Review Article

The Influence of Big Data Analytics on E-Commerce: Case Study of the U.S. and China

Weiqing Zhuang [ID]

School of Internet Economics and Business, Fujian University of Technology, Fuzhou, China

Correspondence should be addressed to Weiqing Zhuang; zmakio@aliyun.com

Big data analytics (BDA) is a wide and deep application in e-commerce, which impacts positively on the global economy, especially the U.S. and China who have done well. This paper seeks to examine the relative influence of theoretical research and practical activities of BDA in e-commerce to explain the differences between the U.S. and China according to the two main literature databases, Web of Science and CNKI, respectively, and by employing other samples that present retail e-commerce sales and the number of some data companies founded in the U.S. and China each year. We further determine the reasons leading to the difference between the U.S. and China in BDA in e-commerce, which can help managers devise appropriate business strategies in e-commerce for each of them, and provide a proof of the significant relationship of theoretical research and practical activities in BDA in e-commerce. In addition, the variables related to big data companies show a moderation effect rather than mediating effect relative to the practice of theoretical research in e-commerce in the United States, but they show a moderate effect and mediating effects in China. The results of this study help clarify doubts regarding the development of China's e-commerce. Moreover, three orientations in e-commerce using BDA and the use of quantum computing in e-commerce to solve existing e-commerce problems are explored to provide better evidence for decision-making that could be valuable in future research.

1. Introduction

Big data are a frontier topic for researchers [1] and have always influenced academic research [2]. However, there are six debates regarding the aspects of the big data literature, including different approaches to big data analytics (BDA), artificial intelligence, big data capability, big data-driven business organizing, big data access, and social risks of big data value realization [3]. The severer is the privacy trust crisis in the era of big data, both in the field of enterprise services and some public services; moreover, there are also doubts about data since individuals may have inaccurate information [4]. More precisely, the key point of the debate and crisis regarding big data is how to use big data well and in a correct manner that is frank and honest, agreeing at the outset to focus only on what really matters [5]. For example, considerations include two things: how to adjust algorithms to ever-changing conditions [6] and how to evade the negative effect of "big data hubris" [7].

Hence, general statistical techniques and computational algorithms are issues that require different tools to manage big data sets [8], particularly with respect to different qualities of information disclosure for different purposes. BDA requires understanding the relationships among features and the explored data [9]. It has evolved from the statistical techniques for data mining from the 1970s to business intelligence (BI) 3.0 today [10]. Another reason that prompts us to focus on this area is that China accounted for 23.1 percent of the total online retail sales in 2017, while the United States ranked behind the UK, South Korea, and Denmark, only sharing 9 percent. Data come from the Statista website, https://www.statista.com/statistics/255083/online-sales-as-share-of-total-retail-sales-in-selected-countries/. Obviously, e-commerce marketing not only depends on information

technology or the population, but it is likely to reflect on the oneness of value and perception that the U.S. lacks. Therefore, the situation of the United States and China regarding research on and practice of BDA in e-commerce is worthy of discussion.

Akter and Wamba [11] fully demonstrate that BDA in e-commerce, as an emerging field since 2006, exhibits strategy-led analytics and has sustainable value-driven facets for businesses involved in organization management, goods sales, production management, data quality, IT infrastructure and its security, HRM, overarching values, and so on. Additionally, Manyika et al. [12] propose five major contributions that big data can promote to businesses: creativity, performance, consumer behavior, decision-making, and innovative business models. In short, a series of new issues in BDA in e-commerce should be focused on, including how to determine the first-rank relationship among a commodity's dynamic pricing, dynamic subsidizing, and cost to e-commerce parties, and how to enact good policies with BDA in dynamic pricing to buyers that depend on a buyer's past good or bad actions, including comments, sales returns, and sharing. Actually, goods prices have not been adjusted to a per-buyer basis in e-commerce firms, while their expenditures differ in completing the transaction. These aspects require a determination of whether there are any respective methods and experiences for the U.S. and China to solve these issues. This paper summarizes the relevant literature and applications as follows.

1.1. U.S. Similarly, there are also many examples that occur in practice. Data breaches occur daily in the American society; Taylor [13] provides a list of 17 breaches, such as the Yahoo 500 million user accounts attacked in 2013, Adult Friend Finder in 2016, eBay in 2014, Equifax in 2017, Heartland Payment Systems in 2008, Target Stores in 2013, TJX Companies in 2006, Uber in 2016, JP Morgan Chase in 2014, US OPM in 2012, Sony's PlayStation Network in 2011, Anthem in 2015, RSA Security in 2011, Stuxnet in 2010, VeriSign in 2010, Home Depot in 2014, and Adobe in 2013. Of course, all of these occurrences are just the tip of the iceberg, as Facebook's data crisis has made Americans worried about and disgusted with big data once again. To some extent, Americans pay more attention to the safety and rationality of data use may be due to their cultural and cognitive characteristics [14], which is somewhat different from China [15, 16].

1.2. China. By first searching "big data" and then add "business" to the Chinese literature database "CNKI" to classify papers by title, only 34 papers were found, and no paper was retrieved when also searching for "crisis" in these 34 papers. Moreover, some critics state that big data in China are overhyped because companies are more interested in using big data to attract media and investors. Data come from the China government website; "How Baidu, Tencent, and Alibaba are leading the way in China's big data revolution," http://www.scmp.com/tech/innovation/article/1852141/how-baidu-tencent-and-alibaba-are-leading-way-chinas-big-data. However, increasing cases of stealing and

trafficking of personal information have been reported in China, covering hundreds of millions of items in transportation, logistics, health care, social networking, and banking. For example, 300 million user information leaks from Shunfeng Express and 500 million user data leakages from Huazhu Group or Wanhao Group have occurred.

Regardless of how serious the criticisms are, a notable example shows that Alibaba predicted the dynamic change of income from impoverished people through mining and analyzing their transaction data on its e-commerce platform and then helped the Chinese government to target poverty alleviation. Of course, since this process is not accurate enough to identify each impoverished person's income, Alibaba has cooperated with the government to develop a platform of "Internet plus targeted poverty alleviation," which connects several public service data and also clarifies the main reason why certain people are still poor, including illness that leads to poverty or unemployment, disasters, laziness, etc.

This paper is composed of four sections and focuses on two topics: BDA in e-commerce for the U.S. and China in academic research and in practice. The second section presents a discussion of empirical analysis on the connection of BDA in e-commerce in theory and practice for the United States and China. The third section (Conclusions) sums up the comparison of BDA in e-commerce between the U.S. and China in theoretical research and in practice. The last section explores avenues for future research and application. This paper focuses on comparing BDA in e-commerce between the United States and China by examining their level of theoretical research and practical application by quantifying the literature and the market and on explaining the status of e-commerce development and the main factors that influence it in the U.S. and China.

2. Discussion

There are several issues, such as the relationship between academic researches and practical activities of BDA in e-commerce, the difference in the relationship between the U.S. and China, and how academic research affects practical activity on BDA in e-commerce for U.S. and China severally, all of which should be discussed in depth after detailed literature review and the analysis above. It widely adopts a statistical analysis method for investigating the effect that theoretical research promotes dramatically practical activities in BDA in e-commerce. Specifically, what follows builds regression models of academic research and its application to find out the action mechanism of BDA applied in e-commerce and then has a comparison of the moderating and mediating effects of BDA between U.S. and China and has an investigation on lag consideration of academic research response to practical application in BDA for e-commerce.

2.1. Data Acquisition. The relevant literature for BDA in e-commerce between U.S. and China, collected by two main literature databases, Web of Science and CNKI, respectively, is aimed at U.S. and China, during the period from 1990 to

TABLE 1: Retail e-commerce sales in the U.S. and China.

Year	U.S.				China					
	Total retail sales ($ billion)	Retail e-commerce sales ($ billion)	Proportion of e-commerce in total retail	Growth rate of retail e-commerce	Total retail sales (¥ billion)	Retail e-commerce sales (¥ billion)	Total retail sales** ($ billion)	Retail e-commerce sales** ($ billion)	Proportion of e-commerce in total retail	Growth rate of retail e-commerce
2017	5076.00	453.46	8.93%	16.54%	36626.20	7175.10	5420.64	1061.91	19.59%	39.17%
2016	4856.33	389.11	8.01%	14.39%	33231.63	5155.57	5002.46	776.08	15.51%	32.97%
2015	4725.99	340.16	7.20%	13.98%	30093.08	3877.32	4789.13	617.05	12.88%	38.98%
2014	4639.44	298.44	6.43%	14.47%	27189.61	2789.80	4415.24	453.03	10.26%	49.70%
2013	4458.45	260.72	5.85%	13.23%	23780.99	1863.60	3865.39	302.91	7.84%	42.15%
2012	4302.23	230.26	5.35%	15.38%	21030.70	1311.00	3332.67	207.75	6.23%	67.52%
2011	4102.95	199.56	4.86%	17.45%	18391.86	782.60	2844.60	121.04	4.26%	53.72%
2010	3818.05	169.92	4.45%	16.78%	15699.84	509.10	2319.04	75.20	3.24%	96.87%
2009	3612.47	145.51	4.03%	2.76%	13267.84	258.60	1942.32	37.86	1.95%	105.73%
2008	3935.32	141.59	3.60%	3.75%	10848.77	125.70	1560.35	18.08	1.16%	124.06%
2007	3995.18	136.47	3.42%	20.41%	8921.00	56.10	1171.77	7.37	0.63%	163.38%
2006	3871.57	113.33	2.93%	24.02%	7641.00	21.30	958.37	2.67	0.28%	10.36%
2005	3689.28	91.39	2.48%	25.88%	6717.66	19.30	819.73	2.36	0.29%	141.25%
2004	3473.05	72.60	2.09%	27.02%	5950.10	8.00	719.16	0.97	0.13%	105.13%
2003	3262.73	57.16	1.75%	28.08%	5251.63	3.90	634.47	0.47	0.07%	129.41%
2002	3128.55	44.62	1.43%	30.25%	4813.59	1.70	581.57	0.21	0.04%	183.33%
2001	3062.27	34.26	1.12%	24.09%	4305.54	0.60	520.06	0.07	0.01%	—
2000	2983.28	27.61	0.93%	90.53%	3910.57	—	472.37	—	—	—
1999	2803.09	14.49	0.52%	190.73%	3564.79	—	430.64	—	—	—
1998	2581.76	4.98	0.19%	—	3337.81	—	402.11	—	—	—

Notes: (1) the data from 2017 U.S. total retail sales and retail e-commerce sales are from the Digitalcommerce360 website (https://www.digitalcommerce360.com/article/us-ecommerce-sales/); other years are from the United States Census Bureau website (https://www.census.gov/); the data for 2017 China total retail sales and retail e-commerce sales are from the China National Bureau of Statistics website and ECRC website (http://www.stats.gov.cn/tjsj/zxfb/201802/t20180228_1585631.html and http://www.100ec.cn/zt/17wlls/); other years are from Yue Hongfei, National Report on E-Commerce Development in China. Inclusive and Sustainable Industrial Development Working Paper Series WP 17, 2017. United Nations Industrial Development Organization. (2) **The yearly average exchange rate for USD (US Dollar) to CNY (Chinese Yuan) during the period from 1990 to 2017 shown on the website https://www.ofx.com/en-us/forex-news/historical-exchange-rates/yearly-average-rates/. (3) Date (GMT), (4) rate, (5) date (GMT), (6) rate, (7) date (GMT), (8) rate, (9) 31-Dec-2017, (10) 6.756806, (11) 31-Dec-2007, (12) 7.613239, (13) 31-Dec-1997, (14) 8.319331, (15) 31-Dec-2016, (16) 6.643058, (17) 31-Dec-2006, (18) 7.972895, (19) 31-Dec-1996, (20) 8.338875, (21) 31-Dec-2015, (22) 6.283627, (23) 31-Dec-2005, (24) 8.19495, (25) 31-Dec-1995, (26) 8.370025, (27) 31-Dec-2014, (28) 6.158134, (29) 31-Dec-2004, (30) 8.273679, (31) 31-Dec-1994, (32) 8.639665, (33) 31-Dec-2013, (34) 6.152292, (35) 31-Dec-2003, (36) 8.277176, (37) 31-Dec-1993, (38) 5.779529, (39) 31-Dec-2012, (40) 6.310468, (41) 31-Dec-2002, (42) 8.276877, (43) 31-Dec-1992, (44) 5.52057, (45) 31-Dec-2011, (46) 6.46553, (47) 31-Dec-2001, (48) 8.278869, (49) 31-Dec-1991, (50) 5.333729, (51) 31-Dec-2010, (52) 6.769961, (53) 31-Dec-2000, (54) 8.278676, (55) 31-Dec-1990, (56) 4.792069, (57) 31-Dec-2009, (58) 6.830938, (59) 31-Dec-1999, (60) 8.277917, (61) 1 unit of USD = X units of CNY, (62) 31-Dec-2008, (63) 6.952764, (64) 31-Dec-1998, 8.300753.

TABLE 2: Listing of a limited number of data companies in the U.S. and China by founded year.

Founded year	Number of data companies in the U.S.					Number of data companies in China				
	Total	Data/ technology	Business analytics	Industrial application	Research/ consulting	Total	Data/ technology	Business analytics	Industrial application	Research/ consulting
2016						106	13	17	40	36
2015	1	0	0	0	1	221	35	31	128	27
2014	8	2	0	6	0	230	34	35	140	21
2013	30	8	3	18	1	148	27	28	80	13
2012	41	5	3	32	1	129	31	19	71	8
2011	51	9	1	37	4	110	23	25	58	4
2010	50	10	4	35	1	79	20	10	46	3
2009	32	7	1	23	1	76	18	10	43	5
2008	26	5	2	14	5	66	15	17	28	6
2007	28	8	1	18	1	57	16	17	22	2
2006	21	3	1	15	2	49	11	13	20	5
2005	17	2	1	14	0	47	15	7	22	3
2004	13	2	0	10	1	30	9	3	18	0
2003	11	4	2	5	0	42	18	3	20	1
2002	6	1	0	5	0	27	7	6	12	2
2001	10	1	2	3	4	30	11	1	17	1
2000	16	6	1	7	2	38	13	6	17	2
1999	10	3	1	6	0	22	9	5	7	1
1998	10	1	1	7	1	19	7	2	10	0
1997	4	0	0	4	0	12	2	4	6	0
1996	4	1	0	3	0	1	1	0	0	0
1995	6	2	0	4	0	10	3	4	3	0
1994	4	0	1	3	0	3	2	0	1	0
1993	5	1	0	2	2	4	1	1	2	0
1992	2	2	0	0	0	4	2	0	2	0
1991	0	0	0	0	0	0	0	0	0	0
1990	2	0	1	1	0	0	0	0	0	0
1980s	27	3	3	17	4	3	2	0	1	0
1970s	18	4	6	6	2	0	0	0	0	0
1960s	16	4	1	5	6	0	0	0	0	0
1950s	5	1	2	2	0	2	1	0	0	1
Total	474	95	38	302	39	1565	346	264	814	141

Notes: the data of the number of data companies in the U.S. are from the OpenData500 website (http://www.opendata500.com/us/list/), and China's data are from the Data Technology Industry Innovation Institute (a report published a listing of 1574 big data companies in 2017); the category of data companies both in the U.S. and China is definitely displayed according to their initial data.

2017. As a whole, we present the results of three stages of searching for subject terms classified by title from several literature databases.

Table 1 presents the growth of retail e-commerce sales in the U.S. and China, and it is observed that the speed of development in the U.S. is significantly different from that in China. China's e-commerce market saw high growth in he past and in the present which will continue to rise in the future. It will lead to a tremendous market in the e-commerce industry applying big data compared with the U.S.

Table 2 lists a certain (limited) number of data companies in the U.S. and China divided by year, which is subject to the difficulty of obtaining the complete information of big data companies. This table collects and organizes data from "OpenData500" for the U.S. and "Data Technology Industry Innovation Institute" for China and can be used to perform a correlation analysis on the issue of theoretical research in BDA in e-commerce to guide practices for the U.S. and China.

2.2. Variable Setting and Data Disposal. Conducting an empirical analysis of the connection of BDA in e-commerce in theory and practice, such as in comparing the United States and China, requires variables to be set up to represent various subjects for literature retrieval and practical activities. The details are shown in Table 3. For instance, searching the subject term of "E-Commerce" in

Table 3: Variable setting.

(a)

Searching subject term	Literature database	Variable	Searching subject term	Literature database	Variable	Searching subject term	Literature database	Variable	Searching subject term	Literature database	Variable
"E-Commerce"	ProQuest	X01	"Business Intelligence Analytics"	WoS	X11	"Apriori"	WoS	X21	"Clustering Algorithm" and "Big Data"	WoS	X31
	WoS	X02		CNKI	X12		CNKI	X22		CNKI	X32
	CNKI (all)	X03	"Big Data Model"	WoS	X13	"k-means" and "Big Data"	WoS	X23	"Cloud" and "Big Data"	WoS	X33
	CNKI	X04		CNKI	X14		CNKI	X24		CNKI	X34
	CNKI (master's & doctoral dissertation)	X05	"Big Data Algorithm"	WoS	X15	"SVM" and "Big Data"	WoS	X25	"Regression" and "Big Data"	WoS	X35
"E-Commerce" and then "Big data"	ProQuest	X06		CNKI	X16		CNKI	X26		CNKI	X36
	WoS	X07	"Hadoop"	WoS	X17	"Machine Learning" and "Big Data"	WoS	X27	"Decision Analytics" and "Big Data"	WoS	X37
	CNKI (all)	X08		CNKI	X18		CNKI	X28		CNKI	X38
	CNKI	X09	"MapReduce"	WoS	X19	"Deep Learning" and "Big Data"	WoS	X29	"Optimization" and "Big Data"	WoS	X39
	CNKI (master's & doctoral dissertation)	X10		CNKI	X20		CNKI	X30		CNKI	X40
"Genetic Algorithm" and "Big Data"	WoS	X41	"Classifier" and "Big Data"	WoS	X49	"Online Consumer Behavior"	WoS	X57	"Cloud Computing & E-Commerce"	WoS	X65
	CNKI	X42		CNKI	X50		CNKI	X58		CNKI	X66
"Neural Networks" and "Big Data"	WoS	X43	"Social Network" and "Big Data"	WoS	X51	"Online Consumer Behavior & Big Data"	WoS	X59	"Artificial Intelligence & Big Data & E-Commerce"	WoS	X67
	CNKI	X44		CNKI	X52		CNKI	X60		CNKI	X68
"Text Analysis" and "Big Data"	WoS	X45	"Prediction Model" and "Big Data"	WoS	X53	"Internet of Things & E-Commerce"	WoS	X61	"Quantum Computing"	WoS	X69
	CNKI	X46		CNKI	X54		CNKI	X62		CNKI	X70
"Association Rules" and "Big Data"	WoS	X47	"E-commerce & Big Data Analytics"	WoS	X55	"Mobile Technology & E-Commerce"	WoS	X63			
	CNKI	X48		CNKI	X56		CNKI	X64			

(b)

Object	Content	Variable	Object	Content	Variable	Object	Classification	Variable	Object	Classification	Variable
U.S.	Total retail sales	Y01	China	Total retail sales	Y05	Founded number of data companies in the U.S.	Total	Y09	Founded number of data companies in China	Total	Y14
	Retail e-commerce sales	Y02		Retail e-commerce sales	Y06		Data/technology	Y10		Data/technology	Y15
	Proportion of e-commerce in total retail	Y03		Proportion of e-commerce in total retail	Y07		Business analytics	Y11		Business analytics	Y16
	Growth rate of retail e-commerce	Y04		Growth rate of retail e-commerce	Y08		Industrial application	Y2		Industrial application	Y17
							Research/consulting	Y13		Research/consulting	Y18

Notes: the default of "WoS" stands for "WoS (Core Collection)" and "CNKI" stands for "CNKI (periodical)."

TABLE 4: Partial listing of retail e-commerce sales in the U.S. and China.

Year	U.S.				China			
	Total retail sales ($ billion)	Retail e-commerce sales ($ billion)	Proportion of e-commerce in total retail	Growth rate of retail e-commerce	Total retail sales (¥ billion)	Retail e-commerce sales (¥ billion)	Proportion of e-commerce in total retail	Growth rate of retail e-commerce
2001	3062.27	34.26	1.12%	24.09%	4305.54	0.60	0.01%	154.54%
2000	2983.28	27.61	0.93%	90.53%	3910.57	0.235719337	6.02775E − 05	162.54%
1999	2803.09	14.49	0.52%	190.73%	3564.79	0.089784161	2.51864E − 05	170.55%
1998	2581.76	4.98	0.19%	75.26%	3337.81	0.033185792	9.94239E − 06	178.56%
1997	2578.32	2.84149264	0.110%	79.68%	3125.29	0.011913337	3.81191E − 06	186.56%
1996	2460.17	1.581418432	0.064%	84.10%	2836.02	0.004157362	1.46591E − 06	194.57%
1995	2342.02	0.858999692	0.037%	88.52%	2361.38	0.001411333	5.97673E − 07	202.58%
1994	2223.87	0.455654409	0.020%	92.95%	1862.29	0.000466433	2.50462E − 07	210.58%
1993	2105.72	0.236151546	0.011%	97.37%	1427.04	0.000150181	1.0524E − 07	218.59%
1992	1987.57	0.11964916	0.006%	101.79%	1099.37	4.71393E − 05	4.28785E − 08	226.60%
1991	1869.42	0.059293899	0.003%	106.21%	941.56	1.44334E − 05	1.53292E − 08	234.60%
1990	1751.27	0.028754134	0.002%	110.63%	830.01	4.31362E − 06	5.19706E − 09	242.61%

Notes: the underlined values are simulated values.

TABLE 5: Partial listing of a limited number of data companies in the U.S. and China by founded year.

Founded year	Number of data companies in the U.S.					Number of data companies in China				
	Total	Data/ technology	Business analytics	Industrial application	Research/ consulting	Total	Data/ technology	Business analytics	Industrial application	Research/ consulting
2017	34	7	2	24	2	154	29	26	83	17
2016	33	6	2	23	2	106	13	17	40	36
2015	1	0	0	0	1	221	35	31	128	27

Notes: these underlined values are simulated values.

the ProQuest database is denoted by "X01," and the total retail sales in the U.S. are denoted by "Y01."

In addition, time series data for "X01" to "X70" and "Y01" to "Y08" are needed for considering further disposals because some variables are rare or data are missing. The method of disposal for these situations is the following: (1) the range of time series data from theoretical research is from 1990 to 2017, and retrieval results without data in the literature database are filled by default with zeroes; (2) a small amount of data are available for analysis, such as for variables X06 to X10 and X23 to X56; and (3) data are missing in aspects of retail sales in the U.S. and China from variables Y01 to Y08, which are estimated by the linear trend method at the missing point, as shown in Table 4; additionally, partial simulated values are generated for variables Y01 to Y08, as shown in Table 5. Of course, when determining model variables, we do not consider all variables in various models and select and analyze the main regression model results through repeated tests.

2.3. Descriptive Statistics. Table 6 shows that X06 and X07's Cronbach's alpha value is 0.840, which is greater than 0.5 [17] and indicates consistency for the ProQuest literature database and WoS (Core Collection) regarding the fact that X01 and X02's Cronbach's alpha value is under 0.5, similar to CNKI (all), CNKI (periodical), and CNKI (master's and

doctoral dissertation). Consequently, choosing objects from the WoS (Core Collection) and CNKI (periodical) for a theoretical comparison of the U.S. and China is sound and representative. Most variables are not good for data when their std. deviations are greater than their mean. Except for X01–X05, X12, X21, X22, X57, X63, X64, X68, X69, and X70, the others show a Kolmogorov-Smirnov Sig.<0.05 [18, 19] and have an abnormal distribution. Moreover, several methods, including converting to a normal distribution of the data [20], adopting an appropriate regression model and regression standardized residual test [21], and employing nonparametric tests [22, 23], can be used, as discussed in the next section.

As Table 7 demonstrates, data for practical activities, such as relevant retail e-commerce sales and the number of data companies founded, exhibit abnormal distributions. Only variables Y01, Y03, Y08, and Y015's show a Kolmogorov-Smirnov Sig.>0.05, and the others are lower than 0.05. In addition, the other statistics from these variables are similar to those of theoretical research variables and will be discussed regarding validation in further regression.

2.4. Linear Regression Model. Multiple regression analysis is used to determine the relationship between the dependent variables Y and independent variables X; both are time series variables [24]:

TABLE 6: Summary statistics, Cronbach's alpha values, and test results of normality for the main variables whose measure is theoretical research.

Variable	Mean	S.D.	Cronbach's alpha	N of items	Kolmogorov-Smirnov[a]
X01	1578.61	1999.45	0.123	2	
X02	220.39	171.21			
X03	565.82	546.48			
X04	283.00	282.99	0.877	3	
X05	131.39	172.08			
X06	1.36	2.83	0.842	2	Where X01-X05, X12, X21, X22, X57, X63, X64, X68, X69, X70, Sig.>0.05; others Sig.<0.05.
X07	1.29	3.52			
X08	2.79	7.05			
X09	1.71	4.86	0.831	3	
X10	0.61	1.85			
X11, X13, X15, …, X65, X67, X69			0.889	30	
X12, X14, X16, …, X66, X68, X70			0.889	30	

[a]Lilliefors significance correction.

TABLE 7: Summary statistics and test results of normality for the variables whose measure is practical activities.

Variable	N	Mean	S.D.	Median	Minimum	Maximum	Kolmogorov-Smirnov[a]		Shapiro-Wilk	
							df	Sig.	df	Sig.
Y01	28	3346.30	980.22	3367.89	1751.27	5076.00	28.00	0.200*	28	0.398
Y02	28	115.42	132.25	64.88	0.03	453.46	28.00	0.01	28	0.001
Y03	28	2.71	2.76	1.92	0.00	8.93	28.00	0.054	28	0.004
Y04	28	50.94	46.12	26.45	2.76	190.73	28.00	0	28	0
Y05	28	10680.61	10786.62	5600.87	830.01	36626.20	28.00	0.001	28	0
Y06	28	855.70	1790.16	5.95	0.00	7175.10	28.00	0	28	0
Y07	28	3.01	5.36	0.10	0.00	19.59	28.00	0	28	0
Y08	28	134.52	70.87	147.90	10.36	242.61	28.00	0.200*	28	0.098
Y09	28	16.96	15.09	10.50	0	51	28.00	0.011	28	0.004
Y10	28	3.43	3.10	2	0	10	28.00	0.002	28	0.007
Y11	28	1.07	1.09	1	0	4	28.00	0	28	0.001
Y12	28	11.39	11.00	6.50	0	37	28.00	0.001	28	0.002
Y13	28	1.11	1.37	1	0	5	28.00	0	28	0
Y14	28	61.21	64.84	40	0	230	28.00	0.018	28	0.001
Y15	28	13.29	10.74	12	0	35	28.00	0.200*	28	0.042
Y16	28	10.36	10.66	6	0	35	28.00	0.008	28	0.002
Y17	28	32.00	37.70	19	0	140	28.00	0	28	0
Y18	28	5.61	9.11	2	0	36	28	0	28	0

[a]Lilliefors significance correction; *This is a lower bound of true significance.

$$y_t = \alpha_0 + \beta_i x_{it} + e_t, \tag{1}$$

where y_t denotes the tth year observation of the dependent variable and x_{it} is a column vector of observations on i independent of the tth year. Four model specification techniques are used to select the variables in a regression model, all possible regression, forward selection, backward elimination, and stepwise regression, showed by Jomnonkwao et al. [24].

2.5. Regression Models of the U.S. Putting Theoretical Research into Practice in E-Commerce

2.5.1. Linear Regression Model for Retail Sales with Theoretical Research Variables. First, an investigation of the promoting effect of theoretical research on retail sales from these normal distribution variables, which include X02, X21, X57, X63, and X69, is conducted by means of running a stepwise regression [25, 26] in SPSS. And it is found the probability of F to enter ≤0.05 [27] of X63, X69, and X21.

TABLE 8: Model summary for Y01[e].

Model	R	R square	Adjusted R square	Std. error of the estimate	Durbin-Watson
1	0.922[a]	0.849	0.843	387.859	
2	0.939[b]	0.882	0.873	349.711	
3	0.950[c]	0.903	0.891	323.319	
4	0.950[d]	0.903	0.895	317.882	1.604

[a]Predictors: (constant), X63; [b]predictors: (constant), X63, X69; [c]predictors: (constant), X63, X69, X21; [d]predictors: (constant), X69, X21; [e]dependent variable: Y01. X21: searching subject term "Apriori" in the WoS (Core Collection). X63: searching subject term "Mobile Technology & E-Commerce" in the WoS (Core Collection). X69: searching subject term "Quantum Computing" in the WoS (Core Collection). Y01: total retail sales for the U.S.

Four models have an excellent fit, with all achieving $R^2 > 0.8$ for evaluating the dependent variable Y01 and the three independent variables, as shown in Table 8. In Table 9, the dependent variable and independent variables have a linear relation, where all Sig. of the F statistics [28] are less than 0.01, and they are available for the predictive analysis adopted by models 1, 2, and 4, where the tolerance values are the same in the multiple linear regression model and their VIF values are less than five, as shown in Table 10, which indicates there is no collinearity [29] between the independent variables, which can also be confirmed in Table 11 by the different eigenvalue and its variance proportions for the independent variables. Moreover, the standardized residuals of the regression are normally distributed [30].

Distinctly, theoretical research in the U.S. promotes dramatically practical activities in e-commerce. For example, we can more clearly understand the role of theoretical research in driving the development of retail e-commerce, from the information provided in Tables 12 and 13, in which the literature on "E-Commerce" and "E-Commerce" and "Big data" has positive effects on retail e-commerce sales in practice.

2.5.2. Investigation of the Moderation and Mediation Effects of Variables of Data Companies Founded in the U.S.
While running several stepwise regressions for variables of the founded number of data companies and some of the random theoretical research variables, it was found that these variables do not fit very well. For example, by selecting X2, X7, X13, X27, X29, X33, X55, X57, X59, X61, X63, X65, X67, and X69 as independent variables in accordance with the theoretical analysis, where the dependent variables are Y09, Y10, Y11, Y12, and Y13 in sequence, multiple linear regression models are constructed. And the results show that only the independent variable X02 was retained in the regression model along with the dependent variable, such as Y09, Y10, Y11, Y12, or Y13. However, the largest R square in these models is 0.407, which is less than 0.5. Likewise, the results show that only dependent variable Y12 is significantly related to X33 and X55 (model 2 in Table 14) while selecting X7, X13, X27, X29, X33, X55, X57, X59, X61, X63, X65, X67, and X69 as independent variables, and R square is 0.787. In summary, the variables representing the number of data

TABLE 9: ANOVA for Y01[a].

Model		Sum of squares	df	Mean square	F	Sig.
1	Regression	22031078.223	1	22031078.223	146.449	0.000[b]
	Residual	3911305.145	26	150434.813		
	Total	25942383.368	27			
2	Regression	22884937.837	2	11442468.919	93.562	0.000[c]
	Residual	3057445.531	25	122297.821		
	Total	25942383.368	27			
3	Regression	23433533.702	3	7811177.901	74.723	0.000[d]
	Residual	2508849.666	24	104535.403		
	Total	25942383.368	27			
4	Regression	23416156.434	2	11708078.217	115.865	0.000[e]
	Residual	2526226.934	25	101049.077		
	Total	25942383.368	27			

[a]Dependent variable: Y01; [b]predictors: (constant), X63; [c]predictors: (constant), X63, X69; [d]predictors: (constant), X63, X69, X21; [e]predictors: (constant), X69, X21.

companies founded do not have a good linear relation to the theoretical research variables. Therefore, their moderation and mediation effects [31, 32] will be investigated in the following section.

Here, models of two regression equations are presented, one of which is made up of the independent variables X63 and X69, moderator variable Y13, and dependent variable Y02, and the other model has the interaction term X69Y13 added. It is determined whether the moderator variable has an effect on the relationship between independent variables and dependent variables or not by judging the significance of the R square change (Sig. F change $= 0.032 < 0.05$ in Table 15), which indicates that the data companies founded (research/consulting company) in the U.S. played a moderating role [31, 32] in the theoretical research work of "e-commerce and information technology" in promoting retail e-commerce sales in practice.

Then, we test the mediation effects [31, 32] of the data companies founded variables in the U.S. as an example, such that X02 is the independent variable (shown in Table 12), Y02 is the dependent variable, and one of the data companies founded variables (Y09–Y13) is the mediating variable. The first step is the regression of Y02 on X02, which has a regression standardized coefficient of 0.797 (Sig. $= 0.000 < 0.05$, $R^2 = 0.635$); the second step is running the linear regression of the independent variable X02 and the dependent variables as one of Y09–Y13, for which all of the regression coefficients are found to be significant, less than 0.05; the last step is building a linear regression of the independent variable X02 and adding one of Y09–Y13, for which the dependent variable is Y02, and all of Y09–Y13 are found to be not significant in this regression model. Hence, a further Sobel test to judge whether variables Y09–Y13 enjoy the mediation effect or not should be performed. The results of the Sobel test shown in Table 16 indicate that the data company variables are not significant regarding their

TABLE 10: Coefficients for Y01[a].

Model		Unstandardized coefficients		Standardized coefficients	t	Sig.	Correlation			Collinearity statistics	
		B	Std. error	Beta			Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	2421.955	105.862		22.878	0.000					
	X63	43.208	3.570	0.922	12.102	0.000	0.922	0.922	0.922	1.000	1.000
2	(Constant)	2196.773	127.959		17.168	0.000					
	X63	32.697	5.118	0.697	6.389	0.000	0.922	0.788	0.439	0.396	2.527
	X69	7.940	3.005	0.288	2.642	0.014	0.830	0.467	0.181	0.396	2.527
3	(Constant)	2129.276	121.916		17.465	0.000					
	X63	5.252	12.881	0.112	0.408	0.687	0.922	0.083	0.026	0.053	18.729
	X69	11.969	3.288	0.435	3.640	0.001	0.830	0.596	0.231	0.283	3.540
	X21	26.882	11.735	0.502	2.291	0.031	0.875	0.424	0.145	0.084	11.918
4	(Constant)	2114.929	114.764		18.428	0.000					
	X69	12.960	2.179	0.471	5.948	0.000	0.830	0.765	0.371	0.622	1.608
	X21	31.332	4.238	0.585	7.393	0.000	0.875	0.828	0.461	0.622	1.608

[a]Dependent variable: Y01, model 3 indicates collinearity.

TABLE 11: Collinearity diagnostics for Y01[a].

Model	Dimension	Eigenvalue	Condition index	Variance proportions			
				(Constant)	X63	X69	X21
1	1	1.722	1.000	0.14	0.14		
	2	0.278	2.486	0.86	0.86		
2	1	2.648	1.000	0.03	0.02	0.01	
	2	0.282	3.067	0.55	0.28	0.00	
	3	0.070	6.145	0.42	0.70	0.98	
3	1	3.456	1.000	0.02	0.00	0.01	0.00
	2	0.416	2.883	0.33	0.01	0.01	0.03
	3	0.112	5.546	0.52	0.00	0.40	0.08
	4	0.015	15.046	0.14	0.99	0.59	0.89
4	1	2.548	1.000	0.03		0.02	0.04
	2	0.341	2.732	0.34		0.00	0.59
	3	0.111	4.794	0.63		0.97	0.37

[a]Dependent variable: Y01.

TABLE 12: Coefficients for Y02[a].

Model		Unstandardized coefficients		Standardized coefficients	t	Sig.	Correlation			Collinearity statistics	
		B	Std. error	Beta			Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	-20.263	25.362		-0.799	0.432					
	X02	0.616	0.091	0.797	6.729	0.000	0.797	0.797	0.797	1.000	1.000
2	(Constant)	-0.255	17.475		-0.015	0.988					
	X02	0.410	0.072	0.530	5.714	0.000	0.797	0.753	0.457	0.742	1.347
	X07	19.772	3.490	0.526	5.665	0.000	0.795	0.750	0.453	0.742	1.347

[a]Dependent variable: Y02, where model 1 R^2 is 0.635 and model 2 is 0.840; low VIF values indicate low collinearity; the standardized residuals are approximately normally distributed. X02: searching subject term "E-Commerce" in the WoS (Core Collection). X07: searching subject term "E-Commerce" and then "Big data" in the WoS (Core Collection). Y02: retail e-commerce sales.

TABLE 13: Coefficients for Y03[a].

Model		Unstandardized coefficients		Standardized coefficients	t	Sig.	Correlation			Collinearity statistics	
		B	Std. error	Beta			Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	0.419	0.204		2.055	0.050					
	X21	0.144	0.009	0.958	16.985	0.000	0.958	0.958	0.958	1.000	1.000
2	(Constant)	0.251	0.158		1.588	0.125					
	X21	0.093	0.013	0.617	7.162	0.000	0.958	0.820	0.304	0.244	4.103
	X22	0.027	0.006	0.392	4.557	0.000	0.929	0.674	0.194	0.244	4.103
3	(Constant)	0.048	0.138		0.347	0.732					
	X21	0.045	0.016	0.295	2.704	0.012	0.958	0.483	0.093	0.098	10.172
	X22	0.024	0.005	0.344	4.877	0.000	0.929	0.706	0.167	0.236	4.240
	X63	0.051	0.013	0.387	3.805	0.001	0.956	0.613	0.130	0.114	8.793

[a]Dependent variable: Y03, where model 1 R^2 is 0.917, model 2 is 0.955, and model 3 is 0.972; low VIF values indicate low collinearity; the standardized residuals are approximately normally distributed. X21: searching subject term "Apriori" in the WoS (Core Collection). X22: searching subject term "Apriori" in CNKI (periodical). X63: searching subject term "Mobile Technology & E-Commerce" in the WoS (Core Collection). Y03: proportion of e-commerce in total retail for U.S.

TABLE 14: Coefficients for Y12[a].

Model		Unstandardized coefficients		Standardized coefficients	t	Sig.	Correlation			Collinearity statistics	
		B	Std. error	Beta			Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	6.201	2.251		2.755	0.011					
	X57	0.702	0.195	0.577	3.604	0.001	0.577	0.577	0.577	1.000	1.000
2	(Constant)	2.223	1.406		1.581	0.126					
	X57	2.041	0.215	1.678	9.498	0.000	0.577	0.885	0.876	0.273	3.667
	X33	-0.261	0.036	-1.290	-7.305	0.000	0.140	-0.825	-0.674	0.273	3.667
3	(Constant)	2.316	1.300		1.782	0.087					
	X57	2.028	0.199	1.667	10.213	0.000	0.577	0.902	0.870	0.272	3.670
	X33	-0.331	0.045	-1.632	-7.401	0.000	0.140	-0.834	-0.631	0.149	6.698
	X55	0.846	0.367	0.402	2.303	0.030	0.230	0.425	0.196	0.238	4.196

[a]Dependent variable: Y12, where model 1 R^2 is 0.333, model 2 is 0.787, and model 3 is 0.826; model 3 indicates collinearity; the standardized residuals are approximately normally distributed. X33: searching subject term "Cloud" and "Big Data" in the WoS (Core Collection). X55: searching subject term "E-commerce & Big Data Analytics" in CNKI (periodical). X57: searching subject term "Online Consumer Behavior" classified by the field of "title" in the WoS (Core Collection). Y12: founded number of data companies in the U.S. classified as industrial application.

mediation effects on the given regression for the fields of theoretical research and practical activities.

2.6. Regression Models of China Putting Theoretical Research into Practice in E-Commerce

2.6.1. Linear Regression Model for Retail Sales with Theoretical Research Variables. The analysis procedure for China is the same as that performed for the U.S. First, the regression is run of the normally distributed variables Y08 and X04, X12, X22, X64, X68, and X70 by means of stepwise regression. And then, the two models are found to have an excellent fit, with all achieving $R^2 > 0.7$, as shown in Table 17. The relationship between X68 and Y08 is positive, but their relationship with X04 is negative. Similarly, China's theoretical research promotes dramatically practical activities in e-commerce. For example, Table 18 shows that the literature on "E-Commerce," "Business Intelligence Analytics,"

"Mobile Technology and E-Commerce," "Artificial Intelligence and Big Data and E-Commerce," "Quantum Computing," etc., all have positive effects on retail e-commerce sales in practice.

2.6.2. Investigation of the Moderating and Mediating Effects of the Variables of Data Companies Founded in China. While the results of running several stepwise regressions for the variables of the founded number of data companies (Y14, Y15, Y16, Y17, and Y18) and some of the random theoretical research variables (X4, X9, X14, X28, X30, X34, X56, X58, X60, X62, X64, X66, X68, and X70) are the same as those for the U.S., the results indicate that China's theoretical research promotes data companies that are founded in an obvious and direct manner, which can be seen in Table 19. This observation is opposite to that in the U.S. in this regard.

Next, we performed an investigation of the moderation and mediation effects [31, 32] of data companies. First, the

TABLE 15: Model summary for the moderation effect[d].

Model	R	R square	Adjusted R square	Std. error of the estimate	Change statistics					Durbin-Watson
					R square change	F change	df1	df2	Sig. F change	
1	0.949[a]	0.901	0.897	42.48186	0.901	235.664	1	26	0.000	
2	0.959[b]	0.921	0.914	38.73184	0.020	6.278	1	25	0.019	
3	0.973[c]	0.946	0.934	33.92368	0.026	3.530	3	22	0.032	1.783

[a]Predictors: (constant), X63; [b]predictors: (constant), X63, X69; [c]predictors: (constant), X63, X69, Y13, X69 × Y13, X63 × Y13, and the Sig. of interaction term X69 × Y13 in the regression model is 0.025, less than 0.05; [d]dependent variable: Y02; X63: searching subject term "Mobile Technology & E-Commerce" in the WoS (Core Collection). X69: searching subject term "Quantum Computing" in the WoS (Core Collection). Y13: founded number of data companies in the U.S. classified as research/consulting. Y02: retail e-commerce sales for the U.S.

TABLE 16: Sobel test for the mediation effects.

						Mediation variable								
	Y09			Y10			Y11			Y12			Y13	
Input	Sobel test		Input	Sobel test		Input	Sobel test		Input	Sobel test		Input	Sobel test	
a	0.056	Test statistic	0.726	a	0.011	Test statistic	0.196	a	0.003	Test statistic	0.379	a	0.038	Test statistic	1.039	a	0.004	Test statistic	-1.079
b	1.001	Std. error	0.077	b	1.304	Std. error	0.073	b	6.309	Std. error	0.050	b	1.906	Std. error	0.070	b	-14.736	Std. error	0.055
S_a	0.013	P value	0.468	S_a	0.003	P value	0.845	S_a	0.001	P value	0.704	S_a	0.010	P value	0.299	S_a	0.001	P value	0.280
S_b	1.360			S_b	6.650			S_b	16.496			S_b	1.765			S_b	13.145		

a = raw (unstandardized) regression coefficient for the association between X02 and mediator; s_a = standard error of a; b = raw coefficient for the association between the mediator and the Y02 (when the X02 is also a predictor of the Y02); s_b = standard error of b; calculation for the Sobel test from online, http://quantpsy.org/sobel/sobel.htm. Y09: founded number of data companies in the U.S. (total); Y10: founded number of data companies in the U.S. classified as data/technology; Y11: founded number of data companies in the U.S. classified as business analytics; Y12: founded number of data companies in the U.S. classified as industrial application; Y13: founded number of data companies in the U.S. classified as research/consulting.

TABLE 17: Coefficients for Y08[a].

Model		Unstandardized coefficients		Standardized coefficients	t	Sig.	Correlation			Collinearity statistics	
		B	Std. error	Beta			Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	196.791	9.301		21.159	0.000					
	X04	-0.220	0.023	-0.879	-9.383	0.000	-0.879	-0.879	-0.879	1.000	1.000
2	(Constant)	199.906	8.734		22.887	0.000					
	X04	-0.317	0.048	-1.268	-6.631	0.000	-0.879	-0.798	-0.576	0.206	4.844
	X68	4.175	1.828	0.437	2.284	0.031	-0.693	0.416	0.198	0.206	4.844

[a]Dependent variable: Y08, where model 1 R^2 is 0.772 and model 2 is 0.811; low VIF values indicate low collinearity; the standardized residuals are approximately normally distributed. X04: searching subject term "E-Commerce" in CNKI (periodical); X68: searching subject term "Artificial Intelligence & Big Data & E-Commerce" in CNKI (periodical); Y08: growth rate of retail e-commerce for China.

moderation of data company variables is judged in the two regression equations, one of which is composed of the dependent variables X04, X12, and X68; moderator variable Y15; and independent variable Y06, and the other equation has the added interaction term, X04Y15, X12Y15, or X68Y15. Then, the decision regarding whether the moderator variable has an effect on the relationship between the independent variables and dependent variables or not is made according to the significance of the R square change. The X04Y15 regression model indicates that the R square change is valid (Sig. F change = 0.000 < 0.05, regression coefficient Sig. = 0.000 < 0.05), the X12Y15 is Sig.F change

= 0.032 < 0.05 and regression coefficient Sig. = 0.010 < 0.05, and the X68Y15 is Sig. F Change = 0.000 < 0.05 and regression coefficient Sig. = 0.000 < 0.05, which indicates that the data companies founded variable (data/technology) in China has a moderating effect on theoretical research work promoting retail e-commerce sales in practice. We also test the mediation effects of the variables of data companies founded in China as an example, using X04 as the independent variable, Y06 as the dependent variable, and one of the data companies founded variables (Y14–Y18) as the mediating variable. The first step is the regression of Y06 on X04, which has a standardized regression coefficient of 0.842

TABLE 18: Regression models for China's theoretical research and practical activities in e-commerce.

Dependent variable	Independent variable					Statistics
	X04	X12	X64	X68	X70	
Y05	1.153*** $t = 11.358$ Sig. = 0.000				-0.221* $t = -2.1788$ Sig. = 0.039	$R^2 = 0.936$, both VIF = 4.035, $F = 183.264$ (Sig.<0.001), the standardized residuals are approximately normally distributed
Y06	0.451* $t = -2.670$ Sig. = 0.013	-0.343*** $t = -3.647$ Sig. = 0.001		0.695*** $t = 4.480$ Sig. = 0.000		$R^2 = 0.881$, VIF (X04) = 5.763, VIF (X12) = 1.787, VIF (X68) = 4.851, $F = 59.268$ (Sig.<0.001), the standardized residuals are approximately normally distributed
Y07	1.497*** $t = 8.762$ Sig. = 0.000		-0.669*** $t = -3.912$ Sig. = 0.001			$R^2 = 0.872$, both VIF = 5.686, $F = 84.834$ (Sig.<0.001), the standardized residuals are approximately normally distributed

*Sig.<0.05, **Sig.<0.01, and ***Sig.<0.001; low VIF values indicate low collinearity; all coefficients are standardized coefficients. X04: searching subject term "E-Commerce" in CNKI (periodical). X12: searching subject term "Business Intelligence Analytics" in CNKI (periodical). X64: searching subject term "Mobile Technology & E-Commerce" in CNKI (periodical). X68: searching subject term "Artificial Intelligence & Big Data & E-Commerce" in CNKI (periodical). X70: searching subject term "Quantum Computing" in CNKI (periodical). Y05: total retail sales for China. Y06: retail e-commerce sales for China. Y07: proportion of e-commerce in total retail for China.

Table 19: Regression models for China's theoretical research and data companies founded.

Dependent variable	Independent variable							Statistics
	X04	X09	X28	X60	X64	X66	X68	
Y14					0.418*** $t = 4.522$ Sig. = 0.000	0.596*** $t = 6.438$ Sig. = 0.000		$R^2 = 0.911$, both VIF = 2.413, $F = 128.417$ (sig.<0.001), the standardized residuals are approximately normally distributed
Y15	1.136*** $t = 12.353$ Sig. = 0.000			-0.480*** $t = -5.223$ Sig. = 0.000				$R^2 = 0.865$, both VIF = 1.566, $F = 80.108$ (Sig.<0.001), the standardized residuals are approximately normally distributed
Y16					0.530*** $t = 4.435$ Sig. = 0.000	0.452*** $t = 3.780$ Sig. = 0.001		$R^2 = 0.852$, both VIF = 2.413, $F = 71.915$ (Sig.<0.001), the standardized residuals are approximately normally distributed
Y17		-0.210* $t = -2.368$ Sig. = 0.026			0.351*** $t = 3.203$ Sig. = 0.004	0.755*** $t = 6.194$ Sig. = 0.000		$R^2 = 0.881$, VIF (X09) = 1.590, VIF (X64) = 2.413, VIF (X66) = 2.992, $F = 59.053$ (Sig.<0.001), the standardized residuals are approximately normally distributed
Y18			1.069*** $t = 12.345$ Sig. = 0.000		0.418*** $t = 6.796$ Sig. = 0.000		-0.469*** $t = -4.863$ Sig. = 0.000	$R^2 = 0.958$, VIF (X28) = 4.321, VIF (X64) = 2.180, VIF (X68) = 5.353, $F = 183.898$ (Sig.<0.001), the standardized residuals are approximately normally distributed

*Sig.<0.05, **Sig.<0.01, and ***Sig.<0.001; low VIF values indicate low collinearity; all coefficients are standardized coefficients; X04: searching subject term "E-Commerce" in CNKI (periodical); X09: searching subject term "E-Commerce" and then "Big data" in CNKI (periodical); X28: searching subject term "Machine Learning" and "Big Data" in CNKI (periodical); X60: searching subject term "Online Consumer Behavior & Big Data" classified by the field of "topic" in CNKI (periodical); X64: searching subject term "Mobile Technology & E-Commerce" in CNKI (periodical); X66: searching subject term "Cloud Computing & E-Commerce" in CNKI (periodical); X68: searching subject term "Artificial Intelligence & Big Data & E-Commerce" in CNKI (periodical); Y14: founded number of data companies in China (total); Y15: founded number of data companies in China classified as data/technology; Y16: founded number of data companies in China classified as business analytics; Y17: founded number of data companies in China classified as research/consulting; Y18: founded number of data companies in China classified as industrial application; Y18: founded number of data companies in China classified as research/consulting.

TABLE 20: Coefficients for Y06[a].

Model		Unstandardized coefficients		Standardized coefficients	t	Sig.	Correlation			Collinearity statistics	
		B	Std. error	Beta			Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	-89.898	208.579		-0.431	0.670					
	Y18	168.642	19.750	0.859	8.539	0.000	0.859	0.859	0.859	1.000	1.000
2	(Constant)	-440.918	243.502		-1.811	0.082					
	Y18	100.833	34.095	0.513	2.957	0.007	0.859	0.509	0.274	0.286	3.500
	X04	2.584	1.098	0.408	2.353	0.027	0.842	0.426	0.218	0.286	3.500

[a]Dependent variable: Y06, where model 1 R^2 is 0.737 and model 2 is 0.785; low VIF values indicate low collinearity; the standardized residuals are approximately normally distributed. X04: searching subject term "E-Commerce" in CNKI (periodical). Y18: founded number of data companies in China classified as research/consulting. Y06: retail e-commerce sales for China.
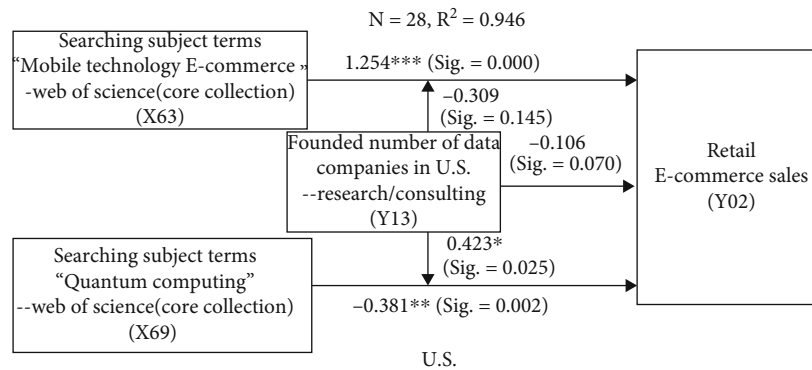
FIGURE 1: Moderating effect of data companies founded in the U.S.
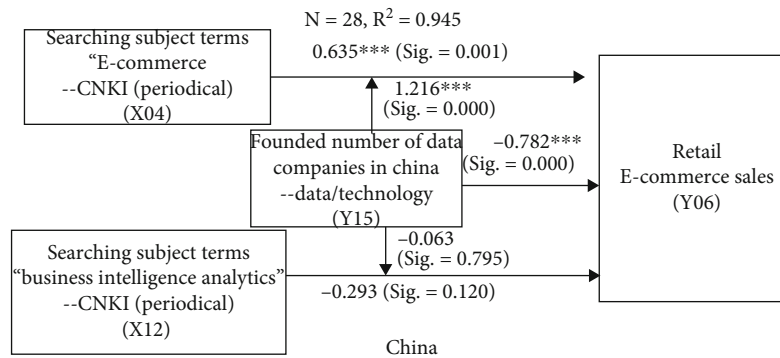
FIGURE 2: Moderating effect of data companies founded in China.

(Sig. = 0.000 < 0.05, R^2 = 0.710); the second step is running a linear regression of the independent variable X04 and dependent variables from one of Y14–Y18, which shows that all of the regression coefficients are significant, less than 0.05. The last step is building a linear regression of the independent variable X04 and adding one of Y14–Y18, for which the dependent variable is Y06, and it is found that only Y18 (Sig. = 0.007 < 0.05) and X04 (Sig. = 0.027 < 0.05) are simultaneously significant in this regression model (shown in Table 20). As a consequence, data companies in China play a mediating role in putting these fields of theoretical research into practice in e-commerce.

2.7. Comparison of the Moderating and Mediating Effects of Data Companies Founded between the U.S. and China. In Figures 1 and 2, we can see that both the U.S. and China show significance in moderating the relationship between theoretical research and practice in e-commerce using BDA, as tested by the moderate variable "founded number of data companies." In the U.S., the variable "founded number of data companies" has a moderating effect on the model of the correlation of "searching subject terms 'Quantum Computing'—WoS (Core Collection) (X69)" and "retail e-commerce sales (Y02)," such that this relationship is negative (X69: -0.381 (Sig. = 0.002 < 0.05); Y13: 0.423
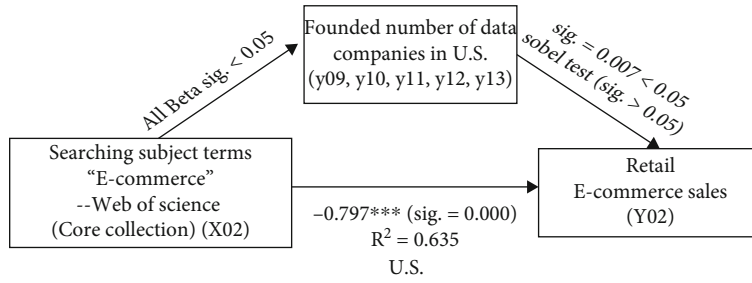
FIGURE 3: Mediating effect of data companies founded in the U.S.

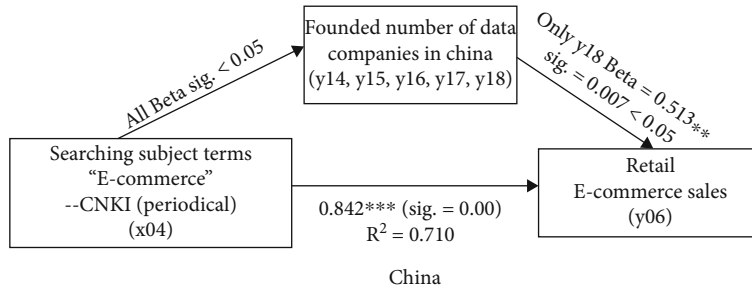

FIGURE 4: Mediating effect of data companies founded in China.

TABLE 21: Lag order selection criteria.

Variable	Lag	LogL	LR	FPE	AIC	SC	HQ	Lag order selection	Observed object
X02	0	-162.121	NA	27223.100	13.050	13.098	13.063		
	1	-139.237	42.107*	4728.674*	11.299*	11.396*	11.326*	○	
	2	-138.811	0.749	4955.199	11.345	11.491	11.385		
	3	-138.625	0.313	5297.007	11.410	11.605	11.464		U.S.
X69	0	-121.768	NA	1078.827	9.821	9.870	9.835		
	1	-108.126	25.101	392.517	8.810	8.908	8.837		
	2	-103.429	8.267*	292.255*	8.514*	8.661*	8.555*	○	
	3	-103.429	0.000	317.104	8.594	8.789	8.648		
X04	0	-175.910	NA	82037.320	14.153	14.202	14.166		
	1	-130.324	83.878*	2317.735*	10.586*	10.683*	10.613*	○	
	2	-130.303	0.036	2508.760	10.664	10.811	10.705		
	3	-130.278	0.043	2716.554	10.742	10.937	10.796		China
X70	0	-83.390	NA	50.069	6.751	6.800	6.765		
	1	-73.001	19.117*	23.630	6.000	6.098	6.027		
	2	-71.283	3.022	22.330*	5.943*	6.089*	5.983*	○	
	3	-70.835	0.753	23.376	5.987	6.182	6.041		

∗ indicates lag order selected by the criterion. LR: sequential modified LR test statistic (each test at 5% level). FPE: final prediction error; AIC: Akaike information criterion; SC: Schwarz information criterion; HQ: Hannan-Quinn information criterion. Sample: 1990-2017. X02: searching subject term "E-Commerce" in the WoS (Core Collection). X04: searching subject term "E-Commerce" in CNKI (periodical). X69: searching subject term "Quantum Computing" in the WoS (Core Collection). X70: searching subject term "Quantum Computing" in CNKI (periodical).

(Sig. = 0.025 < 0.05)), which means that it does not moderate the correlation of "searching subject terms 'Mobile Technology and E-Commerce'—WoS (Core Collection) (X63)" and "retail e-commerce sales (Y02)" (Y13: 0.309 (Sig. = 0.145 > 0.05)). This finding is similar to that in China, where the variable "founded number of data companies" has a moderating effect on the model of the correlation of "searching subject terms 'E-Commerce'—CNKI (periodical) (X04)" and "retail e-commerce sales (Y06)," such that if the relationship is positive (X04: 0.635 (Sig. = 0.001 < 0.05); Y15: 1.216 (Sig. = 0.000 < 0.05)), it has a direct significant negative correlation in relation to "retail e-commerce sales

TABLE 22: Comparison of linear regression models involving lag variables.

Dependent variable	Independent variable			
	X02	X02 (-1)	X69	X69 (-2)
Y02	0.616***	0.624***	2.410***	2.376***
	S.E. = 0.091	S.E. = 0.099	S.E. = 0.554	S.E. = 0.614
	t-statistic = 6.729	t-statistic = 6.282	t-statistic = 4.348	t-statistic = 3.870
	R-squared = 0.635	R-squared = 0.612	R-squared = 0.421	R-squared = 0.384
	F-statistic = 45.277	F-statistic = 39.462	F-statistic = 18.908	F-statistic = 14.981
	X04	X04 (-1)	X70	X70 (-2)
Y06	5.328***	5.868***	170.442***	150.349**
	S.E. = 0.669	S.E. = 0.798	S.E. = 35.220	S.E. = 53.494
	t-statistic = 7.969	t-statistic = 7.357	t-statistic = 4.925	t-statistic = 2.811
	R-squared = 0.710	R-squared = 0.684	R-squared = 0.483	R-squared = 0.248
	F-statistic = 63.507	F-statistic = 54.120	F-statistic = 24.251	F-statistic = 7.899

**Sig.<0.01; ***Sig.<0.001. X02: searching subject term "E-Commerce" in the WoS (Core Collection). X04: searching subject term "E-Commerce" in CNKI (periodical). X69: searching subject term "Quantum Computing" in the WoS (Core Collection). X70: searching subject term "Quantum Computing" in CNKI (periodical). Y02: retail e-commerce sales for U.S. Y06: retail e-commerce sales for China.

FIGURE 5: Trend of big data analytics in e-commerce. Website: https://www.amazon.com/Magformers-Magnetic-Building-Educational-Construction/dp/B06XJLGWST?ref_=Oct_DLandingS_PC_NA_NA#customerReviews.

(Y06)" (Y15: -0.782 (Sig. = 0.000 < 0.05)). However, it does not work in the correlation of "searching subject terms 'Business Intelligence Analytics'—CNKI (periodical) (X12)" (X12: -0.293 (Sig. = 0.120 > 0.05)) and "retail e-commerce sales (Y02)" (Y15: -0.063 (Sig. = 0.795 > 0.05)).

The model presented in Figure 3 assumes a three-variable system, which has a direct and significant relationship between "retail e-commerce sales (Y02)" and "searching subject terms 'E-Commerce'—WoS (Core Collection) (X02)," and the mediator variable "founded number of data companies in the U.S." is introduced to the model. However, this path between Y02 and X02 becomes nonsignificant because for "the number of existing data companies in the United States," it plays an important role in promoting the theoretical research of e-commerce in practice. However, data companies in China shown in Figure 4 are limited to

these types of "research/consulting" data companies and have a vivid mediating effect on the relationship between theoretical research works of BDA in e-commerce and practical producing activities in e-commerce.

2.8. Lag Consideration in Evaluating Theoretical Research Response to Practical Application in E-Commerce. In general, this theoretical research puts into practice needs a certain lag to accomplish the task. Here, a selection of the variables X02 and X69 for the U.S. and X04 and X70 for China is made to test the linear relation between their lag and the retail e-commerce sales. First, the goal is to determine the lag order among the independent variables selected from Table 21, which shows the six criteria [33], and the results of the lag order selection for X02 and X69 and X04 and X70 are a lag order of one and two and one and two, respectively. Next, we construct linear regression models involving the lag variables to determine whether the involved lag variables in the regression models fit better or not. The answer is certainly not. No matter whether the U.S. or China is investigated, R square degrades, as shown in Table 22. These findings indicate that we should not consider the effects of lag on evaluating the theoretical research response to practical applications in e-commerce, which was also demonstrated in previous sections as observed by the nonlag variables in the empirical studies.

3. Conclusions

The rapid growth of e-commerce has benefited not only the evolution of data science over the past two decades but also the boom of big data from various sources. This is what makes China and the United States the largest e-commerce markets and why China accounts for more than the United States in e-commerce sales. Ultimately, we can determine the reasons leading to the difference between the U.S. and China regarding this point. One of the reasons is the institutional differences and commercial value, which makes Chinese society's perception of BDA in e-commerce more acceptable than that of the United States. Another reason involves the theoretical research works on BDA in e-commerce in China, which have attracted slightly more extensive attention than that observed in the U.S. and involved a comparison of literature databases, indicating that proof of a significant relationship between theoretical research and practical activities in BDA in e-commerce could be attained. In addition, in the United States, with regard to the relationship of putting theoretical research into practice, the variables of the data company show moderate but no mediating effect. However, in China, the mediating effects of this relationship have been explained. These results help clarify doubts regarding the development of China e-commerce, which even exceeds that of the U.S. today, in view of the theoretical and practical comparison of BDA in e-commerce between them.

3.1. Avenues for Future Research and Practice. Regardless of whether the U.S. or China is considered, the theoretical research work is deeply impressing and has propelled practi-

cal application of BDA in e-commerce. However, big data hubris and algorithm dynamics issues may contribute to analysis mistakes [7] because of human subjective prejudice, technological objective limitations, and the need to enhance artificial intelligence by processing data more efficiently for e-commerce transactions. We expect that e-commerce activities concerning the seller, buyer, platform provider, etc., would entail self-learning actively through their own generated data, and then, extraction by others (such as the buyer and platform provider) of the critical information would serve as a combination of commodities for improving the quality of sales and service, particularly for increasing the transparency and credibility of goods to attract purchases. For example, currently, a product from Amazon online is mostly displayed with its price and functions; however, it is anticipated that product information, manufacturer information, seller information, customer information, and even extra payment information or more will be shown in the future, as shown in Figure 5, such that the desire to buy and recommend the right goods for purchasers can be reinforced. Therefore, there are three future research orientations in e-commerce using BDA. First, data originating from e-commerce activities will be considered valuable and tradable resources after being processed by BDA, and either the seller, buyer, or platform provider can enact pricing dynamically with his or her data for sale. In addition, the data trading market and its pricing mechanisms in e-commerce will be researched and widely put into use. Second, a new rule for dynamic pricing for each customer developed by applying BDA in e-commerce can be envisioned, such that a product would sell for a different price on a per-customer basis, enlisting every seller, buyer, and platform provider to accomplish each expectation or revenue maximization. Third, puzzling relationships among purchasing behaviors and consumer habits [34], consumer habits and personalities [35, 36], consumer personalities and the growth environment [37, 38] can be unraveled by using BDA for deep learning in e-commerce trading.

In addition, it is expected that the mixture of large data resources and new technologies will challenge many existing e-commerce problems and find out a better solution. A series of new issues should be focused on, such as quantum computing in e-commerce [39], in which theoretical research works serve to observably promote retail sales, both in the U.S. and China (as seen in Tables 9 and 17). Ronald [40] considers the potential impact that the nascent technology of quantum computing may have on e-commerce, more specifically, designing "encrypt" information in such a way to ensure that an e-commerce trade is safe, offering significant speed-ups for faster search and optimization in the big data age, and implementing the quantum cheque transaction in a quantum-networked banking system [41]. As a BDA concept, quantum machine learning could enable machine learning that is faster than that of classical computers for calculating and analyzing e-commerce activities in the big data age [42]. In short, quantum computing in e-commerce is a crucial theoretical research topic and has practical application both for the U.S. and China at present and in the future.

In future applications, we should encourage data companies to devote efforts to big data business issues required for e-commerce because data companies play a moderation or mediation role in putting theoretical research into practice in e-commerce.

Conflicts of Interest

The author declares that they have no conflicts of interest.

Acknowledgments

References

[1] C. L. Philip Chen and C. Y. Zhang, "Data-intensive applications, challenges, techniques and technologies: a survey on big data," *Information Sciences*, vol. 275, pp. 314–347, 2014.

[2] S. Sakr, A. Liu, D. M. Batista, and M. Alomari, "A survey of large scale data management approaches in cloud environments," *IEEE Communications Surveys & Tutorials*, vol. 13, no. 3, pp. 311–336, 2011.

[3] W. A. Günther, M. H. Rezazade Mehrizi, M. Huysman, and F. Feldberg, "Debating big data: a literature review on realizing value from big data," *Journal of Strategic Information Systems*, vol. 26, no. 3, pp. 191–209, 2017.

[4] Z. Wang and Q. Yu, "Privacy trust crisis of personal data in China in the era of big data: the survey and countermeasures," *Computer Law & Security Review*, vol. 31, no. 6, pp. 782–792, 2015.

[5] P. Ohm, "The underwhelming benefits of big data," *University of Pennsylvania Law Review Online*, vol. 161, no. 1, pp. 339–346, 2013.

[6] S. Marcel, "Digital pharmacovigilance and disease surveillance: combining traditional and big-data systems for better public health," *The Journal of Infectious Diseases*, vol. 214, no. 4, pp. 399–403, 2016.

[7] D. Lazer, R. Kennedy, G. King, and A. Vespignani, "The parable of Google flu: traps in big data analysis," *Science*, vol. 343, no. 6176, pp. 1203–1205, 2014.

[8] R. V. Hal, "Big data: new tricks for econometrics," *Journal of Economic Perspectives*, vol. 28, no. 2, pp. 3–28, 2014.

[9] O. Ahmed, B. Fatima-Zahra, A. L. Ayoub, and B. Samir, "Big data technologies: a survey," *Journal of King Saud University—Computer and Information Sciences*, vol. 30, no. 4, pp. 431–448, 2018.

[10] A. Deepak, K. Niraj, and P. K. John, "Understanding big data analytics capabilities in supply chain management: unravelling the issues, challenges and implications for practice," *Transportation Research Part E: Logistics and Transportation Review*, vol. 114, pp. 416–436, 2017.

[11] S. Akter and S. F. Wamba, "Big data analytics in e-commerce: a systematic review and agenda for future research," *Electronic Markets*, vol. 26, no. 2, pp. 173–194, 2016.

[12] J. Manyika, M. Chui, B. Brown et al., *Big Data: The Next Frontier for Innovation, Competition, and Productivity*, McKinsey Global Institute, 2011, June 2018, http://www.mckinsey.com/Insights/MGI/Research/Technology_and_Innovation/Big_data_The_next_frontier_for_innovation.

[13] A. Taylor, *The 17 Biggest Data Breaches of the 21st Century*, CSO, 2018, July 2018, https://www.csoonline.com/article/2130877/data-breach/the-biggest-data-breaches-of-the-21st-century.html.

[14] Q. Bin, S. J. Chen, and S. Q. Sun, "Cultural differences in e-commerce," *Journal of Global Information Management*, vol. 11, no. 2, pp. 48–55, 2003.

[15] M. B. Schmidt, A. C. Johnston, K. P. Arnett, J. Q. Chen, and S. Li, "A cross-cultural comparison of U.S. and Chinese computer security awareness," *Journal of Global Information Management*, vol. 16, no. 2, pp. 91–103, 2008.

[16] H. Dai and P. C. Palvi, "Mobile commerce adoption in China and the United States," *ACM SIGMIS Database: the DATABASE for Advances in Information Systems*, vol. 40, no. 4, pp. 43–61, 2009.

[17] J. C. Nunnally and I. H. Bernstein, *Psychometric Theory (McGraw-Hill Series in Psychology)*, McGraw-Hill, New York, 1994.

[18] Y. H. Chan, "Biostatistics 102: quantitative data-parametric & non-parametric tests," *Singapore Medical Journal*, vol. 44, no. 8, pp. 391–396, 2003.

[19] A. Ghasemi and S. Zahediasl, "Normality tests for statistical analysis: a guide for non-statisticians," *International Journal of Endocrinology and Metabolism*, vol. 10, no. 2, pp. 486–489, 2012.

[20] D. G. Goring and V. I. Nikora, "Despiking acoustic doppler velocimeter data," *Journal of Hydraulic Engineering*, vol. 128, no. 1, pp. 117–126, 2002.

[21] R. L. Brown, J. Durbin, and J. M. Evans, "Techniques for testing the constancy of regression relationships over time," *Journal of the Royal Statistical Society Series B (Methodological)*, vol. 37, no. 2, pp. 149–163, 1975.

[22] B. M. Henry, "Nonparametric tests against trend," *Econometrica*, vol. 13, no. 3, pp. 245–259, 1945.

[23] A. N. Pettitt, "A non-parametric approach to the change-point problem," *Journal of the Royal Statistical Society Series C (Applied Statistics)*, vol. 28, no. 2, pp. 126–135, 1979.

[24] S. Jomnonkwao, S. Uttra, and V. Ratanavaraha, "Forecasting road traffic deaths in Thailand: applications of time-series, curve estimation, multiple linear regression, and path analysis models," *Sustainability*, vol. 12, no. 1, p. 395, 2020.

[25] R. B. Bendel and A. A. Afifi, "Comparison of stopping rules in forward "stepwise" regression," *Journal of the American Statistical Association*, vol. 72, no. 357, pp. 46–53, 1977.

[26] L. Wilkinson, "Tests of significance in stepwise regression," *Psychological Bulletin*, vol. 86, no. 1, pp. 168–174, 1979.

[27] C. Kontoghiorghes, "Reconceptualizing the learning transfer conceptual framework: empirical validation of a new systemic model," *International Journal of Training and Development*, vol. 8, no. 3, pp. 210–221, 2004.

[28] P. T. Pope and J. T. Webster, "The use of an F-statistic in stepwise regression procedures," *Technometrics*, vol. 14, no. 2, pp. 327–340, 1972.

[29] D. A. Belsley, E. Kuh, and R. E. Welsch, *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*, John Wiley & Sons, New York, 2005.

[30] C. M. Jarque and A. K. Bera, "A test for normality of observations and regression residuals," *International Statistical Review/Revue Internationale de Statistique*, vol. 55, no. 2, pp. 163–172, 1987.

[31] A. J. Fairchild and D. P. MacKinnon, "A general model for testing mediation and moderation effects," *Prevention Science*, vol. 10, no. 2, pp. 87–99, 2009.

[32] J. R. Edwards and L. S. Lambert, "Methods for integrating moderation and mediation: a general analytical framework using moderated path analysis," *Psychological Methods*, vol. 12, no. 1, pp. 1–22, 2007.

[33] V. Ivanov and L. Kilian, "A practitioner's guide to lag order selection for VAR impulse response analysis," *Studies in Nonlinear Dynamics & Econometrics*, vol. 9, no. 1, 2005.

[34] M. F. Ji and W. Wood, "Purchase and consumption habits: not necessarily what you intend," *Journal of Consumer Psychology*, vol. 17, no. 4, pp. 261–276, 2007.

[35] W. Wood, "Habit in personality and social psychology," *Personality and Social Psychology Review*, vol. 21, no. 4, pp. 389–403, 2017.

[36] H. K. Harold, "Personality and consumer behavior: a review," *Journal of Marketing Research*, vol. 8, no. 4, pp. 409–418, 1971.

[37] D. Scott and F. K. Willits, "Environmental attitudes and behavior: a Pennsylvania survey," *Environment and Behavior*, vol. 26, no. 2, pp. 239–260, 1994.

[38] S. Bhate, "One world, one environment, one vision: are we close to achieving this? An exploratory study of consumer environmental behaviour across three countries," *Journal of Consumer Behaviour: An International Research Review*, vol. 2, no. 2, pp. 169–184, 2002.

[39] K. Thapliyal and A. Pathak, "Quantum e-commerce: a comparative study of possible protocols for online shopping and other tasks related to e-commerce," 2018, July 2018, https://arxiv.org/pdf/1807.08199.pdf.

[40] D. W. Ronald, "The potential impact of quantum computers on society," *Ethics and Information Technology*, vol. 19, no. 4, pp. 271–276, 2017.

[41] B. K. Behera, A. Banerjee, and P. K. Panigrahi, "Experimental realization of quantum cheque using a five-qubit quantum computer," *Quantum Information Processing*, vol. 16, no. 12, p. 312, 2017.

[42] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, "Quantum machine learning," *Nature*, vol. 549, no. 7671, pp. 195–202, 2017.