can cause riders to go to extreme lengths to complete their algorithm-designated food delivery times. Riders thus face the results of illegal traffic laws and even casualties, with nearly 10,000 rider violations of traffic laws and 155 casualties during a seven-month period in just one city, Chengdu.

Chapter III: Interview

As required by the ADJRP team, one of the primary research activities we should do is to conduct a series of semi-structured interviews with interviewees who fall into one of the categories: policy maker, developer and impacted community. The purpose of the interview is to further engage stakeholders in reviewing the preliminary guidance and give feedback on its relevance and usability in the context of China.

Selection and invitation of interviewees

During the research period, we were able to successfully invite 12 interviewees to participate in our 1-hour long semi-structured. The anonymized demographic information about the 12 interviewees is listed below in Table 1.

Interviewee No.	Description	Stakeholder group	Gender	Age Group
1	A program lead on urban digitization of International organization	Policy Maker	M	30-40
2	A Public Governance Professor currently working in Denmark and previously lived and worked in China	Policy Maker	F	30-40
3	An Expert at state-owned think tank advising gov on data and AI regulation	Policy Maker	M	40-50
4	A Communication Professor who has worked on gender bias in AI era	Policy Maker	M	30-40
5	A researcher working for NGO specialized in Digital Policy in HongKong	Policy Maker, Impacted Community	M	30-40

6	A law professor leading human rights center	Policy Maker , Impacted Community	M	40-50
7	Founder of a Foundation focusing on digital rights and worked on gender bias and woman rights in AI era	Policy Maker , Impacted Community	M	40-50
8	A Senior program lead at a well-known NGO working on education rights of mobile population	Impacted Community	M	30-40
9	A Specialist at NGO working on Digital/information Accessibility	Impacted Community	F	30-40
10	A founder of Environment NGO who worked as software engineer and manager before	Impacted Community , Developer	M	40-50
11	Data Science Lead at a Foreign Market Consulting Company who also worked for a state-owned large AI and data company before	Developer	M	30-40
12	Data Consultant for city government in Taiwan who has city planning and design background and living and working experience in mainland	Developer	M	30-40

All interviewees are either representatives of policymakers, developers or impacted community or as proxy whose job is relevant to policy making, data innovation or campaigning for impacted community (such as NGO). It is also important to note that the impacted community can be loosely defined as it can refer to users of data innovation in the context of China. When selecting interviewees, we also tried to be as inclusive as possible: Geographically, we cover interviewees from mainland but also from Taiwan and from Hongkong whose life experience and values could be very different from mainland; Professionally, we invite interviewees from organizations that are large versus small and state-owned versus foreign.

All interviewees were contacted via message app (i.e. Wechat in Chinese) or email and sent a written invitation detailing: (1) the background of the research including who we are, why we are doing this and what we are going to produce at the end (2) a brief summary of the preliminary guide (as described in chapter 1: internal assessment) and a copy of the guide (3) a list of interview questions for reference (please refer to Annex A) (4) the interview will be recorded and their identity will be anonymized in data sharing and final report. Then each Interviewee agreed on a time and interview method (i.e. offline meeting or online meeting) with our team in advance.

Focus of Interview

As we have mentioned in Chapter II internal assessment, we highlighted our concerns about the usability of current guidance surrounding the following points:

- How do people understand and define data justice?
- What type of data injustices cases are discussed by the citizens in China? any difference from other countries?
- Can Participation approach work in the context of China?
- What new perspective or approach can we identify in China and introduce back to the global community?

Protocol of interview

We have two researchers trained and tasked to conduct interviews. One of the researchers is the primary interviewer and is responsible for briefing the interviewee at the beginning of each interview to obtain the interviewee's consent for recording and participation. After obtaining the consent, our team initiated the recording and started the conversation with the interviewee based upon the pre-drafted interview questions.

Please note that the interview is designed as semi-structured interviews and researchers were told to let conversation flow rather than strictly following the question lists.

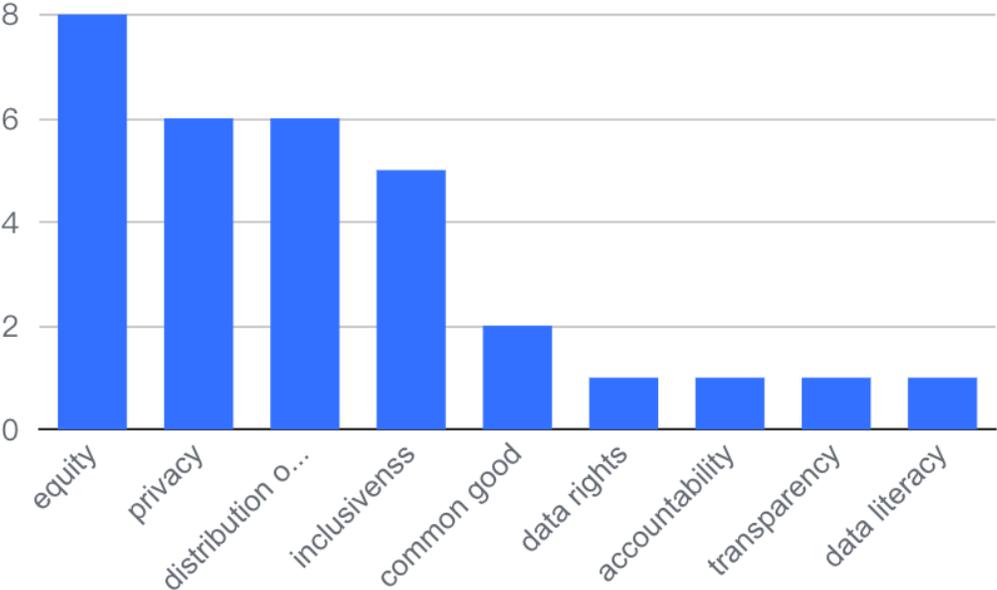
Findings of Interview

Understandings of Data Justice

The first part of the interview asks interviewees to name three keywords, either relevant concepts to data justice or relevant case studies. We here first present an overall statistical analysis of keywords mentioned by interviewees then unpack their interpretation and further conversation.

Overview of Keywords

50% of interviewees mentioned this is the very first time they hear about the term data justice. The rest of them, though not encountering the term for the first time, believe they do not have in-depth understanding of the concept.



When asked to name three relevant concepts or events, "Equity" (N=8) tops the list of keywords and it is followed by "Privacy"(N=6) as well as "distribution of data and its benefits" (N=6). The "Inclusiveness" (N=5) is the third. We also noted that interviewees mentioned "common good/tech for good", "accountability", "transparency" and "data literacy".

Unpacking conversations on Keywords

Equity: human rights perspective versus consumer perspective

When talking about "Equity", we found interviewees often associated it with the particular case about big data discriminatory pricing. In other words, it seems much more related to consumer rights rather than traditional human rights.

In addition, the "Equity" can be understood as fairness in the relations of data production. Interviewee #2 expanded the point and question: "there are different people involved in the process of data production and the Equity is about whether the relationship between them is fair and equal. We need to explore answers to key questions such as Who has the right to control data and who gains the benefit."

Finally Interviewee #9, a frontline expert on digital accessibility, especially for disabled people, warned there is no absolute equity or fairness in reality. And in certain cases, achieving equity for one group may actually create new inequity.

Privacy: the concern of surveillance and chaos on facial recognition

The second toppest keyword is Privacy and Interviewees mentioned it usually also linked it with "surveillance" or specific case on abuse of facial recognition in China. Interviewees did not expand too much on the point of privacy, but we generally get the point for three reasons:

(1) Everyone in China experienced crazy spam text messages or calls at a certain time due to some staff working for mobile companies illegally selling data to black market. Thus privacy is one biggest concern when people hear about data use.

(2) The increasingly installed cameras and facial recognition systems is the second recent event that makes everyone annoyed. Especially when many neighborhood communities ask residents to register face and use face recognition to enter into the neighborhood, that really angers people. And finally the government released new regulations to stop such actions.

(3) The linkage between data and privacy is further enhanced due to the quick development of personal information protection and relevant law in China in recent two years.

Distribution of data and its benefits: rise of interest in individual data rights

In recent years, because of the hot discussion on data monetization and lining up data exchanges, more and more people start developing interest in exploring how can they gain economic benefits from own data including asking big platforms to pay for accessing and using personal data. This is exactly what interviewee #6 raised during the interview that "if say data is used for commercial purposes and especially on how to make people spend more money than economically, whether individuals should be entitled to benefit sharing and how such a system should be designed?"

The new relationship between individual and country or company surrounding one's individual data benefits is exactly the new relation of product as interviewee #9 mentioned earlier on how Equity plays a role in fairly dealing with these relationships. #9 also linked the idea with the concept "welfare state" as she currently lives in Denmark and posed a question "whether there should be new relationship defined between citizens and the country on (sharing data benefits) "

But the discussion can also be beyond simply individual interest because how data as important resource is distributed also may transform the relationship (or let's power structure) between government and companies as interviewee #3 pointed out and interviewee #7 further commented that " due to the nature of data, it is easier to create an oligarch than ever before. It is clear that this is a global challenge and it creates new differences and divergence between big tech platforms and governments.... So we are seeing the US Biden administration working on antitrust and the same thing is happening in China. "

Inclusiveness: a conversation around digital divide

People who are more familiar with human rights issues mention inclusiveness to call for a more diverse but also inclusive society. Interview #9 mentioned the example that most Map Apps by default use a female voice for navigation guidance while never thinking about the fact that women are not born to serve others. The example is used to warn that we should be careful about what data may suggest, while the reality is far more diverse than what data depicts. Furthermore, #9 also emphasized that inclusiveness should mean tolerating differences and creating room for diversity instead of trying to integrate differences to become new one and get rid of diversity.

Interviewee #12, as a data scientist himself, interpreted the inclusiveness from a technical perspective:"always try to capture the full sample reflecting the whole picture". He later expanded the point by sharing a specific case on how a small county in Taiwan got bad reputation during the Pandemic due to non-inclusive collection of data. "You only deploy sensors or human sensors to places where the population is much larger due to limitation of resources. So places like the county can only receive little resources for data collection which can not reflect the whole picture and then can lead to simple but wrong label result. This is what my friend coins as 'sensing bias' ".

Reflection: Is it really about data?

During the interview process, we also noticed interviewees who work in social issues or human rights issues commonly raise a point that in their view, justice and data are separate and independent. When asked to name three keywords, interviewee #8 mentioned:" I don't think I have keywords, but rather simply break data justice down into two separate words: justice and data. In my view, data is neutral, but how justice is defined is the problem. The decision-maker's view on justice determines how data is used and whether the data innovation does good or not. It is important that we can reach consensus on what justice means in a specific social context." However, as pointed out by interviewee #4 the narrative on justice could be quite dynamic thus "it requires a process of negotiation to determine what justice means as there is no objective

standard but only based upon subjective understandings. Also, we should notice views on justice can change over time"

" Data justice should still be about how we achieve justice itself Or let's say achieve common good" Interviewee #6, a human-rights law expert, further added: "No matter how datafication the society will be, it is important to use data to stop injustice and unfairness or at least not make things worse. and if it even can improve the situation or fill gaps then it would be ideal."

Reflecting upon the role of data in achieving justice, interviewee #10 who works on using data for environmental issues, pointed out that "Data is just a tool and how to use it depends on who use it and what values the person holds. In China I believe it is our government and the communist party who defines what justice means and points the right direction for data use. "

In the end, it might be like what interviewee #1 pointed out that data justice can be divided into two groups: the first group is about how data itself is justice or injustice and the other group is how data helps amplify or reduce social (in)justice.

Data (in)Justice Stories in China

When asking interviewees to mention tangible examples or stories they think are relevant to data justice, the most common stories are the big data discriminatory pricing (on the point of equity) and abuse of facial recognition (on the point of privacy).

Interestingly, interviewees also mentioned several new stories they think represent data justice or injustice. We summarize them here for reference:

Invisible Groups in the Population Census

The population census is conducted every 10 years in China. Two interviewees mentioned how their community got left out in the census and became invisible in the statistics. One group is LGBT whose marriage is not legally recognized in China and thus for two gay partners, they have no option to express their marriage status in the Census as it only provides three options in the survey listing: married, single and divorced. This resonates with what "Equity" and "Identity" pillars point out.

Another group impacted by the Census system is the mobile population who leave their home town to big cities for work. Children who move from one place to another with their parents are currently mis-recorded in the Census system for various reasons including technical reasons and other social governance reasons. And because such mobile populations are not reflected in the local Census statistics, education and health resources are not well allocated to serve for those populations and cause new social problems.

Cheating Environmental Sensors

Another interesting story shared by interviewees is on how some companies and even local governments try to cheat environmental sensors to create fake data and in turn create fake environmental justice.

There were reported many different ways to cheat sensors. One story is on a highway, the sensor used to monitor the dust index here collects data regularly every day at 10am and 2pm. The local government thus deliberately sends cleaning vehicles to drive to the corresponding road section 10 minutes before 10am or 2pm and spray cold water to temporarily suppress the dust, thereby reducing the sensor reading. A similar story is that a company uses ice water to irrigate the emission monitoring sensors of the production chimney, thereby creating false data and deceiving government entities that remotely monitor the corporate environment through data.

Can Participation work in China?

One of the key research questions we want to explore is whether Participation this pillar or approach can work in China? And how?

In our interview, we found quite diverse opinions and perspectives on the participation approach, but all seem to be positive, especially on increasing people's participation in regulating big tech platforms.

Speaking about why participation matters, interviewee #12 who is from Taiwan stated that "no matter how participation works, it is important to have such a mechanism in place to build up trust with citizens and let them know they have the right to participate." In addition, Interviewee #1 talked about individual rights to control and participate from a perspective of Web evolution: "we are seeing the rise of Web3.0, of which the decentralized governance of data resources is a key idea", he further expanded that, "... Allowing users to participate and be engaged in the process of sharing data benefits is exactly aligned with the 'Towards Common prosperity by 2035' agenda currently promoted by the central government, which calls for a more fair and even distribution of welfare."

It seems then that "Participation" is essential but do we have foundations in China to make it work? Our conversations with interviewees then further build upon this question and depicts a much clear picture about how Participation may work in China.

Foundation: Legal and Culture Environment is being improved

Interviewees analyzed how China's legal and cultural environment is suitable for the Participation approach and how the environment is being improved to enable Participation.

Strong Legal Foundation

When further unpacking government's regulation and legal development on driving data for good or justice, interviewee #6 emphasized that "A bottom line should be clearly defined by the government on what we can do and can not do with data. Since 2003 or 2004, the UN started holding international corporations responsible for human rights and passed a set of guiding principles for business and human rights. Correspondingly, corporations have set up its own internal policies to regulate their own work to comply with international standards..... And even have human rights impact assessments to ensure business activities meets minimum standards set by the UN."

In addition to the bottom line, interviewees also suggested the importance of regulatory policy in enforcing the law. As explained by Interviewee #10 "In China we have a big government and it means we should have a strong regulatory environment We should have laws to fine injustice and make corporations feel hurt. They will change their behavior only when they and their investors really lose money".

The last point raised by interviewee #9 is on how to keep government regulatory policy aligned with what is promoted by the wider non-profit sector and industry. She gave an example on pushing forward the accessibility agenda in China:" Last year our government issued a new policy to push corporations to create accessible services and websites for elders. This "design for elders" movement is strong now in China but it is strange that it gets separated from the traditional accessibility agenda and creates new problem during implementation." she further added:" It is important that government policy allows room for interpretation and flexibility so corporations can better work on it."

Shift Social Narrative

Strong legal development and enforcement is one side of the story, while the other side is shifting the narratives to how business should be done and what customers and investors value.

Interviewee # 1 and #10 both mentioned the example of how ESG (Environment, Social and Governance) is promoted and adopted by corporations. "The promotion of ESG first is just a civic demand and then is promoted and accepted by the media and investors, so it creates an atmosphere where corporations must follow up the standard and take responsibility. " shared by interviewee #1, ".....Now it is the same. We should have such a culture promoting principles of open,transparent and fairness for data use. "

Big Techs in China are aware of the ever growing demands from citizens and governments on regulating their use of technology, especially data and AI. As we reported in the internal assessment, Tencent, who runs wechat and dominates social network service in China ,already announced to change its corporate strategy to "Tech for Good" and expect to spend 50 Billions RMB during the first phase.

"We observed the trends of taking action on 'Tech for good' by Big Tech firms since 2020 when the Ministry of Science and Technology pushes the new wave of setting up internal ethics committees in tech corporations..... Understanding of the new policy and actions taken are quite diverse across corporations..... Some create new positions to push tech ethics internally and organize internal training for staff, especially including courses on social justice and issues such as gender equity, while some are more conservative and need more time to take actions." Interviewee #7 shared what they observed during interaction with big tech firms, and further added:" It maybe takes another one or two years for local tech firms to realize its their social responsibility to ethically use data and use data for social good. And also they will be more driven and motivated by customer's requirements and demands rather than are forced by government regulations."

Scaffolding: Capacity building and Intermediaries

Speaking about how to enable more participation, it is commonly agreed that both those who consume data innovations and who produce data innovations should build their own capacity. It is also noted that when individuals have no capacity or energy to participate, then they will need professional intermediaries such as NGO to represent and help them. And there is also a trend to create new types of organizations such as data trust to represent and help individuals to defend their own rights. We unpack our conversations with interviewees on the following three sub-points:

Ethics and Tech for good: A New Lesson inside Corporations

To enable corporations to adopt the approach of participation, one of the important elements is to develop internal policy and have engineers and other relevant staff trained on the topic of data ethics and data for good so they have the awareness but also capacities to implement participation. When talking about this kind of capacity building inside corporations, we spoke with interviewee #10, #11 and #12 on their view of how to achieve capacity building inside different types of corporations for data justice.

Interviewee #11, who currently leads a data team in a foreign consultancy firm, first reflected on what they are currently experimenting:" Having our data scientists check and make sure our process of data collection and analysis sticks to the values of non- discrimination and data for common good is also quite new to us. At the current stage, we simply have a checklist and make sure it's part of our standard workflow. " Then he raised the difficulties in standardizing the process: " However, as data innovation is quite unique in every case, we should deal with such checklist also case by case and there is no one size for all. What we can achieve now is to make it a standard requirement for all our data scientists and make sure they are aware of such ethical data and data for good requirements."

"It is important to cultivate the sense of whom your data innovation serves for..... This is the foundational question one should ask when thinking about making data innovation or design innovations. " Interviewee #10 commented on the abilities he thinks engineers or product managers should have while making data innovations:" It is also important to make sure the team

shares the same values on what means good. The core algorithm or analysis of data is especially important and should be built on top of shared good values."

Interviewee #12 further reflected on:" It is noted that for marketing purposes the use of data is often exaggerated. It is also important to hold our marketing staff accountable and challenge their daily narratives on magic of data."

Finally interviewee #10 concluded about what types of corporations are easier to build internal capacities, he said:" I believe state-owned companies and listed corporations are better in building up internal capacity. State-owned companies may take more social responsibilities as it's owned by governments while listed companies value their own brand and public image so they are more easier to care about their own staff capacities in using data for good."

Building Data Literacy : Possible but also Difficult

When asked what they think prevents "participation" works, interviewees often mentioned the fact that data innovation is quite technical and "even you are engaged and given the opportunity to share your thoughts but whether it is something useful for technical people to correct the bias remains a question. And even worse, we are told now that algorithm works like a black box and you may only understand whatever it outputs then you have no idea about how to contribute to improve or correct it. " (as pointed out by interviewee #6)

Then is it possible to build up individual users' capacity? It is possible but also face real challenges as pointed out by interviewee #9 that "some (disabled people) really desire to participate, but the fact is almost 80-90 percent of them were not given enough opportunity to receive higher education.....and it becomes even harder for them to understand how data works" Interviewee #2 added that:" on the one hand, it is about knowledge; on the other hand, it is also about having the opportunity to experience data innovation. If the person has little access to data innovation to get live experience, then the person has no ways to contribute and participate.... (In conclusion) one's chance to participate depends on one's knowledge and experience of data innovation.``

A Demand for Intermediaries

Building on top of the difficulties of building ordinary people's capacity, it is well pointed out by interviewees that we'll need professional organizations and groups to represent individuals to participate in the process of data governance.

Interviewee #3 commented that "Participation is for third-party organizations who have data capacity. Education is important, but it is also noted that only professionals can make real contributions." And what can real contribution mean? Interviewee #9 pointed that such organizations should not only represent the users and impact communities but also have the capacity to translate different languages used by different groups i.e. Speaking technical terms

with data innovations experts while using simple terms to communicate with users. It is also important for such organizations to promote the idea of data justice and encourage wide participation. interviewee #2 emphasized that such intermediaries also need to be able to balance the interest of multi-stakeholders to maintain the stability of the triangle. If intermediaries lose trust from any party, then it will make things even worse.

Interviewee #8 further expanded the role of such intermediaries: "we also need such intermediaries to offer technical help to discover or identify problems in data innovation as it's quite professional work like hackers. So it becomes a triangle relationship: professional organizations identify problems and engage impacted communities in defending own rights with data innovation providers. " he further added: " and speaking about evidence ,such as in the case of big data discriminatory pricing, it is impossible for a customer to find real evidence to prove it while a professional organization has the capacity to dig into the issue and obtain real evidence to prove it is inequity. "

Then do we have any existing intermediaries? Or do such intermediaries have the right opportunities and capacity to participate on behalf of individuals? Interviewees seem to have mixed views on it but generally it should be quite dependent on what type of social issues you are working on.

Finally, one interesting comment from interviewee #2 questioned why we don't explore new types of professional groups in handling data rights: "Can we have a new type of organization? The recent discussion in Germany is building up so -called data co-op or data trust. And it is imported into Denmark now.the power structure is unbalanced between individuals and big tech firms so we should need a collective group to coordinate interest among individuals and defend their rights collectively."

What can China offer to the world on Data Justice?

We deliberately asked interviewees what they think are unique or special in China on the topic of data justice that may be useful for western countries to reflect upon or draw a lesson from.

One common feature pointed out by interviewees about China is "Big Government". China has a history where people trust and rely on its government, so the government becomes quite strong and acts like a parent to rule everything about the society.

Interviewee #12 mentioned that "Big government tends to prepare very well before taking actions and that is why big government is likely to utilize big data way much better than so called democratic government and tends to collect as much data as possible." several interviewee commonly mentioned the example of health passport app used by chinese government which does not explicitly ask for consent to use personal sensitive data but it is able to help shanghai local government to narrow the scope of quarantine and only list the 20 meter square bubble tea shop as a medium risk area. And it seems much more justice than quarantining everybody in that district or block.

Interviewee #1 then argued and raised an interesting point that the Chinese government, under certain emergency public events, may ask citizens to compromise individual privacy rights which could be quite unacceptable under the spirit of GDPR. However, if it in exchange results in better data-driven governance and help government to minimize its quarantine actions and guarantee freedom for a wider group of people then it should be called data justice. This may be a good example for western countries to rethink how to balance privacy protection and wider social benefits.

Chapter IV: Workshop

Following up the interview and being required by Alan Turing, we ran the required workshop engaging participants in further reviewing the preliminary guide with group discussions focusing on specific case studies.

Workshop Task Design

The workshop is designed to explore the usability of the preliminary guide further through collaborative discussion and analysis by groups of participants. As we reported in internal assessment and interview, it seems that rather than focusing on specific gender group, social class group or racial group, the so-called impacted community in China could be as general as possible to include anyone (i.e. user) because data and digital innovations are much more widely embedded into everyday life in China. So when thinking about the identity of participants, we decide to define the potential participant we are going to recruit as an impacted community or general user. And of course, it is also easier for us to attract and recruit enough participants for an hours-long workshop.

Workshop activities are designed based upon the engagement workbook shared by the Alan Turing Institute with the aim to further explore issues we identify at the internal assessment and interview stages. Four tasks are finally designed for the workshop and simple summaries are described below in the Table. The final workshop worksheet for print is available as Annex B for reference. More detailed content about each activity along with the workshop outputs are reported in the Results section.

	Activity Content	Activity Form	Activity Purpose
Task 1	List Three Keywords associated with Data Justice	Individual work then group discussion to reach consent for group result sharing	Further Confirm or Contrast findings in internal assessment and interview
Task 2	Examine the relevance of a list of case studies and Contribute new case studies	Group discussion to review existing case studies and contribute new studies if any	Review case studies covered by our internal assessment, interviews and Preparatory Material and solicit participants views on relevance of those case studies to Data Justice

Task 3	Assess the Usability and Relevance of Pillars in the context of China	Group discussion to review materials of Pillars summary and reach consent on the usability and relevance of pillars given the context of China in the form of ranking score	Solicit overall feedback from participants on the usability of Pillars
Task 4	Assess the Usability and Relevance of Pillars of Particular Case Study in the context of China	Group discussion to review materials of Pillars summary and Questions and reach consent on the usability and relevance of pillars to the case study selected by the group in the form of ranking score	Solicit particular feedback from participants on the usability of Pillars in the context of a particular case study as identified in internal assessment

Workshop Protocol

The workshop is designed for 3-hour long and is conducted by following the protocol as described below:

(1) Participants are organized into groups with 5-6 people per group. Each group is given a workshop sheet (A3 size double sided and instructions are already printed on it), a package of pens, 5-6 copies of preliminary guide with summary of pillars in Chinese, white papers for note taking and 2-3 packs off colored sticky notes and a white board.

(2) The facilitator opens up the workshop by giving an overview of the workshop and explaining to participants how their group work will be used in the research work.

(3) Groups are asked to individually do Task 1: thinking about three keywords associated with the concept of "Data Justice" then summarizing group findings on the workshop sheet. The task is given 5 mins for individual work and 10 mins for group discussion.

(4) Groups are asked to share their keywords and group discussion. Each group has 5 mins.

(5) Facilitator shares the research done by Alan Turing and early findings from internal assessment and interview on the definition of data justice.

(6) Groups then are asked to spend 10 mins on task 2: checking a list of cases whether they are related to data justice or not. Then groups are given a further 10 mins to write 1-3 additional cases they may know.

(7) Facilitator then walks through the list and asks for any disagreement. Groups then are encouraged to share additional cases.

(8) Facilitator further gives a 15 mins talk on pillars identified by Alan Turing Institute.

(9) Groups are asked to do Task 3: assessing the relevance and usability(feasibility) of each pillar in the context of China. The task is given 15 mins for group discussion and each group should have one person reporting their conclusion and discussion within 5 mins.

(10) Facilitator then introduces three classical case studies identified by the research team and asks each group to spend 25 mins completing task 4 which is composed of 4 sub-tasks.

(11) Each group is finally given 10 mins to share their result and discussion of task 4.

(12) Facilitator wraps up workshop discussion and gives participants information about how to follow up research progress and when to expect final outputs.

Workshop Delivery and Participants

Due to the COVID-19 situation, it posed challenges in organizing the workshop and securing attendance of participants. We finally had to run the workshop twice to satisfy the requirements of having 25 participants. In the end, we had one workshop with 11 participants (all offline) and the other with 16 (6 through online methods and are assigned into one group).

Participants are considered as General User or Impacted community in our workshop but we also asked each of them to self-identify themselves in the registration form. Overall we have 27 participants (5 groups), with 56% are female (N=15), 100% are university-educated (N=27) and 52% are Master or above(N =14 and 6 are PhD)| 29% participants (N=8) additionally self-identify as Developer and 15% consider themselves also as Policy Maker (N=4). Participants' occupations are quite diverse covering both non-tech jobs such as journalist, gender-focused social worker, art student as well as tech jobs such as data analyst, data-focused entrepreneur, and tech-focused lawyer. 19% participants (N=5) reported that they heard about the term Data Justice before the workshop.

Workshop Results

In this section we report the results or outputs of each activity based upon the written workshop sheets collected back after workshops as well as notes taken down by the event assistant during workshops.

Task1: Three Keywords

The 5 groups reported quite diverse keywords: the most mentioned one is transparency (4 out of 5) and then the second group (N=2) is composed of "Privacy", "Equity" and "Distribution of data and benefits". Finally, Power(N=1) and Participation (N=1) were mentioned.

It is a bit surprising to us that "Transparency" emerges from group discussion as the top one keyword which is not the case in interviews. The "transparency" keyword can be further unpacked to cover two points: transparency of algorithm(or data use) and transparency of who is using data. The groups mentioned their shared concern about being blind in the digital age while data-powered algorithms seem to control their life. It is vital ,as pointed out by participants, to establish new mechanisms to allow users to access more information about who has access to personal data and what has been done with data.

In the second group of keywords, "Privacy" and "Equity" again are associated with stories on Facial Recognition and Big Data Discriminatory Pricing. One common thing as we observed in the interview is only people who worked in the space of human rights may link data justice with specific social justice issues such as women rights.

"Distribution of data and benefits" in workshop discussion is primarily associated with the recent data monetization progress in China especially Shanghai just launched its own data exchange. People who brought it up work in data-relevant capacity so they have interest in exploring more on whether individuals can benefit either socially or economically from their own data and then question whether they are entitled to share profits from data exchange.

Task2: Case Studies

A list of pre-selected stories based upon our internal assessment and preparatory material is printed on the workshop sheet with a summary or a title and the workshop facilitator walked the workshop participants through the list by unpacking each stories. Then the groups were given time to collaboratively review which story is not relevant to data justice based upon their group understanding.

The stories with the overall results are summarized below:

Overall Results	Summary or Title (printed on the sheet)	Story Narrative
All Agree it's relevant	#1 大数据杀熟 big data discriminatory pricing	Big data discriminatory pricing is one of the common cases we identified in the internal assessment as well as interview. One of the famous stories is based upon a study conducted by a Fudan University professor whose team tested the price estimated by the Didi platform for the same route and same journey on different branded phones and it turned out the price is significantly higher if you call the taxi on the iPhone. More information about the story can be accessed: https://kr-asia.com/researchers-took-over-800-trips-using-chinese-ride-hailing-apps-heres-what-they-found
All Agree it's relevant	#2 困在算法中的外卖员 Delivery workers who are trapped in the algorithm	The story is one of the hottest social news in 2020 discussing how platforms use algorithms controlling the speed and efficiency of food delivery by its workers without considering the complexity of reality and showing care and tolerance to its workers . More information can be accessed at: https://chuangcn.org/2020/11/delivery-renwu-translation/
All Agree it's relevant	#3 利用算法筛选简历 Using algorithm to make auto-decisions on screening candidates	A recent trend is that algorithms are employed to automatically screen candidates during the hiring process. And it is reported that in some cases females are discriminated against for certain jobs.
2 voted it is not relevant	#4 政府开放土地数据使得富人更快占据土地 Rich people can take	The story is based upon the Bhoomi India case covered in preparatory material on the point of "open data under conditions of unequal capabilities"

	<p>advantage of open land data to control more land resources</p>	
<p>3 are unsure about it</p>	<p>#5 人口普查中没有相应的LGBT群体的真实取向和婚姻状态的选项</p> <p>No right marriage status options for LGBT in China's Population Census</p>	<p>In China's Population Census it is only given three options for marriage status: single, divorced and Married. It is also noted that Gay marriage is not legal in China. Therefore there is no right option for cases such as two gays staying together as Partner (not married though). Then LGBT groups are kind of invisible in the final output of the Census.</p>
<p>All Agree</p>	<p>#6 健康码自动归集使用个人数据自动计算健康状态</p> <p>Health Passport app uses personal data to calculate your current health status</p>	<p>During COVID-19 Pandemic, Chinese government runs a data-powered health passport app to use personal data including your GPS and mobile data to automatically calculate whether you are safe or health and shows your status based upon a 'green-yellow-red' scheme. The app does not explicitly ask for your consent and there is no opt-out option.</p>
<p>All Agree</p>	<p>#7 通过透明自治的数据合作社是的临床数据得以被研究人员安全使用加速新药研发</p> <p>Establishing a transparent and participatory data coop to allow secure patient data access to facilitate the development of new drugs.</p>	<p>A Data Coop or Data Collaborative is a way for individuals to collectively govern and make decisions on how their data should be accessed and used. One example is to make patient data accessible and usable by drug companies securely through such data coop and allow individuals to also benefit from such data donation.</p>

Looking at the results ,it is no surprise to us that people can reach positive consensus on #1, #2, and #6 which are also identified in our internal assessment and mentioned multiple times by interviewees. The cases #3 and #7 include examples reported in western countries and interestingly all groups agree they are relevant cases, though some of them did mention such examples may never come to their mind if it is not listed here.

One interesting discussion on #1 is one group believe sometimes personalized prices based upon big data can also mean good. The group shared their thoughts that the problem was not about using big data to personalize the price but who benefits from the personalization: the customer or the platform? If the platform can personalize the price to allow people who are poorer to get cheap taxi prices, then is it a good thing?

Regarding Case #4, the groups who voted false explained that they do not think it is a problem of data justice because social resources such as education or natural resources like water are also widely made accessible to all citizens, but there are also differences in taking advantage of those resources among different groups of people due to capacity or other reasons. Those groups thus consider Case#4 should not be considered as a data justice problem but rather a more general social justice issue.

The same rationale also was applied to the judgment in Case #5 where the groups who voted false believe it is a problem caused by social justice and the way to address it should be more focused on changing the social system than focusing on data .

Task3: Relevance and Usability of Pillars in the context of China

In Task3, the main output we required is to ask each group to discuss and then assess the relevance of each pillar and its usability in the context of China. The assessment is done by assigning each pillar with two numerical scores ranging from 1 (low relevance or usability) to 5 (high relevance or usability).

We calculate the average score for each pillar and present the group's score as well as average score below:

	Power	Equity	Identity	Knowledge	Access	Participation
Relevance-G1	5	5	3	5	5	5
Relevance-G2	5	4	3	5	4	4
Relevance-G3	5	5	5	5	5	5
Relevance-G4	5	5	3	3	4	4

Relevance-G5	4	3	5	3	2	2
Relevance-Avg	4.8	4.4	3.8	4.2	4	4
Usability-G1	5	1	3	5	2	1
Usability-G2	4	4	3	3	3	2
Usability-G3	2	4	1	4	3	4
Usability-G4	4	4	3	3	2	2
Usability-G5	1	4	5	4	3	1
Usability-Avg	3.2	3.4	3	3.8	2.6	2

Relevance

Regarding the relevance of Pillars to China's social, political and economic context, we observed that the Pillar "Identity" receives the lowest score. This finding resonates with what we discussed in the Internal Assessment that "identity" probably is not one of the most perceived or observed social issues in today's China and that's why groups score it the lowest and perceive it's irrelevant in China. The two groups scoring "Identity" a 5 both have members who work full-time or volunteer in the space of LGBT or women's rights.

In contrast, "Power" is scored as the most relevant pillar. In China, we are a society of big government, and people generally believe and rely on the government to govern the society. Therefore, workshop groups tend to agree that the discussion of power in China should focus more on how to empower the government to make stronger influence and regulate the behavior of corporations, which is exactly what we found through desktop research during the internal assessment. The Chinese government is continuously strengthening the legal system through strong administrative force and exercising restraint on the abuse of data innovation by big tech platforms.

On the other hand, the groups shared that power is worth discussing because we see that the increasing power of large technology companies, especially between consumers and platforms, labor and platforms, constitutes a new power inequality and imbalance. On the one hand, of course, we have to consider how the Chinese government regulates platforms in the context of big government, but on the other hand, we also seem to need to consider how to strengthen individual power or the ability to defend rights, and transforming the current power structure may be one way.

Usability

The first observation we made is none of the pillars receiving a score higher than 4. It may suggest an overall negative attitude towards the usability of six pillars in China.

It is no surprise to us that the lowest average score is assigned to the pillar "Participation" as we already hypothesized that "Participation" can hardly work in China as citizens generally do not have sense or interest in participating. Groups' sharing also validate what we think and it is added that even though citizens may have interest, they may not have the right skills and capacities to make a real contribution in the participation process. However, it is worthy to note that most groups also highlight the importance of multi-stakeholder governance and then dream that someone or some professional organization can represent themselves in participating in rulemaking and auditing data practices.

Compared to other pillars, "Knowledge" was scored the highest. As groups explained, "Knowledge" is a neutral concept: it is not too political like "participation and Access" that could be hard to be implemented in the context while it is also not focused on human rights like "Identity and Equity" that seems not directly relevant to the wider public in China.

Task4: Relevance and Usability of Pillars in Selected Case Studies

In Task4, we ask participants to assess the relevance and usability of pillars in the given three case studies: "Big data discriminatory pricing", "Delivery workers who are trapped in the algorithm" and Health Passport apps. Please note participants were told to also read reflection questions in this task.

When selecting case studies here, we deliberately selected the three case studies as each of them dealing with one kind of relationship: customer versus big platform, labor versus big platform, and citizen versus government. We expect to see how participants may assess the relevance and usability of power and participation in different settings of relationships.

The table below summarizes the scores and average scores for each case study as well as the overall situation. Please note that each cell contains a pair of scores: relevance and usability.

	Power	Equity	Identity	Knowledge	Access	Participation
G1	5,5	1,2	5,3	5,1	5,1	1,5
G2	4,2	5,1	5,4	4,3	3,3	3,2
G3	3,2	5,5	3,3	4,4	2,2	4,5
G4	1,3	5,3	4,4	3,4	3,5	2,5
G5	2,1	5,5	4,1	2,4	3,2	2,3
Case 1	3,2.6	4.2,3.2	4.2,3	3.6,3.2	3.2,2.6	2.4,4
G1	5,2	1,2	5,1	1,1	5,1	1,5
G2	4,1	4,3	3,3	3,2	4,3	4,4
G3	5,3	4,3	4,3	4,2	4,2	4,2
G4	1,3	4,1	5,4	4,4	3,5	2,5
G5	1,4	2,3	1,3	4,4	3,5	5,5
Case 2	3.2,2.6	3,2.4	3.6,2.8	3.2,3.6	3.8,3.2	3.2,4.2
G1	5,1	5,5	5,1	5,1	1,1	1,1
G2	5,3	4,4	4,2	3,3	4,2	4,1

G3	5,2	3,3	4,3	5,2	5,1	4,1
G4	5,5	3,3	3,3	4,4	4,4	4,4
G5	5,3	2,4	1,4	1,4	4,3	3,3
Case 3	5,2.8	3.4,3.8	3.4,2.8	3.6,2.8	3.6,2.2	3.2,2
Overall	3.7,2.7	3.5,3.7	3.7,2.9	3.5,3.2	3.5,2.7	2.9,3.4

Score Analysis

If we first look at the overall average score representing groups' analysis of six pillars against the three case studies, we found that "power" and "identity" are the most relevant pillars while "participation" is the least. It is different from what we found in Task 3. When asking why groups treat "identity" differently in two tasks, one group argued "When we talk about 'identity' in general, it seems to be irrelevant to our life and more relevant to specific minorities, such as LGBT or disabled people. But in specific cases, it's easier for us to step into the shoes of those impacted and then discuss power structures, so this may be why we think 'identity' is more relevant in specific cases."

Regarding usability of pillars, the overall score distribution is quite similar to what we observed in Task 3. It is interesting though that "Participation" pillar is scored lowest in terms of relevance but received quite high score in terms of usability especially in Case 1 and 2. It may suggest that people consider it more feasible to implement the idea of participation in corporations rather than in governments.

Reflection Questions

We also asked participants to pick up one specific case study and read carefully about reflection questions then pick up one to three questions they believe are most useful. Groups' responses to this question are quite diverse and among each group it is also very hard to reach consensus as each participant has their own perspective. Here we do not focus on specific questions but report our observations across groups about which pillar is mentioned more than others regarding the usefulness of its reflection questions.

One of the pillars is "Power" pillar. Participants shared that questions under Power Pillars help them better reflect upon the asymmetry power relation that they may not realize deeply before reading it. It is also interesting as one participant shared "I never thought about changing the power structure. Is it possible? But reading the question drives me to think a lot about it and I ask myself why not?"

The other one is "Access" Pillar as questions under this pillar raised two interesting points: one is whether you have enough information about the data innovation itself and two whether you have access to benefits of data innovation. The former is due to the recent rapid development of data rights protection and artificial intelligence regulation in China. People consciously hear a lot of concepts such as algorithm transparency on the news, so they will resonate with the reflection question. As for the access and sharing of data benefits, some groups mentioned this is the first time they have learned about it and know the possibility through the reflection question. They feel that this is a wake up call for them that there is such a possibility to further discuss benefits sharing.

Chapter V: Discussion and Suggestions

In this research work, we tried to unpack what data justice means in China and analyze the usability of six pillars as came up with by Alan Turing Institute in the social, political and economical context of China. We interviewed 12 people representing policy makers, developers and impacted communities and ran workshops with 25 ordinary people who are more likely to be users or impacted community members of data innovations. Our findings were reported in Chapter III and IV and in this chapter we discuss what our findings may suggest and also report our suggestions to further research or take action on data justice.

Focus of Data Justice in China

In our research, we found that the focus of data justice in China as depicted by case studies and associated keywords is on the relationship between users and big tech platforms. The users could be consumers who receive service or purchase goods on platforms or laborers who work on big tech platforms to deliver food, provide taxi service and so on.

Thus, discussing data justice in China is usually related to consumer rights or labor rights rather than traditional human rights issues such as racial or gender rights. We also observed that case studies on consumer or labor rights both involve tangible financial loss and this is how Chinese people generally perceive data injustice.

That suggests that if we want to further promote the idea of data justice, it is already easy to engage ordinary people in paying attention to data justice issues because ordinary people already have real life knowledge or experience of data injustice cases such as "big data discriminatory pricing".

However, speaking data injustice beyond "big data discriminatory pricing" such as "invisible groups in census", it may be less known to ordinary people and only people who work on human rights issues and have decent knowledge on data may have such awareness. Therefore, It is important that we can expand the narrative of data justice in China to make ordinary people be more aware of those human rights related cases; On the other hand, we also should take actions to engage people who are working on human rights issues but have little knowledge of data in understanding how data may amplify or reduce the human rights issues and help them understand how to respond to such new data challenge in their human rights work.

Landing Six Pillars Framework in China

In our interviews and workshops, we tried to brief participants about the six pillar framework in as much detail as possible. Given the responses from interviews and workshops, we believe the six pillars as identified by Alan Turing and its associated reflection questions are relevant to the context of China, and secondly they are usable and feasible in China though people may have specific doubts on each of them given China's special political and social system.

We first noticed that the "Identity" pillar is inconsistently scored during our workshop tasks and it suggests that people generally do not feel "identity" is connected to people's daily life because social issues on identity e.g. typically racial or gender identity are less concerned in China. However, when people are given specific context (i.e. a case study) , they feel much easier to put themselves into the shoes of the weak party. And when people feel they are in an unbalanced power relationship and as the weak party, they realize the "identity" is a problem. Thus, it indicates that it may be a good idea to link the power pillar with the identity pillar to better help people reflect and take actions.

"Power" pillar is consistently considered as a relevant but also usable pillar across interviews and workshops. However, it is worth noting that people are less concerned about the power structure between citizens and governments or probably from the other angle that they do not think it is possible to make any changes.

The relevant pillar "Participation" however received quite mixed scores on both relevance and usability. It is noted that direct participation by citizens or users may not be feasible due to the fact that data innovation involves quite high technical elements and it prevents people from understanding and getting involved.

Professional intermediaries who have knowledge and skills on dealing with data issues are therefore expected to represent people to participate and defend their data rights. Such intermediaries can be existing NGOs or media, or be new types of organizations similar to the idea of data coop or data trust which can collectively represent individuals to govern data and data rights. Thinking about the "distribution of data and data benefits" keywords mentioned in interviews and workshops, it may be a good idea to also explore how to leverage such data trust and data coop ideas in addressing such concerns on distribution of data benefits.

Suggestions

- Based upon what we found in people's understanding of data justice in China, we suggest that the data justice guide may consider creating two versions for two types of users: the one described as an impacted community who are an impacted group in social issues and the other one is as general consumers. And somehow for different types of users, the pillar of equity and identity could mean very different things.
- It may be a good idea to have a more comprehensive guide for ordinary people who have zero knowledge about data and data rights to learn data justice but also quick basics from the guidance.

- It would also be ideal to have case studies organized in "identity" groups which are more relevant to the reader's group to build emotional links between cases and readers to let them quickly understand why identity is an issue.
- It may be a good idea to add a new stakeholder group: intermediaries and create action and reflection guides for this group to take actions to defend users' data rights. It would be helpful to include a resource kit for intermediaries to know who may help empower them and what tools are already there for them to re-use.
- It is interesting to discuss whether fair distribution of data benefits or fair sharing of economic benefits generated by data is one of many focuses of data justice agenda and explore how the narratives around economic benefits sharing are different from the narratives around data justice as in social justice or human rights.
- It may be worth exploring the idea of what data for common good means under the framework of data justice and debate on how technology or data user plays a role in this process. One argument based upon China's practice is that the extreme fine utilization of data may achieve more freedom and benefit the wider collective community while at the cost of compromising a certain level of individual privacy interest.

Annex A: Interview Question

Q1：当你听到数据正义时，你觉得它是什么意思？三个最为相关的关键字或关键短语是什么？

Q1：What do you think data justice means when you hear it? What are the three most relevant keywords or key phrases?

Q2：你能想到的数据正义的案例是什么？有没有你自身（你所在社群群体）亲身感受的案例？

Q2：What cases of data justice can you think of? Are there any cases that you (your community group) have personally experienced?

Q3：你认为当前数据正义最大的问题是什么？是社会不公平被数据扩大/放大，还是数据创造了新的不公平？

Q3：What do you think is the biggest problem with data justice? Is social injustice magnified/amplified by data, or is data creating new injustice?

Q4：政策制定者、应用开发者/提供者、用户（特别是特殊群体）这是指南中涉及三类群体，你觉得除了这些角色外，数据正义还和谁有关？这些角色又如何受数据正义所影响或影响数据正义？

Q4：Policy makers, application developers/providers, and users (especially special groups). These are the three groups involved in the guide. In addition to these roles, who do you think data justice is related to? How are these roles affected by or affect data justice?

Q5：您自身认为您可以扮演什么样的角色？您个人可以为达成数据正义做些什么？

Q5：What role do you think you can play? What can you personally do to achieve data justice?

Q6：就你来看，数据不公正由多大程度由数据采集和应用的信息不透明所致？我们应当推动哪些信息的进一步透明？

Q6：In your opinion, to what extent is data injustice caused by the lack of transparency in data collection and application information? What information should we promote for further transparency?

Q7：一种理论是，让受影响群体能够全程参与到数据创新中（数据标准制定、数据采集、数据处理、数据应用等环节），可以使得数据被应用时能够更好的具备包容度，达到更为正义的效果。就你来看，参与性本身的可实现性多大？用户有能力去参与吗？在公共机构的数据创新和私营机构的数据创新中，用户的参与可实现性会有何不同？

Q7 : One theory is that allowing affected groups to participate in the whole process of data innovation (data standard setting, data collection, data processing, data application, etc.) can make the data more inclusive and achieve more just results when it is applied. In your opinion, how achievable is participation itself? Do users have the ability to participate? In the data innovation of public institutions and private institutions, how will the user's participation be different?

Q8：我们的公共法律或私营机构的内部体制能够现在鼓励参与性吗？这些法律或内部机制体制的可重塑性大吗？或者说去做出改变的难点在哪里？

Q8 : Can our public laws or internal institutions of private institutions now encourage participation? Are these laws or internal institutions highly removable? Or what are the difficulties in making changes?

Q9：当前遇到数据不公正的问题时，有方法去纠错、投诉、改进吗？（救济的渠道）能否结合具体的案例来说明？

Q9 : Is there a way to correct, complain, and improve the current problem of data unfairness? (The channel of relief) Can it be explained with specific cases?

Q10：个体（或特殊群体）面对机构时有足够力量去博弈吗？我们现有的法律、制度、内部治理体系有多大弹性去改变？

Q10 : Do individuals (or special groups) have enough power to play games against institutions? How flexible are our existing laws, institutions, and internal governance systems to change?

Q11：就你来看当我们谈论数据正义，特别是如何去确保数据的采集、应用公正公平，除了需注意权利的不对等、信息的不对称、问责的缺失，还有什么关键点、问题或路径是我们要注意的？

Q11 : As far as you are concerned, when we talk about data justice, especially how to ensure the fairness and fairness of data collection and application, in addition to the need to pay

attention to the inequality of rights, the asymmetry of information, and the lack of accountability, what are the key points, Problems or paths we should pay attention to?

Q12 : 对于你去维护自身数据权益，或者通过你的工作职能去确保数据正义，除了提升你对数据正义的理解、帮助你去注意到（或反思）类似上述权利对等、信息对称等要点外，你觉得还需要什么样的工具或方法？

Q12 : For you to safeguard your own data rights and interests, or to ensure data justice through your work functions, in addition to improving your understanding of data justice, helping you to notice (or reflect on) the above-mentioned points such as equal rights and information symmetry, What other tools or methods do you think are needed?

Annex B: Workshop Activity Sheet

Data Justice Workshop Activity Sheet (1 of 2)

Task1 : Definition of Data Justice
 谈及 Data Justice 你第一时间能够联想的三个词汇是
 (the three words associated to Data Justice)

Task2 : Case Studies and the Definition
 下列哪些案例事件你认为可能符合数据正义/不正义的定义：(which are about data justice)

- 大数据杀熟
- 困在算法中的外卖员
- 利用算法筛选简历
- 政府开放土地数据使得富人更快占据土地
- 人口普查中没有对应 LGBT 群体的真实取向和婚姻状态的选项
- 健康码调度个人数据自动计算健康状态
- 通过建立透明自治的数据合作社使得临床数据得以被研究人员安全使用加速新药研发

你能想到的其他体现数据正义 或 不正义的案例有？
 (any other cases?)

(1) _____

(2) _____

反面还有

Task3 : Pillars Usability Analysis: Context
 (1) 对于 Alan Turing 所总结的 Pillars, 哪一项或几项 pillar 超出你的预期？为什么？
 (which pillar(s) surprises you and Why?)

(2) 你认为有什么新的维度应当被补充进 Pillars 当中？
 (anything missing in current pillars)

(3) 综合我们的国情和你的认知, 你认为现有的 Pillars 模型各维度, 和中国的相关性和落地性是怎样的 (1-5 分, 5 最相关、最落地)
 (score the relevance and usability of each pillar)

	权力	公平	身份	知识	访问	参与
相关性						
落地性						

Data Justice Workshop Activity Sheet (2 of 2)

Task4 : Pillars Usability Analysis : Case Studies

案例1：滴滴大数据杀熟
 案例概述：复旦大学孙金云团队实际测试 800 余次打车，实证使用的手机品牌差异会导致订单价格差异，品牌越贵，车费越贵。

案例2：困在算法中的外卖员
 案例概述：外卖员的送餐时长由平台通过过往时间、路况、天气等因素算法计算得出并用于惩罚外卖员超时，美团已公开其算法基础逻辑。

案例3：健康码
 案例概述：政府采用基站、身份等综合市民个人数据，自动计算健康状态，部分地区拟对数据采集做常态化和功能增强被喊停，健康码背后数据使用和算法未曾披露过，如发生错误有一定的救济机制和反馈机制。

(1) 阅读各 pillar 的综述和反思问题，找出您认为和这个案例相关的 Pillar，并对这些 Pillars 做两个数值分：第一组，按照 Pillar 和案例的相关性给分 (1-5, 5 最相关)，第二组，按案例的落地性可能给分 (1-5, 5 最能落地)

	权力	公平	身份	知识	访问	参与
案例 1						
案例 2						
案例 3						

(2) 挑选一个案例，分析其设计的政策制定者 (Policy Maker)、开发者 (Developer) 和用户 (Impacted Community) 三个群体，重新审视各 Pillars 及其反思问题的落地性和相关性？

PM :
 Developer :
 Impacted Community :

	权力	公平	身份	知识	访问	参与
PM						
D						
IC						

(3) 基于上述案例，阅读其相关 Pillars 的综述与反思问题，挑选出您觉得最启发您或最重要的 3 个反思问题 (填写其编号即可)

案例：_____

问题：_____, _____, _____

(4) 针对上述挑选的案例，阐明是否有未能被 Pillars 所覆盖的维度或考量：