An Ethical Framework for Big Data and Smart Cities

Victor Chang

Artificial Intelligence and Information Systems Research Group, School of Computing, Engineering and Digital Technologies, Teesside University, Middlesbrough, UK

V.Chang@tees.ac.uk/ic.victor.chang@gmail.com

Abstract

This paper presents an ethical framework for Big Data and Smart cities, focusing on contemporary ethical and non-ethical issues in big data analytics applications in smart cities and public transportation systems. The framework provides reviews and analysis of ethical and emerging issues and provides a summary of recommendations and discussions for four emerging areas. By reviewing recent studies on both the technological development and emerging ethical problems in the emerging industries, this paper seeks to find and raise public awareness of ethical issues lying in urban big data analytics and public transportation systems. In order to deal with emerging issues, four recommendations have been explained and subsequently, two areas of discussion have been described in detail to support the ethical framework. This paper addresses emerging issues and their ethical concerns for big data and smart cities. Possible recommendations and solutions have been demonstrated to promote the competency of companies and organizations in this big data era. How the ethical framework can be used by six smart cities have been described. Our findings and analysis for big data for high growth, innovation and core competencies and validity of the ethical framework have been justified.

Keywords: Ethical framework; big data; smart cities; ethics for big data.

1 Introduction

Due to the contemporary development of ICTs (Information and Communications Technology), the "big-bang" style of big data has provided us an enormous amount of digitized and fresh data, particularly about decisions and interactions between people, organizations and companies. Disruptive technologies are technologies that can use high-tech and can offer transformation to the ways we work. These include the Internet of Things (IoT), big data, artificial intelligence (AI) and so on (Abdel-Basset et al., 2020). The rising importance of big data and analytics has become a subject of focus for disruptive technologies. Turning these big data into an understanding of real-world information can help us make decisions and improve the performance in e-commerce, market intelligence, e-government, science and technology, smart health, security and public safety sectors (Hsinchun & Chiang, 2012). Especially in urban areas, the adoption of disruptive technologies with IoT focus (Scuotto et al., 2016) such as CCTV (Closed-Circuit Television) cameras, ANPR (Automatic Number Plate Recognition) (Kitchin, 2014) and wearable sensors have tremendously increased the life quality and safety of citizens.

While smart urbanism is constructed upon big data, issues arise when practitioners and researchers adopt these emerging technologies. Privacy of our citizens should be the first concern since companies nowadays can accumulate a massive amount of information about people (Martinez-Pelleste & Solanas, 2013). Through persistent scanning and recording, information of citizens about consumption, transportation, healthcare and transactions can be processed, analyzed and stored to enhance the development of smart cities (Wu, 2017). Particularly when collaborations happen between multiple companies and governments, information can be transferred, shared and even auctioned from one party to another. Moreover, the adoption of analytics can make sensitive information to be revealed easily (Kitchin, 2016). Consequently, protecting confidential information becomes extremely important and challenging.

Data analytics can provide a way to present processed outputs of Big Data with visualization and graphical formats. Data analytics can collect data from real-world environments to digitize and analyze processes and environments. Data analytics helps managers to transform data from the field into insightful information and to support decision processes. Data analytics as a management tool can lead to process optimization and increased efficiency in multiple departments and industries (Albright, 2017). This paper focuses on the impact of data analytics in smart cities, including their public transportation systems.

Insights are gained by reviewing research papers on data analytics from Europe, China, South and North America from the last four years. Additionally, papers on data privacy, security and analytics regulation will be reviewed for technical guidance. The conclusions towards a sustainable and ethical framework are correlated between the current big data analytics practices of industry players and the state-of-the-art user privacy protection guidelines led by leading research bodies. The purpose of this paper is to raise awareness of existing ethical and related issues for disruptive technologies, particularly big data and smart cities, with the recommendations of the ethical framework. By following guidelines that avoid contemporary issues in big data, companies and organizations can promote their new technology revolution competency.

2 Literature Review

2.1 Smart cities

A smart city originally developed by IBM illustrates a harmonized city where ICT-driven technology enhances city services such as business, transportation, health care, communication and energy supplies (Batty et al., 2012). In 2010, 130 cities worldwide accepted the 'smarter cities challenge' proposed by IBM. An investigation driven by Alizadeh (2017) reveals that e-government, transportation and environment are the three most important areas in participating cities' desire. With fast-growing emerging big data analytics, Kitchin (2014) summarizes that a data-rich city can use sophisticated data analytics to understand, monitor, regulate, and plan. However, controversial ethical issues and criticisms arise due to the fast growth of big urban data. Taylor (2016) argues that the legibility and detailed human mobility information-driven by smart devices and big data can lead to overreaction to conflict, migration, and the right balance is required between privacy and safety. Caragliu et al. (2011) present the case for smart cities in Europe. They did an urban audit survey between 2003 and 2006. Based on their data, they performed data analysis for European cities. They compared the proportion of employment in culture and entertainment industries, multimodal accessibility, and the length of the public transport network versus GDP per head in 2004 for 100 European cities. Interestingly some cities were well above the average regression lines, such as Frankfurt am Main and Dusseldorf. However, their measurements and analysis were not presented in detail. Based on their work, public transportation is an emerging topic for smart cities. In China, Wu *et al.* (2018) propose several risks, including information safety, emergency response and technology restrictions in constructing smart cities with Chinese features. They attempt to see if the development and establishment in China can have some extent of correlation to the development of the public transport sector in smart cities.

2.2 Competencies of big data analytics in smart cities

Although the term 'smart cities' is extensively discussed throughout the years, the definition of a smart city itself is still ambiguous. In previous kinds of literature, many words appear to be interchangeable for 'smart' such as 'intelligent', 'digital' (Alizadeh, 2017). As a transdisciplinary area, smart cities are often interconnected with various subjects such as urban science, ICT (Angelidou, 2014), IoT (Scuotto et al., 2016), health care (Aborokbah et al., 2018). Additionally, smart cities include technologies like artificial intelligence (Srivastava et al., 2017), transportation (Masek et al., 2016) and even government regulations (Hartswood et al., 2014).

Glasmeier and Christopherson (2015) define smart cities in a subtle and futuristic way as it has two attributes: coordinating fragmented urban facilities using new technology and calling forth a new reality for residents living. For more common explanations, Khatoun and Zeadally (2016) define 'smart' as being able to promote the efficiency of urban planning using diverse technologies. Despite having the variability of definitions, a deduced emphasis of smart cities should be the adoption of 'ICT', and the goal to enhance the quality of life for citizens. As shown in Figure 1, the comprehensive smart city model can be described by Khatoun and Zeadally (2016) as a complex system including human, infrastructure and process sections.

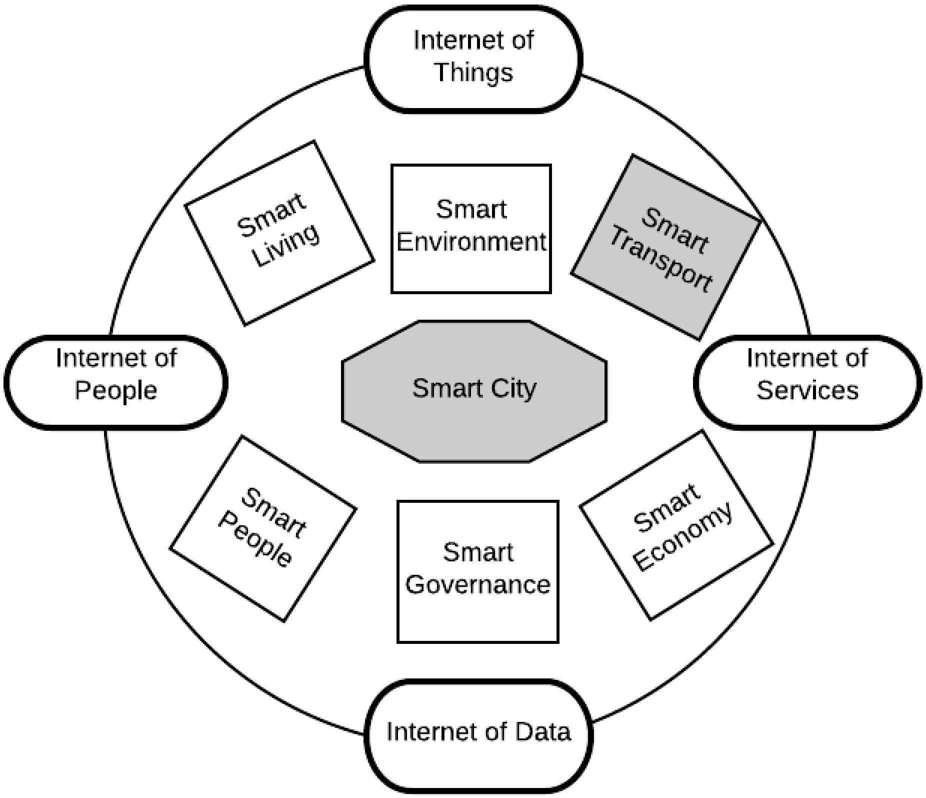

Figure 1 A Smart City Model (Khatoun and Zeadally, 2016)

When big data technology is on offer, smart cities' construction has been raised to a new level. The traditional datasets such as national census, government record of personal details, geographical information are primarily based on samples with limited variables and time scales. However, big data breaks that barrier since it can capture a large real-time scale of datasets generated by sensor-based devices, cameras, RFID tags, mobile phones through wireless networks (Khatoun & Zeadally, 2016). These datasets have mainly four attributes, as described by IBM. They are namely volume (consists of terabytes or petabytes of data); variety (either structured, unstructured, temporally or spatially); velocity (data are collected in real-time) and veracity (very accurate to reveal the nature of a person or object) (Herschel & Miori, 2017). This is widely adopted in Singapore and London. In Singapore, they have transport systems that can take pictures of all the speeding cars or cars breaking the rules on the road since 2001 (Ibrahim, 2003). Additionally, they also have medical systems, which they plan to have data stored and shared among hospitals in Singapore. For London, it is one of the largest European smart cities with a comprehensive plan for data management, collection and analysis for different services and sectors available in London (Angelidou, 2016).

2.3 Public transportation in smart cities and data privacy

Utilizing real-time analytics towards these data helps city governments regulate and dispatch social resources. A typical example of using analytics in the transportation section is Public Transportation. Global cities are getting increasingly large and dense, which leads to congested automotive traffic, traffic jams and increasing air pollution. All of these can decrease the quality of urban life. While increasing reliance on Public Transportation (PT) systems instead of automotive traffic can relieve urban areas from these negative influences (Meng & Fanxin, 2017). This paper defines its PT systems as stand-alone and/or connected bus, tram and train systems, spatially restricted to one metropolitan area.

While aiming to interpret privacy better, Solove (2006) provides a taxonomy that consists of four aspects, which are helpful when investigating privacy harms driven by data. Solove (2013) proceeds his work by considering the prevailing big data technology and illustrates the conflict between consent and emerging data technology. By applying the taxonomy of privacy (Solove, 2013) into smart cities, Kitchin (2016) demonstrates an overview of ethical issues beneath the prosperity of smart cities. These include 'datafication' and privacy, surveillance, inferencing privacy harms, anonymization and de-identification, and data repurposing. In this paper, the concept of smart cities and analytics is introduced. Additionally, ethical issues with instances will be discussed under the framework provided by Solove (2006). Cases of real-world examples can be provided to illustrate the importance of ethical issues under smart city construction.

3 An Ethical Framework for Big Data

This section presents our proposed ethical framework for Big Data, which aims to consolidate recommendations for emerging issues and concerns.

3.1 Competencies of big data analytics in Public Transportation

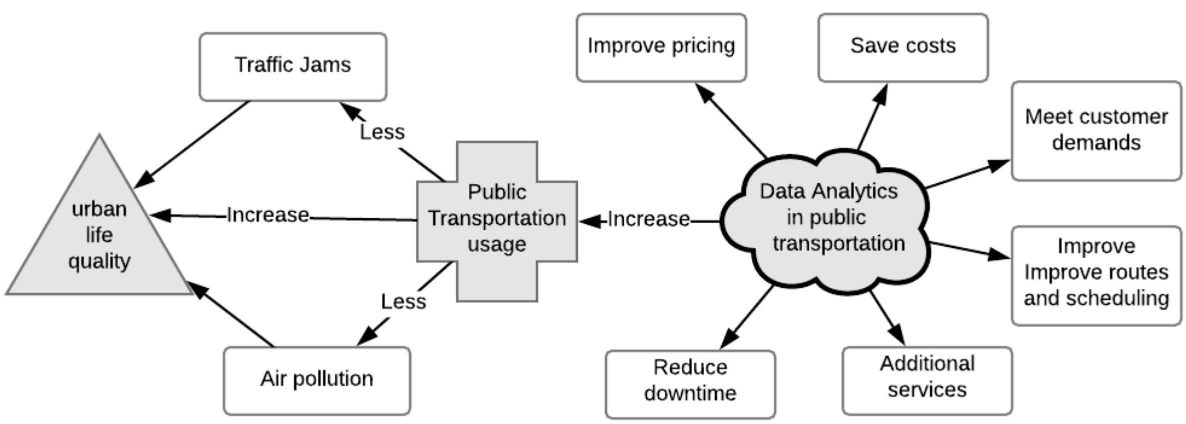

Figure 2 Data Analytics in Public Transportation

The ethical framework should offer six major benefits: improving pricing, saving costs, meeting customer demands, improving routes and schedules, offering additional services, and reducing downtime. This can be represented in Figure 2, which shows the data analytics and its relations in public transportation (PT). In PT, data analytics analyzes customer origin-destination patterns, machine operations, customer top-up histories, time patterns, and payments. Moreover, PT companies use these data for inter-correlation with external data sets to further draw insights on customer time-spatial habits (Ma et al., 2017). The collected data is used to increase efficiency for numerous internal operations (Stelzer, 2016) and align PT operations with customer demands (Lee & Hickmann, 2014). Subsequently, public transport (PT) companies, city planners, and government are interested in applying analytics for PT and increasing its urban application.

3.2 Analysis of ethical concerns

While not all data analytics has necessarily ethical problems, our framework should differentiate what needs ethical concerns. Figure 3 explains the type of data and their respective ethical concerns. The data in PT operations collected using temperature, humidity and vibration sensors, cameras and manual scripts are used to improve safety as well as maintenance, repair and operational efficiency (Thaduri et al., 2015). For data collection and analysis of data based on machinery data, no ethical concerns were found in the literature and it was excluded from the scope. However, data concerned with the customer and citizen data may hold potential ethical threats.

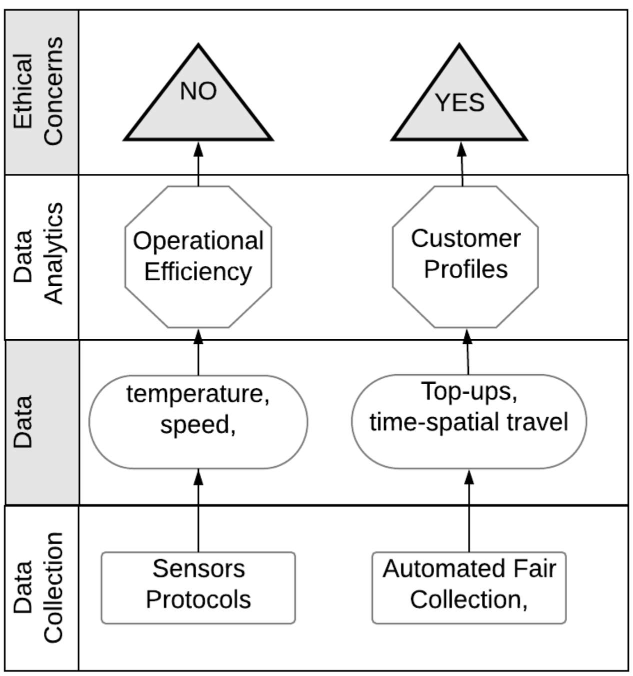

Figure 3 What kind of data raises ethical concern?

3.3 Ethical and emerging issues in big data analytics

Herschel and Miori (2017) define ethics as 'an analysis of conduct that can cause benefit or harm to other people'. Generally, **ethics is a code of conduct for people to decide what is right and what is wrong, which is beyond the law. Good ethical theories or frameworks guide individuals and organizations to make logical, reasoned and persuasive decisions.**

In 2012, as reported by Duhigg (2012), a retail store called 'Target' extracted implicit information about a high-school girl beyond her father's knowledge. The father discovered coupons for baby clothes and cribs sent by the retailer to his daughter, who was then found pregnant. With the information collected both by the retailer and by purchase from other companies, the algorithmic analytics of the company could confirm the pregnancy even without the permission of the girl and sent off product coupons.

With data analytics, companies learn more about customers and can provide more personalized services. However, having too much information about a person creates privacy issues. This is particularly true when data has been collected without the consent of the person. Figure 4 can provide a simplified but useful illustration of data analytics and ethical decisions. Analytics are capable of massive amounts of tasks, but the legal and ethical parts of these tasks are important since ethical decisions ought to be observed for humanities. That is the reason to discuss ethical issues in big data analytics and smart cities in this paper.

With increasing opportunities for data analytics to create new business benefits, insights and opportunities, the ethical application of these data analytic techniques as well become increasingly important. Ethics can always be relevant when a party has to decide between two different decisions. If a situation has only one possible solution, no decision making and subsequently, no ethical considerations are required (Center for Business Ethics, 1992). Most consumer-focused data analytic operations are a trade-off between corporation benefit and customer privacy, making ethical consideration a required business challenge.

Vidgen *et al.* (2017) executed a Delphi study and three in-depth interviews with data analytic experts on their biggest challenges to adopt data analytics practices. One aspect mentioned by the data analytic experts was ethical behavior (Vidgen et al., 2017). The main expert's concerns were losing the customer's trust or the value of their brand image due to unethical perceived operations. Brand value and customer relationships are an increasingly valuable asset in modern business models (Tasci & Asli, 2016). Therefore, business leaders aim to protect them from malicious news caused by their data analytics operations. The article by Vidgen *et al.* (2017) calls for internal

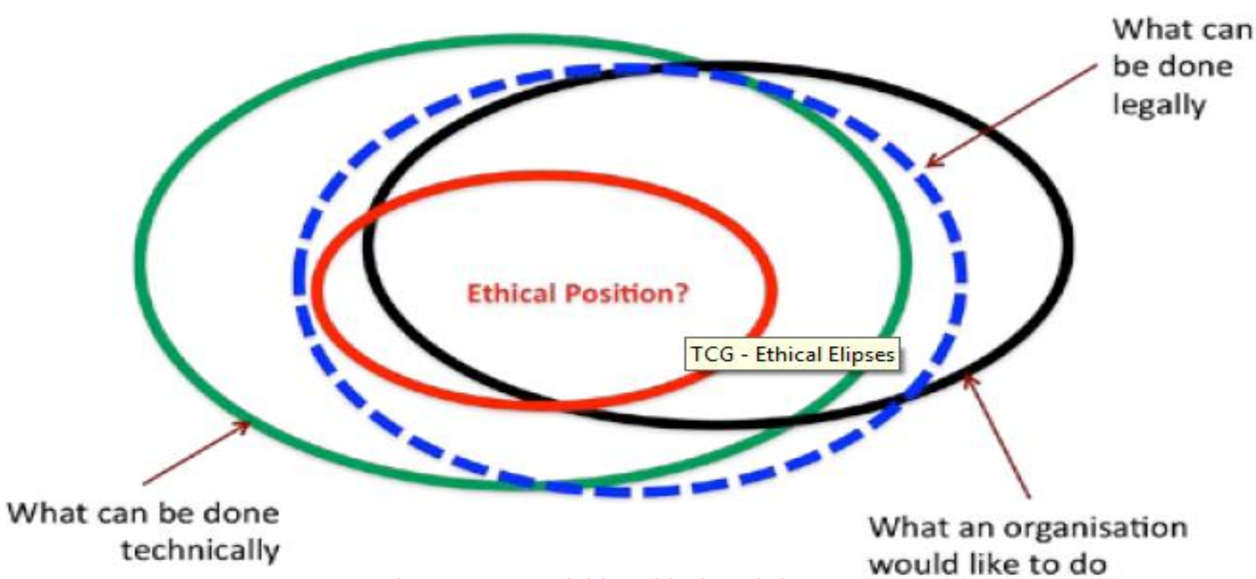

Figure 4 IBM Model for Ethical Analytics

ethics committees to oversee the companies' data analytics operations as a solution to keep data analytic operations aligned with ethical values.

This task of keeping analytics to be ethical-compliant becomes increasingly challenging because the legal framework surrounding data analytics operations are often un-mature. Moreover, the un-mature legal frameworks are not necessarily in-line with customer ethical values (Vidgen et al., 2017). Besides, data analytics operations have pressure 1) from management to meet business goals and 2) from engineers to stay within the technologic possibilities. Therefore, companies have to spend a considerable amount of time to design their Da operations ethically. As shown in Figure 4, IBM presented a data analytics development framework in 2014, considering stakeholders from the government, developers, managers, and society to design data analytics processes. Ethical concerns can be positioned between technical, organizational and legal requirements. They can be managed well in the process of developing and delivering analytics services.

In this big data era, companies and organizations with core competencies can appropriately regulate their decision-making process to be legal and ethical (Chessell & Mandy, 2014). Although the moral position is flexible and can be chosen by companies and organizations themselves according to the external factors, ethical decision-making promotes reputation and trust from end-users so that companies and organizations appear more competitive.

4 Four Emerging Issues the Ethical Framework

In this section, four emerging issues have been described as they are commonly experienced in all smart cities, including countries with difficulties implementing due to lower ethical requirements. Recommendations to improve are then explained further.

4.1 Emerging Issue 1 of the Ethical Framework: Privacy and surveillance

As a fundamental human right, privacy is to optionally reveal oneself to other people and the rest of the world. According to Solove's (2006) taxonomy, the privacy of a person can be harmed through three stages: information collection, information processing and information dissemination or invasion, each of which has distinct damages with others. These privacy perspectives help understand how emerging urban big data and smart city technology affects privacy (Solove, 2006). In this section, these privacy issues embedded in contemporary smart city

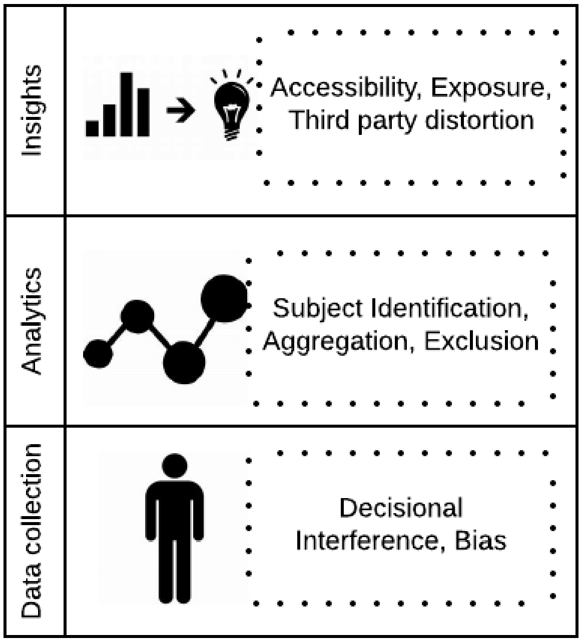

Figure 5 Definition of Privacy in Data Analytics

constructions will be demonstrated. Figure 5 shows the definition of privacy in data analytics involved in three aspects: data collection, analytics and insights. Specific focuses and tasks are contained for each section. To further expand on these, concerns and issues will be described in the following sections.

4.1.1 Information collection, surveillance & absence of consent

The term 'datafication' refers to a trend to turn many aspects of our life into computerized data. With the development of ICT and big data, the use of unique identification information is intensively increasing for people to access services. Every day in our life, we use our names, passwords, accounts with detailed personal information such as addresses, emails, phone numbers, credit card numbers and transportation cards to achieve services. The intuition behind datafication is that it is almost impossible for us to live without leaving traces and footprints in our everyday lives. A large amount of data regarding our own information and activities can be collected, even without being aware.

With the rapid growth of ICT and big data, especially wireless mobile technology, performing surveillance and geo-surveillance (tracking the geographical location of someone) on individuals is becoming easy and effortless nowadays. Most cities are filled with remote-controlled cameras that can arbitrarily move to track and even recognize (Williams, 2015) pedestrians. Toll cameras, road cameras, ANPR (Automatic Number Plate Recognition) and wireless systems constitute public transport networks that can record every vehicle with its behaviors and routes. Similarly, smartphones regularly communicate their locations with base stations, GPSs and their MAC address can be identified by Wireless network providers (Vincent, 2013). Indeed, these surveillance actions have enhanced many aspects of our lives, such as public safety and transportation. Suspects have nowhere to hide with intensive cameras spread across streets. Congestions are relieved by controlling traffic lights through traffic control centers, and traffic rules violations can be discovered at a low cost. However, it can be controversial when surveillance and geo-surveillance approaches are intensively adapted since citizens will have private behaviors exposed and always under the monitor.

A study also shows Alipay's current legislation is still insufficient to protect customer payment information (Liu, 2015). As one of the largest e-commerce and mobile payment providers worldwide, Alipay provides convenient payment methods through mobile phones. However, terabytes of individual payment information are generated every day and they can be easily used to analyze and predict customer behaviors and preferences.

Collecting and using information about a person usually needs the consent of himself/herself. However, as big urban data and smart city technologies emerge, requests for consent have diminishingly absent when governments and companies have begun to collect and process personal data. The reason behind the consent absence may lie in big data technology itself. Solove (2013) indicates the number of privacy agreements is simply too large for individuals to police under the circumstance of big data. If each interaction of data collection, data use and dataset merger needs an agreement, even if someone wants to confirm his terms and agreements himself, it would be impossible for him to finish all of them due to high volume and diversity. Consequently, people are gradually losing rights and control towards their data under big urban data and smart cities' age.

4.1.2 Information processing and re-identification

With urban big data, companies and governments continuously process data generated in our daily lives, which can be converted into insights and understandings of individuals and organizations through data analytics (Dameri & Rosenthal-Sabroux, 2014). While these data create economic profits for data owners, private information such as habits, lifestyles, preferences can be revealed easily. Additionally, when datasets are aggregated, merged through re-arranging and purchasing, the insights that a data scientist can get from becomes tremendously detailed. The term - personally identifiable information or PII refers to information that leads to inferences about individuals with urban big data modeling. For example, identifying a person who frequently visits a gay bar generates an inference that the person should be highly likely to be gay, which would be sensitive and controversial to know (Kitchin, 2016). Especially if the information has been disseminated in the third stage of privacy, the destructive impact may destroy humans.

Anonymity can be considered as one of the fundamental approaches to protect personal privacies. While most companies claim to have a severe regulation on anonymity, they never disclose they have ways to distinguish an anonymous person from others (Barocas & Nissenbaum, 2014). By using PII-like 'with whom they interact the most', 'to where do they go' and 'how much do they usually spend every day', an anonymous person can be distinguished out of the crowd even without those unique persistent identifiers (e.g., names, ages, identification card ID, date of birth, and so forth). After all, knowing every behavior about an individual except for his name is equivalent to knowing everything about him. This increases the risk of society since individuals' information can always be revealed by companies with data; people might have no sense about the connotation behind it.

4.1.3 Information dissemination

In the era of big data, technologies have enhanced the speed and breadth of information dissemination. When Social Network Sites (SNSs) like Facebook (Luarn et al., 2014) come forth, personal networks on SNSs not only provide information and interaction between users but also promote the efficiency of information dissemination (Bakshy et al., 2011). Fast information feed reduces information inequality, but at the same time produces ethical issues. According to Solove (2006), dissemination harms privacy when confidentiality is breached, and information can be disclosed or exposed to the public. For example, if the police disclose its surveillance camera information of someone famous to the public, a problem would occur when the public knows the information, especially if the information is sensitive.

Moreover, big data analytics may increase the conflict between the use of personal information and individual privacy. Who should control information about you? Who should access it? Who can use it? The answer is not always explicit. For example, medical monitoring devices accumulate information about patient health; mobile device apps collect information about personal information, location, etc. Can those kinds of information be shared or traded? To answer these questions, laws and regulations on data ownership and properties need to be discussed. There are already a few laws that acknowledge personal information as property in the UK, but further regulations are required (Rees, 2014). However, in China, the issue of personal data ownership is far from being mature (Lu, 2005). Regulations should be updated to protect personal privacy. One example is explicitly regulating rights and permission about the use of personal information by service providers and any third parties.

4.1.4 Invasion

Invasion of the personal sphere happens when personal information can be utilized to interfere with personal life, which influences decision-making (Solove, 2006). Intrusion in the smart city mainly involves spam, junk mail, unintended telemarketing, which harasses individual life. More seriously, blackmail may happen when criminals utilize private information through telephone or even by physical contact. As a form of invasion into the private sphere driven by ICT and big data technologies, cyberbullying can cause irreparable damage to another person's reputation and, at the same time, result in severe psychological trauma (Ong, 2015). When privacy harm comes to this final stage, a data subject can encounter irretrievable loss.

4.2 Emerging Issue 2 of the Ethical Framework: Data integrity

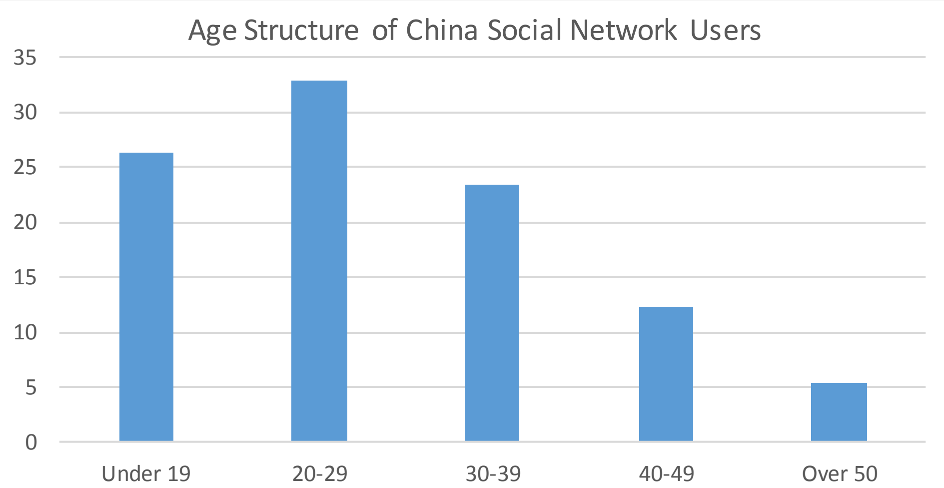

Figure 6. Age Structure of China Social Network Users in 2015 (CNNIC, 2015)

Some researchers regard 'veracity' as one of the big data's feathers. For example, Liu *et al.* (2016) argue that big data may cause errors due to three aspects, namely inauthentic data collection, information incompleteness and representativeness problems. Firstly, business companies collect data differently from scientific researchers in which authenticity and credibility issues occur frequently. For example, social network companies such as Weibo, Facebook, and Twitter have 'zombie' or 'faked' accounts operated by computers, which brought economic profit. Therefore, these accounts not only provide noise for social network analytics but also create biased results. Secondly, some big data are fine in volume but bad at completeness. For example, the data fields of some mobile phone data are limited, and user IDs are pseudo so that socio-economic values cannot be discovered. Moreover, some geographical location information is inaccurate since the locations recorded are not exactly the real ones but the location of the mobile phone tower (Gao et al., 2013). Additionally, technically some machine learning algorithms can encounter overfitting issues, thus resulting in an inaccurate prediction. Thirdly, big data may fail to represent the majority of people since data are collected based on a limited social group. Consequently, sampling bias still exists, and the data cannot represent the entire population despite the volume of big data (Boyd & Crawford, 2012). Figure 6 shows a report conducted by CNNIC reveals that over 80 percent of social media users are over 40 years old in 2015 (CNNIC, 2015). Consequently, young people comprise the majority of social network users, far from being a representative dataset for the entire population. If the dataset is used for scientific uses, the result is brought may not be applied to the entire population.

4.3 Emerging Issue 3 of the Ethical Framework: Social equity and the aging population

While the development of smart cities has benefited the majority of the citizens clearly, there are other groups of people who do not get benefits from those emerging technologies. As Wei (2015) discussed in her research, issues can occur when the disadvantaged groups are facing smart cities' development. The disadvantaged groups mainly include the aged or mid-aged population and people with low education. The fundamental feature of the disadvantaged group is that they face difficulties learning those new technologies and facilities in smart cities, either physiologically or psychologically. Since the construction of smart cities highly depends on ICTs, it would be hard for those people to learn emerging technologies, especially when they are not even able to use computers or smartphones. Consequently, people with physical or mental disadvantages may get little from technology development and be isolated from the new smart city lifestyle (Hwang and Choe, 2014).

A typical example of this ethical issue is the use of China railway online ticketing platform "12306.cn" since 2011. The platform has successfully enhanced the experience of purchasing tickets for lots of people. However, this platform becomes inaccessible for those people who are not able to use the internet. According to a Chinese government report, there are 485 million net users in 2011 in China (CNNIC, 2014). Compared to the Chinese population of 1.3 billion, around 63% of Chinese citizens could not use the online ticketing platform to buy railway tickets by themselves. They must have their friends or families to help them or use traditional approaches like offline purchases. Excluding those who have to manage to learn to use the internet, there should be quite a lot of people who can still not use the new platform. This makes citizens of different ages inequity of information and raises social sorting (Skouby et al., 2014).

4.4 Emerging issue 4 of the Ethical Framework: Public Transportation

In line with the privacy taxonomy by Solove (2006), ethical issues can arise during the data collection as well as data analysis. In the following chapter, the different data analytics layers are outlined specifically for smart public transportation (PT) and their ethical concerns.

4.4.1 Data Collection in Public Transportation

The more data is available and meaningfully correlated, the smaller is the share of unexplained phenomena and the more robust data analytics becomes (Pries, 2015). According to literature by Kim *et al.*, user data currently monitored by PT companies include their chosen fair charges, travel origins, travel time, choice of means, travel destinations and frequency of travel. The importance of the width of data collected will be further highlighted in the next chapter, discussing the data analysis techniques applied (Kim et al., 2017).

The span of data collection has been increased in recent years by applying new data collection technologies, called automated fare collection (AFC). AFC automatically collects fair price, origin-destination and time aspects. The cards for AFC differ between personalizing and non-personalized cards (Bahamonde et al., 2014). PT companies tend to foster the usage of personalized cards to increase their demographic data collection by offering discounts to personalized card owners (Lee & Hickmann, 2014). AFC operates through near field communication (NFC) cards, which are topped up and used by the customer to log in and out at his travel locations, automatically generating data sent to the PT information system (Fraga-Lamas, 2017).

However, due to various reasons, not all PT networks, not even within Europe, are utilizing AFC data collection. The PT companies, not applying the newest technology, rely on traditional means of data collection, including manual transcripts, human visual observations and customer interviews (Stelzer, 2016). While the new data collection technology has strong advantages over traditional methods in the amount and consistency of data gathered, it has drawbacks in customer privacy.

First, the AFC data collection method does not explicitly ask for customer consent for the continuous collection of customer data. Moreover, it does not educate the customers of the intended application of his data. This deficit increases in relevance when customers are disclosing private data through personalized AFC cards. Compared with AFC cards, traditional methods, such as face-to-face interviews, are more transparent means for data collection (Stelzer, 2016).

4.4.2 Data Analytics in Public Transportation

When collecting sufficient data for their data analytics operations, PT companies need to decide on the strategic fit for their analysis methods (Albright, 2017). The variation of different data analysis processes can briefly be separated into Analytical Processing (AP) and Transactional Processing (TP). AP uses holistic, cross-disciplinary and long-term data to forecast future or analyze past processes, such as designing a new railroad track or analyzing customer needs during peak hours. TP, on the other hand, deals with ad-hoc requests, such as current speed, number of passengers and expected arrival time of a train, which can be used to make short-term decisions on the current PT operations (Parekh & Arpit, 2013).

This document focused on AP, as they are the larger and more critical data analytics operation (Avoine et al., 2014). When PT companies are looking for data AP, they are mainly interested in matching their operations to the customer's needs in terms of service frequency (Stelzer, 2016), origin-destination needs (Lee & Hickmann, 2014) and fair pricing (Ma et al., 2017).

The most general data analytics case in PT the blending with behaviors. For example, a user has entered the PT system with the destination. He or she is leaving to observe the frequency of time-spatial patterns and investigate the consistency of chosen travel routes. The travel information gives the PT company the potential to adapt its routes to the most favorable destinations, align its service frequency with the service demand during peak hours and times of average demand and adjust prices for times and areas with higher demand (Kim et al., 2017).

Such data analytics can be carried out on the group and the individual level. When executed on the group level, individuals need to have little privacy concerns related to their user data. Analysis techniques marked by literature as more critical for ethical concerns are presented in the following section.

5 Recommendations of the Ethical Framework: To promote competency in making data-driven decisions

The ethical framework has been described to contain all the factors for big data development in high-growth, core competencies and innovation. These include the advancement of technology, competencies of big data analytics and innovation. Recommendations in this section can contribute to further developing the ethical framework that can promote users' and organizations' competency and improve their awareness, privacy and ethical concerns and practices. These recommendations are crucial for countries with lower ethical requirements or concerns, since they are valid and related.

5.1 Public awareness of privacy protection

The fundamental strategy of the ethical framework is to raise public awareness of privacy concerns by identifying and addressing ethical issues driven by big urban data and smart city technologies. Nowadays, big data economy has come into a culture that citizens tend to give out their data in expectation of a return for tangible assets (e.g., services, knowledge, or monetary rewards). However, the situation is that consent or notice of private data collection is pervasively absent in the data analytics process (Kitchin, 2016), and the 'return' is compulsory with no alternative. Moreover, the value of data most often exists in the data sharing process when data is monetized and traded between parties, which can be against individual interests. Consequently, ethical practice is required to make data users and communities aware of (1) the data collected from them, (2) additional information inferred from them, (3) the control and accessibility they have, and finally (4) any benefits from the data subsequently used. In the process of achieving these four aims, consent, full transparency and data service availability are required by data holders. Otherwise, this ethical vision is still far to reach. However, the goal of achieving an ethical practice should start by having a pervasive public awareness of information protection. Ethical issues can be addressed when people are aware of entities' behaviors towards their data and knowing what potential harm lies in their behaviors.

Next, according to data analytics regulation theory, consumers should be required to give their consent for data analytics. However, giving informed consent to data analytics activities becomes increasingly difficult for consumers, as the data collection and analytics technique becomes ubiquitous and less apparent. Therefore, current privacy regulations based on purpose specification, limited usage principle and the notice and consent model need to be extended towards flexible opt-in and opt-out options (Mantelero, 2014).

5.2 Enriching user datasets with external demographic and geographic data

The ethical framework can allow PT companies to emphasize the benefit of enriching or cross-analyzing their data with additional demographic datasets, from government data banks or by consent disclosed information for various reasons. On the one hand, it allows them to better know their customers and better cater to their needs (Goulet-Langlois et al., 2016) and on the other hand, it makes their available data more lucrative for third-party usage (Ma et al., 2017).

As shown in Figure 7, customer privacy-sensitive data is involved with four aspects: sales to external, internal usage, own data collection and those acquired by the third-party data. The most common datasets combined with PT user data are demographic datasets. In research presented by Avoine *et al.* (2014), independent researchers could estimate a statistically significant chance of belonging to a specific social group by combining government, education, or business registration records. Moreover, researchers were able to disclose their full identities when adding the government IDs linked to personified AFC cards, highlighting potential threats of the technology (Avoine et al., 2014).

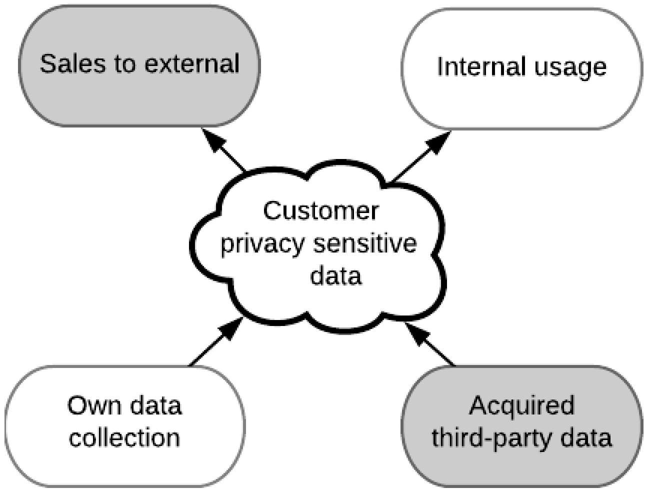

Figure 7 External data interaction

Potential de-anonymization becomes an even more serious threat to privacy when user data is combined with spatial data, such as available from government records. Researchers could statistically significantly connect PT access or exit points with the intended activity at the location, such as school, work, or other tasks (Lee & Hickmann, 2014).

When analyzing and protecting at the group level, the fatal consequences are limited; however, as outlined above, the de-anonymization of PT data is a real threat. Lee and Hickmann (2014) further displayed in their research how analysts, based on the PT enriched personalized, spatial and time data, could estimate the work and living place of individuals and analyze as well as forecast their absence from home and workplace. As well, they were able to determine the locations close to the next PT station (Lee & Hickmann, 2014). As discussed, the threat of losing privacy through PT user data is drastically increased if the individuals decide to use personified AFCs, which can be directly linked to vast official databases such as health, tax and housing records (Bahamonde et al., 2014).

The benefits from geographically enriched PT data are to improve urban living quality by finding mismatches in supported routes between living and working areas (Ma et al., 2017). Nevertheless, the possibility for enriched PT data to roughly allocate living space and determine the pattern in home absence and attendance offers strong privacy concerns, similar to the ones presented by Borsboom online at Pleaserobme.com.

Thieves and other illicit motivated can use such discrete information to exploit PT user's habits. The same threat has been found in extensive social network usage (Borsboom et al., 2010). While such activities cannot be expected to be carried out by PT companies, it can never be fully canceled out that data informally enters the wrong hands. Additionally, there have been cases where PT companies were looking to utilize their data analytics operations for additional revenues by selling insights to external companies. Further control over the data is lost, as discussed in the next sub-section.

5.3 Storage, management and security of public transportation data

Most papers outlined the needs and benefits of business analytics to draw personalized customer profiles better and cater to customer needs. However, PT companies can generate additional income from offering their business analytics services to third parties. Potential benefits listed by Ma *et al.* (2017) include marketing analytics services for retail companies interested in the exact demographics, passing stores in or around the PT station, by providing individual or grouped demographic information of its customers. PT companies can help retailers to find the right spots for their shops and help them specialize in their marketing tactics according to the PT station users.

Figure 8 shows threats from illicit sources that can negative to businesses explained as follows. While marketing analytics makes a lot of sense from a business perspective, it can be perceived as privacy infringement by PT users (Ma et al., 2017). Marketing analytics is discussed by research from other papers and outside the scope of this document. An ethical dilemma happens when PT companies decide to use their data analytics operations to generate income from third parties. Whether the analytics operations should remain in the hand of the PT company, which merely consulates the third party on its own findings, or whether the PT company should sell the raw data to the third party, in which case it loses the control of eventual illicit future processing (Ma et al., 2017). A well-known case is presented by Avoine *et al. (2014)*, in which the largest Japanese railway company JR was in discussions with the Japanese electronics manufacturer Hitachi to sell PT user data to the company in 2013. The deal was not executed, but it shows the economic potential of PT data sales to external parties (Avoine et al., 2014). The threats from third party data analytics on PT user data was discussed during multiple Japanese articles after the public opinion refused the JR-Hitachi case.

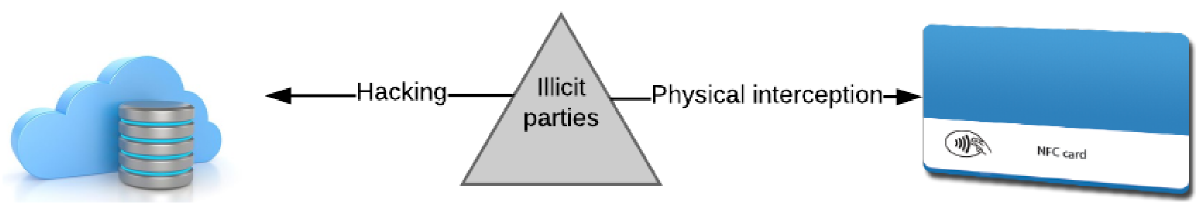

Figure 8 Threats from Illicit sources

Potential causes include control over individual travel behaviors, to test for correlated activities, as it was the case with one Tokyo Metro employee, who stalked his lover across the metro system (japanprobe, 2012). Other articles point to institutional users, who might be interested in controlling PT users without their consent, up to government surveillance (Newell, 2014).

However, stopping PT companies from selling user data to third parties does not entirely eliminate the risk of PT user data, reaching the hands of ill-intended analysts. The main threat of data leakage is the retrieval of data from AFC cards, which can be easily read out by spyware due to the weak hash codes, as presented by Kim *et al.* (2017). Moreover, the more data is stored on the AFC card, the more user data is exposed to potential theft.

If PT companies would store TP data only for short periods and only store AP data long-term, the amount of data potentially retrieved by malicious hardware can be reduced. AP data such as access and exit points and route choices, on the other hand, need to be long-term stored, to maintain analysis capabilities (Kim et al., 2017). TP data can consist of sensitive information such as card balance, top-up amount, payment method and does, unlike AP data, not need to be stored for long-term analysis (Parekh & Arpit, 2013).

The danger of data interception is even aggravated, as PT companies send all data transactions between AFC card and reader bundled. Therefore, potential attackers can steal all data in one set, increasing the harm to user privacy (Ma et al., 2017).

5.4 Ethical data analytics outside their own profitability

PT is a role model to other industries in terms of data analytics for business and process-oriented applications. However, it does not yet use data analytics outside its own business interest. If PT companies can apply their data analytic operations to help the society at large, it can further improve its utilization and improve its public image (Chessell & Mandy, 2014). Therefore, companies and governments should as well look to exploit humanitarian and charitable opportunities based on their existing data resources. The ethical framework can provide guidelines for everyone to follow. Suggestions are as follows. First, all businesses, institutors and governments should follow GDPR closely and carefully. Second, for any services that require privacy and anonymity or manage a large amount of user and public data, it needs consent from the users. Third, for highly sensitive data such as patient records or personal information of residents living in the smart cities, it should be managed and approved by the appropriate government departments and also the regional or national ethical committee. These practices can be adopted and regulated in cities with stricter control and a larger population such as Shanghai and Jakarta to balance the needs for GDPR, privacy and regulation.

6. Discussion

This section describes the benefits of the ethical framework.

6.1 Regulations on data ownership and data usage

Since the social transformation driven by technologies is rapid, regulations on big data privacy and ethical issues are still initializing (Rees, 2014). Rees (2014) argues that the law should adopt a proper approach to regulate ethical issues on data, which proprietary of data should be clearly defined. For example, data ownership should be to the data subject, and companies or governments who collect these data only have the right to use them. A data subject should have the right to control his/her data by authorization, usage, modification, and deletion. The right to opt and control one's own data should be essential to protect personal information and informed consent during data collection and usage processes.

Regulations on data usage are essential because data can be easily collected and used in a contemporary big data environment. This can be crucial for cities with large populations such as

Singapore and London and relevant for cities with stricter controls such as Shanghai. That is why the ethical framework can work with regulations as follows. First, the purpose of data collection should be explicitly stated by the data collector. For example, researchers should report the actual usage of the information collected to the data subjects so that data subjects are aware of the benefits, drawbacks and potential risks of data collection. Moreover, data collector should not collect those data which are irrelevant with the purpose of collection. Second, the right to data usage with big data should be explicitly regulated. With emerging technologies such as data mining and text mining, companies' and governments' decisions are supported more on massive data and big data analytics. Those technologies have intensively raised the safety risk of information and data subjects. Laws and regulations should be made on how deep the power of public entities (companies and governments) can step into those private areas and what kind of information should be shared by the public. Consequently, specific data protection institutes should be established to deal with issues brought by big data technologies.

6.2 Big data for high growth, innovation and core competencies

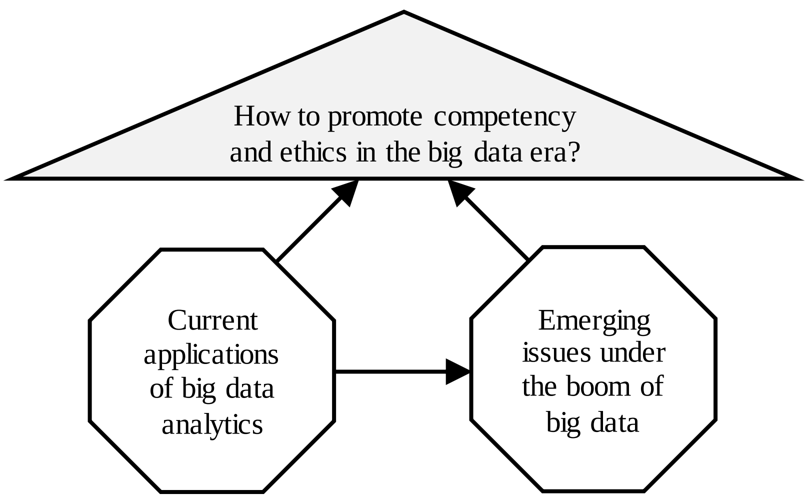

Figure 9 Triangular relationship of the Ethical Framework

The ethical framework can describe how our analysis can contribute to big data for high growth, innovation and core competencies. Figure 9 shows the triangular relationship of the ethical framework: current applications of big data analytics need to be developed and evolved. Hence, emerging issues have been raised. When both aspects are developed further, they can contribute to the competency and ethics of big data.

Additionally, smart cities fall into core competencies to enable better access to services, shorter response time to get our services completed and a platform for having more opportunities in businesses, transportation, education and healthcare. The high growth sector includes public transport since it can change how we use, interact, pay and receive goods and services. Innovation for technologies and methods involved in data analytics has been used to allow quicker access, a more comprehensive understanding of data usage, and better service quality. However, what has been missing in the existing literature is the rising privacy and ethical concerns due to all these changes. These destructive changes and issues can develop different types of challenges to be resolved. The ethical framework has discussed different points to minimize negative impacts by privacy and ethical concerns in Sections 3 and 4 and improve the user experience of adopting data analytics and public transportation. Figure 8 shows an ethical framework to help us maintain our work on the positive aspects and reduce the possibilities to get into the negative aspects of this.

Any individual, organization, third-party company and government should think and act very carefully on these three areas and revise their policies and practices if there are areas in breach or potential breach of ethical concerns. Current services and applications can help us improve efficiency and reduce costs of operation, but when we deal with personal data, we have to follow all the compliance and regulations in place.

6.3 Ethical Framework for COVID-19

Developed countries have currently developed apps and services to track the status of COVID-19 and nearby COVID-19 infected patients. This allows users to know any potential risk around them without revealing any personal identifier and information of the infected person (Dar et al., 2020). The ethical framework can be maintained by keeping the personal information of each user anonymous while revealing the crucial information posing risks to public health via contract tracing apps (Ahmed et al., 2020). The ethical framework can provide guidelines and recommendations for big data collection, usage, processing, analysis and management. This practice will be very relevant for any smart city in the world to maintain public health and safety without breaching regulations, privacy and ethics.

6.4 How the Ethical Framework can be used by smart cities

In this section, how the ethical framework can be used by six smart cities in the next few years, particularly for medical, transportation, education and COVID-19 emergency. A summary can be presented in Table 1.

Table 1: How the ethical framework can be used by smart cities

Cities and Areas to be applied	Medical	Transportation	Education	COVID-19 emergency
Singapore	It has the combined use of teleconsultation and controlled medical services to keep patients' data safe.	The government can manage all the data with the ERP (Electronic Road Pricing).	Face-to-face learning with social distancing is in place. This can be further developed.	Use apps to track status in real-time. Identities are anonymous.
London	NHS is to co-ordinate the patient care, record management and privacy.	The government manages all the data via CCTV and speed cameras mainly to track speeding and illegal parking.	Face-to-face learning with social distancing and online distance learning are in place. This can be further developed.	Use NHS apps to track status in real-time. Identities can be anonymous.

Paris	Teleconsultation is used to reduce transmission. Privacy can be further improved.	Similar to London but with lower percentages of CCTV. More focus on terrorism alert.	Face-to-face learning with social distancing and online distance learning are in place. This can be further developed.	Use apps to track status in real-time. Identities can be anonymous.
New York	Smart medical systems are available but not effective in controlling COVID-19. Privacy and ethical issues can be consolidated.	The government manages all the data via CCTV and speed cameras mainly to track speeding and illegal parking.	Distance learning with social distancing is in place. This can be further developed.	Use COVID Alert NY apps to track status in real-time. Identities can be anonymous.
Shanghai	Hospitals have patients;' records. Hospitals can improve ethical practices together with the government.	The government owns the data. Ethical practices can be consolidated.	Face-to-face learning with social distancing is in place. This can be further developed.	The government has health green checks and apps to track status since the start of the outbreak. Ethical practices can be made clearer.
Jakarta	It is in the process of developing smart medical systems.	It is in the process of developing smart transportation systems.	Have some forms of face-to-face learning with social distancing and online distance learning.	The government has developed apps to track status in real-time. Ethical practices can be made clearer.

For medical areas, the ethical framework can be effectively used for Singapore, London and Paris. The focus is to blend data privacy and ethical requirement with medical practices. Patients' records can be kept safe and anonymous and the practices can fulfill the ethical requirements recommended by the framework. New York is not coping well with an increasing number of COVID-19 patients and the ethical challenges are more towards the priorities for treatments. Shanghai has developed robust patient records with the government, but the ethical requirements can be made clearer. Jakarta is still under the development of smart medical systems.

For transportation, it is apparent that governments of smart cities take full control of data ownership and privacy. It is unclear what types of ethical practices and regulations are in place, but the policy and regulations can be blended with legal enforcement. The only exception is Jakarta, which is developing its smart transport systems. For education, all the smart cities are in the process of developing smooth delivery of lessons, such as face-to-face learning with social distancing and online distance education. Ethical practices can be enhanced by setting strict policies for not showing pictures and videos of children under the age of 18. All users on online learning platforms can follow regulations closely.

For the COVID-19 emergency, all the national and city governments of the smart cities have developed apps for their citizens to track COVID-19 status, news and updates in real-time. Even though it is unlikely to have full identities anonymous, apps from Singapore, London, Paris and New York can protect user identities from the systems. Although the security level in Shanghai is high, it is unclear to know what ethical practices are. For Jakarta, they have apps but do not seem to describe much on ethical practices. In summary, the ethical framework can be used as guidelines and recommendations for smart cities above. Citizens and policymakers can check and update before finalizing their ethical regulations. Policymakers can also improve their ethical practices regularly to ensure privacy, fairness and regulation for every user and citizen.

7 Conclusion

This paper proposed an ethical framework and provided a careful balance of benefits and risks driven by disruptive technologies for big data and smart cities. While big urban data and smart cities offered benefits for citizens in diverse areas such as public safety (Mendonca, 2016), transportation (Masek et al., 2016), health care (Aborokbah et al., 2017), education (Mothukuri et al., 2017), and governance (Hartswood et al., 2014), issues and concerns emerged as well. With a prudent consideration of the ethical problems and consequences driven by smart cities and big data analytics, the ethical framework synthesized pervasive discussions on individual privacy, data integrity and social equity. The recommendation to address ethical issues mainly arose from individual and government aspects. Individuals should increase public awareness of private information while policymakers could boost the pace of legislation on these emerging areas. The ethical challenges have become increasingly critical when technology and urban development advance. Those issues were not as serious in the last decade since it was difficult to obtain personal data. However, personal and sensitive data could be more easily obtained through third-party companies and service providers in recent years. The careful handling of data in compliance with privacy and data protection should be exercised, monitored and checked regularly. To ensure long-term success, a structured and organized ethical framework was developed, improved and disseminated along the way, and allowed all the policymakers to implement it for individual, organizational and societal levels.

During the development of smart cities and big data analytics, it was important to acknowledge that related issues and concerns should be addressed. Data analytics competencies in smart cities could be promoted when solutions were adopted to stabilize the progression process. While this paper has discussed the relevant ethical and related issues and recommendations on future directions, further studies are required in this area to ensure all stakeholders can benefit from this technology revolution, particularly big data for high-growth, innovation and core competencies. We also demonstrate how the ethical framework can be used by six smart cities in detail, focusing on medical, transportation, education and COVID-19 emergency, so that policymakers can effectively blend and improve their ethical practices and policies.

Future work will include three streams. First, it will focus on the study of smart cities in Singapore and London and the review of its services for public health, transportation, education and e-government, since these four areas can get a very high number of users. Second, the review of how the ethical framework can be used and applied in smart cities. Third, the improvement of smart cities and their service after the adoption of the ethical framework.

Acknowledgment

This research is supported by VC Research (VCR 0000003).

7 References

Abdel-Basset, M., Chang, V., & Nabeeh, N. A. (2020). An intelligent framework using disruptive technologies for COVID-19 analysis. *Technological Forecasting and Social Change*, 120431.

Aborokbah, M. M., Al-Mutairi, S., Sangaiah, A. K., & Samuel, O. W. (2018). Adaptive context aware decision computing paradigm for intensive health care delivery in smart cities—A case analysis. *Sustainable cities and society*, *41*, 919-924.

Adame, T., Bel, A., Carreras, A., Melià-Seguí, J., Oliver, M., & Pous, R. (2018). CUIDATS: An RFID–WSN hybrid monitoring system for smart health care environments. *Future Generation Computer Systems*, *78*, 602-615.

Ahmed, N., Michelin, R. A., Xue, W., Ruj, S., Malaney, R., Kanhere, S. S., ... & Jha, S. K. (2020). A survey of covid-19 contact tracing apps. IEEE Access, 8, 134577-134601.

Albright, S. C., Winston, W., & Zappe, C. (2010). *Data analysis and decision making*. Cengage Learning.

Alizadeh, T. (2017). An investigation of IBM's Smarter Cites Challenge: What do participating cities want?. *Cities*, *63*, 70-80.

Angelidou, M. (2014). Smart city policies: A spatial approach. *Cities, 41*, S3-S11.

Angelidou, M. (2016). Four European smart city strategies. Int'l J. Soc. Sci. Stud., 4, 18.

Avoine, G., Calderoni, L., Delvaux, J., Maio, D., & Palmieri, P. (2014). Passengers information in public transport and privacy: Can anonymous tickets prevent tracking?. *International Journal of Information Management*, *34*(5), 682-688.

Bahamonde, J., Hevia, A., Font, G., Bustos-Jiménez, J., & Montero, C. (2014). Mining private information from public data: The Transantiago Case. *IEEE Pervasive Computing*, *13*(2), 37-43.

Bakshy, E., Hofman, J. M., Mason, W. A., & Watts, D. J. (2011, February). Everyone's an influencer: quantifying influence on twitter. In *Proceedings of the fourth ACM international conference on Web search and data mining* (pp. 65-74). ACM.

Barocas, S., & Nissenbaum, H. (2014). Big data's end run around procedural privacy protections. *Communications of the ACM*, *57*(11), 31-33.

Batty, M., Axhausen, K. W., Giannotti, F., Pozdnoukhov, A., Bazzani, A., Wachowicz, M., ... & Portugali, Y. (2012). Smart cities of the future. *The European Physical Journal Special Topics*, *214*(1), 481-518.

Borsboom, B., van Amstel, B. & Groeneveld, F. (2010). *Please Rob Me.* [Online]

Available at: http://pleaserobme.com/

[Accessed 21 November 2020].

Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, communication & society*, *15*(5), 662-679.

Caragliu, A., Del Bo, C., & Nijkamp, P. (2011). Smart cities in Europe. *Journal of urban technology*, *18*(2), 65-82.

Center for Business Ethics. (1992). Instilling ethical values in large corporations. *Journal of Business Ethics*, 863-867.

Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: from big data to big impact. *MIS quarterly*, 1165-1188.

Chen, M., Pang, M. & Zhang, M. (2016). Study on departre plan of intelligent public transportation scheduling for urban suburbs based on MAST. *Journal of Hebei university of technology,* p. 80.

Chessell & Mandy, 2014. *Ethics for Big Data and analytics,* s.l.: IBM.

CNNIC, 2014. *China internet network development status statistical report,* s.l.: China Internet Information Center .

CNNIC, 2015. *Research report on social media user behaviors*, s.l.: CNNIC.

Dameri, R. P., & Rosenthal-Sabroux, C. (Eds.). (2014). *Smart city: How to create public and economic value with high technology in urban space*. Springer.

Dar, A. B., Lone, A. H., Zahoor, S., Khan, A. A., & Naaz, R. (2020). Applicability of mobile contact tracing in fighting pandemic (COVID-19): Issues, challenges and solutions. Computer Science Review, 100307.

Dodge, M., & Kitchin, R. (2007). The automatic management of drivers and driving spaces. *Geoforum*, *38*(2), 264-275.

Duhigg, 2012. *The New York Times Magazine*. [Online] available at: http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html?pagewanted=1&_r=1&hp

[Accessed 3 November 2020].

Fraga-Lamas, P., Fernández-Caramés, T. M., & Castedo, L. (2017). Towards the Internet of Smart Trains: A Review on Industrial IoT-Connected Railways. *Sensors*, *17*(6), 1457.

Gao, S., Liu, Y., Wang, Y., & Ma, X. (2013). Discovering Spatial Interaction Communities from Mobile Phone D ata. *Transactions in GIS*, *17*(3), 463-481.

Gardham, 2015. *Herald Scotland*. [Online] available at: http://www.heraldscotland.com/news/13215304.Controversial_face_recognition_software_is_being_used_by_Police_Scotland__the_force_confirms/ [Accessed 9 November 2020].

Glasmeier, A., & Christopherson, S. (2015). Thinking about smart cities.

Goulet-Langlois, G., Koutsopoulos, H. N., & Zhao, J. (2016). Inferring patterns in the multi-week activity sequences of public transport users. *Transportation Research Part C: Emerging Technologies*, *64*, 1-16.

Graham, S. (2011). *Cities under siege: The new military urbanism*. Verso Books.

Hartswood, M., Grimpe, B., Jirotka, M., & Anderson, S. (2014). Towards the ethical governance of smart society. In *Social Collective Intelligence* (pp. 3-30). Springer, Cham.

Hwang, J. S., & Choe, Y. H. (2014). Smart Cities Seoul: a case study ITU-T Technology Watch Report February 2013, ITU.

Herschel, R., & Miori, V. M. (2017). Ethics & Big Data. *Technology in Society*, *49*, 31-36.

Ibrahim, M. F. (2003). Improvements and integration of a public transport system: the case of Singapore. Cities, 20(3), 205-216.

japanprobe, 2012. *Pasmo cards used to probe extramartial affairs*. [Online] available at: http://www.japanprobe.com/2012/04/04/pasmo-cards-used-to-prove-extramarital-affairs/ [Accessed 31 October 2018].

Khatoun, R., & Zeadally, S. (2016). Smart cities: concepts, architectures, research opportunities. *Communications of the ACM*, *59*(8), 46-57.

Kim, J., Corcoran, J., & Papamanolis, M. (2017). Route choice stickiness of public transport passengers: Measuring habitual bus ridership behaviour using smart card data. *Transportation Research Part C: Emerging Technologies*, *83*, 146-164.

King, N. J., & Raja, V. T. (2012). Protecting the privacy and security of sensitive customer data in the cloud. *Computer Law & Security Review*, *28*(3), 308-319.

Kitchin, R. (2014). The real-time city? Big data and smart urbanism. *GeoJournal*, *79*(1), 1-14.

Kitchin, R. (2016). The ethics of smart cities and urban science. *Phil. Trans. R. Soc. A*, *374*(2083), 20160115.

Lee, S. G., & Hickman, M. (2014). Trip purpose inference using automated fare collection data. *Public Transport*, *6*(1-2), 1-20.

Liu, Y. (2015). Consumer protection in mobile payments in China. *Computer law and security review*, 31(5), 679-688.

Liu, J., Li, J., Li, W., & Wu, J. (2016). Rethinking big data: A review on the data quality and usage issues. *ISPRS Journal of Photogrammetry and Remote Sensing*, *115*, 134-142.

Lu, YH (2005). Privacy and data privacy issues in contemporary China. Ethics and Information Technology, 7(1), 7-15.

Luarn, P., Yang, J. C., & Chiu, Y. P. (2014). The network effect on information dissemination on social network sites. *Computers in Human Behavior*, *37*, 1-8.

Mantelero, A. (2014). The future of consumer data protection in the EU Re-thinking the "notice and consent" paradigm in the new era of predictive analytics. *Computer Law & Security Review*, *30*(6), 643-660.

Martínez-Ballesté, A., Pérez-Martínez, P. A., & Solanas, A. (2013). The pursuit of citizens' privacy: a privacy-aware smart city is possible. *IEEE Communications Magazine*, *51*(6), 136-141.

Masek, P., Masek, J., Frantik, P., Fujdiak, R., Ometov, A., Hosek, J., ... & Misurec, J. (2016). A harmonized perspective on transportation management in smart cities: The novel IoT-driven environment for road traffic modeling. *Sensors*, *16*(11), 1872.

Ma, X., Liu, C., Wen, H., Wang, Y., & Wu, Y. J. (2017). Understanding commuting patterns using transit smart card data. *Journal of Transport Geography*, *58*, 135-145.

Mendonça, M., Moreira, B., Coelho, J., Cacho, N., Lopes, F., Cavalcante, E., ... & Moura, B. (2016, September). Improving public safety at fingertips: A smart city experience. In *Smart Cities Conference (ISC2), 2016 IEEE International* (pp. 1-6). IEEE.

Meng, F., Liu, G., Yang, Z., Casazza, M., Cui, S., & Ulgiati, S. (2017). Energy efficiency of urban transportation system in Xiamen, China. An integrated approach. *Applied Energy*, *186*, 234-248.

Mothukuri, U. K., Reddy, B. V., Reddy, P. N., Gutti, S., Mandula, K., Parupalli, R., ... & Magesh, E. (2017, August). Improvisation of learning experience using learning analytics in eLearning. In *E-Learning & E-Learning Technologies (ELELTECH), 2017 5th National Conference on* (pp. 1-6). IEEE.

Newell, B. C. (2014). Technopolicing, surveillance, and citizen oversight: A neorepublican theory of liberty and information control. Government Information Quarterly, 31(3), 421-431.

Ong, R. (2015). Cyber-bullying and young people: How Hong Kong keeps the new playground safe. *Computer Law & Security Review*, *31*(5), 668-678.

Arpit, P. (2013). Introduction on Data Warehouse with OLTP and OLAP. *International Journal Of Engineering And Computer Science*, *2*(08).

Pries, K. H. D. R., 2015. *Big Data Analytics - A practical guide for managers.* Boca Raton, Florida: CRC Press.

Raza, A. (2016, August). LTE network strategy for Smart City Public Safety. In *Emerging Technologies and Innovative Business Practices for the Transformation of Societies (EmergiTech), IEEE International Conference on* (pp. 34-37). IEEE.

Rees, C. (2014). Who owns our data?. *Computer Law & Security Review*, *30*(1), 75-79.

Scuotto, V., Ferraris, A., & Bresciani, S. (2016). Internet of Things: Applications and challenges in smart cities: a case study of IBM smart city projects. *Business Process Management Journal*, *22*(2), 357-367.

Skouby, K. E., Kivimäki, A., Haukiputo, L., Lynggaard, P., & Windekilde, I. M. (2014, May). Smart cities and the aging population. In The 32nd Meeting of WWRF.

Solove, D. J. (2005). A taxonomy of privacy. *U. Pa. L. Rev.*, *154*, 477.

Solove, D. J. (2012). Introduction: Privacy self-management and the consent dilemma. *Harv. L. Rev.*, *126*, 1880.

Srivastava, S., Bisht, A., & Narayan, N. (2017, January). Safety and security in smart cities using artificial intelligence—A review. In *Cloud Computing, Data Science & Engineering-Confluence, 2017 7th International Conference on* (pp. 130-133). IEEE.

Stelzer, A., Englert, F., Hörold, S., & Mayas, C. (2016). Improving service quality in public transportation systems using automated customer feedback. *Transportation Research Part E: Logistics and Transportation Review*, *89*, 259-271.

Tasci, A. D. (2016). A critical review of consumer value and its complex relationships in the consumer-based brand equity network. *Journal of Destination Marketing & Management*, *5*(3), 171-191.

Thaduri, A., Galar, D., & Kumar, U. (2015). Railway assets: A potential domain for big data analytics. *Procedia Computer Science*, *53*, 457-467.

Vidgen, R., Shaw, S., & Grant, D. B. (2017). Management challenges in creating value from business analytics. *European Journal of Operational Research*, *261*(2), 626-639.

Vincent, J. (2013). *Independent*. [Online] available at: http://www.independent.co.uk/life-style/gadgets-and-tech/news/updated-londons-bins-are-tracking-your-smartphone-8754924.html [Accessed 11 November 2020].

Williams, B. (2015). *New York Times*. [Online] available at: https://www.nytimes.com/2015/08/13/us/facial-recognition-software-moves-from-overseas-wars-to-local-police.html [Accessed 11 November 2020]

Wu, Y., Zhang, W., Shen, J., Mo, Z., & Peng, Y. (2018). Smart city with Chinese characteristics against the background of big data: Idea, action and risk. Journal of Cleaner Production, 173, 60-66.