



Università
Ca' Foscari
Venezia

Master's Degree Programme

In Languages, Economics and Institutions of Asia and North Africa

“Second Cycle (D.M.) 270/2004”

Final Thesis

The Exploitation of Big Data Technologies in Chinese Retail

Towards the Creation of a Regulatory Framework for Privacy Protection.

Supervisor

Ch. Prof. Renzo Riccardo Cavalieri

Assistant supervisor

PhD Student Diego Todaro

Graduand

Federica Stortoni

Matriculation number 856935

Academic Year

2019 / 2020

前言

过去 30 年中，数字技术的发展导致数据和信息社会的相关性以及在线信息的可用性不断提高。该网络构成了一个新的数据矿山（无论是个人数据还是其他数据），它们的提取可以创造价值 and 新的知识水平。恰恰是处理个人数据的现代技术可以实现公共和私人参与者的上述目标。一方面，数据处理在信息社会的发展中起着关键作用；另一方面，它对保护隐私和个人数据提出了复杂的挑战。讨论的主题是大数据现象。

这项工作旨在分析应用于中国零售业的大数据现象对保护消费者的隐私和个人数据带来的挑战，之前确定了可以突出现象主要特征的大数据定义并关注特别是关于通过大数据分析技术进行数据处理的作用。

投资中国零售业的数字革命是史无前例的：2013 年，中国电子商务在规模和利润方面均已超过美国，但仍保持着世界纪录。中国电子商务生态系统是一个创新孵化器。多年来，由于智能手机、电子支付和面部识别支付、二维码的出现，以及包括 KOL 和 KOC 在内的新影响力者的诞生，它对传统的购买方式进行了许多改变，如移动商务的发展。

对于中国消费者而言，零售领域的创新也在发生：他们不再只从零售商那里寻求实质性产品，而是希望拥有一种既能娱乐又有趣的体验，并能增加购买过程的价值。这种新现象由阿里巴巴创始人马云（Jack Ma）定义，名称为“Entertainmerce”，对公司而言意味着对营销和与客户沟通技术的改变。实际上，零售商必须根据消费者需求调整报价，根据每个人的喜好和口味定制产品。通过这种方式，可以更改其业务方法，从而为生产和产品沟通创建“以客户为中心”的模型。

本文的第一章着重分析了中国零售界发生的技术变革。提供其当前特征的概述，并描述从简单的客户到公司生产和决策过程不可或缺的部分，消费者角色的转变。此外，我还将在本章中强调技术领域的三大巨头（百度，阿里巴巴，腾讯）在中国经济中的作用：由于使用了分析技术，这三家公司有机会充分了解客户的消费历程，并能够实时利用各种接触点上可用的大量信息，例如社交媒体，物联网，电子商务平台等。

大数据代表零售业的未来，要在这个不断发展的数字现实中生存并取得成功，再加上人工智能的使用，改变了中国零售业的世界，使阿里巴巴，百度和腾讯等公司成为了零售业的领导者。由于其庞大的注册用户数量，每日流量和交易量，其行业严重依赖于消费者的个人数据来开发量身定制的创新服务，并跻身全球十大互联网公司之列。我将特别描述腾讯微信社交平台的现象。实际上，这标志着“多合一应用程序”系统的出现，成为一种数字超级应用程序，用户可以在不离开该应用程序的情况下进行各种活动。这个新系统令人担忧的方面是，无数用户数据和信息集中在一家公司手中，这加剧了中国人对在线隐私保护的日益关注。

因此，大数据就像一把双刃剑，这使社区受益于它们的使用，但隐私也很容易受到侵犯。大数据是社会经历过的最大的隐私和数据保护挑战之一：如此之多的个人信息（PI）从来没有像现在这样自由。这就是为什么当今对于了解大数据的功能以及公司可以利用它们的用途极为重要，以使用户了解这些新技术的风险和收益并动员起来保护其新技术。个人信息，意识到他们在网上共享的内容。

因此，在第二章中，将详细分析公司利用大数据技术及其功能的现象，为该术语提供详尽的定义并描述其主要特征。零售商如何利用通过各种接触点收集的这些信息，以及它们可能给零售商和消费者带来的利益和风险。确实，尽管零售商需要为合格的人员和充足的技术资源进行巨额投资，但为了获得更好的战略决策，更好的操作流程控制以及对客户的了解和降低成本，消费者却获得了独特且个性化的服务。根据他们的喜好提供服务，从而促进他们的购买过程。

但是要付出的代价是什么？他们的隐私被侵犯了。

因此，第三章将重点讨论隐私问题：首先，简要介绍一下中国对隐私的看法，因为中国是儒家传统，由共产党统治数十年，所以与西方国家有所不同。如果提供中国的文化起源，则对隐私的感知会受到儒家价值观所描绘的严格等级制社会的影响，这种社会期望个人尊重家庭单位的需求；此外，历史历史因素也影响了个人与社会的关系，因为毛泽东时代的中共政府想要建立一个集体社会。

社会信用体系的问题代表了习近平目前在政治领导层的社会控制中与过去的连续性的一个例子，并说明了中国害羞的隐私不同于西方独立和权力概念的原因。决定权（无论是否共享个人信息），但更多地被认为是保护个人信息免受人员和公司侵害，而不是受到政府干预。

最后，将提供对该问题固有的当前监管框架的描述。

在中华人民共和国，还没有全面的数据保护法。自 2017 年起实施的《中华人民共和国网络安全法》是有关保护个人数据和数据安全的最全面的法律。尽管它遵循其他国际立法（例如欧洲 GDPR 或美国的《消费者隐私法案》）的类似原则，但它仍然具有一些中国特色，确实引起了国内（但主要是外国企业）对解释的认识和怀疑。

政府正在表现出维护公民权利的兴趣，因为各种法律（例如《电子商务法》）中还有其他与隐私保护相关的实施法规和条款，这些法规和条款是仍在发展中的复杂立法框架的一部分。

INTRODUCTION

Since 1979 China has undergone profound changes that have modified its economic structure forever. After decades of fast-paced economic growth, in the 2000s the increasing penetration of the Internet in China, especially mobile Internet, has led to an unprecedented digital revolution, so much that the retail sector has been projected into the online world, giving life to a unique commercial reality: Chinese e-commerce already in 2013 surpassed that of the United States both for size and profits, grabbing the world record.

With the digitization of purchasing processes there have been numerous innovations: just think of the development of mobile commerce thanks to the advent of smartphones, e-payments and facial-recognition payments, QR codes, the birth of new influencers including KOLs and KOCs, online shopping events like Alibaba's hugely popular Single's Day and more. All this meant that even the Chinese consumer underwent a radical change: in an era in which everything is spectacularized and appearance has become the axiom of human existence, consumers no longer seek a simple material product from retailers but rather they ask for an experience that entertains them and that is customized according to their tastes and preferences. Only in this way companies today could hope to establish a bond of trust with the Chinese consumer.

However, to achieve this result, retailers have had to adapt by changing their traditional business and marketing strategies as well, creating a purely "customer-centered" model for both production and product communication. All this with the indispensable help of analytical technologies that have allowed them to fully understand their customers by adapting the offer to the demand and enabling them to leverage the big amount of data and information available from multiple touchpoints, such as Social Media, IoT, and e-commerce platforms, in real-time.

Big Data, represents the future of Retail Industry to survive and succeed in this ever-advancing digital reality, and combined with the use of Artificial Intelligence have changed the world of Chinese retail, allowing companies such as Alibaba, Baidu and Tencent, leaders in the sector and whose businesses depend heavily on consumers' personal data, to develop tailor-made and innovative services and to be in the ranking of the world's top ten internet companies, thanks to their huge number of registered users, daily traffic and trading volume.

Therefore, the thesis aims to analyze the phenomenon of the increasingly indispensable use of Big Data for companies in the retail world, due to the need to anticipate and satisfy the demands of their consumers, as Big Data have become essential tools for the acquisition of valuable information on the habits, tastes and characteristics of each user. In the first chapter an overview of the situation of the retail sector in China, which stands out in the world for its rapid and outstanding digital evolution, will be provided, describing briefly its innovative structure and focusing at last on the very unique phenomenon brought about by the Tencent-owned platform

WeChat: the “all-in-one app” or “Super-app” system, which is tailor-made for Chinese consumers, as they rarely use the search engine or directly go to the official website to make their purchases, enabling them to experience many services optimizing their time, and providing huge amount of insights to the company. In this way also raising awareness of personal privacy protection, as so much power is concentrated in just a few hands.

In the second chapter the function and phenomenon of companies leveraging Big Data technologies will be analyzed in detail, providing respectively the definition of personal data and sensitive personal data, how these information collected through the various touchpoints (social networks and e-commerce platforms) are exploited and the benefits and risks that they can cause to both retailers and consumers. Indeed, while retailers need to make a huge investment both for qualified personnel and adequate technological resources, in order to gain better strategic decisions, improved control of operation processes, as well as better understanding of customers and cost reduction, the latter receive a unique and personalized service, in line with their preferences, thus facilitating their purchase process.

But what is the tradeoff? The violation of their privacy.

Therefore, the third chapter will focus on the issue of privacy: at first, a brief introduction will be made on the perception of privacy in China, which differs from the Western world as China is a country with Confucian heritage and governed under decades of Communist leadership. Firstly, a brief analysis of the concept of privacy will be taken starting from a cultural point of view, underlying the submission of the needs of individual respect to the family unit under Confucian values and strictly hierarchical society; moreover, the historical roots have influenced the individual self, as the government of Mao, the stringent principles of the CCP were about to establish a collectivistic society, in which individuals were deprived

from private property right and the commitment to the public or to the greater good of the country was an expected virtue.

The analysis continues with an hint of the social control exercised by the present political leadership of Xi Jinping, which designates a kind of continuity with socialist and Confucian values, as demonstrated by the launch of the Social Credit System, internet censorship and propaganda, and provides an idea of the influence that may have on the concept of privacy, considered in China more as a protection of personal information from people and companies rather than from government interference.

At last, a description of the current regulatory framework inherent to this issue will be provided. Although a comprehensive data protection law in the People's Republic of China does not exist yet, the Government is demonstrating its interest in safeguarding citizens' rights: the rules relating to the protection of personal data and data security are best collected under the 2017 Cybersecurity Law of, even though many other implementing regulations and articles in various laws, such as the E-commerce Law, take part of a complex legislative framework still under development.

Table of Contents

前言	1
Introduction	4
Chapter 1 DIGITAL TRANSFORMATION IN CHINA AND THE EVOLUTION OF E-COMMERCE RETAIL.....	9
1.1 E-commerce market overview	9
1.1.1 Chinese consumers seek “Retailtainment” and a personalized shopping experience	10
1.1.2 Internet in China and Firewall	12
1.1.3 Main technological players: Baidu, Alibaba & Tencent	15
1.2 Mobile commerce and Social Commerce	17
1.2.1 The Chinese Super app: WeChat	19
Chapter 2 THE USE OF BIG DATA FOR COMMERCIAL PURPOSES	22
2.1 Definition of Big Data	23
2.1.1 Personal data & Sensitive personal data.....	27
2.2 How Retail companies use Big Data: customers’ profiling	28
2.2.1 Collection & Processing of data	29
2.2.2 Analysis & Targeting of data	34
2.3 Pros	38
2.3.1 Pros for companies	38
2.3.2 Pros for consumers	43
2.4 Cons and challenges.....	45

Chapter 3 CHINESE LEGAL FRAMEWORK FOR INTERNET PRIVACY PROTECTION	49
3.1 The concept of privacy in Chinese society: Are Chinese consumers concerned about privacy?	49
3.1.1 Cultural point of view (Confucian values)	50
3.1.2 Historical background	52
3.1.3 Social control	53
3.2 There is not a unique and comprehensive law for the privacy issue	53
3.2.1 The Cybersecurity Law of the People’s Republic of China	54
3.2.2 Implementing Regulations and Guidelines	59
Conclusions	64
References	68
List of Laws and Regulations	76

CHAPTER 1 – DIGITAL TRANSFORMATION IN CHINA AND THE EVOLUTION OF E-COMMERCE RETAIL

1.1 E-commerce Market Overview

In 1978, China started its economic reform and Open-door policy. Since then, its economy has been developing rapidly, and the national income and per capita disposable income have been increasing steadily. As a result, Chinese residents were getting richer and becoming more able to spend and to consume, giving the retail market a strong boost.¹

Simultaneously, responding to the new consumption patterns and transforming peoples' consumption needs, the retail sector in urban China has undergone profound structural changes in the past 15–20 years.

Rising consumerism in urban China has featured the increasing complexity of consumer culture, which has been accompanied by status-seeking behaviors and a desire for fashion, stylish, and positional goods that are not just for practical use. For more and more urban residents, shopping is no longer an unavoidable search for daily necessities but has become a social pastime and an increasingly important avenue for leisure.

Meanwhile, technological advancements in China provide foundations for new retail formats to flourish: China's improved physical infrastructure enables substantial development in distribution and logistics while developments in technology have fueled the growth in the availability of information online, the proliferation of social media platforms, and the growth of e-commerce. All these factors have led to consumers increasingly spending their time online and discovering the convenience of online shopping.²

In 2013 Chinese retail market has overcome for the first time the giant US market, ranking first and placing ahead of its main competitors like the UK and Japanese markets and, with the most Internet users of any country, China became the world's largest and fastest growing e-commerce market.³

¹ Teresa LAM, Christy LI, Echo GONG, *The Changing Face of China's Retail Market*, Fung Business Intelligence Centre, September 2014, pp. 4-8.

² Richard D., Chen Y., Gordon O., James M., Michael C., Elsie C., *China's E-tail Revolution: Online Shopping as a Catalyst for Growth*, McKinsey Global Institute, March 2013, pp.27-33.

³ , Mel F ZUBERI, Daniel RAJARATNAM, "Measuring retail performance in an omni-channel world", *Journal of Marketing Channels*, 26, 25th March 2020, Pages 120-126.

The Chinese population’s growing wealth and the rapid development of e-commerce have driven the country’s epic retail boom. Indeed, the Chinese costumer’s market, in all its vastness and vibrancy, is no longer stranger to anyone: in 2019, the total retail sales of consumer goods reached 41,164.9 billion yuan, a growth of 8.0 percent over the previous year. An analysis on different areas, conducted by the National Bureau of Statistics of China, showed that the retail sales of consumer goods in urban areas stood at 35,131.7 billion yuan, up by 7.9 percent, and that in rural areas reached 6,033.2 billion yuan, up by 9.0 percent.⁴



Figure 1 Total Retail Sales of Consumer Goods 2015-2019

1.1.1 Chinese Consumers Seek “Retailtainment” and a Personalized Shopping Experience

Today that the world has entered a loop of rapid change and innovation because of globalization and rising competition among companies, the relationship between consumers and sellers has changed as well, in accordance with the new trends of the moment.

Actually, the new trend that has been shaping the retail market is typically consumer-centered, meaning that companies and brands are focusing on consumers’ needs, requests and care

⁴ National Bureau of Statistics of China, *Statistical Communiqué of the People's Republic of China on the 2019 National Economic and Social Development*, February 28, 2020. http://www.stats.gov.cn/english/PressRelease/202002/t20200228_1728917.html

about their feedbacks that they can easily capture through a deep analysis of the traces they left on various online and offline touchpoints. Companies feed themselves with sensitive data coming from their clients and use them to adapt and perfectionate a specific offer for each single individual.

While in the past companies were used to promote their products using marketing campaigns hoping to reach the largest number of people possible and convince them to buy their products; nowadays we are witnessing a radical change mainly due to the evolution of Chinese consumers, eager to receive high-quality services and most importantly a shopping experience that could entertain them and could be adapted and personalized according to their personal tastes and peculiarities.⁵

The category of population that mostly embraced this phenomenon is the generation of the so called “post-millennials”, *linglinghou* 零零后 in Chinese, the smartphone-wielding group of youngsters born in and after the year 2000. In China they form a key pillar of consumer spending and are most likely to spend on digital entertainment, including videos, music as well as information, followed by food, travel and e-commerce, as they look at the purchase process as an amusement. They were born in an era where everything is spectacularized and appearance is the axiom of their existence; that explains the reason why they seek fun from the world and pretend to live an exclusive experience while enjoying a service.

According to the South China Morning Post (SCMP), this generation represents the driving force for the explosive growth of short-videos and live-streaming apps, like the most popular one Douyin 抖音, known as TikTok outside mainland China, where both users and brands are allowed to share their contents through a fifteen-seconds short-video accompanied by music and special effects and which is having great success especially among its younger users (80 percent of them aged under 30).⁶

⁵ McKinsey China Digital Consumer Trends 2019

<https://www.mckinsey.com/~/media/mckinsey/featured%20insights/china/china%20digital%20consumer%20trends%20in%202019/china-digital-consumer-trends-in-2019.ashx>

⁶ Zen SOO, *China's 'post-millennial' generation is different. Here's how*, in “South China Morning Post”, August 17, 2020, <https://www.scmp.com/tech/innovation/article/2160003/chinas-post-millennial-generation-different-hereshow>

Known with the acronym of “Retailtainment”⁷ or with the term coined by Alibaba, “Entertainmerce”⁸, the new e-commerce model highlights the increasing expectations of customers from brands. As a matter of facts, modern consumers demand more from brands than ever before: they expect personalized messaging and offers and want to buy from brands that align with their unique personalities and value-sets. Perhaps most importantly they no longer have patience for disparate and irrelevant experiences.

In an increasingly competitive marketplace, brands are focusing on creating more impactful relationships with their consumers because, if they hope to succeed, they must engage their audiences more effectively, promoting loyalty. Marketers have understood the need to take a more customer-centric approach, and to get into the granular detail of their customers’ behavior online, in order to create better long-lasting relations with them, creating also trust in the brands.

1.1.2 Internet in China and Firewall

The retail sector has experienced a progressive change over the years due to the rise of the role and use of technologies and its development has been symbiotic to consumers’ behavior and changing in needs. The advent of the Internet has deeply modified the structure of its retail market, causing the shift of the shopping experience from the brick-and-mortar shop to the online world, where every limit- of time and space- have been removed.

China is a market that has that embraced the internet like no other, thanks to the incentives of Chinese government to implement the quality and availability of the Internet all over the Country, in a bid to build up China’s strength in cyberspace and IT. On March 24th 2015, Premier Li Keqiang first encouraged the State Council to introduce the “Internet Plus” (互联网+) action plan, aiming to spoor innovation and economic growth, through the integration

⁷ *Retailtainment* is a marketing concept introduced by an American sociologist, George Ritzer, in 1999; he defined it as “*the use of sound, ambience, emotion and activity to get customers interested in the merchandise and in the mood to buy.*” (from: <https://medium.com/@megaXcoin/retailtainment-retail-is-notabout-selling-products-9a169c344701>)

⁸ *Entertainmerce* is a term used to define a new form of e-commerce, where the e-commerce model merges with entertainment. It was coined by the Chinese firm Alibaba when talking about its “see now, buy now” fashion show on Tmall platform, where consumers could buy in real time the latest designer clothes before the runway show even finishes.

of the mobile internet, cloud computing, Big Data and the Internet of Things with modern manufacturing, in order to encourage the healthy development of e-commerce, industrial networks and internet banking, and to guide internet-based companies to increase their presence in the international market. In other words, the main goal of the “Internet Plus” strategy is to make Internet a fundamental pillar of economic and social development by 2025.⁹ As per this objective, China has kept providing more access to the Internet increasing the network speed and reducing access charges, ensuring a high level of security while surfing the net, thus achieving remarkable results and a great consensus among the population.¹⁰

Indeed, according to the 44th Statistical Report on Internet Development in China provided by the *China Internet Network Information Center (CNNIC)*, China enjoys the highest concentration of internet users of the world, covering almost the total netizens of the country.¹¹

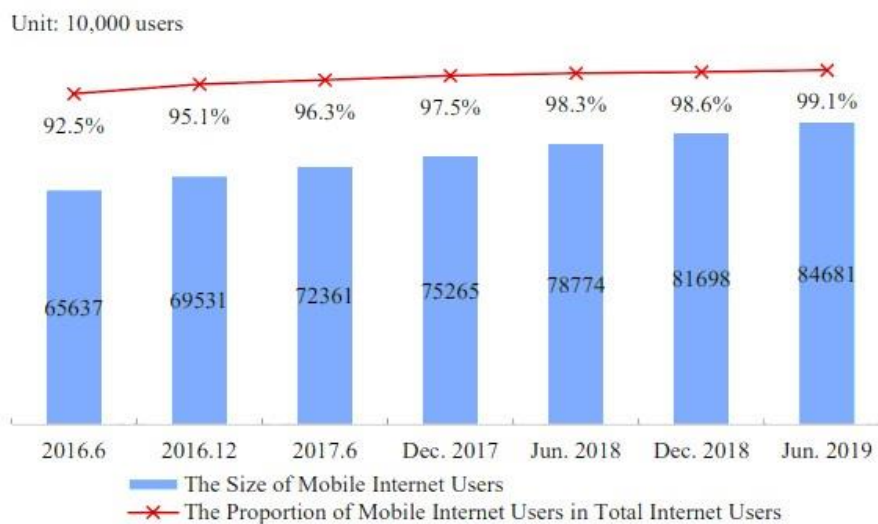


Figure 2 The Size of Mobile Internet Users in China and Its Proportion in Internet Users (Source: CNNIC Statistical Survey on Internet Development in China)

⁹ Jingsu WANG, The State Council issued the "Guiding Opinions" on Actively Promoting the "Internet +" Action, in "Xinhua net", July 4th, 2015.

¹⁰ WU Hequan 邬贺铨, *shi shenme rang women de shenghuo ruci bianjie - zhongguo hulianwang fazhan de 25 nian* 是什么让我们的生活如此便捷——中国互联网发展的 25 年(What makes our lives so convenient - 25 years of development of the Internet in China), *纵横*, Issue 10, pp. 20-23, 2019. <http://www.cqvip.com/QK/81392X/201910/7003149296.html>

¹¹ "As of June 2019, China had 854.49 million netizens, up by 25.98 million from the end of 2018, and its Internet penetration had reached 61.2%, up 1.6 percentage point over the end of 2018." From: CNNIC 44th Statistical internet Survey.

Among all the Internet users, the proportion of those accessing the Internet through mobile phones increased from 98.6% at the end of 2018 to 99.1%, reaching 846.81 million mobile Internet users in 2019.¹²

As data explain, mobile internet has one of the fastest penetrations in China. Its adoption is more than three times as compared to the US.¹³ These data show the popularity of the Internet in the country and the development it has undergone in few years.

But it is important to notice that the popularity of Chinese platforms and websites is a phenomenon encouraged by the fact that for China the internet is more like an *intranet*¹⁴. Indeed, it is largely walled off from the Western world by the incredible complex system of filters and blocks, represented by the Great Firewall¹⁵. From a political point of view, the Great Firewall censors content or sites the Communist Party that could mine the national security of the country, feeding the risk of protests or act of rebellion against the Government, but despite its political aspect, it has also functioned as an economic shield for the state to reserve its domestic market for home-grown players. These companies have grown so much that have gained global competitiveness as well, becoming sources of innovation themselves.¹⁶ The main example of the huge step made by Chinese platforms in terms of innovative technologies and services is the Tencent owned super app WeChat, that will be further analyzed in chapter 1.2.1.

¹² Ibidem.

¹³ Pradeep MENON, *New Retail: How Big Data and AI is Fueling Retail Transformation*, in “Medium.com”, August 6, 2019, <https://medium.com/@rpradeepmenon/new-retail-how-big-data-and-ai-are-fuelling-retailtransformation-76c81105c82d>

¹⁴ A national intranet is an Internet protocol-based walled garden network maintained by a nation state as a national substitute for the global Internet, with the aim of controlling and monitoring the communications of its inhabitants, as well as restricting their access to outside media. (from https://en.wikipedia.org/wiki/National_intranet)

¹⁵ *The Great Firewall* is the combination of legislative actions and technologies enforced by the People's Republic of China to regulate the Internet domestically. Its role in Internet censorship in China is to block access to selected foreign websites and to slow down cross-border internet traffic. The effect includes: limiting access to foreign information sources, blocking foreign internet tools (e.g. Google search, Facebook, Twitter, Wikipedia, and others) and mobile apps, and requiring foreign companies to adapt to domestic regulations. (from https://en.wikipedia.org/wiki/Great_Firewall)

¹⁶ Hong SHEN, China's Tech Giants: Baidu, Alibaba, Tencent, in “Digital Asia”, Panorama, pp.33-34. https://www.kas.de/documents/288143/4843367/panorama_digital_asia_v3b_HongShen.pdf/a21ab7b98e37-acfa-19a3-955a5881088f

1.1.3 Main Technological Players: Baidu, Alibaba & Tencent

With the rise of a group of powerful internet companies the Chinese internet has become the center of global attention.

Therefore, some big companies have been the direct beneficiaries of the persistent government support and censorship policy and, taking advantage of this digitalization phenomenon, have made this trend their basic value for their businesses. Named with the acronym of BAT, which stands for Baidu, Alibaba and Tencent, these three companies are leaders in search engines, retail e-commerce market, and social media and mobile gaming respectively, and have changed the way of doing shopping by boasting the use of new technologies and big data analysis at the core of their business strategies.

However, BATs are more than just search, e-commerce or social media companies. Indeed, over the past two decades, they have each developed an extremely complicated digital empire, evolving into multifaceted tech platforms.¹⁷

Considering for example the case of Alibaba (阿里巴巴集团) company, although in its early stages described itself as “the largest online and mobile commerce company in the world”¹⁸, ecommerce represents a part of its market. Actually, Alibaba’s massive corporate system constitutes not only its core in commerce (both online and offline), but also the supporting layer of logistics, payment and finance, cloud computing and consumer services, and the outermost layer that extends from media and entertainment to healthcare and even automobile manufacturing. The Company does not just focus on its flagship sector but aims to play important roles in other fields as well, making “innovation” the main value of the company.

Its main businesses in the retail sector, including China’s largest consumer-to-consumer (C2C) mobile commerce platform, Taobao 淘宝网, and Tmall 天猫, which has become the world’s largest third-party online and business-to-consumer (B2C) mobile commerce platform for brands and retailers, are guided by the same desire for innovation, providing consumers with engaging and personalized shopping experiences, optimized by big data analytics.

¹⁷ *Supra* note 16, pp.34-36.

¹⁸ Alibaba Group, SEC Form F-1/A, 100.

Alibaba gained enormous popularity among people and that is due to its entertaining and captivating tools, such as live-streaming sponsorships, gaming and live shows. Indeed, since 2009 Alibaba has kept scoring extraordinary results from its main 24h event held on the 11th of November every year, known as the Single's Day or Double Eleven¹⁹.

In the battle of e-commerce shopping holidays, Alibaba's Single's Day has overcome the Black Friday and Cyber Monday in terms of online sales. This year, Alibaba Double Eleven achieved another sales record- of 38.4 billion US dollars in 24 hours, a 26% increase over 2018's total of 30.08 billion US dollars, selling more than half of what Amazon sells in an entire quarter, confirming again its position as the largest online shopping day in the world.²⁰

Finally, it is worth mentioning that with more than 1 trillion US dollars of Gross Merchandise Volume (GMV) transacted in the Alibaba Digital Economy for fiscal year 2020²¹, Chinese ecommerce giant Alibaba is not only an intermediary connecting companies and customers in China and all over the world, but it is the main driving force behind the success of e-commerce.

In a similar vein, Tencent 腾讯, since 1998 has invested heavily in talent and technological innovation, actively participating in the development of the Internet industry. Chinese communications and social platforms Weixin 微信, also known as WeChat, and QQ connect users with each other, with digital content and daily life services in just a few clicks. The adoption of its high-performance advertising platform has helped brands and marketers reach out to hundreds of millions of consumers in China, enhancing the communication between consumers and retailers, and the birth of a new, innovative way of doing shopping: The Social Commerce.

¹⁹ The celebration began among students at China's Nanjing University around 1993 as a sort of anti-Valentine's Day. The most widely accepted theory is that four male students ("all single men") discussed how they could break away from the monotony of having no significant other and agreed that November 11 would be a day of events and celebrations in honor of being single. Thanks to Alibaba, Singles' Day is now a holiday celebrated in China on November 11 when unmarried people commemorate the occasion by treating themselves to gifts and presents.

<https://www.investopedia.com/terms/s/singles-day.asp>

²⁰ Sergei KLEBNIKOV, *Alibaba's 11/11 Singles' Day By The Numbers: A Record \$38 Billion Haul*, in "Forbes", November 11, 2019, <https://www.forbes.com/sites/sergeiklebnikov/2019/11/11/alibabas-1111singles-day-by-the-numbers-a-record-38-billion-haul/#678a3b272772>

²¹ Alibaba Group: Fiscal Year 2020 Annual Report.

1.2 Mobile Commerce and Social Commerce

Chinese users rarely use search engines to access online platforms, instead they land there through a myriad of contacts or touchpoints, that they basically catch through social media apps and mostly while using their mobile phones. They are not used to get directly on a specific brand's website to get to know its products, instead they usually get there when doing something else, like chatting with friends, watching videos or playing games. As shown in the image below, most of them spend their time on social media platforms and just 4 percent of them go directly on the web for shopping purposes.

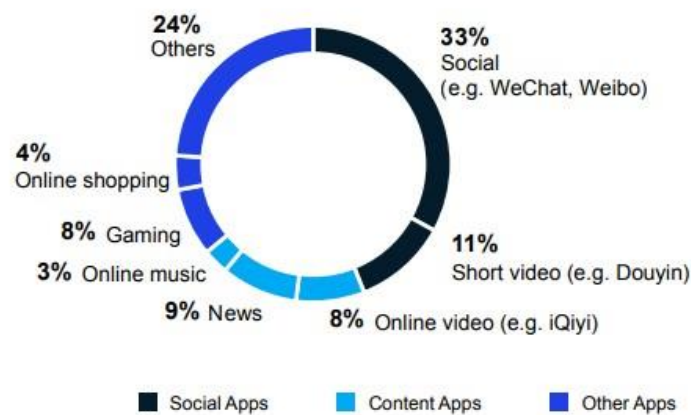


Figure 3 Percentage of Time Spent by Digital Consumers on Different Online Applications & Activities
(Source: QuestMobile; McKinsey China Digital Consumer Trends 2019.)

Social platforms act as link for buyers and sellers, as brands are converting the awareness they generate on social media into purchases, and awareness is converting into sales.

Therefore, as of June 2019, 99.1 percent of people accessed the internet through their mobile devices and almost 56 percent of e-commerce sales are completed on a mobile device, compared to the 35 percent of the US. This means that Chinese e-commerce took on the characteristics of a mobile commerce²², given the popularity of purchases made through

²² Mobile commerce refers to e-commerce conducted over mobile devices through wireless networks; while Social commerce uses social media for business activities through social interactions and users' contributions;

smartphone devices. Indeed, almost three out of every four of China's e-commerce transactions are carried out on a mobile device, generating \$873.3 billion in sales.²³

It is precisely from this trend that China experienced the birth of a new way of doing shopping, through the use of social applications, becoming itself a social experience.

As a matter of fact, Chinese consumers are voracious users of social media. According to a research led by the McKinsey Consulting Group, they spend as much as 44 percent of their time on social media apps, the majority of which, 33 percent, is spent on social applications such as WeChat and Weibo's microblogging service. Another 11 percent of their time is spent watching, sharing, and creating short videos on apps such as the immensely popular Douyin and over-the-top video streaming services like Tencent Video.²⁴

Social platforms are emerging as drivers of impulse shopping and creating incremental demand. Social interactions, including interactions with KOLs, and recommendations from contacts motivate consumers to buy products, demonstrating the increasing importance of authentic communication with customers and the fundamental element of trust in the purchasing process.

As social media exercises greater sway over consumer decisions, marketers are gradually shifting resources and attention away from traditional marketing tools to social based marketing channels and formats, such as user-generated content and short videos.

Nowadays companies can easily monitor and engage with their users, as they communicate with consumers at a faster pace and in real time, collecting and analyzing data that they then feedback into their social media marketing plans. In this way, companies' role has shifted into a subordinate of the buyer, who has become the main actor in the retail sector, as companies deeply depend on their data to perfectionate their offer and services.

²³ JP Morgan, *E-commerce payment trends: China*, in "J.P. Morgan.com", 2019, <https://www.jpmorgan.com/merchant-services/insights/reports/china>

²⁴ China Digital Consumer Trends 2019, September 2019, <https://www.mckinsey.com/~media/mckinsey/featured%20insights/china/china%20digital%20consumer%20trends%20in%202019/china-digital-consumer-trends-in-2019.ashx>

1.2.1 The Chinese Super App Model: WeChat

For most emerging markets in the Asia-Pacific region, including China, people access the internet primarily through their mobile phones. As smartphone penetration grew alongside mobile internet penetration, mobile phones developed more use cases and became more embedded in people's daily lives. This "mobile-first" culture has driven app developers to the creation of dedicated software which could capture a variety of services and tasks in the simplest way possible: they have developed the "all in one app" model.

Nowadays, mobile users in China are accustomed to doing multiple, often seemingly unrelated functions, with just a single app. Indeed, the widely popular lifestyle super-apps have become the real key drivers of e/m-commerce in the country, having fundamentally changed the way shoppers research, organize, spend and shop in their daily lives, and becoming indispensable to users and crucial for companies to lock users into their ecosystems and keep their eyeballs glued to their products.

The greatest example of a super-app is represented by the Tencent-owned instant-messaging application Weixin 微信, internationally known as WeChat. Created in 2011, in few years the WeChat ecosystem has become an integral part of the daily life of Chinese people and beyond, given its nature of being a multifunctional platform. Actually, besides being a messaging application with which you can send messages, make video calls, send voicemails and so on, through WeChat users also have the possibility to follow individuals and companies, make purchases online, hail taxis, conduct financial investments, play games, access public services such as booking a visit to the doctor and much more. But it is not the variety of activities users can do on the platform that make it so powerful, it is the fact that they can do it all within one app.

This "all in one app" system is possible because of a collaboration between WeChat and thirdparty companies, allowing them to build up their own mini-applications inside the WeChat ecosystem. Officially called Mini Programs in China, these "apps within an app" allow mobile users to use services of third-party companies without downloading separate apps. Indeed, launched in January 2017 to facilitate the promotion of content and products to consumers, these "apps in the app" or mini-apps are very similar to applications but lighter (they can weigh up to 10MB) and thanks to these features, the mini-programs make

navigation and user experience smoother and faster, giving brands a platform to leverage social elements to build consumer awareness and engagement.²⁵

This explains why companies in China mainly relies on WeChat for their marketing campaigns: logging in the platform with a company official account, brands can get a lot of benefits especially as regards their advertisement strategies; through the platform's facilities, they can easily have contact with customers and create real advertising campaigns sending push notifications about promotions or events or direct their followers to their official websites and connect mini-programs. In addition, a super-app as WeChat enables companies to have access to a massive amount of users' personal data captured and stored by Tencent, like for instance their purchase habits, their geographical information, their preferences, interests, interactions with people and so on. This helps businesses to filter out the most valuable target of consumers from the over 1 billion users and create a personalized WeChat advertisement specifically targeted for each group.

For advertisers, this system is a real miracle but looking to users' point of view, it could be problematic as incorporating multiple services in one app raises concerns about handling, storage and use of personal data. Someone could think about the fact that concentrating so much data and power in so few hands could lay the groundwork for an Orwellian world, where companies and the government can easily track every move you make. Collecting a staggering volume of personal data, through WeChat it is possible to find out what you talk about, who you talk about it with, what you read, where you go, why you are going there, who is there, how you spend your money while you are online or offline as well, and the list goes on indefinitely, but nothing of this features could be possible without companies most longed weapon: Big Data.

²⁵ Jing ZHANG, Xiǎo chéngxù qiào dòng dà shāngjī 小程序撬动大商机 (Mini-Programs Leverage Big Opportunities), in "Xinhua.net", August 14, 2020. http://www.xinhuanet.com/2020-08/14/c_1126365632.htm

Starting from the 2010s until now, social media platforms became mainstream allowing the entrance of China in the era of an hyperconnected reality guided by the use of big data analytics and precision marketing tools.

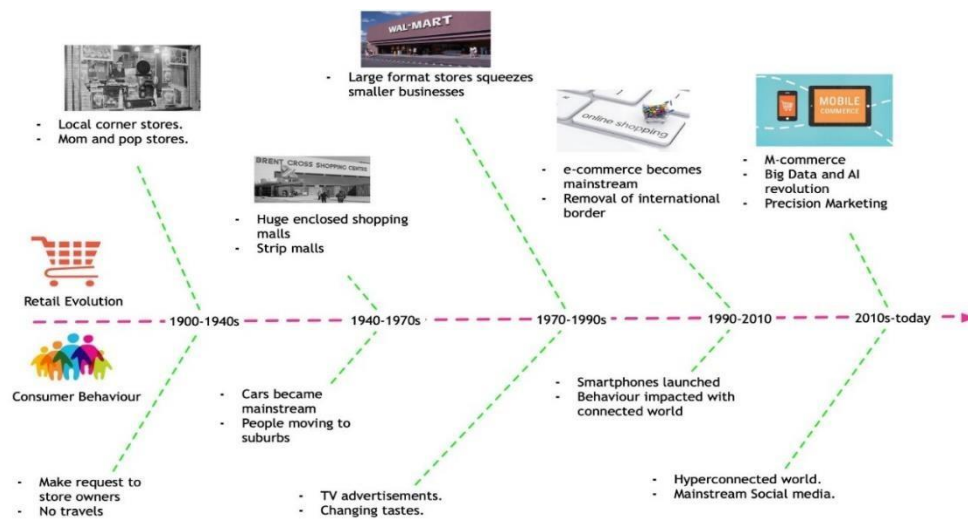


Figure 4 The Evolution of Chinese Retail (Source: Medium.com)

CHAPTER 2 – THE USE OF BIG DATA FOR COMMERCIAL PURPOSES

“If we know where to look, we can see the new emerging.”

Martin Luther King

In the modern era dominated by social media and by an intense virtual presence, information represents a real gold mine. Big Data (BD), as defined by IBM company, is a term applied to data sets whose main characteristics consist on high velocity, high variety and high volume and whose size or type is beyond the ability of traditional relational databases to capture, manage and process the data with low latency.²⁶ BD have become a mass phenomenon created by the enormous amount of information available online and offline which, once processed and analyzed in order to bring to light information and knowledge that cannot be identified a priori, place the company in an advantageous position compared to those who do not have access to it.

As for the increase in the availability of information, it is a consequence of two phenomena: datafication and the Internet of Things (IoT). The first is considered the fuel of big data since it allows data to be extracted and correlated. It differs from digitization, which is the process of converting analog elements into digital format. Datafication, on the other hand, refers to the process of converting a certain phenomenon into a quantitative form, so that it can be tabulated and analyzed. This phenomenon will lead to a sea change: the possibility of using all the data available, instead of a series of restricted samples. Datafication is covering the most disparate areas, in the case of social networks for example, Facebook and QQ transforms relationships into data, Twitter has made it possible to date feelings through a rather immediate system and LinkedIn has professional experiences by transforming the information acquired into forecasts on our present and future, advising us who we might know or the work we might consider. The Internet of Things, on the other hand, allows the collection of data on the status and functioning of objects, through special sensors placed on the objects themselves (examples are GPS systems).

Much personal information such as family members, work units, marriage status, online shopping records, Internet records and so on in real life are not necessarily of practical

²⁶ IBM, What is Big Data Analytics?, in: <https://www.ibm.com/analytics/hadoop/big-data-analytics>

significance, but in large data background, they have economic value. Indeed, analyzing them, BD in the retail sector can facilitate consumers comparison, predict the price trend of products, and help the user to identify the best time to buy. In facts, through the association of **Data Mining**²⁷, it is easy to find a lot of privacy information in the surface of mass data, which brings a new problem to the information security and privacy protection. Therefore, about 80 percent of the organizations have their own database, such as Baidu and Tencent, and these data are mostly in the form of digital, being easy to spread out. What is relevant here is that once spread out, it is irreversible.²⁸

Big Data is like a double-edged sword, which makes the community has benefited from their use, but privacy is also easy to be invaded. Big Data represents one of the biggest challenges to privacy and data protection society has ever seen: never before have so much Personal Information (PI) been available so freely to so many. That is the reason why it is extremely important to understand the features of BD and the use companies could make of them, in order for users to protect their personal information becoming conscious of what they are sharing while online.

2.1 Definition of Big Data

As stated by *The Economist*, BD represent the “digital oil” of the society, being to this century what oil was to the previous one: a driving force for growth and change.²⁹

There are plenty of definitions of the term Big Data, which does not only refer to extremely large datasets, but also mainly consists on the extraction of knowledge and insights from unstructured data. In other words, we could say that more than data mining, Big Data consist more about *data refining*: the transformation of unstructured facts and figures into fresh

²⁷ Data mining is a process of discovering patterns in large data sets involving methods at the intersection of machine learning, statistics, and database systems. Data mining is an interdisciplinary subfield of computer science and statistics with an overall goal to extract information (with intelligent methods) from a data set and transform the information into a comprehensible structure for further use. www.wikipedia.org

²⁸ Hui ZHAO and Haoxin DONG, Research on Personal Privacy Protection of China in the Era of Big Data, in “Open Journal of Social Sciences”, 5, June 16, 2017, pages 139-145. <https://doi.org/10.4236/jss.2017.56012>

²⁹ Wijnand JONGEN, The End of Online Shopping: The Future of New Retail in an Always-Connected World, Nubiz, Singapore, 2018, pages 11-27.

insights, decisions and value. Indeed, the ability to unlock hidden value within this huge amount of data is vital for companies that wish to be competitive in today’s fast paced world.

During the past twenty years, different approaches have been proposed to describe their peculiarities: starting from the first model of the analyst *Douglas Laney*, from the information technology research and advisory company *Gartner*, released in 2001, which attributed to the BD three main characteristics, known as the 3Vs (Velocity, Volume and Variety), a simple model to define the new data generated by the increase in information sources and more generally by the evolution of technologies. More recently big-data practitioners and thought leaders have proposed additional Vs, further extending the traditional ones with the introduction of a fourth V, that of Veracity, and then a fifth, that of Value.³⁰

- **Volume:** It refers to the huge amount and size of all types of data generated by different heterogeneous sources that cannot be managed by traditional databases but need to be organized and analyzed.

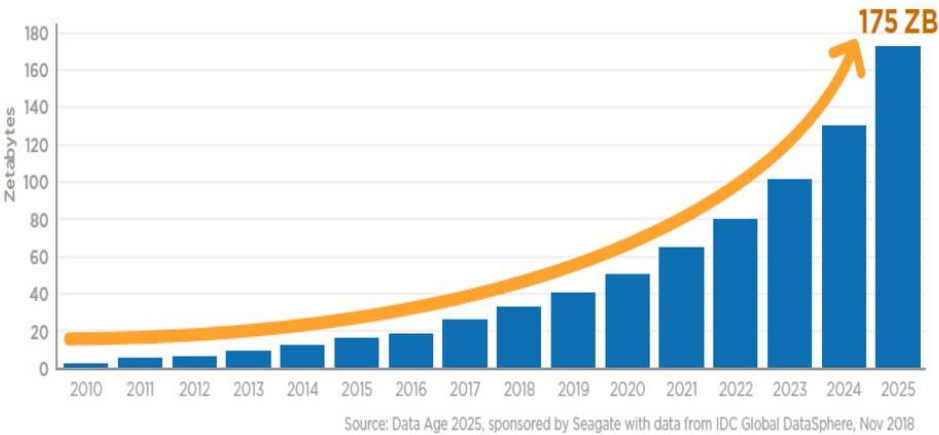


Figure 5 Data Growth Prediction

- **Velocity:** It refers to the speed with which data is generated and the fact that data must be quickly collected, processed, stored and analyzed. An example of what velocity means could be the fact that every 60 seconds almost 204 million emails are

³⁰ Francis X. DIEBOLD, On the Origin(s) and Development of the Term 'Big Data', PIER Working Paper No. 12-037, September 21, 2012, Available at <http://dx.doi.org/10.2139/ssrn.2152421>

sent, and 72 hours of YouTube videos and 216.000 Instagram posts are uploaded, meaning for companies a huge challenge to collect and process in real-time such an amount of information.

- **Variety:** It refers to the different types of data: *Structured* (It is the traditional data which is organized and conforms to the formal structure of data. This type of data is relatively easy to enter, store, query, and analyze and can be stored in a relational database. Examples of it could be a bank statement containing date, time, amount etc.), *Semi-structured* (does not conform to the formal structure of data. Examples of this type could be log files, JSON files, Sensor data, csv files etc.) and *Unstructured* (This type of data is more difficult to sort and extract value from as it is not an organized data and does not fit into rows and columns structure of a relational database. Examples could be text files, emails, images, videos, voicemails, audio files and more.) data, due to different sources of data generated either by humans or by machines.

- **Veracity:** It refers to the quality of data collected and the peculiarity of some data that could be uncertain or imprecise due to their inconsistency, incompleteness or ambiguities. For instance, social media data are inherently uncertain. If source data is not correct, analyses will be worthless and can lead to statistical errors and misinterpretations of the collected information. As the world moves toward automated decision-making, where computers make choices instead of humans, it becomes imperative that organizations be able to trust the quality of the data. The very challenge consists therefore in the conversion of Big Data into “Right Data”, so in useful data for strategic planning.³¹

- **Value:** it is considered the most important aspect of Big Data and refers to the process of identifying a high hidden value within many different and rapidly growing data, that could generate competitive advantage for a company.

³¹ Michael WALKER, Data Veracity, November 28, 2012. www.datasciencecentral.com

But extracting value from unstructured data is easily said than done, as IBM company explains when defining Big Data.³² Indeed, in the definition provided by IBM, reference is made to the inadequacy of traditional means to cope with the management of this large amount of data.

Big data have always existed, but the birth of new data sources and the development of devices for their collection have meant that they have reached significant dimensions, to the point of requiring new tools for their management. According to IBM, 2.5 quintillion (2.5×10^{18}) bytes of data are created every day, so much that 90 percent of the data in the world today has been created in the last few years. In the Digital Era, each day 500 million tweets and 294 billion emails are sent, 5 billion online searches are made, 45 billion of messages are created on WeChat³³ and it is estimated that 463 exabytes of data will be created each day globally by 2025, equal to 212.765.957 DVDs per day.³⁴

Regarding the amount of data, the growth is exponential, if we think that from the beginning of the digital age until 2003 the human being had generated 5 exabytes and that nowadays, five exabytes are generated every two days (Eric Schmidt - Executive Chairman, Google). Furthermore, the type of data that is recorded has been considerably enriched and therefore the management of the recording of the same has necessarily had to adapt, revolutionizing its architecture, as explained in the chart below, where the scale of data units shows the evolution of the main unit of the data, the bit, which has rapidly evolved in a long and complex list of multiples, reaching its highest level, that of the yottabyte, as a consequence to the rise of Big Data in the world.

³² *Supra* note 26

³³ Tencent Global, Tencent's 2019 Fourth Quarter and Annual Results, March 18, 2019. www.weixin.qq.com

³⁴ Jeff DESJARDINS, How Much Data Is Generated Each Day, April 17, 2019 www.weforum.org

Abbreviation	Unit	Value	Size (in bytes)
b	bit	0 or 1	1/8 of a byte
B	bytes	8 bits	1 byte
KB	kilobytes	1,000 bytes	1,000 bytes
MB	megabyte	1,000 ² bytes	1,000,000 bytes
GB	gigabyte	1,000 ³ bytes	1,000,000,000 bytes
TB	terabyte	1,000 ⁴ bytes	1,000,000,000,000 bytes
PB	petabyte	1,000 ⁵ bytes	1,000,000,000,000,000 bytes
EB	exabyte	1,000 ⁶ bytes	1,000,000,000,000,000,000 bytes
ZB	zettabyte	1,000 ⁷ bytes	1,000,000,000,000,000,000,000 bytes
YB	yottabyte	1,000 ⁸ bytes	1,000,000,000,000,000,000,000,000 bytes

Figure 6 Scale of Data Units (Source: Weforum.org)

2.1.1 Personal Data & Sensitive Personal Data

A distinction of data occurs also when talking about the category of information that could be extrapolated from the analysis of data. According to the European *General Data Protection Regulation* (GDPR), it is possible to distinguish two types of data, generated from the analysis of the raw digital footprints people leave online: Personal Data and Sensitive Personal Data.

According to the EU GDPR article 4.1, the definition of personal data describes them as

any information relating to an identified or identifiable natural person or ‘data subject’; an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person.³⁵

This kind of data is more commonly collected since apps and websites often need these details to run payments or maintain subscriptions.

While sensitive personal information include information (or opinion) about an individual's health or genetics, as well as their racial or ethnic origin, political opinions, membership of a political association, membership of a professional association, trade association or trade union, religious beliefs or affiliations, philosophical beliefs, sexual preferences or practices

³⁵ General Data Protection Regulation (GDPR), Chapter 1, Art. 4.1, May 25, 2018. <https://eur-lex.europa.eu/legalcontent/IT/TXT/HTML/?uri=CELEX:32016R0679>

or criminal record.³⁶ This type of information is delicate as if revealed they could expose an individual to discrimination or harassment, as happens in China, especially in the Xinjiang province, where about 2.5 million people, the Uyghur Muslim minority, are constantly monitored and a track of their sensitive personal information is detected and analyzed by Chinese authorities in order to preserve China's national security and prevent possible terrorist attacks.³⁷

2.2 How Retail Companies Use Big Data: Customers' Profiling.

"Customers buy for their reasons, not yours."

Orvel Ray Wilson, 1984

Today data and information form a fundamental basis for decision making processes and problem solving, effective and efficient management of big data can guarantee companies the possibility of making choices in accordance with more accurate and reliable analyses, especially considering the strong dynamism of the environment in which companies find themselves operating: the primary challenge for today's retailer is that the modern customer is quality and brands conscious as well as compare for services provided to them by different outlets at the comfort of home with a single click.

Consumers are also undergoing a shift in perceived value, putting less emphasis on traditional "transactional value" based consumption, thereby changing their focus on more "personal values". This greatly affects and led to a change in retailers' business and marketing strategy, moving from traditional mass-media, mass-marketing towards a more personalized one. To reach that goal, however, companies and retailers could not continue to rely on their traditional ways of gathering data, i.e., market surveys and feedback forms. Big data, therefore, is increasingly favored as the "new" standard.

³⁶ Australian Privacy Act 1988, Part III. www.legislation.gov.au

³⁷ Erin HANDLEY, China's Mass Surveillance of Uyghur Muslims in Xinjiang Province Revealed in Data Security Flaw, in "www.abc.net.au", February 18, 2019.

In view of this change, companies have had to adapt their marketing strategies, modifying them to the needs of each individual, capturing their preferences and opinions from the various online touchpoints.

Today, companies' need of information requires customers profiling tools as the best way to gather the insights needed to identify, segment and define their target audience, getting closer to their consumers and building a strong and mutual relationship. The Internet allows marketers to obtain detailed customer-related information and enables sellers to individualize their marketing instruments on a large scale and in real time. Moreover, since data are already in digital form, they can be used directly and quickly in companies' analytical processes, which makes it possible to shape interaction individually and responsively, as a function of observed behaviors.

2.2.1 Collection & Processing of Data

While in the past the information needed were in the form of analogical sources, nowadays, in the digital era, they are mostly available online, consequently to their transformation in digital ones. Indeed what is relevant now is the fact that the collection process of data has shifted from the analogical world towards the digital landscape, where everybody is entitled to have access to the mass information available on the web that users leave, consciously or unconsciously, in the form of digital footprints in different touchpoints: social media, web browsers, e-mails, videos, pictures, messages, and so on.

The rise of digital and mobile communication has made the world become more connected, networked, and traceable and has typically led to the availability of such large-scale data sets. Consequently, the abundance of resources and sources has meant that traditional analysis techniques and technologies have given way to new, more precise analytical systems. It is the skills and technologies of Analytics that transform raw data into valuable information for business decision makers: it is now possible to gain competitive advantage thanks to timely and more informed decisions. The data collection and management process has changed, technologies supporting the data life cycle evolved and new skills are developed for data enhancement.

- Where and How Are Data Collected?

Today's retailers have a spectrum of data available to them that come from both internal and external sources.

A company could leverage the data it produces and gets directly from its own firewall, the so called first-party data, which includes information collected straight from customers, in the form of behavioral data, like people's clicks, views, comments, purchases and other interactions with the brand; subscription data, which allows a company to learn more about who is interested in its brand and content; social data and survey data which provide invaluable insights into a customer's feeling and perception of the company. Retailers could collect first-party data straight from their audience and are considered high-quality and highly relevant as coming straight from the source, they are likely to be accurate and often readily available to retailers, meaning a cost reduction in the gathering process of information respect to the use of second-party or third-party data.

Indeed, a retailer may decide to rely on other data providers to get a more complete picture and be able to gain a broader understanding of the market. The possibility to purchase second-party data, which simply includes another entity's first-party data, allows marketers to supplement their first-party data with customers' insights of another company who operates in the same or similar market, thus developing and enlarging their clientele, reaching new audiences.³⁸

The combination of first- and second-party data with third-party data allows an even broader and deeper targeting scale as the latter are collected from many different sources by a Data Management Platform (DMP) or data providers, that aggregate and organize information collected from various touchpoints, usually thanks to Cookies (special browsers with source code modified for enhanced data collection), and then sell them to companies who need them to empower their customers' understanding.

³⁸ Digital Transformation through Data: A Guide for Retailers to Drive Value with Data, May 2019, www.deloitte.com

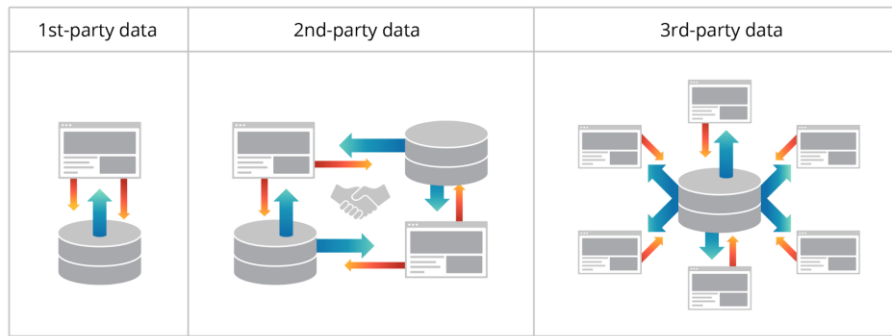


Figure 7 The Collection of First-, Second- and Third-Party Data. (Source:Piwik.pro)

- Where Are Big Data Stored?

After data are gathered, they need to be stored into databases or storage services for further processing.

In the early days of data management, the relational database was the primary method that companies used to collect, store, and analyze data. Relational databases, also known as Relational Database Management Systems (DBMSes), offered a way for companies to store and analyze highly structured data about their customers using Structured Query Language (SQL). For many years, relational databases were enough for companies' needs as the amount of data that needed to be stored was relatively small, and relational databases were simple and reliable. But with the rise of the internet and of the increasing amount of data, relational databases have been progressively replaced with data warehouses, which became the most dominant architecture for big companies from the late 90's.³⁹

The concept of data warehousing was first introduced in 1970s by *Bill Inmon* and then further developed by IBM researchers *Paul Murphy* and *Barry Devlin* in 1988 and consists on the electronic storage of a large amount of information, in the form of structured data. Indeed, what distinguishes data warehouses from other storage tools, is its Extract, Transform and Load (ETL) process through which raw data coming from different sources are firstly collected and then subjected to a *data cleansing* process, where they are corrected from any errors and then converted and translated into the warehouse required format, according to the goal set before the data are stored. In a Data Warehouse System, starting from a set of raw

³⁹ History and Evolution of Data Lakes, www.databricks.com

data, these are structured and processed through a so-called “Schema-on-write” approach, where first the structure of the database that must host the data is defined, then the data is written inside the predefined structure and, when they are taken for analysis, they are returned in the predefined format.

This model is expensive and time-consuming as it requires data to be transformed into structured data, that could be then easily analyzed and at disposal for company’s needs.⁴⁰

Coined in 2010 by *James Dixon*, the most recent Data Lake model represents a valid alternative of data storage.

A Data Lake is a new working method that simplifies and enhances the storage, management and analysis of Big Data, using data from diversified and inhomogeneous sources, in their native format, including both structured, semi-structured and unstructured data. This is a new working method because the systems historically used to store, process and analyze data were defined and structured according to the final use that is expected to be made of them, like in the case of Data Warehouses. A Data Lake type system, conversely, adopts a so-called *Schema-onread* approach, in which the data are acquired in their native raw format according to policies that standardize, for the different types of data, methods, times and rules for entering data in the data lake.

Each element is associated with an identifier and a set of *metadata*⁴¹ that qualify it so that, when it is necessary to access the data to search for a specific result, the data lake can be queried to return all the relevant data. It is the analysis question that determines the selection of data from which to draw information, and the search is not limited to a database set up for that type of analysis, but accesses all available information, regardless of the source that generated it.

So, the main difference between the two models regards the feature of data lake in which the schema and data requirements are not defined until the data is queried.

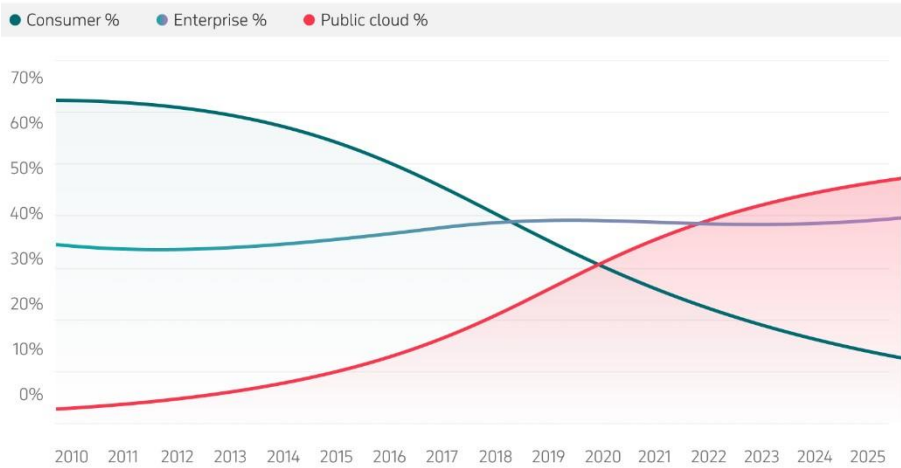
Today, companies may decide whether to archive this information in in-house servers or in the Cloud, before performing accurate analysis and derive results to make critical business decisions.

⁴⁰ Jake FRANKENFIELD, Data Warehousing, June 28, 2020. www.investopedia.com

⁴¹ Metadata describes other data. It provides information about a certain item's content. (From: <https://techterms.com/definition/metadata>)

While the first requires more efforts in terms of technological infrastructures and money and is completely under the control of the enterprise, the latter is owned and managed by a third party paying a monthly fee, and mainly consists of a vast network of remote servers around the globe which are hooked together and meant to operate as a single ecosystem. An example of a Cloud service provider is Alibaba’s Alibaba Cloud, also known as Aliyun 阿里云, which represents the main provider of cloud computing services to online businesses and Alibaba’s own ecommerce ecosystem.⁴²

These servers are designed to store and manage data or deliver a service such as web mail or social media and leave freedom to access information from any internet-capable device anytime. Both public and private cloud systems are widely used, in response to the increase in size of big data, either by consumers nor by companies to store and manage their data and is a developing phenomenon, as the graph below explains.



Data source: Data Age 2025, sponsored by Seagate with data from IDC Global DataSphere, Nov 2018

Figure 8 Where Are Big Data Stored?

- How Are Data Processed?

After data are collected, they are taken to the next stage of the analytical process: they must be processed for the creation of customers’ profiles.

⁴² <https://cn.aliyun.com/>

Data preparation is an important issue for both data warehousing and data mining, as real-world data tends to be incomplete, noisy, and inconsistent. Data preparation includes data cleaning, data integration, data transformation, and data reduction.

The processing stage consists on the extraction of useful information from unstructured data for supporting and providing decisions, and it starts with *data cleaning*, the filtering out of nonrelevant information, redundant, incomplete or incorrect data. Secondly, the information need to undergone a *data integration* process, which is used when data is gathered from various data sources and are combined to form consistent data (combines data from multiple sources into a coherent data store); they are then ready for the *data transformation* stage, which is used to convert the raw data into a specified format according to the need of the chosen analytical model; Data reduction techniques such as data cube aggregation, dimension reduction and data compression and so on could be used to get a reduced representation of the data in order to quicken and ease the analysis of them, at the same time minimizing the loss of information content.⁴³

2.2.2 Analysis & Targeting of Data

After the process of data collection and data cleaning, raw data have been cleaned and transformed to the required data and ready for the next step of data analysis. Data analysis is a process of exploring, cleaning, transforming and modeling data which aims to discover valid information and support decision-making (Vaughn, 2008).

In order to grasp useful value, analysts must be able to understand what to look for and how to correctly ask the data for the necessary information, so as to be able to choose the most suitable tools and to program the analysis software which, without a precise human intervention, could not act effectively.

The strategy a company may implement depends on the goal it aspires to reach and the possibilities it has to sustain the costs and the technological tools required for each data analytics process, which are reduced to four categories: descriptive, diagnostic, predictive and prescriptive analytics (Porter and Heppelmann, 2015). Descriptive analytics regards the interpretation of historical data for a better understanding of the changes that have occurred

⁴³ Data Pre-processing & Mining Algorithm, Knowledge & Data Mining & Preprocessing, 3rd edition, Han & Kamber.

in the business during a set period of time, like for example, year-over-year pricing changes, month-over-month sales growth, the number of users, or the total revenue per subscriber. However it gives valuable insights into the past, its findings simply signal that something is wrong or right without explaining why, so this kind of analytics is often combined with other types of data analytics in order to get a more comprehensive and exhaustive understanding of an issue; diagnostic analytics focuses more on the reason why something happened to get a more insightful understanding of a particular issue, like for instance whether a company's promotional campaign was successful or not; predictive analytics helps retailers make forecasts about what is likely to happen in the future, based on historical data, algorithms and external data. It makes use of the findings of descriptive and diagnostic analytics to predict future trends, the accuracy of which highly depends on data quality and stability of the situation, so it requires a continuous optimization; the last step in data analysis is prescription analytics which gives a prescription about the solutions to eliminate a future problem or the action to take to get full advantage of a promising trend also showing the implications related to each decision option.

Each type of analytics allows retailers to understand the value of each information and suggest solutions to optimize the performance of a given product, service or process. Yet, to accomplish this task, there is no single technology that encompasses big data analytics, but several tools that if combined provide a deep support to retailers for their decision-making process.

Artificial Intelligence (AI) and its subsets, Machine Learning and Deep Learning, are emerging as important tools for market design. Therefore, the improved forecasting and management that AI and machine learning algorithms provide, help retailers better anticipate consumer demand and producer supply as well as help target products and activities to finer segmented markets.

Using Natural Language Processing (NLP), a branch of AI involved in making computers understand the way humans naturally talk and type online, turning these information into structured data for further analysis, is one of the favorite method used by marketplaces to mine the data generated by text data like messages, emails, comments and feedbacks in order to better predict the kind of features that customers value. This method is particularly useful and leveraged by Alibaba's Taobao marketplace, which understood the potential of consumers' feedback as early as March 2012 when it began investing in a "Rebate-for-

Feedback" (RFF) feature, which consists in rewarding customers with a coupon or special offer after they leave a relevant comment about their purchase on the online platform.⁴⁴

Another relevant application of deep learning system is that of image recognition. This type of tool has become indispensable in today's world since the rising in power of social networks where it has been demonstrated how people are more likely to learn and communicate through pictures or videos instead of writing in words. Thence, text messages have experienced a slowdown to leave the stage to the rising popularity of voice messages instead. This trend had significant repercussion in the analytical field, as new disciplines and tools were designed, such as image classification or facial recognition systems, speech recognition, also known as Automatic speech Recognition (ASR), computer speech recognition or Speech To Text (STT), and video content analysis.

Facial recognition is also described as a Biometric Artificial Intelligence based application as it permits the identification of a person analyzing patterns based on her facial texture and shape through a comparison with faces within a database. It is widely used in security systems and is gaining popularity also in banking systems, enabling for example clients to enter a shop, take whatever is needed and then leave without pass from the cashier. Actually, China is promoting facial payment technologies since its mobile payment system has already widely permeated the society. Alipay, the financial arm of E-commerce giant Alibaba, is investing in the sector and recently launched its "Smile-to-pay" system, which allows customers to process their payment simply by smiling at a screen where a 3D camera then scans their faces, which are linked to a bank account, to verify their identity.⁴⁵

This system makes the entire shopping process much faster, erasing long queues at pay cash or giving customers the possibility to not take with them any cash nor smartphones. However, this tool is still more firmly established online than in the physical world. Indeed, many social media and search engines have implemented the image classification analysis. According to the first, it is an adaptation since in social networks, consumers meet and communicate with

⁴⁴ Lingfang LI, Steven TADELIS, Xiaolan ZHOU, Buying Reputation as a Signal of Quality: Evidence from an Online Marketplace, March 24, 2017.

⁴⁵ Tom BRENNAN, Alipay Launches 'Smile to Pay' for Commercial Use in China, September 1, 2017. www.alizila.com

other network members by uploading videos or pictures that reflect their activities, interests and opinions. Since these pictures are assumed to contain valuable visual information they could be well suited to provide a basis for lifestyle segmentation, being a precious element for identifying market segments with similar lifestyles, thanks to the automatic tagging system used by *Facebook* since 2010 for example, which recognizes the identity of a person from a picture and links it immediately to the user profile.⁴⁶

So social media are gaining power since they detain a huge amount of data about their users' lives and retail companies are taking advantage from this treasure of data, paying these platforms for having access to the precious insights that they could get from them. Therefore, WeChat has a dense network of third-party companies having registered their official account in the platform in order to get the most out of it and build a stronger relationship and communication with possible consumers.

But image recognition plays an essential role in e-commerce enhancing customer experience. Indeed, this technology facilitates users in their search for a specific product, and companies easily understand what consumers are looking for. Alibaba Cloud launched its Image Search engine in February 2018 and implemented the system in its retail platform, Taobao, where this intelligent image service based on machine learning and deep learning, allows users to search for a product simply by uploading a picture of it without any textual product description and get a quick response in milliseconds. The system will then select a list of related or similar products after a attentive analysis of the shapes, sizes and colors of the item, being able to identify the brand and product names. This represents a valuable tool as it liberates users from finding the right tag or words to describe the product they want and to fasten the entire searching process, while it helps retailers in cross-selling and recommendation process, as if they do not have the exact product that a customer is looking for, they could show related or similar items that might satisfy consumers' expectation as well.⁴⁷

⁴⁶ Srinivas NARAYANAN, An Update About Face Recognition on Facebook, September 3, 2019, <https://about.fb.com/news/2019/09/update-face-recognition/>

⁴⁷ Alibaba Clouder, Finding Similar Images with Alibaba Cloud Image Search, July 6, 2018, https://www.alibabacloud.com/blog/finding-similar-images-with-alibaba-cloud-image-search_593805

Summing up, AI and machine learning cannot only help predict a customer's intent, but given the large heterogeneity on consumer tastes, AI can help a marketplace or retailer better segment the many customers into groups that can be better served with tailored information.

The information derived from the analysis of the data permit marketers to target their consumers and get a complete and efficient marketing strategy done, and, in this way beginning to get revenues from the process. Based on the results, these companies develop customer lifetime value models and prioritize segments.

2.3 Pros

There are many thinkers about the rising phenomenon of Big Data Analytics (BDA) and its economic, political and social implications. According to 19th-century German sociologist *Max Weber*, any increase in knowledge and more powerful technology have both positive and negative sides: as applied to BD, the implication is that there are benefits with greater control over the world, but also dangers with increased surveillance and threats to privacy.

Here in this chapter the main positive features as well as negative ones are going to be presented and discussed, in order to better understand this phenomenon, from both retailers' and customers' point of view.

2.3.1 Pros for Companies

Personal Information is the lifeblood of most digital businesses today. McKinsey estimates that a retailer using big data to the full has the potential to increase its operating margin by more than 60 percent,⁴⁸ and it has been proved to be a precious tool for helping companies on any step of the customer journey and any step of the supply chain. Yet, Big Data analytics has improved companies' work and apported many benefits in term of their marketing, logistic and customer-service management.

⁴⁸ McKinsey Global Institute, Big Data: The Next Frontier for Innovation, Competition, and Productivity, May 1, 2011. <https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/big-data-the-next-frontierfor-innovation>



Figure 9 Main Benefits of Big Data in Retail (Source: BARC)

- Better Strategic Decisions

- Predictive Trends and Sales

The data collected from consumers allows companies, like Alibaba, to see where customer needs exist—and meet them—before even those customers do.⁴⁹

With the introduction of Big data in retail, retailers are now able to better understand their customers' behavior and make future decisions on what will actually work best for their company. Knowing customers' purchasing habits and consequently their interests and lifestyle, gives marketers a big privilege as they are enabled to figure out what will be the next purchase or the right kind of product to propose to them. Therefore, this high interest on customers' needs stimulates retail companies to implement "consumer-centered" strategies instead of traditional "brand and product-centered" approaches, in this way building an end-to-end consumer value chain, perfectly aligning to the real clients' demand.⁵⁰

- Improved Supply Chain Management

Indeed, using the insights they gain from customers' habits, retailers are now able to predict the demand for a product and what competitors are doing in real time and to adjust prices accordingly. As demand changes throughout the life of a product, analytics help companies to see when that is happening and, thence, decide which type of product fits the trends the most and whether to raise or to lower the price of it to reflect the demand as it changes, which

⁴⁹ Alizila staff, What Does The Customer Want? Big Data Knows, July 11, 2017. <https://www.alizila.com/whatdo-you-need-big-data-knnetrepreneur-ows/>

⁵⁰ Deloitte, Serial Studies of "Future of Consumer"- Big Data-driven Consumer Insights, January 2020.

could possibly bring revenues on the product pack up. This aspect of big data allows them to stop wasting money on dead end products and fill the “shelves” with things people really want. Additionally, this process improves the organization of the supply chain and help companies forecasts on inventory allocation and optimization in terms of product assortment of quantities, in this way also notably reducing costs.⁵¹

- Improved Control of Operation Processes

The existence of online platforms in which to carry out personal data collection activities represents for companies the tool for the realization of potentially opportunistic behaviors while for users it constitutes a new element of vulnerability in the control of privacy. The market study activities carried out online form part of the corporate digital marketing strategy, in which respect for the needs and values of the users of the customer network is fundamental in order to improve the commercial relationship with demand and avoid putting the entire brand reputation at risk.

As already stated, Big data gives companies an opportunity to analyze interactions across multiple channels and decode shopping patterns, thence enabling them to accurately predict what each customer is going to buy before the customer even realizes that he is ready for that purchase. This capability aims at improving marketing strategies, reducing the waste of time previously required to target a specific cluster of customers without the right knowledge. Indeed, while in the past the decision-making capabilities in the retail industry were limited due to the lack of insights, everything has changed with the advent of BD. Now retailers can seamlessly track the performance of supply chain and other marketing activities in real-time and make informed decisions.⁵²

⁵¹ Daniel GUTIERREZ, How Retailers Are Using Big Data to Their Advantage, January 21, 2017.

<https://insidebigdata.com/2017/01/21/how-retailers-are-using-big-data-to-their-advantage/#:~:text=With%20the%20introduction%20of%20big,customer%20will%20buy%20next%20time>

⁵² Four Big Ways in wich Big Data Can Transform Retail Industry, September 16, 2018.

<https://www.getsmartcoders.com/blog/4-big-ways-in-which-big-data-can-transform-retail-industry/>

This phenomenon allows retailers to gain control over the entire business process: through the study of the many traces left online, they can arrange a *remarketing*⁵³ strategy, suggesting products in line with customers' past searches and purchases, thus optimizing their marketing Return On Investment (ROI).

- Better Understanding of Customers

“The business that succeeds the most is the one that understands its customers the best”, explained the renowned email marketing guru *Andre Chaperon*. This sentence emphasizes the power of BDA to build a strong relationship based on trust with customers, considered as the essence of the whole process.

A new recent trend in retail demonstrates that people are more willing to align themselves with brands that understand their needs, recognize them and connect with them on a human level. In fact consumers have grown more and more immune to advertisements that are not targeted directly at them and claim for a personalized and self-focused service from marketers.

By focusing on customers' needs and providing the personalized experience they crave, brands can create resilient emotional bonds that lead to **loyalty**. In fact, according to *Accenture*, delivering a good shopping experience improves customer satisfaction, repeat purchases, customer loyalty, customer referrals, revenues and customer engagement.

The engagement with customers is of vital importance especially in the Chinese market, where 79 percent of Chinese customers consider positive interactions with brands on social media have driven them to endorse the brand more, and 71 percent of customers have spent more as a result.⁵⁴

Relationship-based marketing approaches are proliferating in Chinese retail leaving the stage to new ways of promoting a brand's product, like the rising influencing power of the Key Opinion Consumers (KOCs). Thanks to the fact that consumers would rather receive advice from everyday consumers respect to that of a company whose main aim is selling its product to them, KOCs are considered as more authentic and credible because they are ordinary

⁵³ Remarketing (also known as retargeting) is a feature that allows companies customize their display ads campaign for users who have previously visited their sites, and tailor their ads, to these visitors when they browse the web and use apps.

⁵⁴ Michael CHENG, eCommerce in China- the future is already here, 2017, www.pwccn.com

people who offer impartial product reviews, and consumers feel somehow more connected to them by sharing the same purchasing experience.⁵⁵

Another important trend that is arousing customers' curiosity is that of vertical marketplaces, which naturally bind like-minded consumers together leading to strong social communities centered around a given sector or product. For instance, *gegejia.com* (格家网络)⁵⁵, a snack shopping e-commerce platform, is an example of successful vertical market in which users share tips and recommendations for snacks products. These kinds of platforms are helping users making their buying decisions as they provide better content and more meaningful experiences in some areas, and offer a community where like-minded customers can interact and where the relationship with other users who share the same interests represents the main positive aspect, both for users and retailers. Indeed, according to a survey by McKinsey in 2019, two-thirds of China's digital consumers use the information they collect from vertical websites to influence their purchasing decisions.

- Boost Innovation

Big Data Analytics is used by various firms to create new products and services for their customers. Companies through Big Data analyze different customers' opinions and perceptions about their products, gaining insights about what they are lacking and what are the significant things to be kept in mind while developing any new product. This system helps them developing new products according to customers' requirements and thinking beyond the ordinary.

The innovation process is no longer a blind or casual act of companies which hope to reach the greatest number of people, but rather it has now become a well-studied and organized process, in which retailers have all the necessary data available to create tailored products for every type of customer.

This consumer-to-manufacturer (C2M) model is really popular in China, as it allows manufacturers and retailers to provide consumers with a tailored product at a low, thus reducing links in the supply chain while cutting expenses, like that for market research,

⁵⁵ Coresight Research, Retail 2020: 10 Trends for China E-Commerce, January 20, 2020. www.coresight.com

bringing new growth points for themselves while better addressing user demands. The involvement of consumers in the innovation process is expected to become a new driver in shaping China's ecommerce landscape, in which through the application of AI-powered data analytics mass customization has become a reality in the country. ⁵⁶

JD company has launched its C2M initiative, to provide better insights of its consumers for the benefit of brands on its platform. JD has helped Nestlé better tailor their wafer bar products to the Chinese market: throughout a deep analysis of customers' feedbacks and purchase habits, JD found that consumers mostly buy this snacks for trips and family outings, and were complaining about the size of the wafer bars and packages, as they preferred a larger package and a smaller wafer bars in a variety of flavors to be able to share more easily. Thanks to this information, JD cooperate with Nestlé in the design of a new, larger packs containing more mini-sized bars of different flavors. In addition, the package was given the shape of a backpack to position the product to consumers as an ideal choice for outings. This partnership led to the increase of 70 percent of sales for Nestlé in the period from the 1st to the 18th of June respect to the same period of the previous month. ⁵⁷

2.3.2 Pros for Users

- Overcoming The "Paradox of Choice"

Today consumers are overwhelmed by the quantity of information available online and offline and have less time to spend for scrolling down to catch the most suitable product for themselves.

Therefore, resulting in the adaptation of companies' marketing strategies, from a hard selling to a softer one, meaning for consumers a more customer-centric approach and a customized experience.

⁵⁶ Emma LEE, Jill SHEN, China's Data-Based C2M Model to Drive E-Commerce Forward, November 12, 2019, in www.technode.com

⁵⁷ Cats and Campers: How JD and Nestlé Better Slake the Appetites of both through Insight. From Wafer Bars to Cat Food, JD Helps Nestlé Win in China, August 13, 2019, in www.jdcorporateblog.com

According to a study conducted by BCG, customers increasingly prefer a shopping experience that is easy and fast and that helps them make purchase decisions. This phenomenon is explained by the American psychologist Barry Schwartz, stating that too many choices create consumer anxiety.⁵⁸

Indeed, if shoppers are offered too many options, they might feel confusion over what to buy and therefore not purchase anything. The abundance of products listed randomly on an ecommerce platform for example, instead of increasing customers satisfaction because of the variety of options, makes them heighten their expectations, creates a sense of anxiety and fear of not making the right purchase decision in consumers' mind, leading to inconclusive results. What helps users solve the risk of choice paralysis or buyer's remorse is the simplification and personalization of each customer' homepage of marketing campaign. In this way, users are helped narrowing options and choosing the products that best fit their needs, thanks to companies' understanding of each user personal information and preferences through online data analysis, providing tailor-made recommendations and special offers which facilitate and speed up their purchasing decision journey.

- Personalization

Customers are no longer passive and now actively interact with services and content. Top-down communication from services is no longer enough as users demand tailor-made solutions and a personalized customer experience. Thanks to Customer Relationship Management (CRM) systems and behavioral analytics solutions, retailers have the ability to collect and track an incredible amount of users' information, such as preferences, buying patterns, and personal information as well, like their ID, age and gender, and use all of these insights to adapt their marketing strategies and their offer to their customers in real-time.

As a result, they can create personalized experiences that make customers feel valued, considered and welcomed because they are being matched with the most relevant content for them, and are diversified from the mass of a company's customers. Personalization makes them feel like they are unique and ease their entire buying process. Actually they get special communication tailored according to their interests that results in the recommendation of

⁵⁸ Barry SCHWARTZ, *The Paradox of Choice: Why More Is Less*, New York, Ecco, 2004.

products according to their own tastes and based on their past queries (action-to-item) or related to past purchases (item-to-item) or even based on what users with similar previous purchases have bought (user-to-user). In this way, the purchase process becomes like users were accompanied by the most attentive of personal shoppers.⁵⁹

2.4 Cons and Challenges

Although big data can help companies achieve competitive advantage over its rivals through many aspects, and customers get the most suitable and tailor-made service, big data analytics still face a variety of challenges.

According to a survey conducted by McKinsey & Company regarding the use of data and analytics, it has been shown that the main barriers and challenges for companies in this field could be broken into three categories: strategy, leadership and talent; organizational structure and processes; and technology infrastructure.

One of the main challenges in China, as in the world, when talking about adapting to the smart retail revolution, is the shortage of people with deep analysis skills and managers who are able to understand and manage the data and lead the team promptly. A company needs to revolutionize its culture to keep up with the digital retail, placing innovation, constant learning and technology at the core of its structure.

Even though technology and big data analytics systems are fundamental for the creation of a strong analytical business strategy, human capital remains a crucial element in the realization of the full potential of data and analytics that companies should not underestimate. Four roles are indispensable for the success of a company: data architects, who design data systems and related processes; data engineers, who deal with the scaling of data solutions and the creation of new products; data scientists, the experts in the analysis of data using increasingly sophisticated techniques to develop insights; and “business translators” who, having both technical and business knowledge, are able to translate analytical data into value and profit for the company. Without the presence of these four characters a company could face some

⁵⁹ Charlie BRAITHWAITE, Customer Data in Retail 2019: Key Research Findings, May 7, 2019, www.clickz.com

problems when dealing with data, actually, with today's large amount of data available, analysts, especially if with limited analytical skills, could commit false associations between the data which would lead to serious management consequences and misinterpretations. In fact, as McKinsey Global Institute has reported, the chief challenge for companies today in implementing analytics is more in their ability to attract and retain data professionals.

Another challenge is represented by the need of suitable technological infrastructures that allow the acquisition, storage and analysis of the huge amount of data in order to fully exploit their potential. Retailers require expensive software and huge computational infrastructure and are asked a huge investment either in their R&D departments or in the setting up of some partnerships or collaborations with third party companies, as Nestlé did in 2016 with the Chinese giant Alibaba, being the first company to use Alibaba's big data platform to personalize the shopping experience on Tmall or like Mars, which collaborated with Alibaba for the creation of a new chili-infused Snickers bar, based on the insights released from its database.⁶⁰

Most brands and retailers in China need to cooperate with third-party companies or experts in data analysis as they are struggling with massive challenges of data fragmentation, quality and completeness. Indeed, more data availability also means more false and unreliable information, that could cause confusion and frustration to retailers or gear the business to the wrong direction⁶¹, and as the data available increases, the personal information that will be available to companies and to hackers' interest or anyone who can review the data will also increase, thus identifying a rising concern relating to **privacy**.

Thus, one of the most important challenges facing the new technological scenarios is the protection of users' privacy and personal data, as there are still some basic problems left unsolved: security in passwords, encryption or access permissions, and mobile device applications that do not encrypt communications.

According to Gartner, big data has become mainstream in modern society and companies within retail, insurance, travel and banking, cannot help but develop increasingly intelligent

⁶⁰ Jenny W. HSU, Alibaba Now Product Designer, as Well as Seller, For Brands, April 19, 2018, www.alizila.com

⁶¹ Steffi NOEL, AI and Big Data in Retail in China, Daxue Talks n. 25, December 26, 2019. <https://daxueconsulting.com/ai-big-data-retail-china/>

algorithms on consumer behavior which translate consumers' individual preferences into the production of specific goods and services. So, personality traits and recent buying patterns can both predict what could be interesting next for the consumer and bring great benefits to users' daily lives by making their domestic activities easier and time efficient. Yet what raises concerns is the fact that this big amount of data is controlled and stored by just a few hands, like the three Chinese giants Baidu, Alibaba and Tencent, that in this way are gaining more and more power and the monopoly of the market. Therefore, the main threat of big data is summarized in the statement of the German data entrepreneur Yvonne Hofstetter "Sie Wissen Alles" – They know everything, relating to the fear of many critics of big data about the potential use of them to patronize and manipulate people as well as business models and whole societies, that could lead to a sociological problem as more than 800 million of Chinese users are constantly under the eye of companies which then share their insights with the government.

In fact, another thorny issue for big data analytics efforts is complying with government regulations. Much of the information included in companies' big data stores is sensitive or personal, and that means the firm may need to ensure that they are meeting industry standards or government requirements when handling and storing the data. In China this aspect is extremely relevant as private companies are asked to share users' personal data with the authorities in compliance with China's regulatory environment such as the most recent Cybersecurity Law, in order to ensure national security standards.

An example of this is the ban of Facebook from the Chinese market in 2009, after refusing to share with the authorities the conversations and information about the activists involved in the Xinjiang riots who were communicating through this platform, thus failing to adapt to Chinese government's prerogatives.⁶²

So, the privacy issue does not only relate to the way companies manage the data collected, but, more importantly, it refers to the government's use and protection of its citizens' information. Indeed, unlike some European countries' governments, the Chinese government is exempted from the weak privacy protection regulation and holds an extremely huge amount of data on Chinese citizens, thanks also to the cooperation network between hi-tech companies and the state: iFlytek(科大讯飞), a leading company in speech recognition

⁶² Kristina ZUCCHI, Why Facebook Is Banned in China & How to Access It, October 22, 2019. www.investopedia.com

and natural language processing technologies since 1999⁶³, is an example of the close bond between central government and companies in key sectors. Together with the BATs, in 2017 it became part of the “AI Team” of the State, so that the CEO and founder of the company *Liu Qingfeng* became also delegate of the National People’s Congress, collaborating with the State by providing free access to the information contained in its databases in the name of the protection of the national security and social surveillance purposes. In fact, iFlytech helps in the build of the current digital surveillance system applied by the government both in its “smart cities” and online world, as explained by Maya Wang in the 2017’s report conducted by the Human Rights Watch, when stating that the company’s tools are an essential part of the party’s plan to “build a digital totalitarian state” thanks to the leverage of iFlytek’s voice biometric technologies which make “tracking and identifying individuals possible”.⁶⁴

This control system rises the feeling of bewilderment among Chinese citizens and a general confusion as it is not clear if the object of protection is the individual or the national safety against the individual: the current security measures are based on the presumption of guilt, indeed, everyone is presumed to be a danger to public safety and needs to pass through increasingly stringent security checks without exception; in this way, people feel they are over-controlled and mistrusted by the state.⁶⁵

In the next chapter the privacy issue in China will be deeply analyzed to better understand Chinese people’ perception of this topic and government’s measures to protect their citizens’ personal privacy.

⁶³ iFlytek is a well-known intelligent speech and artificial intelligence publicly listed company in the Asia Pacific Region. In 2017 the MIT Technology Review named it to its list of the world’s 50 smartest companies, as it actively promotes the development of AI products with a visions that voice computing will penetrate every sphere of society, aligning with the Chinese Communist Party’s vision for a surveillance state. For a more understanding of the company’s profile see: <https://www.iflytek.com/en/usr/uploads/2020/04/3271692900.pdf>

⁶⁴ Mara HVISTENDAHL, How a Chinese AI Giant Made Chatting-and Surveillance-Easy, in “Wired.com”, October 5,2020. <https://www.wired.com/story/iflytek-china-ai-giant-voice-chatting-surveillance/>

⁶⁵ LAO Dongyan 劳东燕, *renlian shibie jishu de yinyou* 人脸识别技术的隐忧. (The hidden worries of facial recognition technology), October 31, 2019. <https://xw.qq.com/cmsid/20191230A08BFU00> translation available at https://docs.google.com/document/d/18L4FuiUjGN5Y2j_4-VnYi2U116KWEIaG4pxVp78x-ss/edit#heading=h.x7nrlj7mw63v

3.1 The Concept of Privacy in Chinese Society: Are Chinese Consumers Concerned about Privacy?

Privacy is a common value that exists in every culture, but the interpretation and the expression of this concept is different worldwide.

The first definition of privacy dates back to 1890 and is attributable to American jurists Samuel D. Warren and Louis Brandeis, who in their article “*The Right to Privacy*” linked the term privacy with the “right to be left alone”.⁶⁶ Thus, the idea of privacy today was developed out of a Western moral and legal framework that presupposes the primacy of individual rights against collective values, except in the most extreme circumstances.

Western societies considers privacy as a right belonging to the individual and people concern about its protection is very high, as exemplified in the graph below, in which a study conducted by the Boston Consulting Group (BCG) company has shown that the country whose consumers expressed the lowest concern about sharing personal information online is China.⁶⁴

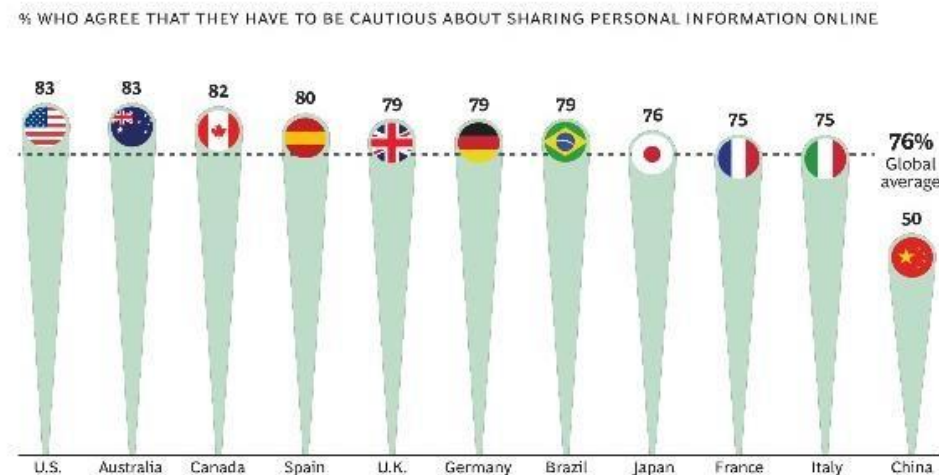


Figure 10 Countries' concern about sharing personal information online (Source: Boston Consulting Group)

⁶⁶ Samuel D. WARREN & Louis D. BRANDEIS, *The Right to Privacy*, 4, HARV. L REV., 1890, p.193.

⁶⁴ BCG, 2013 Global Consumer Sentiment Survey.

This graph could support the common belief that considers the Chinese population as "asleep" or "inert" to social control policies and indifferent to asserting their privacy rights.⁶⁷ However, these beliefs are not entirely true as the term "privacy" is understood differently in China as a result of a long cultural and historical tradition. Indeed, the reason of these findings may be linked to a different perception of the term 'privacy' which, while in the West has individualistic connotations, in China assumes a more collectivistic meaning.

In China, the perception of privacy has been developing for some years and is becoming increasingly important nowadays, especially regarding the online world. In fact, starting from 28th December 2012, China's Standing Committee of the National People's Congress issuing the *Decision on Strengthening Network Information Protection*⁶⁸ raised for the first time the problem of the protection of Chinese netizens but above all the need to strengthen the protection of information in the networks, demonstrating the rising awareness of citizens of sharing their data online. This decision represented the basis of Chinese Government's interest in the issue and determined the beginning of the development of subsequent laws and regulations related to the Internet privacy.

In this chapter, particular attention will be given to the analysis of the process that has led Chinese people as well as the Government and companies to reconsider the value of privacy, starting from an overview of both the cultural background, in which Confucianism and its values still permeates Chinese tradition and society, and historical, referring to the principles of the Chinese Communist Party and the imperatives of sharing and common collectivity, providing at last a brief presentation of the present political and sociological situation. In this way, outlining the evolution of the regulatory framework for the protection of privacy rights and internet privacy and the existing limits and problems.

3.1.1 Cultural Point of View (Confucian Values)

The Chinese concept of privacy differs greatly from the Western one, as it is not intended as a right of the single individual but needs to be understood in collectivistic terms. The Chinese

⁶⁷ Harrison JACOBS, Chinese People Don't Care about Privacy on the Internet, in "Business Insider", June 26, 2018. <https://www.businessinsider.com/why-china-chinese-people-dont-care-about-privacy-2018-6?IR=T>

⁶⁸ Committee of the National People's Congress, Decision on Strengthening Protection of Online Information (全国人大常委会关于加强网络信息保护的決定 *quanguo renda changweihui guanyu jiaqiang wangluo xixi baohu de jueding*), December 28, 2012. http://www.gov.cn/jrzq/2012-12/28/content_2301231.htm

Confucian tradition, in fact, established at the base of society a strictly hierarchical system⁶⁹ in which the needs of the individual was subordinate to that of the family unit, the community, and the country.

Thence, the concepts of public and private themselves needed to be considered as related to the family unit with respect to the public, rather than a relationship between the public and the individual.

In addition, the definition of privacy is often confused with portrait, reputation and personal dignity, relating it to the concept of “face” (*mianzi* 面子), still representing a sensitive issue in modern society.⁷⁰ As already said, privacy at first was considered as a prevention over the shame, in fact since 1980 it was pointed with the term *yīn sī* (阴私) meaning shameful or embarrassing private affair, while just recently it has acquired the proper meaning of *yīn sī* (隐私, privacy), aligning with the shared global definition of the term as defined by Westin (1967) who argued that privacy is “the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others”.⁶⁹

Permeated by the Confucian ideas of shame and deference to the community hierarchy, people’s main concern about the disclosure of personal information is related to the fear of being judged by others thus leading to a public shame and a loss of face, not only limited to the single individual but which could damage the reputation of his/her entire family. This means for Chinese netizens that their idea of privacy is connected to their role and duty towards the society more than to their personal right as human beings, and so does the regulations which supports the individual’s role in the community rather than protecting the individual against the community as in the West.⁷¹

⁶⁹ Confucianism is based on a set of ethical and philosophical teachings stressing the love for humanity and family, loyalty to government, and harmony in thought and conduct with one’s defined role in society. See, e.g., Judith A. BERLING, Confucianism, Focus on Asian Studies, Vol. II, No. 1: Asian Religions, 1982, pp. 5-7.

⁷⁰ In Chinese culture it is possible to distinguish two kinds of face: social face, based on one’s achievements and dignity, and moral face, based on a person’s feelings of shame. For a better understanding See Hu, “The Chinese Concepts of ‘Face.’”, American Anthropologist, vol. 46, no. 1, 1944, pp. 45, www.jstor.org/stable/662926 ⁶⁹ Hao WANG, Protecting Privacy in China, A Research on China’s Privacy Standards and the Possibility of Establishing the Right to Privacy and the Information Privacy Protection Legislation in Modern China, SpringerVerlag Berlin Heidelberg, 2011, p.34.

⁷¹ Bonnie S. MCDOUGALL, Privacy in Modern China, History Compass, (casa ed. Wiley), Vol. 2, Issue 1, 2005, pp. 1-8. <https://doi.org/10.1111/j.1478-0542.2004.00097.x>

In China national interests are used to prevail over personal interests, and the difficult balance of privacy and security has a heavy weight on the side of security and community. Therefore, China's current legal system does not afford significant privacy protections against government intrusion but provides a restriction only in the case of private and corporates' actions against people's rights.⁷²

3.1.2 Historical Background

Despite China's current digital and economic vibrancy, Confucian and Socialist influences made the commitment to the public or to the greater good of the country as an expected virtue. Therefore, during the Communist regime in 1949, both urban and rural Chinese residents have been placed under peer surveillance in the name of social good.⁷³

For many years the Communist principles shaped people's perception of the term "private" as it aimed to build a country mainly based on collectivism rather than individual pursuits, where people were considered as equal and the collectivization of private property forced people living in communes, meaning a common lifestyle and the deprivation of the possibility to own both tangible and intangible things, like the right to privacy for instance.⁷⁴

Some changes have begun to appear when China adopted its reform and Open-up Policy in 1978, which enabled people to pursue their own career and life paths and just recently, privacy have been re-evaluated as the pursuit of the new Chinese middle class thanks to the shift promoted by the Government from a planned to a market economy and from nationalization to privatization.⁷⁵

⁷² Ibidem. p.51.

⁷³ Tao FU, An Examination of Privacy in the Socio-Technological Context of Big Data and the Socio-Cultural Context of China, Southern Illinois University Carbondale, 2015, p.6.

⁷⁴ Tiffany LI, Jill BRONFMAN, Zhou ZHOU, Saving Face, Unfolding the Screen of Chinese Privacy Law, Boston University School of Law, 2018, p.6.

⁷⁵ Inundated with the global consumerism, together with the privatization of Chinese economy, Chinese society was polarized, and new social strata were formed. Since lifestyle had become the measure for one's subjectivity, there was a considerable desire for privacy. (Tao FU, An Examination of Privacy in the Socio-Technological Context of Big Data and the Socio-Cultural Context of China, Southern Illinois University Carbondale, 2015, p.71)

3.1.3 Social Control

Despite the change brought about by Chinese historical and cultural background, the right to privacy nevertheless seems to be subordinated to the needs of the central government. In fact, the Chinese government exercises profound control over society, justified by the need to maintain national stability and thus favoring the country's economic development.

Explicative examples of the social control policy could be the Great Firewall, as a means of controlling content and activities on the web, or the more recent Social Credit System, which together with the facial-recognition cameras surveillance network, implies a strict social control over Chinese society, considered by many critics as the embodiment of the world depicted in the Orwellian novel *1984* or described by political scientist Sebastian Heilmann as a ‘new digital Leninism’.⁷⁶

In this context permeated by Government’s extensive surveillance and control over social and market behaviors, Chinese people’s perception of privacy is altered. In fact, a study conducted by the KPMG corporation in 2017 privacy report⁷⁷, showed that Chinese consumers are much more likely to share their information online with companies, especially in exchange for personalized and improved services, while they show no concern regarding the compliance of these businesses with the Government. Consequently, even the regulatory system regulating the protection of consumers' personal information excludes the Chinese government from all limitation and is mostly aimed at protecting against companies and criminal activities.

3.2 There is not a unique and comprehensive law for the privacy issue

Since the post-Mao era, Chinese people began to concern more about their personal information security and so did the Government which gradually expressed its interest in protecting people’s personal information disclosure, as a result of the urge to regulate its Internet-driven economy.

⁷⁶ Sebastian HEILMANN, Big Data Reshapes China’s Approach to Governance, in “Financial Times”, September 29, 2017. <https://www.ft.com/content/43170fd2-a46d-11e7-b797-b61809486fe2>

⁷⁷ “Although 60 percent of the respondents think personalized billboard advertising is cool, 39 percent are extremely concerned about the way companies handle and use their personal data”, KPMG, Crossing the Line – Staying on the right side of consumer privacy, in “KPMG International”, p. 12, January 2017.

Indeed, given the incredible technological increase of the internet in the Chinese territory and the development of an extremely efficient and functional digital retail system, many problems related to security arose, such as in the case of hacker attacks or illegal uses or dispersion of users' personal and financial data.⁷⁸

Hence, a national-level cybersecurity law was a key priority for the Chinese government in order to better protect consumers' privacy, but also for companies in order not to suffer from potential leak risks and enjoy a more trustworthy reputation in the eyes of consumers. This need led to the emanation in 2017 of the *Cybersecurity Law*, representing today's main law concerning the regulation of online activities both for institutions and users.

However, even though Government's significant strides to create new laws governing data privacy and confidentiality of personal information, Chinese privacy law is still undeveloped and is still in its early stages of implementation. A comprehensive and uniform national privacy law does not exist yet, and the existing regulations and governmental guidance include similar elements of international norms but even some Chinese-specific characteristics, such as the exemption from the applicability of the law as what concerns the government's influence and matters of public interest.⁷⁹

3.2.1 The Cybersecurity Law of the People's Republic of China

The first law aiming to regulate the management of information and security online is the recent *Zhonghua renmin gongheguo wangluo anquanfa* 中华人民共和国网络安全法, Cybersecurity Law of the People's Republic of China⁷⁹ (hereinafter "Cybersecurity Law"),

⁷⁸ According to a report conducted by China's CNNIC, in 2016, 70.5 percent of netizens encountered cyber security incidents (CNNIC, Statistical Report of Internet Development in China, 2017, pp. 105-109). While in 2019 still high is the number of websites tampered with hacker attacks (185,573). CNNIC, Statistical Report of Internet Development in China, April 2020, pp. 75-80. <https://cnnic.com.cn/IDR/ReportDownloads/>

⁷⁹ JIANG Lin 姜琳, *Lun wangluo jiaoyi zhong xiaofei zhe quanyi de baohu* 论网络交易中消费者权益的保护 (On the Protection of Consumer Rights and Interests in Online Transactions), in "CNKI.net", 2018. kns.cnki.net.

⁷⁹ Cybersecurity Law of the People's Republic of China, *Zhonghua Renmin Gongheguo Wangluo Anquan Fa* 中华人民共和国网络安全法, 2016, in Chinese: http://www.gov.cn/xinwen/2016-11/07/content_5129723.htm

which was adopted at the 24th Session of the Standing Committee of the 12th National People's Congress on 7 November 2016 and is effective since the 1st of June 2017.

Composed by seven chapters and seventy-nine articles, this law was “enacted in order to protect cyber security, safeguard cyberspace sovereignty and national security, social public interests, protect the legitimate rights and interests of citizens, legal persons, and other organizations, and promote the healthy development of economic and social informatization”⁸⁰ (Art. 1) and for the management of the network security inside the territory of the PRC⁸¹ (Art. 2). Moreover, it adopts measures to monitor, defend and deal with cybersecurity risks and threats originating within and outside the PRC, protects critical information infrastructure from attacks, intrusions, interference and damage, punishes illegal and criminal activities on the Internet in accordance with the law and maintains Internet security and order.⁸² (Art. 5)

For the first time in Chinese law history, a proper definition of “personal information” 个人信息 is provided, defined in chapter seven Article 76.5 as “all kinds of information, recorded electronically or through other means, that taken alone or together with other information, is sufficient to identify a natural person's identity, including, inter alia, name and surname, date of birth, identification number, biometric data, address and phone number”.⁸³

Even though it is not entirely focused on the issue of users’ privacy, these kinds of information are safeguarded under a set of articles which provide obligations for “network operators” 网络运营者⁸⁴ and at the same time aiming for more transparency in the market and the prevention of information leaking cases. Thence, the principle of confidentiality is clearly stated, as Article 40 includes a duty for network operators to improve users’ information

⁸⁰ 中华人民共和国网络安全法 *Zhonghua Renmin Gongheguo Wangluo Anquan Fa*, Cybersecurity Law of the People’s Republic of China, 2016, Art. 1.

A translation of selective articles taken into consideration in this thesis may be found at: <http://www.lawinfochina.com/display.aspx?id=22826&lib=law&SearchKeyword=cybersecurity%20law&SearchCKeyword=>

⁸¹ Ibidem. Art. 2

⁸² Ibidem. Art. 5

⁸³ Ibidem. Art.76.5

⁸⁴ Network operator refers to the owner, manager and network service provider of the network. (Ibidem. Art. 76.3)

protection systems⁸⁵ and Articles 44 and 45 stress on their liability of keeping the information, legally collected, safe and confidential⁸⁶; they ought to pay attention to specific guidelines also following the principles of lawfulness 合法, fairness 正当 and necessity 必要 and before collecting users' personal data, they are required to state in detail their purpose, method and scope both for the collection, processing and the use of the insights captured online, in order to obtain users' consent for it (Art. 41)⁸⁷. In addition, information shall be inherent to the type of service they provide and shall not be disclosed, destroyed or shared with others without the consent of the user being collected.

The safety of personal information is a fundamental priority, as Article 42 underlines, network operators have the responsibility to ensure the prevention of information leakage, damage or loss, but in the event that they fail in doing so, they should immediately take remedial measures and promptly notify users and report to relevant competent authorities.⁸⁸

This law draws upon the relative provisions of the *General Data Protection Regulation* (GDPR) of the European Union (2016), as well as the U.S. *Consumer Privacy Protection Act of 2015*, and establishes the international principles for the collection, use and handling of personal information of network operators, such as the principles of openness, informed consent, clear purpose, and limitation of purpose.

Given these provisions, Chinese netizens should feel relieved, being enhanced with the right to be informed whether a company wants to have access to or make use of their personal data (Art. 22) and if they find out that the collected information are wrong or violate the laws of the agreement by and between the operator and the user, they could make appeal to their right to be forgotten⁸⁹, demanding for the correction or deletion of his/her personal information

⁸⁵ “网络运营者应当对其收集的用户信息严格保密，并建立健全用户信息保护制度” (Network operators shall strictly keep confidential the user information they collect and establish and improve user information protection systems.) Ibidem. Art. 40

⁸⁶ Article 44 of the CSL recites: “No individual or organization may steal or obtain personal information in other illegal ways, or illegally sell or illegally provide personal information to others.” *Supra* note 16 Art. 44. While Article 45 states: “Departments and their staff that are responsible for network security supervision and management in accordance with the law must strictly keep the personal information, privacy, and business secrets learned in the performance of their duties confidential, and must not disclose, sell or illegally provide them to others.” Ibidem. art. 45.

⁸⁷ Ibidem. Art. 41

⁸⁸ Ibidem. Art. 42

⁸⁹ This provision partially draws upon the right to correction under Article 16 and the right to be forgotten under Article 17 of the European GDPR

<https://eur-lex.europa.eu/legalcontent/EN/TXT/?qid=1601291356387&uri=CELEX:32016R0679>

(Art. 43)⁹⁰. In this way, surfing online becomes a conscious act for consumers which empowers them with the right to decide whether to give their consent for the treatment of their information, and enables them to establish a relationship of trust with the seller. In this win-win process, both offer something to each other: consumers offer their data and preferences, fundamental for retailers' market strategies, while retailers lay bare, offering customers transparency and insights into their business.

However, in exchange with the needed services, Chinese users are imposed to supply network operators with their official personal identity authentication, as exposed in Article 24⁹¹, otherwise network operators who fail in requiring users to provide true identity information or in providing relevant service to users who do not provide true identity information could be fined from 50.000 yuan to 500.000 yuan or may see their business suspended (Art. 61).⁹²

The importance of a real-name identification system is crucial for the Government's philosophy of fostering a "responsible" Internet. In fact, according to Li Gang, dean of the Internet Application College of Shenyang City University, this provision could restrain Internet users to a certain extent, making them responsible for their own remarks, and thus purifying the Internet environment, regulating and tracing netizens' social behavior.⁹³ Nonetheless, this regulation has simultaneously led to fierce controversy as Professor Hu Yong, associate professor at the School of Journalism and Communication at Peking University, criticizes it for the deprivation of users from the right to anonymity, stating that it may harm people's privacy and freedom of speech.⁹⁴

As regards the State's interest in internet security, another important provision is that included in article 37, in which operators of "Critical Information Infrastructure"⁹⁵ 关键信息基础设施

⁹⁰ *Supra* note 16, Art. 43.

⁹¹ *Ibidem*.Art.24.

⁹² *Ibidem*. Art.61.

⁹³ Jyh-An LEE, Ching-Yi LIU, Real-Name Registration Rules and the Fading Digital Anonymity in China, 25 Wash. L. Rev. 1 (2016). Available at: <https://digitalcommons.law.uw.edu/wilj/vol25/iss1/3>

⁹⁴ Yunqiao PU 普韵乔, Wangluo shimingzhi quanmian daolai, ruhe baozhang women de xuni kongjian geng "qingshuang" 网络实名制全面到来, 如何保障我们的虚拟空间更 "清爽"(The network real-name system is coming, how to make our virtual space more "fresh"), in "Xinhua", August 30, 2017. http://www.xinhuanet.com/live/2017-08/30/c_136568575.htm

⁹⁵ A definition of "critical information infrastructure" is not provided in the Cybersecurity Law, but according to the Guidance for National Network Safety Inspection Operation, established in June 2016 by the Office of the Central Leading Group for Cyberspace Affairs, **critical information infrastructure** refers to "operating information systems or industrial control systems that provide network information services to the public or support energy, communications, finance, transportation, utilities and other important industries". In other words, it includes not only governmental websites, but also popular online platforms providing instant

施的运营者(CII) are subject to special requirements in connection with procurement of products and services and cross-border transfer of data. In particular, an operator of CII is subject to *data sovereignty* requirement, prohibiting the transfer abroad of any important data or personal information collected within the PRC territory, and even subjecting to security assessment and government approval all the data that urge to provide overseas due to business needs.⁹⁶ This means for China more control over companies' business, as the latter are required to build up China-based databases instead of transferring the insights in clouds around the world, while an increase in costs and a loss of direct control for foreign multinationals.⁹⁷

Indeed, during the legislative process of this rule, the American Chamber of Commerce led an opposition group against the rule's enactment, fearing that foreigners' businesses related to the data locally stored in China would have been jeopardized by the law.⁹⁸ This protest confirms the fact that the data localization rule does raise several concerns, mostly within foreign firms as there are many parts and statements that appear uncertain to interpret. Therefore, a definition of what "important information" and "critical information infrastructure" does not appear in the document, leaving doubts and uncertainty to both domestic and foreign businesses.⁹⁹

Given the lack of legal framework for privacy issues in China, despite its ambiguities and typically Chinese characteristics, Cybersecurity law is globally perceived as the most

messaging, ecommerce, search engine, email, map, and other services.
<https://doi.org/10.1016/j.clsr.2018.08.007>

⁹⁶ Article 37 of the Cybersecurity Law provides "Personal information and important data collected and generated by operators of critical information infrastructure during operations within the territory of the People's Republic of China shall be stored within the territory. If it is really necessary to provide it overseas due to business needs, the security assessment shall be conducted in accordance with the methods formulated by the national cyberspace administration department in conjunction with the relevant departments of the State Council; where laws and administrative regulations provide otherwise, follow those provisions."
http://www.cac.gov.cn/2016-11/07/c_1119867116_3.htm

⁹⁷ Yang YUAN, China's Cyber Security Law Rattles Multinationals, in "Financial Times", May 30, 2017,
<https://www.ft.com/content/b302269c-44ff-11e7-8519-9f94ee97d996>

⁹⁸ Minjie ZHU 朱敏洁, *Wai mei cheng wangluo anquan fa hui yaoqiu mei qi xiang zhongguo jiao chu yuan daima* 外媒称网络安全法会要求美企向中国交出源代码 (Foreign media said the Cybersecurity Law will require US companies to hand over source code to China), in "Guancha.cn", November 11, 2016.
https://www.guancha.cn/FaZhi/2016_11_08_379835_1.shtml

⁹⁹ John SELBY, Data Localization Laws: Trade Barriers or Legitimate Responses to Cybersecurity Risks, or Both?, in "International Journal of Law and Information Technology", Vol.25, No.3, 2017,
<https://doi.org/10.1093/ijlit/eax010>.

effective safeguard for Chinese citizens' online privacy rights even though it provides a basic legal framework for cyberspace governance in China, to be supplemented by implementing regulations in years to come.

3.2.2 Implementing Regulations and Guidelines

The interest for the protection of privacy began with the issue of the *Decision on Strengthening Online Information Protection* (The Decision), by the Standing Committee of the National

People's Congress, effective from December 28, 2012, which provided general principles for network service providers to protect personal information of Chinese netizens. The Decision, together with the *National Standard of Information Security Technology – Guideline for Personal Information Protection within Information System for Public and Commercial Services* (effective from February 1, 2013) represent the backbone of general data protection rules currently in the PRC.

Even though at present Cybersecurity Law represents China's first national-level law¹⁰⁰ to address cybersecurity and data privacy protection, other important implementing regulations and guidelines have been proposed, issued or revised and are part of a basic legal system established for protecting personal data and which sets out the compliance requirements for market players to follow. Indeed, Since the Cybersecurity Law is a high-level law and does not provide practical guidelines, China's national data protection authorities have been drafting a series of related implementation regulations and national standards in order to reinforce and better flesh out the principles introduced under the Cybersecurity Law and together constituting China's legal regime for cybersecurity and data protection. Some of the most relevant guidelines are listed in the table below:

¹⁰⁰ G. GREENLEAF, and S. LIVINGSTONE, China's Cybersecurity Law – also a data privacy law?, *Privacy Laws & Business International Report*, 144, pp. 1-7, 2016.

Title	Issuer	Date of effectiveness
National Standard of Information Security Technology – <i>Personal Information Security Specification (PI Specification)</i> 信息安全技术 个人信息安全规范	The State Administration for Market Regulation (SAMR) and the Standardization Administration of China (SAC)	Effective since 2018. It will be replaced on October 1, 2020 by the newly revised version issued on March 6 th , 2020.
Guidelines on Internet Personal Information Security Protection 互联网个人信息安全保护指南	PRC's Ministry of Public Security (MPS) 中华人民共和国公安部	Effective from April 19, 2019
(Draft) National Standard of Information Security Technology – Guidelines on Personal Information Security Impact Assessment 信息安全技术-个人信息安全影响评估指南	全国信息安全标准化委员会 National Information Security Standardization Committee	Released on June 11, 2018.
Information Security Technology- Guidelines for Personal Information Protection Within Public and Commercial Services Information Systems (信息安全技术公共及商用服务信息系统个人信息保护指南)	The Standardization Administration of China (SAC)	Effective since February 1, 2013.
Information Security Technology - Guide for De-identifying Personal Information (信息安全技术 - 个人信息去标识化指南),	The Standardization Administration of China (SAC)	effective from 1 March 2020.
Cybersecurity Practices Guidelines – Self-assessment Guide for Collecting Personal Information by Mobile Internet Applications 2020 (网络安全标准实践指南 - 移动互联网应用程序(App)收集使用个人信息自评评估指南)	The Standardization Administration of China (SAC)	Issued and effective since July 22, 2020.
Information Security Technology – Guidelines for Personal Information Notices and Consent (Draft for Comments) (信息安全技术 - 个人信息告知同意指南(征求意见稿)),	The Standardization Administration of China (SAC)	Released on January 20, 2020.

While these guidelines are only technical guides providing detailed guidance on data processing, and thus not legally binding, they are regarded as important references by enterprises. The PI Specification in particular, provides a wider and more exhaustive definition of ‘personal information’ respect to Cybersecurity Law, including “names, dates of birth, identity card numbers, biometric information, addresses, communication records and contents, account passwords, property information, credit information, location data, health and physiological information, transaction data etc.”. In addition, it provides for the first time a precise definition of “personal sensitive information” 个人敏感信息, describing them as “information that, once leaked, illegally provided, or abused, can threaten personal and property security and/or easily cause personal reputation damage, physical and mental health damage or discrimination”. Additionally, PI Specification empowers netizens with the ability to decide whether to receive or not a Personalized Display of search results of products or services and to control the degree and extent to which their personal information can be utilized to generate the Personalized Display.¹⁰¹

Other laws whose selected articles impact on data protection include:

- *The General Provisions of the Civil Law of the People’s Republic of China*¹⁰² which, in Article 111, sets out the basic principle that natural person’s personal data shall be protected by law and enforces the concept of users’ consent to the treatment of their personal data in order for network operators to collect and process the insights, also prohibiting the illegal disclosure, provision and selling or purchasing of personal data from or towards any third parties;
- *The Criminal Law of the People’s Republic of China* (Criminal Law) sets forth offences relating to infringing personal data and privacy as article 253 explains, condemns the refusing to fulfil information network security responsibilities, in article 286, and relates to the offence of stealing, purchasing or illegally disclosing other people’s credit card information, in article 177. As per these issues, further explanation is provided by

¹⁰¹ Lester ROSS, Kenneth ZHOU & Tingting LIU, China Issue New Personal Information Security Specification, in “Wilmerhale.com”, March 24, 2020. <https://www.wilmerhale.com/en/insights/client-alerts/20200324-chinaissues-new-personal-information-security-specification>

¹⁰² Adopted at the 5th session of the 12th national people’s congress on March 15,2017.

the Interpretation of the Supreme People's Court and the Supreme People's Procuratorate on Several Issues concerning the Application of Law in the Handling of Criminal Cases of Infringing on Citizens' Personal Data (“Interpretation on Criminal Cases of Infringing on Citizen’s Personal Data) issued in 2017;

- *E-commerce Law of the People's Republic of China* took effect on January 1, 2019. It fosters consumer data protection (art. 5)¹⁰³ by requiring e-commerce operators to provide absolute transparency and by placing restrictions on abuses of consumer profiling. Indeed, article 18 states that

if an e-commerce operator provides him with search results for goods or services according to the characteristics of the consumer's interests, hobbies and consumption habits, he shall at the same time provide the consumer with the option not to target his or her personal characteristics, and respect and equally protect the legitimate rights and interests of the consumer.¹⁰⁴

The E-commerce Law provides specific obligations for retailers, such as that included in article 19, which warns e-commerce operators to bring goods or services to the attention of consumers in a significant manner without the use of tying goods or services as an option for default consent,¹⁰⁵ or that included in article 24, which empowers consumers with the right to demand for cancellation or correction of their personal data, reciting that e-commerce operators shall immediately take remedies when receiving an application for user information correction or deletion.¹⁰⁶ Article 25 introduces the concept of the compliance of e-commerce operators with competent authorities in case of necessity, providing information on its e-commerce data¹⁰⁷, confirming the concern of some foreign companies about the interference of Chinese government on their business.

¹⁰³ Article 5 of the E-commerce Law recites “The State shall protect consumers' legal rights and interests against infringement. The State shall adopt measures to protect consumers in the legal exercise of their rights and shall safeguard consumers' legal rights and interests.” (E-commerce Law of the People's Republic of China, *zhonghua renmin gongheguo dianzishangwufa 中华人民共和国电子商务法*, Standing Committee of the National People's Congress, effective since January 1, 2019. http://www.cac.gov.cn/2018-09/01/c_1123362506.htm)

¹⁰⁴ Ibidem. Article 18

¹⁰⁵ Ibidem. Article 19

¹⁰⁶ Ibidem. Article 24

¹⁰⁷ Ibidem. Article 25

In conclusion, Chinese privacy law is still fairly underdeveloped and fragmented and could raise ambiguity especially for foreign enterprises and users, mainly regarding the interference of the Government keeping up with its social control policy. The weakness in China's privacy legislation is due both for its legal status and especially for the many loopholes in it: some sizable exemptions are present for the collection and use of data, including when related to security, health, or the vague and flexibly interpretable "significant public interest". In fact, the privacy system in China seems aiming to protect the privacy of consumers, rather than that of citizens as a whole, thus allowing the government to collect personal data whenever it may be necessary or merely useful for its policies.¹⁰⁸

However, the efforts of Chinese government and competent authorities to expand the regulatory framework in recent years are remarkable, as evidenced by the approval of the National People's Congress of the new Chinese Civil Code, which will cover a range of topics on individuals' civil rights, and will provide further legal basis for individuals to protect their personal data and privacy.¹⁰⁹

The Standing Committee of China's National People's Congress (NPCSC) this year, released also a draft of the recent *Data Security Law of the People's Republic of China* (Draft Law)¹¹⁰ on July 3, 2020 and that of the forthcoming *Personal Information Protection Law* which are expected to come into force in 2021, to put China into a path in line with global trends in data protection law and to represent a major step towards the regularization of personal information protection into an integrated, comprehensive framework.

¹⁰⁸ Luciano FLORIDI, et. al., "The Chinese Approach to AI: An Analysis of Policy, Ethics, and Regulation", AI & Soc, 2019, pp.13-14. <https://doi.org/10.1007/s00146-020-00992-2>

¹⁰⁹ Provisions covering the right to privacy and personal information protection are included in the 6th chapter of the section "Personal Rights of the Civil Code 民法典人格权编". <https://npcobserver.com/legislation/civil-code/>

¹¹⁰ Lisa L.S. MARK, China Releases Draft Data Security Law, in "Haynesboone", August 9, 2020. <https://www.haynesboone.com/alerts/china-releases-draft-data-security-law>

CONCLUSIONS

The thesis aimed to underline the need for retailers, which has now become impossible to ignore, to abandon the traditional methods of sales, communication and production of the product, in favor of the adoption of analytical technologies in order to keep up with a world in which change is the only constant and where digital innovation is impacting every sector. By placing particular emphasis on the retail field, China has shown itself ready to face the challenges brought about by digital advancement and the change in the seller-consumer relationship, while in the past the power to decide which products to offer to the mass was exclusively held by sellers. Now, thanks to the advent of the Internet and the development of infrastructures, encouraged by the Chinese government with policies such as the "Internet +" of the State Council and the ambitious "Made in China 2025" project, customers have developed their own decision-making autonomy and more sophisticated tastes, having the possibility of evaluating various offers from an infinite multitude of companies with a simple "click". It is precisely the digitization of the purchasing process that has allowed companies to have an enormous amount of data available from various touchpoints and the fact that consumers are constantly bombarded with so much information from the most disparate products, which has meant that now the trend sees the consumer at the center of the company's decision-making process, the latter more attentive than ever to the needs and requirements of the former. However, nothing would be possible without the exploitation of Big Data and the techniques of analyzing the numerous information which, no longer analogue but now digital, offer a much more complete and at the same time complex picture of the habits, preferences and personal characteristics of consumers. Big Data offers an opportunity for retailers to create added value to their products and, having the ability to anticipate demand and therefore predict trends and sales, to improve the management of their supply chain avoiding waste, waste of time valuable and money.

Within the paper, I have tried to provide the reader with a description of the collection, analysis and storage processes of Big Data, as well as a rather exhaustive definition of what they are and how they work and the surprising evolution they have had over the years, because without understanding this new phenomenon it is easy for consumers to continue helplessly sharing important information unconsciously. Proceeding with the drafting of the text, I noticed the advantages that this innovation can bring: we have seen how companies, through the right application of big data, have the possibility to improve operational

efficiency, reduce costs, improve production performance, customer relations, formulate a communication strategy in real-time, improve and simplify the decision-making process. Companies are therefore faced with a real innovation which, however, also brings with it the most critical aspects since it requires high investments in technologies and highly specialized know-how and can lead to fragile results if based solely and exclusively on data analysis and not on specific studies and knowledge of the sector. The final impact, however, is a competitive advantage, in the long term, which can hardly be recovered by competitors who will not be able to derive the right benefits from the data.

Big data will therefore continue to be highly relevant in the retail market as we move towards a more engaging consumer journey. It is therefore time and duty for actors and stakeholders to create a secure network environment to address the many challenges associated with increasing the power of these technological structures, which if abused, misused or not adequately protected can result in a loss of trust in the company and even more importantly, a violation of the personal sphere of the consumer, undermining one of the fundamental human rights, the right to privacy.

Although the concept of privacy in China is relative and often connected to the concept of "mianzi" and therefore of personal reputation in the eyes of society and not so much as in the Western sense of the term, linked to the right to decide how and to what extent to share its information with third parties, in 2012, in this regard, the Standing Committee of the National People's Congress issued the Decision on Strengthening Online Information Protection, marking the beginning of a growing interest in the regulation and protection of the online network. This passage also marks the awakening of Chinese citizens, asleep for years of Confucian traditions and characterizing historical events, who began to demand greater protection for themselves to deal with the various problems encountered online, including fraud and extortion of personal information. It was then shown that the Chinese people are less sensitive to the risks associated with sharing information with companies, in exchange for personalized and high-performance benefits or services, demonstrating the fact that if they are used and measured well, Big Data and data retention can truly enrich the consumer experience without creating skepticism.

The report shows that the current regulatory framework relating to the protection of privacy is still very fragmented: the most current cultural, historical and social origins mean that a

comprehensive law for the protection of privacy does not exist. Furthermore, the Chinese government is excluded from any obligation of the law and its intervention / influence is justified by the need to ensure a certain stability to the country and preserve national security. Today, as in Confucian society, the public good has priority over the rights of the individual, typical of a purely collectivist Chinese society. Moreover, the laws on the subject have a generalizing structure that raises concern especially in foreign players, especially as regards the limitations envisaged for the export of sensitive data outside the territory of the PRC and for the compliance of foreign countries with different legal systems with the will of the Chinese government. Retailers are required absolute transparency, both towards users and the government, in order to make the users involved in online data retention activities an ethical treatment of personal data, that is, in compliance with the consensual principle of users.

However, the need to regulate the use of the Internet and its resources is a fundamental aspect for the realization of the project by Xi Jinping himself to make AI and Big Data the pillars of Chinese technological power, tools that according to the plan " Made in China 2025 ", will allow the country to be able to confront the world powers and achieve the goal of becoming the world leader in technology by 2030; moreover, as he himself declared in his speech at the National Conference on Cybersecurity and Informatization, in 2018, "Without network security, there will be no national security, there will be no economic and social stability, and the interests of the masses will be difficult to be protected." (in Chinese 没有网络安全就没有国家安全, 就没有经济社会稳定运行, 广大人民群众利益也难以得到保障。).

In conclusion, it is important to reiterate that China lacks a strong legislative framework for the protection of consumer privacy, despite the presence of some guidelines and laws, including the Cybersecurity Law of the People's Republic of China, which represents the closest preventing web threats such as disclosure, illegal sale of citizens' personal information and hacker attacks, the Chinese government is required to make more effort to implement new and more accurate laws in the years to come.

References

Alibaba Cloud, Finding Similar Images with Alibaba Cloud Image Search, in

“Alibabacloud.com”, July 6, 2018.

https://www.alibabacloud.com/blog/finding-similarimages-with-alibaba-cloud-image-search_593805

Alibaba Group: Fiscal Year 2020 Annual Report.

<https://www.alibabagroup.com/reports/fy2020/ar/ebook/en/index.html>

BIRTHWHISTLE, Tom, China’s Next Retail Disruption: End-to-End Value Chain Digitisation, PwC’s Global Consumer Insights Survey 2018 China Report, 2018.

BRAITHWAITE, Charlie, *Customer data in retail 2019: Key research findings*, in

“Clickz.com”, 7 May 2019. <https://www.clickz.com/customer-data-in-retail-2019-keyresearch-findings/235595/>

CAMPBELL, Chris, Top Five Differences between Data Lakes and Data Warehouses, in “Bluegranite.com”, January 26, 2015.

<https://www.bluegranite.com/blog/bid/402596/topfive-differences-between-data-lakes-and-data-warehouses>

CHANG, Wo, GRADY, Nancy, NIST Big Data Interoperability Framework: Definitions, Volume 1, National Institute of Standards and Technology (NIST), September 16, 2015.

<https://datasciencedegree.wisconsin.edu/data-science/what-is-big-data/>

CHENG, Michael, *E-commerce in China-the future is already here*, in “PwC’s total retail

survey 2017”, [https://www.pwccn.com/en/retail-and-consumer/publications/total-retail-](https://www.pwccn.com/en/retail-and-consumer/publications/total-retail-2017-china/total-retail-survey-2017-china-cut.pdf?utm_content=buffer28aaf&utm_medium=social&utm_source=twitter.com&utm_campaign=buffer)

[2017-china/total-retail-survey-2017-china-](https://www.pwccn.com/en/retail-and-consumer/publications/total-retail-2017-china/total-retail-survey-2017-china-cut.pdf?utm_content=buffer28aaf&utm_medium=social&utm_source=twitter.com&utm_campaign=buffer)

[cut.pdf?utm_content=buffer28aaf&utm_medium=social&utm_source=twitter.com&utm_campaign=buffer](https://www.pwccn.com/en/retail-and-consumer/publications/total-retail-2017-china/total-retail-survey-2017-china-cut.pdf?utm_content=buffer28aaf&utm_medium=social&utm_source=twitter.com&utm_campaign=buffer)

CNNIC 44th Statistical Internet Report

CRONE, Sven F., LESSMANN, Stefan, STAHLBOCK, Robert, "The impact of preprocessing on data mining: An evaluation of classifier sensitivity in direct marketing", European Journal of Operational Research, 173, 2006, pages 781-800.

Dataflair team, Big Data in Retail Industry, in “Dataflair.com”, October 18, 2019.

<https://data-flair.training/blogs/big-data-in-retail-industry-real-world-uses-examples/>

DE HERT, Paul & PAPAKONSTANTINOY, Vagelis, The Data Protection Regime in China, in www.europarl.europa.eu, pp. 13-24, 2015.

[https://www.europarl.europa.eu/RegData/etudes/IDAN/2015/536472/IPOL_IDA\(2015\)53_6472_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/IDAN/2015/536472/IPOL_IDA(2015)53_6472_EN.pdf)

Deloitte China, “Big Data: What’s the Big Deal?”, Deloitte Perspective Report, Vol. 4, 2013, pages 1-13.

Deloitte China, Serial Studies of “Future of Consumer”: Big Data Driven Consumer Insights, January 2020. <https://www2.deloitte.com/cn/en/pages/consumer-business/articles/big-data-driven-consumer-insights.html>

Deloitte, Digital Transformation Through Data: a Guide for Retailers to Drive Value with Data, May 2019.

<https://www2.deloitte.com/content/dam/Deloitte/us/Documents/consumer-business/usdigital-transformation-through-data-for-retail.pdf>

DIEBOLD, Francis X., On the Origin(s) and Development of the Term 'Big Data', PIER Working Paper No. 12-037, September 21, 2012. <http://dx.doi.org/10.2139/ssrn.2152421>

E-commerce payment trends: China, in “J.P. Morgan.com”, 2019

<https://www.jpmorgan.com/merchantservices/insights/reports/china>

ERIKSON, Kate, *Anticipating customer needs with data and AI*, in “Clikz.com”, 2 July 2020. <https://www.clickz.com/anticipating-customer-needs-with-data-and-ai/261985/>

FLORIDI, L., et. al., “The Chinese Approach to AI: An Analysis of Policy, Ethics, and Regulation”, *AI & Soc*, 2019, pp.13-14. <https://doi.org/10.1007/s00146-020-00992-2>

FOOTE, Keith D., A Brief History of Data Lakes, in “dataversity.net”, July 2, 2020.

<https://www.dataversity.net/brief-history-data-lakes/>

GetSmartcoders, 4 Big Ways in which Big Data Can Transform Retail Industry, September 2016. <https://www.getsmartcoders.com/blog/4-big-ways-in-which-big-data-can-transform-retail-industry/>

GREENLEAF, Graham & LIVINGSTONE, Scott, China’s Cybersecurity Law – also a data privacy law?, *Privacy Laws & Business International Report*, 144, pp. 1-7, 2016.

GREENLEAF, Graham & LIVINGSTONE, Scott, China's Personal Information Standard: The Long March to a Privacy Law, in "Privacy Law & Business International Report", 150, pp.25-28, 2017.

GUTIERREZ, Daniel, How Retailers Are Using Big Data to Their Advantage, in "Insidebigdata.com", January 21, 2017.

<https://insidebigdata.com/2017/01/21/howretailers-are-using-big-data-to-theiradvantage/#:~:text=With%20the%20introduction%20of%20big,customer%20will%20buy%20next%20time.>

GUTTA, Surya, Data Science: The 5V's of Big Data, in "Medium.com", May 4, 2020.

<https://medium.com/@suryagutta/the-5-vs-of-big-data-2758bfcc51d>

HAN, Jiawei, PEI, Jian and KAMBER, Micheline, Data Mining: Concept and Techniques, 3rd Edition, Morgan Kaufmann, June 2011.

HANDLEY, Erin, China's Mass Surveillance of Uyghur Muslims in Xinjiang Province Revealed in Data Security Flaw, in "www.abc.net.au", February 18, 2019.

HEILMANN, Sebastian, Big Data Reshapes China's Approach to Governance, in "Financial Times", September 29, 2017. <https://www.ft.com/content/43170fd2-a46d-11e7b797-b61809486fe2>

HU, H., "The Chinese Concepts of 'Face.'", American Anthropologist, Vol. 46, No. 1, pp. 45–64, 1944. www.jstor.org/stable/662926

HUA TAN, Kim, JI, Guojun, PENG LIM, Chee & TSENG, Ming-Lang, *Using big data to make better decisions in the digital economy*, in "International Journal of Production Research", Volume 55, pages 4998-5000, 24 May 2017.

IBM, What is Big Data Analytics?, in "ibm.com"

https://www.ibm.com/analytics/hadoop/big-dataanalytics?mhsrc=ibmsearch_a&mhq=what%20is%20big%20data%3F

JACOBS, Harrison, Chinese People Don't Care about Privacy on the Internet, in "Business Insider", June 26, 2018. <https://www.businessinsider.com/why-china-chinese-people-dontcare-about-privacy-2018-6?IR=T>

JIANG, Lin 姜琳, *Lun wangluo jiaoyi zhong xiaofei zhe quanyi de baohu* 论网络交易中消费者权益的保护 (On the Protection of Consumer Rights and Interests in Online Transactions), in "CNKI.net", 2018. kns.cnki.net.

JIANG, Min & FU, King-Wa, Chinese Social Media and Big Data: Big Data, Big Brother, Big Profit?, Policy & Internet, Vol. 10, No. 4, 2018.

<https://onlinelibrary.wiley.com/doi/epdf/10.1002/poi3.187>

JONGEN, Wijnand, *The End of Online Shopping: The future of new retail in an alwaysconnected world*, 2018,

<https://books.google.it/books?id=3lZoDwAAQBAJ&printsec=frontcover&dq=big+data+i+n+china+retail&hl=it&sa=X&ved=2ahUKEwi2kYyt1bbrAhUEKewKHa2bD4E4FBD0A TAGegQICBAC#v=onepage&q=big%20data%20in%20china%20retail&f=false>

KANTARCIOGLU, Murat & FERRARI, Elena, Research Challenges at the Intersection of Big Data, Security and Privacy, in *Frontiers in Big Data*, Vol. 2, No. 1, February 14, 2019. doi: 10.3389/fdata.2019.00001

KLEBNIKOV, Sergei, *Alibaba's 11/11 Singles' Day By The Numbers: A Record \$38 Billion Haul*, in "Forbes", November 11, 2019.

<https://www.forbes.com/sites/sergeiklebnikov/2019/11/11/alibabas-1111-singles-day-bythenumbers-a-record-38-billion-haul/#678a3b272772>

LAO Dongyan 劳东燕, *renlian shibie jishu de yinyou* 人脸识别技术的隐忧. (The hidden worries of facial recognition technology), October 31, 2019.

<https://xw.qq.com/cmsid/20191230A08BFU00>

translation available at https://docs.google.com/document/d/18L4FuiUjGN5Y2j_4-VnYi2U116KWEIaG4pxVp78x-ss/edit#heading=h.x7nrlj7mw63v

LEVINE, Barry, *How retailers struggle between data privacy and personalization*, in "Clickz.com", 14 January 2020. <https://www.clickz.com/how-retailers-struggle-between-data-privacy-and-personalization/259491/>

LI, Lingfang, TADELIS, Steven & ZHOU, Xiaolan, *Buying Reputation as a Signal of Quality: Evidence from an Online Marketplace*, March 24, 2017.

https://editorialexpress.com/cgi-bin/conference/download.cgi?db_name=IIOC2017&paper_id=75

LI, Tiffany, BRONFMAN, Jill, ZHOU, Zhou, *Saving Face; Unfolding the Screen of Chinese Privacy Law*, Boston University School of Law, 2018.

LOTAME, 1st Party Data, 2nd Party Data, 3rd Party Data: What Does It All Mean?, in

“Lotame.com”, August 19, 2019. <https://www.lotame.com/1st-party-2nd-party-3rd-partydata-what-does-it-all-mean/>

Lotus Ruan, *Big Data in China and the battle for privacy*, in “Australian Stratetig Policy Insitute”, 22 June 2018. <https://www.aspi.org.au/report/big-data-china-and-battle-privacy>

LU, Yao-Huai, Privacy and Data Privacy Issues in Contemporary China, *Ethics and Information Technology*, Vol. 7, pp. 7–15, 2005. <https://doi.org/10.1007/s10676-005-0456y>

LUO, Su 罗素, *Wangluo jiaoyi zhong xiaofei zhe quanyi baohu yanjiu* 网络交易中消费者权益保护研究 (“ Research on the Protection of Consumers’ Rights and Interests in Internet Transactions”), in “CNKI.net”, May 1, 2017.

<https://kns.cnki.net/kcms/detail/detail.aspx>

MANYIKA, James, CHUI, Michael, Big Data: The Next Frontier for Innovation, Competition and Productivity, in “McKinsey.com”, May 1, 2011.

<https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/big-data-the-next-frontier-for-innovation>

McDOUGALL, Bonnie, Privacy in Modern China, *History Compass*, Vol.2, pp. 1-8, 2004.

McKinsey China Digital Consumer Trends 2019.

<https://www.mckinsey.com/~media/mckinsey/featured%20insights/china/china%20digital%20consumer%20trends%20in%202019/china-digital-consumer-trends-in-2019.ashx>

McKinsey&Company, “Big Data, Analytics and the Future of Marketing & Sales”, Marketing & Sales, March 2015.

MEGAX, *Retail in China long tail on steroids*, in “Medium.com”, June 12 2019, <https://medium.com/@megaXcoin/retail-in-china-long-tail-on-steroids-100c19850d19>

MEGAX, *Retailtainment: retail is not about selling products*, in “Medium.com”, 13 Feb 2019.

<https://medium.com/@megaXcoin/retailtainment-retail-is-not-about-sellingproducts-9a169c344701>

MENON, Pradeep, *New Retail: How Big Data and AI is Fueling Retail Transformation*, in “Medium.com”, August 6, 2019,

<https://medium.com/@rpradeepmenon/new-retail-howbig-data-and-ai-are-fuelling-retail-transformation-76c81105c82d>

MILGROM, Paul R., TADELIS, Steven, *How Artificial Intelligence and Machine Learning Can Impact Market Design*, National Bureau of Economic Research (NBER), Cambridge, working paper 24282, February 2018

National Bureau of Statistics of China, *Statistical Communiqué of the People's Republic of China on the 2019 National Economic and Social Development*, February 28, 2020.

http://www.stats.gov.cn/english/PressRelease/202002/t20200228_1728917.html

NOEL, Steffi, *AI and Big Data in Retail in China*, in “Daxueconsulting.com”, December 26, 2019. <https://daxueconsulting.com/ai-big-data-retail-china/>

OH, Nathaniel, *A China perspective: how big data is changing the retail landscape*, in “The Xnode.com”, 30 June 2018.

<http://www.thexnode.com/blog/a-china-perspective-how-bigdata-is-changing-the-retail-landscape>

OHLHORST, Frank, *Big Data Analytics: turning Big Data into Big Money*, Hoboken, John Wiley & Sons, Inc., 2012.

PERNOT-LEPLAY, Emmanuel, *China's Approach on Data Privacy Law: A Third Way Between the U.S. and the EU?*, in “the Penn State Journal of Law and International Affairs”, vol. 8.1., 27 March 2020. [China CyberSecurity Law: Comparison with the GDPR & US Laws](#)

PIERANNI, Simone, *Big Data e Intelligenza Artificiale. La Sfida della Cina*, in “il Manifesto”, February 12, 2019.

<http://www.minimaetmoralia.it/wp/big-data-intelligenzaartificiale-la-sfida-della-cina/>

PU, Yunqiao 普韵乔, *Wangluo shimingzhi quanmian daolai, ruhe baozhang women de xuni kongjian geng “qingshuang”* 网络实名制全面到来, 如何保障我们的虚拟空间更“清爽”(The network real-name system is coming, how to make our virtual space more “fresh”), in “Xinhua”, August 30, 2017.

http://www.xinhuanet.com/live/201708/30/c_136568575.htm

RAM, Jiwat, ZHANG, Changyu, and KORONIOS, Andy, “The Implications of Big Data Analytics on Business Intelligence: A Qualitative Study in China”, *Procedia Computer Science*, 87, 2016, pages 221-226.

SCHWARTZ, Berry, *The Paradox of Choice: Why More Is Less*, New York, Ecco, 2004.

SELBY, John, Data Localization Laws: Trade Barriers or Legitimate Responses to Cybersecurity Risks, or Both?, in “*International Journal of Law and Information Technology*”, Vol.25, No.3, 2017, <https://doi.org/10.1093/ijlit/eax010>.

SHEIK, Meyar, *AI Driven Personalization Yields Impressive Retail Roi*, in “Clickz.com”, 19 June 2020. <https://www.clickz.com/ai-driven-personalization-yields-impressive-retailroi/261846/>

SOO, Zen, *China’s ‘post-millennial’ generation is different. Here’s how*, in “South China Morning Post”, August 17, 2020.

<https://www.scmp.com/tech/innovation/article/2160003/chinas-post-millennialgeneration-different-heres-how>

Sources of Big Data: Where Does It Come From?, in “Cloudmoyo.com”, August 18, 2016. <https://www.cloudmoyo.com/blog/data-architecture/what-is-big-data-and-where-it-comesfrom/>

Stirista, Third-Party Data: What it is, its Benefits and Best Use Case, in “Stirista.com”, March 16, 2020. <https://www.stirista.com/third-party-data>

Tao FU, *An Examination of Privacy in the Socio-Technological Context of Big Data and the Socio-Cultural Context of China*, Southern Illinois University Carbondale, 2015. <https://search.proquest.com/docview/1734464489?accountid=17274>)

TOTI, E., *La Legge sull’Ecommerce: Maggiore Sicurezza Informatica anche a Vantaggio degli Operatori Stranieri*, in “Cinalex”, September 7, 2018.

<http://www.cinalex.it/blog/p/lalegge-sulle-commerce-maggiore-sicurezza-informatica-anche-a-vantaggio-deglioperatori-stranieri>

WALKER, Michael, *Data Veracity*, in “Datasciencecentral.com”, November 28, 2012. <https://www.datasciencecentral.com/profiles/blogs/data-veracity>

WANG, Hao, *Protecting Privacy in China, A Research on China's Privacy Standards and the Possibility of Establishing the Right to Privacy and the Information Privacy Protection Legislation in Modern China*, Springer-Verlag Berlin Heidelberg, 2011, p.51.

WANG, Helen H., *Why China is leading us to the future of commerce*, Forbes, April 3 2017.

<https://www.forbes.com/sites/helenwang/2017/04/03/why-china-is-leading-us-to-the-future-of-commerce/#81a87435d8ea>

WARREN, Samuel D. & BRANDEIS, Louis D., *The Right to Privacy*, Harvard Law Review, Vol. 4, No. 5, pp. 193- 220, 1890.

WEINSWIG, Deborah, *Retail 2020: 10 Trends for China E-Commerce*, Coresight Research, January 20, 2020. <https://coresight.com/wp-content/uploads/2020/01/Retail-2020-10-Trends-for-China-E-Commerce-Jan-20-2020-2.pdf>

WESSING, Taylor, *China's New E-commerce Law- Good or Bad News for Business?*, in "Lexology.com", November 30, 2018. [China's New E-commerce Law - Good or Bad News for Business? - Lexology](https://www.lexology.com/library/detail.aspx?l=5111111)

XU, Jinghong, *Evolving Legal Frameworks for Protecting the Right to Internet Privacy in China*, in "China and Cybersecurity: Espionage, Strategy, and Politics in the Digital Domain", (a cura di) J.R. Lindsay, T.M. Cheung & D.S. Reveron, Chapter 10, 2015. doi: [10.1093/acprof:oso/9780190201265.001.0001](https://doi.org/10.1093/acprof:oso/9780190201265.001.0001)

YUSUPU, Aniwa'er 艾尼瓦尔·玉苏普, *Woguo wangluo xiaofei zhe quanyi baohu falu wrnti yanjiu* 我国网络消费者权益保护法律问题研究 (Study on the Legal Issue of the Protection of the Rights and Interests of Online Consumers in China), in "CNKI.net", 2016. <https://kns.cnki.net/kcms/detail/detail.aspx>

Zhao, H. and Dong, H.X. (2017) *Research on Personal Privacy Protection of China in the Era of Big Data*. Open Journal of Social Sciences, 5, pages 139-145. <https://doi.org/10.4236/jss.2017.56012>

ZHENG, Yiran, *Customer-to-Manufactory Business Model Set for Rapid Growth*, in "Chinadaily.com.cn", May 17, 2019. <http://www.chinadaily.com.cn/a/201905/17/WS5cdea56da3104842260bc56b.html>

ZHU, Minjie 朱敏洁, *Wai mei cheng wangluo anquan fa hui yaoqiu mei qi xiang zhongguo jiao chu yuan daima* 外媒称网络安全法会要求美企向中国交出源代码 (Foreign media

said the Cybersecurity Law will require US companies to hand over source code to China), in “Guancha.cn”, November 11, 2016.

https://www.guancha.cn/FaZhi/2016_11_08_379835_1.shtml

ZUBERI, Mel F., RAJARATNAM, Daniel, *Measuring Retail Performance in an Omnichannel World*, Journal of Marketing Channels, 26, 25th March 2020, Pages 120-126. <https://www.tandfonline.com/doi/full/10.1080/1046669X.2020.1741992>

List of Laws and Regulations

- Committee of the National People's Congress, 全国人大常委会关于加强网络信息保护的決定 *quanguo renda changweihui guanyu jiaqiang wangluo xinxi baohu de jue ding* (Decision on Strengthening Protection of Online Information), (2012).
http://www.gov.cn/jrzq/201212/28/content_2301231.htm
- Committee of the National People's Congress. (2013) Decision on Amending the Law of the People's Republic of China on the Protection of Consumer Rights and Interests (2013 Amendment)(中华人民共和国消费者权益保护法 修改 *zhonghua renmin gongheguo xiaofeizhe quanyi baohufá xiugāi*), http://www.gov.cn/jrzq/2013-10/25/content_2515455.htm
- Consumer Privacy Protection Act of 2015 of the United States of America, April 30, 2015.
<https://www.congress.gov/bill/114th-congress/senate-bill/1158/text>
- Cybersecurity Law of the People's Republic of China, *Zhonghua Renmin Gongheguo Wangluo Anquan Fa* 中华人民共和国网络安全法, Standing Committee of the National People's Congress, 2016, in Chinese: http://www.npc.gov.cn/npc/xinwen/2016-11/07/content_2001605.htm. Translation of selected articles is available at: <http://www.lawinfochina.com/display.aspx?id=22826&lib=law&SearchKeyword=cybersecurity%20law&SearchCKeyword=>
- Cyberspace Administration of China, Cyber Security Review Measures, April 27, 2020.
http://www.cac.gov.cn/2020-04/27/c_1589535450769077.htm
- E-commerce Law of the People's Republic of China, *zhonghua renmin gongheguo dianzishangwufa* 中华人民共和国电子商务法, Standing Committee of the National People's Congress, effective since January 1, 2019.
http://www.cac.gov.cn/201809/01/c_1123362506.htm
- General Data Protection Regulation (GDPR) of the European Union, 2016.
<https://eurlex.europa.eu/legal-content/EN/TXT/?qid=1601291356387&uri=CELEX:32016R0679>