
The New Civilization
Upon Data
大數據如何重塑
人類文明、商業形態
和個人世界
數文明

涂子沛◎著



香港中和出版有限公司
www.hkopenpage.com

我們可以改變世界，是因為首先改變了自己，國家亦然。

CONTENTS

目錄

推薦序一 / 林毅夫	XII
------------------	-----

推薦序二 / 孟慶國	XIV
------------------	-----

前言 從幼稚到成熟：我們這個時代的數據革命	1
-----------------------------	---

01 數據平權：新商業文明的衝擊和原罪

心理入侵：大數據“讀心術”誕生了	15
------------------------	----

價格操控：大數據“殺熟”和算法合謀	22
-------------------------	----

數懼的最深層：人工智能三宗罪	27
----------------------	----

新經濟到底新在哪裡：智能商業	33
----------------------	----

金礦如何形成：個人數據的價值困境	36
------------------------	----

數權：互聯網原罪浮出水面	41
--------------------	----

02 無匿名追蹤：天網的隱喻

要是此案在中國，早破了	49
-------------------	----

三體：天網的真正維度	53
------------------	----

互聯網要向天網學習甚麼	57
-------------------	----

霧計算：人工智能的競技主場	61
---------------------	----

以圖搜車：追蹤億萬車輛之軌跡	63
----------------------	----

硬盤和眼藥水為甚麼同時脫銷	69
---------------------	----

03 人臉和人工智能

照片開路：構建身份社會	79
起步矽谷：幾何時代的徘徊	81
機器能否學習：人工智能之爭	84
“不明覺厲”：深度學習的崛起	87
數據田徑場：政府怎樣推動商業創新	97
歷史的意外：“9·11”事件如何拯救失敗	100
無匿名社會：動態識別的前景	103

04 高清晰社會：單粒度治理和籠中人險境

數紋：邁入高清晰社會	111
模糊社會的困境	116
超級檔案的產生：單粒度治理	119
抑制僥倖：中國古代的治國經驗	126
高能個體：人人皆持劍，又皆為劍下人	129
數據即證據：無僥倖天下	131

05 數力：普適記錄如何“掰彎”人性

唐宗宋祖的煩惱	141
愛迪生拓寬記錄的疆域	147
尼克遜之困：白宮錄音小史	150
特朗普的錄音風波	156
視頻直播為甚麼低效	158
普適記錄：上帝的終極武器	160

06 數文明：社會、商業和個人如何被記錄賦能

一個新的發展視角：記錄	171
歷史上中國文明領跑與掉隊的原因	175
三本書在三個大陸的三種命運	180
十計九記：商業文明的進步密碼	185
電梯裡的羞辱	189
全面記錄全面計算：開創數文明時代	192

07 數據新政：建設現代國家的治理體系

數基：世界級創新和本土難題	202
數聯網：數據維度上的整體性政府	208
城市大腦：新時代的南京長江大橋	214
僅此一次：“最多跑一次”如何升級	219
開放攝像頭：天網的未來	222
對數據和算法的現代治理	229

08 人工智能的邊界、風險和未來

一切皆可預測：拉普拉斯之妖	237
量子思維：不確定的現實	240
人是城市中的粒子：測不準	241
數據相對論：普適記錄的極限	245
表情分析：人工智能的邊界	249
轉型智能社會：懷揣瓷器花瓶，進入花花世界	258

結語 第四次浪潮：我們如何再次領先	265
後記 野路無人自還	273
大事記	277
索引	285

推薦序 FOREWORD

數據自古有之，在互聯網出現、普及之後，因為數碼化而記錄、積累成為可供計算機快速提取、分析的大數據。近幾年來，它被廣泛地運用於人類社會的生產、生活、管理和社會治理，成為並列於資本、勞動和自然資源的新的要素稟賦。

這一新稟賦的出現對世界政治、經濟、文化的影響將不亞於 15 世紀末美洲新大陸的發現，值得社會各界認真研究和關注。

我國雖然起步晚於美國等發達國家幾年，但是由於人口規模、經濟體量和快速的發展，在大數據這一稟賦及其開發運用上相比於世界上其他國家具有比較優勢，並且已經湧現了百度、阿里巴巴、騰訊、今日頭條等世界前沿的領先企業和眾多的獨角獸公司，國家也正式提出了“國家大數據戰略”，先後出台了《促進大數據發展行動綱要》《大數據產業發展規劃（2016—2020 年）》等指導性文件，大數據的開發、運用正處於方興未艾的階段。

這本《數文明》為中國第一位系統研究大數據的權威學者涂子沛著的第三本力作，分享了他對大數據給人類、社會和文明的各個層面可能帶來的機遇和挑戰的探索和思考，縱橫歷史，文字雋永，深入淺出，信息量大，涉及面廣，讀之獲益甚多，難以釋卷。

作者預言了數據經濟時代的到來，其觀點獨到，具有前瞻性。毫無疑問，大數據將成為全球經濟的重要組成部分，其對中國經濟的賦能作用也將日益凸顯，但作者也提出，它也可能給經濟社會投下陰影。如何發揮大數據的正向作用、限制其負面影響，我希望看到更多的研究和討論。

林毅夫

北京大學新結構經濟學研究院院長
世界銀行原高級副行長、首席經濟學家

推薦序 FOREWORD

從人類的文明發展史看，以大數據和人工智能為驅動力的技術革命正在引領人類走向新文明時代。而新時代的人工智能，又不能脫離數據，數據可謂人工智能的母體。

因為新技術的普及，引發了新觀念的產生、新模式的出現、新思維的塑造、新治理的構建，它們既成為進入新文明時代的重要標誌，又構成了需要深入探討的重大命題，這些命題日益引起社會各界空前的關注和思考，而其中的主線，就是大數據。

《數文明》是一本應時力作，這本書提出了很多富有啟發性的概念，從量數到據數，從普適記錄到高清社會，從數權到數治，從數基到數聯網等等，體現出了作者對數據文明的深邃思考和敏銳洞察。書中的許多觀點和內容極具思想性，引用的案例生動鮮活，可讀可鑑性很強，閱讀時讓人幾乎不願跳過任何細節。作者縱橫古今，關聯西東的數據時空觀和行文風格，時時讓人掩卷深思，充滿對未來新文明時代的無限遐想。

這是涂子沛先生第三本關於大數據的著作，他的前兩本書，《大數據》《數據之巔》我也很喜歡。據我所知，涂先生的前兩本書都是在國外完成的，但本書的立足點、出發點、關注點很明顯都在中國，它講的多是中國故事和

案例，很多時候直溯中國歷史，從中挖掘現代元素，同時又兼具世界眼光。
我相信，這也是中國夢的另一種表達和講述。

孟慶國

清華大學公共管理學院教授

清華大學國家治理研究院執行院長

前言

從幼稚到成熟： 我們這個時代的數據革命

一個新的數據劃分方法

隨着大數據的興起，“言必稱數據”逐漸成了我們這個時代一個顯著的特徵，但問題是，此數據非彼數據。

今天，所有“記錄”的結果，甚至包括文字，都被統稱為數據。這其中暗含的邏輯是，數據作為一個概念，它的內涵擴大了。傳統意義上的數據是人類對事物進行測量的結果，是作為“量”而存在的數據，可以稱為“量數”；今天的照片、視頻、音頻不是源於測量而是源於對周圍環境的記錄，是作為一種證據、根據而存在的，可以稱為“據數”。

量數和據數，兩者原本風馬牛不相及，但在今天這個時代，它們又有了一個共同

的特點——以“比特”為單位進行存儲。關於大數據的定義，我主張用這樣一個等式較為簡潔、精確地表示：

$$\text{大數據} = \text{傳統的量數} + \text{現代的據數}$$

↑
↑

（量數源於測量，如氣溫 28℃）
（據數源於記錄，如一張照片）

量數雖然比據數更接近“數”，但從歷史上看，據數的出現要早於量數。人類早期對自身活動的記錄，即“史”，就是早期的據數，也可以說，據數是歷史的影子。量數則是在記錄的實踐中慢慢產生的，特別是針對天空、星體、山川等外物的記錄，它們追求精確，於是我們逐漸延伸出測量的工具和行為。一切科學都源於測量，量數是否充沛，決定了科學的種子何時萌芽，決定了科學是否發達，也可以說，量數是科學的母親，其核心要義是精確。

在 16 世紀前後，人類開啟了大航海時代，量數出現了一個高峰。隨着航海儀器的普及，歐洲對土地測量、建築設計、礦山開採、人口統計的需求也應運而生。人類發現，定性描述不足以解決問題，只有更加精確的測量和計算，才能夠滿足科學和管理的需要，這引發了歷史上第一次數據爆炸：量數爆炸。

這也是人類歷史上第一個數據的春天。

進入 20 世紀後，由於計算機、互聯網和智能手機的普及，據數開始爆炸，相較於 16 世紀的量數爆炸，據數爆炸的規模更大。今天大數據的主體，從體量上來看，毫無疑問是據數，即對人類活動及周邊環境越來越多的記錄，或稱“普適記錄”。互聯網平台記錄的，大部分是據數。

和今天的普適記錄相比，過去 5000 年的文明看似浩如煙海，但其實相當有限，史書雖然一本比一本厚，但大部分都聚焦在為數不多的帝王將相身上，關於普通百姓的個體性記錄，在全世界各個國家都少之又少。隨着普適記錄的興起，這種情況正在迅速改變，未來不僅有國家史、社會史、行業史，還會有數量驚人的“個人史”。可供歷史學家研究的資料，將會像雪球一樣越滾越大，其規模將前所未有。

量數對於中國的巨大意義，自不待言，黃仁宇的“數目字管理”在 20 世紀曾經開啟過一場討論，我的前兩本書《大數據》《數據之巔》¹ 重點關注的是量數，尤其是《數據之巔》，它延續了黃先生的討論，而本書更側重據數。

一場還沒有完成的革命

今天的“大數據之熱”，熱的其實是據數，和精確的量數幾乎沒有任何關係。

“據數之熱”並不奇怪，它源於互聯網帶來的變革和它已經展現出來的巨大的財富效應，形形色色的手機應用給了每一個人更加直觀的感受，不管是電商、社交還是自媒體，哪一個不是跟記錄之據數有關呢？

幾乎人類的一切行為，今天都可能、可以被記錄，並被轉化為據數。如果說量數的核心要義是“精確”，那據數的關鍵就在於“清晰”地留據。

因為智能手機的普及，據數已經無處不在，但我們大部分人對據數的理解是膚淺的，迄今為止，以據數為中心的大數據討論主要停留在以下三個層面。

一是精準營銷，即互聯網廣告業。和傳統的廣告業相比，今天的互聯網和智能手機通過記錄消費者不斷產生的數據，可以向終端用戶推送個性化的廣告，這大大提高了行業效率。這是大數據革命在商業領域的起源。這也是目前互聯網企業，無論是中國的 BAT（百度、阿里巴巴、騰訊），還是美國的谷歌、Facebook（臉譜網）、亞馬遜最主要的贏利途徑。

二是商業和社會信用，其主體是金融機構。除了精準營銷，這是利用大數據賺錢的第二個法門，也是我們看到諸多互聯網企業陸續進入金融領域的原因。其商業模式是，通過消費者的消費記錄評估消費者的信用，從後續的金融服務中贏利。例如阿里巴巴旗下的“芝麻信用”和騰訊旗下的“微粒貸”，它們在給消費者打“信用分”的基礎上，向單個消費者提供貸款等金融服務。

這兩種商業模式，都需要通過數據監控消費者在互聯網上的一舉一動，消費者個體因此成為被觀察、被分析、被監測的對象，這就帶出了第三個層面的問題——隱

私。這兩種商業模式的副作用是，我們幾乎每天都能聽到數據泄露的新聞，時不時還會看到因為它而導致的悲劇。

最近，我的一個朋友告訴我，今日頭條向他推送泳衣和泳鏡的廣告，明顯是知道他最近去游泳了，他琢磨了一下，發現唯一可能的原因，是他在下水之前把智能手錶調節到了泳池模式。

毫無疑問，今日頭條打通了這款智能手錶的數據。大眾對這種商業模式有一種矛盾的心理：一方面，我們感到權利受到了侵犯，沒有得到尊重，被出賣了，社會輿論也往往會把它簡單地放在公平的背景下進行考量，一邊倒地反對這種商業模式；另一方面，也許我們真的需要一副泳鏡，不想拒絕這種便利和高效。

這兩種商業模式之所以可行，是因為它們以據數為基礎掌握了消費者的動態，據數的商業和管理價值正是本書論述的重點，但為了行文方便，本書大部分時候仍使用了“數據”一詞，因此本書的“數據”二字實為指代“據數”一義。

這就是大眾眼中的大數據，前兩者為商業利潤而生，而隱私問題屢屢成為社會公共話題。可是，這三個層面僅僅揭開了冰山的一角，難道大數據就是養了幾家大公司，方便了公眾生活，改善了信用評級體系，讓生活更美好或者多了點麻煩這麼簡單？

我認為遠不止於此。

在商業層面，大數據還在進一步深化變革，它所催生的互聯網應用仍在不斷更新迭代。當大數據完全發揮出它的潛力時，其最終形態將是全自動商業，或稱智能商業，商業文明將會被重塑和再造，本書將對此進行闡述。

據數的商業化應用帶動了大數據的興起，但商業只是故事的一個邊角，革命是社會化的，未來我們還會看到智能製造業，它所依靠的還是數據，數據引發的變化還在向社會治理和個人生活領域全面拓進，它涉及社會生活的方方面面，將會推動整個社會進入文明新狀態，改變社會的全貌。一個新的故事正在世界範圍內浮現。我們必須拋開細枝末節，看到那些更深刻的、方向性的東西。數據的力量正在重塑整個社會甚至人類的天性。

我稱之為“數文明”，這是本書的主題。

一種新文明的興起

何謂文明？文明是歷史沉澱下來的，被絕大多數人認可和接受的發明創造、人文精神以及公序良俗的總和。這些集合至少包括了以下要素：語言、文字、工具、道德、信仰、宗教、法律、家族、城邦和國家。

今天的文字，也屬於數據的一種，換句話說，今天的數據，包括了文字，還超越了文字，文字只是數據的一個子集，如果說文字是金子，那數據就是金屬。

就像語言文字曾經塑造了人類的文明一樣，數據這個新工具正在重塑人類的精神、道德甚至宗教信仰，本書將對此展開論述。

人類畢生的追求，莫過於“明”。孔子說：“朝聞道，夕死可矣。”說的是，人早上明白了一個道理，如果晚上死了，那也是值得的，而不是說，早上過上了幸福的生活，那晚上死，也是值得的。明智、明德、明理，做個明白人，追求光明、知識和規律，這應該是人畢生的目標。光明，指的是普照萬物的自然之光，它是地球上萬物存在和生長的前提。在生存的基礎之上，我們把和野蠻相對的、較為先進的狀態稱為文明，顧名思義，這種“明”，應歸功於文字的發明。

而今天，在光明和文明之外，數據帶來了一種新的“明”。因為數據，人類歷史上一些精細的、微妙的、隱性的，甚至曾經難以捕捉表述的關係和知識，在今天都可以變為顯性的關係和知識，清清楚楚地為人類所用；因為數據，人類從來沒有像今天一樣清晰、明白、客觀、精確地認知和管理自己所生活的社會；因為數據，大量的事實可以被還原再現，人類的僥倖心理得到了抑制，人性的幽暗之處得以變得光明，人類正在邁向一個更加文明、安全的時代。

西方的《聖經》裡記錄了一個傳說，從前，天下人都講一樣的語言，都有一樣的口音，人和人之間很容易溝通和交流，於是一群人说：“來吧！我們要建造一座城和一座塔，塔頂通天，為要傳揚我們的名，免得我們分散在全地上。”人類於是聯合起來，開始興建一座通往天堂的高塔，塔越建越高，上帝為此感到不安：“看哪！他們成為一樣的人民，都是一樣的言語，如今既做起這事來，以後他們所要做的事，就沒有不成就的了。”

為了阻止人類的計劃，上帝就讓人類說不同的語言，人類相互之間自此不能有效溝通，通天塔因此失敗，人類各散東西。

今天的數據，無論是量數還是據數，都由比特“0”和“1”組成，它是全世界統一的語言。按照《聖經》的邏輯，建造一座通天塔在今天已經成為可能。

在國家和社會的層面之上，我們將會看到更大的變化。越來越多的個人行為在被記錄，對國家而言，這意味着每一個國民個體、每一輛車甚至每一個其他物體都可以被追蹤。如果懂得使用數據，那麼站在官僚層級的金字塔上，我們的社會將呈現出一種現在就非常清晰而且會越來越清晰的狀態，據數就是這個高清社會的紋理。本書認為，清晰性是有效治理的前提。

本書堅持、發展了黃仁宇的“數目字管理”。我認為，中國近百年來的落後，是源於眾多國民對精確的漠視，在中國的歷史上量數一度匱乏。

和量數相比，據數為國家治理提供了新的工具和手段。就此而言，用好大數據是通向國家治理現代化的最佳路徑。我甚至認為，若論中國，我們的國家治理現代化，捨此途徑，無從抵達。

在中國的歷史上，曾經長期存在德治和法治之爭，德治以仁愛為主，法治倡導峻法，爭了上千年，今天中國要建設法治社會已經是共識，本書要提出的，是數治。數治就是憑藉對數據的有效收集、處理和分析來治理國家以及與之對應的數據治理，數據治理將是國家治理的重要內容。

對國家而言，以大數據為結果的互聯網代表的是新生的力量，它正在重塑傳統社會，比如智能攝像頭被廣泛應用，大量的事實可以被還原再現，人類的僥倖心理得到了抑制，人類的犯罪行為將會大幅減少，數治正在有效地解決人類對安全的根本性需要。

另一方面，大數據、互聯網又帶來挑戰。國家力量已經不可能完全左右互聯網連接起來的有機社會了，相反，越來越多的公共功能在向私人公司、社會機構轉移。想了解商家的信譽，不妨去看看好評差評；想給災區捐點錢，人們第一時間要找的是網絡平台，而不是傳統的慈善機構；政府開發的服務平台點擊率可能不會太高，而同樣

的功能一旦被集成到互聯網公司的平台上，就可能帶來一場公共服務的變革。又比如媒體，現在最大的媒體不是電視台，也不是各大報紙，而是今日頭條。憑藉巨大的數據和流量，商業互聯網公司可以挑戰政府的公權，在這種複雜的局面下，如何重塑公權力的公信力？與互聯網公司緊密合作已經成為一個繞不開的話題。

如何利用這股力量，又約束住這股力量？這股力量可能會重塑整個社會的結構，催生新的政治文明，數治因此是一個具有挑戰性的話題。數治做好了，中國在這個新的時代就會具備“數據優勢”，國家是這樣，企業亦然，各級地方政府也一樣。

一條可靠的成功路徑

數文明不僅和國家、社會相關，也直接關係到個人。

對個人來說，掌握未來發展、演進的方向，知道哪些行業將消失，哪些行業又將興起，這當然非常重要。一個在黑夜中行走的人是走不快、走不遠的，他也無法領略到沿途的風景之美。

我認為，在數文明的時代，通過記錄賦能，個人會成為高能個體，一個具備數據意識、數據頭腦和數據技能的數據公民當然更容易獲得成功。新的時代會改變個人的命運，我們也應該調整我們的價值觀。

記得小時候聽故事，開頭一般都是“很久很久以前，在山的那邊，一個遙遠的地方……”，一句“山的那邊”就將我們的心靈拉到了遠方，山那邊的那個世界總是能讓少年心馳神往。可說到底，這個講故事的手法是通過拉開時間和空間的距離，創造一個可以隨意講述的空間。因為你沒去過、沒見過、沒經歷過，你不懂，講故事的人就有了權威。在古代，歲數大的人、遠航過的水手都被認為是最有智慧的人，因為他們在時間維度上看到了更多，在空間維度上經歷了更多。

也可以概括地說，那是一個未知的世界，未知世界的特點就是“沒有記錄”。

今天，一切都變了。你不了解也無法了解的“山的那邊”已然不存在了，你足不出戶，谷歌地圖和街景就可以把你帶到任何一個地方，每條街道、每幢樓、每扇門都

看得清清楚楚。人們隨時隨地都在記錄，後人可以跨越時間和空間找到今天我們的真相。

唐代詩人元稹是有名的才子。中國有一句千古傳誦的愛情佳句就是他的手筆——“曾經滄海難為水，除卻巫山不是雲”。這句詩是元稹為悼念亡妻韋叢而作的，情深意切，令人動容。可是真實的元稹並非專情之人，元稹娶韋叢之前曾與崔鶯鶯有私情，但為了攀附，棄崔鶯鶯而迎娶了高官之女韋叢。這段歷史，元稹寫進了自己的作品《鶯鶯傳》，即《西廂記》的前身。

元稹曾多次出使地方，每到一處，都留下風流韻事。例如當時的江南女詩人劉采春、蜀中女詩人薛濤都為其所累。薛濤在和元稹勞燕分飛之後，只能書信往來，由此產生了文學史上久負盛名的“薛濤箋”。對方一往情深，元稹卻不斷移情別戀，其紅顏知己皆不得善終，元稹可謂始亂終棄，他的真實形象與其詩文中所呈現的形象大相徑庭。

人無完人，歷史上的每一個王侯將相、名人聖人大抵如此。只是因為記錄的缺失，我們不知道罷了，歷史留下的總是最美的剪影，而正因為記錄的缺乏，後世通過想像和“腦補”又將其神化了。關羽成為武聖，很大程度上要拜各種傳說所賜；王羲之沒有真跡留下來，後人反而有足夠的空間去演繹、神化他的書法。回頭審視，歷史上的造神運動，也都是缺乏真實記錄造成的，當然，其根本原因是當時缺乏記錄的工具和手段。今天，情況已完全改觀，一部手機在手，所有人都可以錄音、拍照、錄像、做筆記，而且可以快速傳播出去，這就是本書定義的普適記錄。

在普適記錄的時代，要神化一個人越來越難。別說是風流才子元稹，哪怕是一個聖人，如果他的身後有一台攝像機跟着，24小時如實地記錄他的行為，我們很快就會發現，外表再光鮮的英雄、聖人、領袖也有陰暗的一面，甚至如小丑一般可笑。從這個角度觀察，聖人與普通人並無太大的差異。

馬雲創業之初的艱難和落魄，現在大家都有所了解。如果沒有留下影像資料，人們很難想像馬雲當初的窘境，眼裡只會看到他現在的光環，把他神化成歷史上所有的傑出人物那樣。正是那些能錄音、錄像的新的記錄工具，幫助我們留住了真實。

記錄可以把偉人還原成普通人，抹去英雄與平民的差別。這不是把歷史虛無化，而恰恰是真實的歷史。中國人有崇拜先祖的傳統，對先祖文明的推崇確保了我們文化的綿延不絕，有其積極意義，但是在數據時代，我們更要看到另一層意義：偉人也是普通人，凡人也能走向成功，不必高山仰止，妄自菲薄。

記錄可以祛魅。祛魅，意味着打破不可知的神秘，凡人也可成功，人人皆有可能成功。

既然人人都可能成功，那在數據時代，一個人究竟如何才能邁向成功？

本書分析了社會和商業的文明史，找出了文明發展的“金線”，而且我認為，這條文明發展的“金線”也同樣適用於個人，可以幫助個人獲得職業上和專業上的成功。

我在本書中回憶了自己剛參加工作時的一段青澀困頓的時光。1996年我在部隊服役，因為不擅長寫材料，屢屢受到頂頭上司的嘲諷和鄙視。這刺激了一個“理工男”學習寫作，我最終領悟到一個快速有效的笨辦法，通過摘錄、分類、不斷溫誦歸納，在很短的時間內，我寫材料的水平大有長進。

我的這段親身經歷證明，人與人之間的差別遠沒有想像中的那麼大。大家都是凡人，人人都有七情六慾，個個都有懵懂無知的少年時。苦心人，天不負，只要肯努力，很多事情，你也能做到。“王侯將相寧有種乎”，這句話你也一樣能勇敢地喊出來。

通往個人專業成功的有效路徑就是記錄。就記錄而言，人腦不如電腦，因為人腦是微分機制，而電腦是積分機制。有效的學習，更需要積分機制。善用記錄和數據，我們就能在成功的道路上獲得能量“加持”。

個人的成功和一個民族的文明自有相通之處，這個相通之處就是記錄。因為普通記錄，個人更容易獲得成功，而推動文明發展的最終動力則是無數國民源源不斷的創新。正因如此，數文明的能量和潛力有可能超越歷史上所有的文明。在我看來，數據正在改變所有那些組成文明的要素，就像支付寶改變了傳統的銀行業，微信改變了傳統的通信行業一樣，數文明在更優越的模式基礎上，將形成新的法則、新的語言、新的公序良俗甚至新的文化和信仰。數據的新力量，就如同農耕之於古代文明，工業革命之於現代文明，數據將催生一種全新的文明形態。

一類新的隱私觀

當然，新文明本身也問題纏身。

當我們撥開大數據表面上的浮雲，立刻就可以看到一個商業逐利和社會控制的世界。互聯網巨頭貪婪地吞噬着大數據的紅利，短短的 10 多年時間裡，在美國和中國都出現了巨無霸型的企業。那些原本匍匐在芸芸眾生腳下的互聯網企業，漸露猙獰。無論在中國，還是在歐洲、美國，個人數據都長期處於被無徵求收集、無授權使用的狀態。2018 年 3 月，Facebook 曝出其用戶數據被不當買賣的醜聞，公眾甚至質疑它已經淪為政治操控的工具。個人向互聯網企業讓渡的數據反過來為互聯網企業的“殺熟”行為提供了便利，無數人的個人生活被圈在一個固定的小天地裡，看個性化推薦的新聞，閱讀個性化定制的消费指南，他們感覺很舒服，事實上，我們出讓的數據正在成為我們的電子腳鐐和枷鎖。

這是新文明的兩個悖論。一方面，數據越清晰、越全面、越真實，就越有利於個性化生產，避免資源浪費，比如精準營銷、個性化頁面、私人定制服務；另一方面，數據又帶來了信息繭房、信息窄化的風險。一方面，大數據要求更加開放甚至是無限制的連接，² 另一方面這又將傷害個人的隱私和權利。互聯網漸成壟斷之勢，而個人的命運像一片孤舟，漂浮在無邊的數據海洋裡，我們以為自己是主人，結果卻成了奴僕，我們以為太陽照常升起，可是新的一天已經跟前一天完全不同了。

那該如何看待新文明的問題？這些問題的存在很容易讓人悲觀、失望，但我不贊成將問題擴大化，那是一葉障目，不見泰山。

文明從來都是在衝突中成長的，沒有哪種文明一開始就很完美。中華文明經歷春秋戰國時期的百家爭鳴、幾百年的戰爭，強行融合，最終才在秦漢兩朝形成大一統的文明形態，而儒家思想的確立更是後來的事。可見，文明不是生來就是其最終形態的，它會磨合、變幻，才能最終為大多數人所接受，成為一種認同、一種信仰。今天信息技術的發展提供了新的可能，但數文明最終的形態是需要大家一起創造的。

事實上，上述悖論也可能很快被破解。比如，通證經濟來臨，區塊鏈技術正日臻

完善，未來一個人的數據很可能並不保存在 Facebook、阿里巴巴、騰訊這些大型互聯網公司，而是保存在一個公共的區塊鏈上，這些企業使用我們的數據都必須經過我們的同意，被區塊鏈記錄。只能說，我們目前所見證的數據革命，還遠遠沒有結束，如果真要說結束，那也只是一個序曲的結束。

即使是隱私問題，隨着人工智能的普及，它也在出現新的態勢，我認為整個人類，無論東方或西方，亟須建立一種新的隱私觀。

例如，被中國大眾頻繁詬病的“大數據殺熟”和“千人千價”，它們是通過算法對數據的自動處理實現的，主觀上它沒有泄露任何人的數據。我們使用今日頭條，它可以根據個人瀏覽點擊的記錄推測出個人喜好，做出精準推送，個人信息看似被泄露了，可是處理這些信息的並不是某個具體的人，而是算法和機器，它們自動運行，自行匹配，人為干預的程度很低。那麼，這種情況該不該簡單地被認定為侵犯個人隱私呢？

這就是新的情況：你的數據都是算法和機器在處理，並沒有被泄露給“人”，在一定程度上，你的隱私並沒有受到“人為”的侵犯。人為泄露個人數據的案例和情況當然還會出現，但我相信會越來越少。我們的數據需不需要對算法和機器保密？這才是一個新的問題。

我們不會介意自然環境在注視或監視我們，那我們是否介意算法和機器注視着我們？或者說，我們應該介意嗎？未來，算法和機器就是我們生活環境的一部分，讓機器了解我們，向機器開放我們的數據，這恐怕是通向智能時代、機器人時代、人機協同時代唯一的選擇。

人類新的隱私觀，其核心是要為商業和公共領域的算法劃定一個使用個人數據的邊界。

數文明的發展和延續，我相信是以百年、千年的時間為單位的，探討數文明，就是思考人類的百年大計、千年大計。不管是國家還是個人，我們需要跟上新文明的演進步伐，否則就可能被新文明所淘汰。本書力圖站在人類文明發展的高度探討數據和人工智能，並從中發現新文明的一些規律。

注釋

- 1 《數據之巔》一書的中文繁體版已由香港中和出版有限公司於 2015 年出版。—— 編者注
- 2 聯接不是簡單的“連接”，而是在共同標識基礎上的有機勾連。本書在提到數據的有機勾連時，統一使用“聯接”一詞。

01

數據平權： 新商業文明的衝擊和原罪

文明始於規則的確定。數文明的變革性力量發源於商業，本章從新商業文明的成就和問題入手，提出數據存在產權問題，就像土地存在所有權問題一樣。數據作為一種新興的資源，它應該是公有還是私有？對這個問題的不同回答，就好像歷史上截然不同的兩種主義，它們代表了新文明的兩條道路。如果數據能夠成為個人資產，它將改變現有的商業格局、個人生活和經濟文明。

我們的文明程度越高，潛在的恐懼就會越深。

2018年3月，全球最大的社交網站 Facebook 被曝出負面新聞：一個名不見經傳的小公司，通過不正當的手段，在 Facebook 網站上獲取了 8700 萬用戶的數據。¹ 這些數據隨後被用於多個國家選舉中的選民分析，2016 年當選的美國總統特朗普就曾經雇用這家公司，這引發了關於數據操縱選舉的批評，直逼特朗普當選的合法性，觸目驚心。

排山倒海式的批評，很容易讓人聯想到中國的微信和今日頭條。

就功能而言，美國的 Facebook 不僅僅是一個和朋友交流的工具，它的用戶還可以接收各種信息推送，包括廣告。準確地說，Facebook 相當於中國的微信和今日頭條的混合體，Facebook 存在和發生的問題，微信和今日頭條其實都可以“照鏡子”。

沸沸揚揚的風波之後，Facebook 決定全面參照歐盟的《通用數據保護條例》（General Data Protection Regulation, GDPR），規定在其平台上向用戶推送信息和廣告的企業，必須保證事先獲得了用戶的授權和准許。

這場風波揭開了一個序幕。我認為，這是即將發生變革的第一步，因為核心的問題並非我們過去一直認為的隱私，一個新的問題正在浮出水面，它就是數據權益。數據權益問題事關大部分互聯網企業現行的商業模式，它不僅是一個法律問題，還是一個公益問題。之所以是公益，是因為很多人還沒有認識到它是一種權益。

今天的互聯網企業必須圍繞數據權益對現行的商業模式進行反思。如果大眾完成了對數據權益的認識升級，那麼互聯網就會開始真正的下半場。在這個下半場，我們

將看到互聯網商業格局的巨大改寫以及大眾和政府對傳統互聯網企業權力的限制。

不可否認的是，互聯網的上半場已經開創了一個新的商業文明。我們都看到了新經濟的興起，對這種興起，最近幾年有很多提法，比如信息經濟、網絡經濟、數字經濟、智能經濟等，我認為其本質是“數據經濟”。

澎湃的數據浪潮正在拍打世界經濟和政治，一浪推一浪，人類正在把越來越多的日常生活決策交給數據和算法。在這個過程中，我們感受到的不僅僅是清涼的“海水”，還有“火焰”，它會突然出現，令我們感到一陣陣灼痛。在文明的躍升和轉折中，人類注定既興奮又迷茫，在人類內心的最深之處，還夾雜着對未知的恐懼。

心理入侵：大數據“讀心術”誕生了

在選舉中大量使用數據的做法，起源於美國第 44 任總統奧巴馬。2008 年，他第一次代表民主黨參選期間，建立了專門的個人競選網站（Barackobama.com），收集了 1300 萬人的個人信息和郵箱地址。在此之前，大規模的營銷和宣傳活動一般使用“信息群發”的方式，但奧巴馬放棄了這種方式。他雇用了一批數據科學家，嘗試通過數據對選民進行分類，向不同類別的選民推送不一樣的定制信息。到 2012 年，奧巴馬競選連任，這時候 Facebook 已經聚集了 8 億用戶，奧巴馬的個人競選網站實現了和 Facebook 聯動。支持者一登錄，就被要求提交自己的 Facebook 賬號，網站會詢問是否可以讀取其 Facebook 上的檔案信息，甚至索要在其社交網頁上發佈消息的授權。

這兩屆總統選舉都以民主黨大獲全勝告終，奧巴馬使用的大數據分析方法也成為教科書級別的經典案例。對民主黨因為技術進步而獲得的強大動員能力，當時的共和黨候選人無不羨慕，我曾經將之概括為“政治技術決定政治成敗”。²

在此之後，共和黨痛定思痛，決心迎頭趕上。個別共和黨“大佬”開始和技術極客交朋友，出資參與他們成立的數據公司，其中有一家就是 Facebook 風波的主角劍橋分析公司（Cambridge Analytica），它在共和黨“大佬”班農（Stephen Bannon）和共和黨金主默瑟（Robert Mercer）的支持下，於 2014 年成立。

這家公司之所以叫劍橋分析，和它的團隊和數據來源有關。劍橋大學心理系有一名年輕的研究員科根（Aleksandr Kogan）。2015 年，科根以學術研究的名義在 Facebook 上推出了一個小小的程序“這是你的數字化生活”（This Is Your Digital Life），它宣稱可以免費提供性格測試，參與者還可以得到一個 5 美元左右的現金紅包。這也是今天一個奇怪的現象，一毛錢掉在地上，很多人都懶得彎腰去撿，但對互聯網上的一毛錢，大眾趨之若鶩，搶得不亦樂乎，最終 32 萬人參與了科根的調查。

是時候揭下互聯網的“免費”外衣了：幾乎所有的互聯網服務都是免費的，但使用者其實付出了對價——我們交出了“數據”。

這也正是科根的用意。他不僅收集性格測試的數據，還在他的程序中隱藏了一個數據爬蟲，一旦用戶使用這個程序，它就把通過這些用戶賬號可以看到的資料和信息，諸如“所在城市”“點讚”“好友”等“扒”下來，保存在自己的服務器上。

32 萬人如何突變成 8700 萬人呢？這個數字被放大了 200 多倍。這是因為 Facebook 精心設置的“缺省”隱私政策，即如果用戶沒有特別表態，他的個人信息就對他所有的朋友開放。放到科根爬蟲的工作場景中，這意味着除非一個人的朋友已經特別聲明自己不願分享信息，否則科根的爬蟲在扒取某人信息的同時，也可以把他朋友的信息一併扒取。於是，雖然參與科根測試的用戶只有區區 32 萬，但憑藉缺省許可，這些“種子”最終開枝散葉，科根裂變式地收集到了 8700 萬用戶的信息。而其代價，僅僅為 160 萬美元。

那 Facebook 對科根的爬蟲知情嗎？

完全知情。科根在短短幾週之內要從 Facebook 服務器上扒下來 8700 萬用戶的數據。這些下載數據的行為，會加重 Facebook 服務器的負擔、觸發安全警報。一般情況下，Facebook 不允許開發人員這樣採集數據，但科根隨後以學術研究為由應對了 Facebook 的問詢。在 Facebook 的平台上，學術研究恰恰是被允許的。

科根之後轉手，8700 萬人的數據在雲端悄無聲息地歸到了劍橋分析公司的名下。科根的行為違反了他的承諾，所謂的學術研究很可能一開始就是幌子，但問題是，即使是 Facebook 也無法再追蹤後續數據的交易，如果不是風波發酵，除了交易的雙方，

又有誰能知道數據被買賣了多少次呢？我們無法監控數據的交易，這也是今天這個時代一個難以解決的突出問題。

Facebook 之所以大意，還有一個重要原因，即這些數據並不“敏感”：它們既不是用戶賬號，也不是密碼，只不過是用戶名、用戶所在地、用戶發帖以及一些貌似無關緊要的“點讚”數據。這些信息在 Facebook 上是公開的，幾乎所有人都看得見。

出乎人們意料的是，劍橋分析公司用這些看似不起眼的公開數據，玩出了大花樣。這家公司的方法，不是從傳統的人口學維度給選民劃分陣營，而是立足心理學給選民分類。這種全新的方法，是在奧巴馬的兩次大選之後產生的。

拓展話題

如何通過數據實現人的心理特質分類

心理學家一直相信，人類的特質會被人類賦予一套詞彙來描述。在所有的語言當中，越重要的特質，就會被賦予越多的詞彙來描述，這被稱為詞彙學假設 (lexical hypothesis)。於是，從詞彙中確認重要的心理特質已成為最近 100 年心理學研究的一種方法。美國心理學家奧爾波特 (Gordon W. Allport, 1897—1967) 率先開始了這項艱苦的工作。他在英語詞典中找到了 17953 個描述人格、心理特質的詞彙，又從中挑選出 4500 個做歸類分析，之後幾組研究團隊接力，歷經幾十年對這些詞彙進行篩選和分析。不同的研究團隊發現，總是有五大類特質最先出現在列表上，這就是後來的大五人格 (OCEAN)。大五人格在不同的實證研究中不斷被重複發現，最後被心理學家公認為人格特質模型。

大五人格把人格特質分為五大類：開放型 (openness)，具有想像、求異和創新的特點；責任型 (conscientiousness)，有自律、謹慎、追求完美的傾向；外向型 (extraversion)，有與人交往的強烈意願；親和型 (agreeableness)，表現出高度的體貼和配合；神經過敏型 (neuroticism)，容易不安和焦慮。

大數據和心理學相結合竟然爆發出如此大的威力。研究人員通過實驗發現，通過社交網站的數據可以判斷一個人的心理特質，其判斷結果甚至比調查訪談這個人的親朋好友還要準確。

也就是說，只要有足夠的社交數據，不用任何人為的介入，計算機和算法就可以自動判別一個人的心理特質，甚至僅僅憑藉“點讚”數據就可以完成，因為沒有無緣無故的愛，每一個“點讚”背後都有原因。如果掌握一個人在 Facebook 上的 10 個“點讚”，算法對他的了解就可能超出他的普通同事；掌握 70 個就可能超過他的朋友；掌握 150 個就可能超過其家庭成員；掌握 300 個就可能超過其最親密的妻子或丈夫。³

劍橋分析公司在這條路上走得更遠。他們把 8700 萬人的社交數據和美國商業市場上 2.2 億人的消費數據進行匹配、組合和串聯，找出誰是誰，然後就性別、年齡、興趣愛好、性格特點、職業專長、政治立場、觀點傾向等上百個維度給選民一一打上標籤，進行心理畫像，建立心理檔案，再通過這些心理檔案開展分析，總結出不同人群的希望點、恐懼點、共鳴點、興奮點、煽情點以及“心魔”所在。

當代大數據“讀心術”就此誕生了。

掌握了一個人的“心魔”，就可以評估一個人最容易受哪種信息的影響，就可以知道信息該如何包裝、如何推送，才能搔到接收者的癢處，潛移默化地影響一個人的選擇和判斷。

回到特朗普的選舉上：對義憤填膺的愛國者，可以推送“雞血”文章，號召他們保衛美國，告訴他們特朗普能讓美國重新偉大；對知識精英，則推送包裝得比較溫和的、理性的、顯得高深的分析，以獲得他們的共鳴；對傾向於特朗普但又不願意捐款、不願意出門投票的人，就告訴他們特朗普面臨着嚴峻的挑戰，如果希拉莉上台，後果將會非常慘痛，鼓動他們出錢出力；對民主黨的支持者，則大量推送和希拉莉家族有關的負面信息，讓他們對希拉莉失望；而對那些容易被煽動起來的網民，就推送各種聳人聽聞的陰謀論和“標題黨”內容，進行情緒刺激和情緒強化。

當然，很多推送的信息是被精心修飾過的事實，有些甚至是由專門的團隊為特定細分群體“炮製”出來的虛假消息，但這些消息迎合了人性的需求，直達人類心理的最深處，而且算法絕不會推送觀點相反的信息，接收者將會沉浸在一種氛圍當中，質證、懷疑的可能性因此大大降低。

相同的配方，熟悉的味道，這些精準投放的宣傳資料，就好像一顆顆用數據調配

的藥丸，被推送給選民反覆服用，之後慢慢地發生作用，最終引導選民做出藥丸配製方預設的政治行動。

從本質看，這是互聯網的精準廣告投放技術在選舉中的應用。

現代廣告業已經有上百年歷史，它和心理學的發展緊密相關。從 20 世紀頭 10 年開始，廣告商就採用統計分析的方法研究如何編輯、呈現信息，希望以情動人、感召大眾掏出錢來。這個過程就好像愛迪生為了找到燈絲的最佳選材，連續做了 1600 多次試驗，最終發現了鎢絲一樣，廣告商通過不斷試錯發現了組織、呈現信息的最佳配方。心理學已經證明，人類雖然有理智，但人性中有很多“漏洞”，人類大部分時候都會被情感左右，一些簡單的伎倆就可以影響、操縱人類的情感。經過 100 多年的研究，廣告業對人性弱點的洞察已經爐火純青，諾貝爾經濟學獎得主阿克洛夫 (George A. Akerlof) 將這種利用人性“漏洞”的廣告稱為“釣愚”。

互聯網從兩個方面大幅提升了傳統廣告的效率。

一是以 A/B 測試的迭代方法加速尋找最佳的“釣愚”配方。

假設一個電商平台有 100 萬個用戶，如果它要推出一件商品的特定廣告，可以推出兩個不同的設計版本，然後在兩個用戶群中進行實驗。假設其中 50 萬用戶看到版本 A，產生了 10 萬購買；另外 50 萬用戶看到版本 B，產生了 12 萬購買，這就說明版本 B 的轉化率更高，通過控制平台，它可以迅速將版本 B 推送給所有用戶。

廣告的版本當然不止 A、B 兩種，其版本的變化有很多維度。拿標題舉例，是長標題還是短標題，是疑問句還是陳述句，是正式語氣還是非正式語氣，是強調賣點 A 還是強調賣點 B，效果會大不相同。即使是同一組文字，其字體、粗細、大小、顏色、背景的不同也會給瀏覽者不同的感受。再比如佈局，是橫幅還是多列，是大圖片還是小圖片，是手繪圖還是實景照等，其中有近乎無數的變化。這意味着 A/B 測試的過程可以不斷重複、優化，直到無法進行任何新的改進，到這個時候，一個趨近於完美的“釣愚”配方就出現了。一旦推出，它就將帶來最高的點擊率和購買率。

要調整、追蹤、對比這樣上百個維度的變化，對人類來說很難完成，但這恰恰是人工智能擅長之處。算法不僅能記錄每一次變化以及每次變化的效果，而且可以自動

地調整、推送這些變化，在上百萬人的大規模人群中不斷重複這項工作，其細緻之處，人類難以企及。

二是給不同的用戶推送不同的廣告。

傳統廣告業最大的痛點就是千人一面。無論是電視還是平面媒體，面對一千個不同的觀眾和讀者，自己的廣告只能是一次一個版本，很明顯，無論如何洞察人性，我們所理解到的大部分人性也只是一種共性，一個廣告不可能適合所有的人群。世界是一個大湖，魚很多，它們有共性，但個性更突出，如果每一個“釣魚”的餌料能夠在共性的基礎上稍微調整一下，讓它適合每一條魚的偏好和口味，那上鉤的魚自然更多。

互聯網可以做到。它通過對消費者一舉一動的監控，分析消費者的動態和偏好，根據消費者的行為和需要設計推送信息，即使對同一個事實和新聞，它也可以根據目標受眾心理特質的不同，變換語氣和語句，疊加不同的優惠券，確定不同人群的最佳推送時間和推送頻次，這就是所謂的精準推送。

研究人類心理的異同，這是科學。把合適的信息在合適的地點、合適的時間推送給合適的人，信息可以動態調整，對象可以動態匹配，這就是技術。就此而言，大數據不僅僅是科學，還是技術，它是科學和技術的混合體。

這也是特朗普在選舉中雇用劍橋分析公司所要完成的主要任務。在特朗普公開的競選支出明細中，劍橋分析公司的名稱赫然在列，特朗普團隊先後 5 次購買該公司的數據服務，共支付了 5912500 美元。⁴ 在 2016 年的選舉中，希拉莉所代表的民主黨最終敗給了從頭到尾都不被“大佬”和精英看好的特朗普，劍橋分析公司給選民推送了甚麼樣的“數據藥丸”？它們究竟對勝選起了多大的作用？這些我們已經無從得知，即使歷史可以重來一遍，不同的“數據藥丸”對不同選民的心理影響也很難被量化、被證明。

在劍橋分析公司的主頁上，這個公司號稱服務過世界五大洲的 100 多場選舉，其中還包括 2016 年英國的脫歐公投。這是 2016 年度全球最大的“黑天鵝事件”，脫歐派以 52% 對 48% 的微弱優勢最終勝出，全世界為之感到意外。雖然“數據藥丸”對選舉的作用大小很難量化，但有一點可以肯定：在一場勢均力敵的、差距很小的選舉中，它會起到關鍵的作用。



圖 1-1 數據驅動選舉：劍橋分析公司的主頁

劍橋分析公司主頁的自我介紹：我們定位你的選民，然後驅使他們行動。我們是用數據驅動競選活動的全球領導者，擁有超過 25 年的經驗，助力了五大洲 100 多場競選活動。僅在美國，我們在總統競選以及國會、州議會選舉中就發揮了關鍵作用。

我不認為 Facebook 風波是傳統的隱私問題。

在 Facebook 最後的道歉聲明中，公司並不承認這是一場“數據失竊”事故，而是強調“沒有保護好用戶的數據”。事實上，這起風波連“泄露”都談不上，因為“點讚”數據幾乎是公開的，數據在整個過程中都是經過用戶同意才被扒取的。如果真要說這是隱私侵犯，也是合理侵犯，但真正的問題卻比隱私侵犯還要嚴重。

真正的問題在於，即使通過公開的數據，互聯網也可以成為影響、操縱、控制他人心理和觀點的媒介工具，這不是隱私侵犯，而是心理入侵、思想入侵、意識入侵。

風波背後還有更多的尷尬和無奈。事實上，Facebook 早就發現了科根在濫用數據。2015 年，Facebook 甚至查封過科根的賬號，並要求他和劍橋分析公司銷毀相關數據，但所謂的銷毀只是一個文件的聲明，由科根在上面打鉤簽字，然後寄回 Facebook 即可。

這些數據當然沒有被銷毀。數據是一種特殊的資源，沒有任何公司和個人能夠有

結語

第四次浪潮：我們如何再次領先

公元 1500 年被很多學者視為世界近代史的開端，美國著名歷史學家斯塔夫里阿諾斯（Leften Stavros Stavrianos，1913—2004）的《全球通史》（*A Global History*）就以公元 1500 年作為重要的歷史分期點，其主要依據是 1500 年前後，世界開始連為一體，標誌性事件就是大航海的興起，它讓歐洲、非洲和美洲聯繫到了一起。

在此之前，中國曾經是一個全面領先的帝國，但這之後，中國逐漸衰落，其中的原因，歷來眾說紛紜。

我從記錄的角度考察，認為導致這次落後的主要原因就是印刷機在歐洲發明。印刷機，可以說是繼文字和紙張發明之後的第三次浪潮，但中國錯過了這一浪潮。人類的歷史有一個著名的現象：當一個重要的機遇出現時，抓住機遇的人從此會擁有更多；而錯失機遇者不僅得不到新東西，連他原來有的優勢也會逐漸喪失，即強者越強，弱者越弱，這就是馬太效應。這種“馬太之痛”在中國一直持續到 20 世紀下半葉。

印刷機之後，西方在 19 世紀 40 年代發明了照相機，在 19 世紀 80 年代發明了留

聲機，人類發現圖像竟然可以留駐，聲音竟然可以重現，這令當時的人類感到難以置信。憑藉這些新的工具，人類可以記錄圖像、聲音，這在文字的基礎上又擴大了記錄的範圍，但這時，因為設備昂貴、攜帶不便，人們要照相就要去照相館，要錄音也要去專業的場所，這些新的記錄手段仍然掌握在組織手裡，沒有普及到個人。

20 世紀的最後 10 年，這種情況開始改觀。由於互聯網的興起，人類的記錄能力得到了空前提高，文字、聲音、圖片、視頻，所有的信息都開始被數據化。2007 年，蘋果公司推出了第一代智能手機 iPhone，隨着智能手機的普及，這些空前強大的記錄能力已經轉移到了每一個普通人的手掌之中，普適記錄的時代大幕就此拉開。這是繼文字、紙張、印刷機之後，記錄的第四次浪潮，幾乎人類所有的生活過程都可以被記錄，這些多樣化的記錄還可以存儲在雲端，萬世可見。

回顧人類的歷史，從農業社會開始，人類早期僅僅是在大地上勞作，靠大地的產出生活，但人類對大地的影響微乎其微；進入工業文明之後，人類在地表架設機器、開採資源，對地表進行改造，創造了玻璃、鋼鐵、水泥、自來水、電、氣等新的材料和基礎設施，城市成了文明的中心，人類的精英在此集聚，在地表之上，城市堪稱一個半人工體；再到信息社會，人類創造了一種新的資源——數據，互聯網成了人類社會沉澱數據的大陸架。在短時間內，數據迅猛增多，而且不斷相聯接，人類正在創造一個嶄新的空間——數據空間，這意味着在宇宙之外又出現了一個空間，在這個新的空間裡，數據是滋生萬物的土壤，人類是它完全的主人。

換句話說，數據空間，是一個純粹的人工世界。

人類從這個新的空間中獲得了能量“加持”。相較於 20 世紀，因為普適記錄的個體化，今天的個體是高能個體，數據空間的能量將逐步被釋放，天下將為之改變。

今天，我們寫下的任何一段話，都可以在互聯網的知識庫中做實時的檢索、對比和計算，這句話和哪一本書中哪一段落相似，你的思考、感悟、結論是新的嗎？前人就此話題有哪些類似的發言、論述，有沒有深入闡述，或只是粗略地提及？當古今中外所有類似的論述，即使是歪評厥詞，都盡收眼底的時候，一個人已經站到了前人的肩膀上，你再往前邁小小一步，就是創新。

2017年元宵節，我和杭州詩人王自亮先生聊天，他送了我一本他新出的詩集，談到他一直想以“貓”為題材寫一首詩，但一直不敢寫，因為法國詩人波德萊爾有一首《貓》，那首詩已經把貓寫到了形神皆具、登峰造極的地步。這讓我想到，一首好詩就是一首“永遠不需要重寫的詩”，一篇好文章、一本好書也是同樣的道理。因為它不可替代，它就是“邊界”。

人類現有的知識是有邊界的，但知識本身沒有邊界。對一個開拓者來說，快速地知道邊界在哪裡是至關重要的，知道了“自己在哪裡，在哪一條邊上，前人已經畫了一道甚麼樣的邊”，就可能“站到前人的肩膀上”，一旦站上去，創新就更容易了。

如果把現有的、所有前人開拓的知識比作一個圓，那在它的邊上、在它邊上的任何一個點上，後人都可能有所突破，隨着這一個一個的邊向外拓展，人類知識的邊界就會擴大。

就此而言，以前的創新，大部分是專業人員才可能做出來的，這是因為“找準這條邊、了解這條邊”，“站到前人的肩膀上”，需要專業的努力，即很高的成本。但今天，因為互聯網的數據沉澱，因為強大的、實時的、低廉的、無處不在的普適記錄，這個門檻降低了，普通人，即大眾，也可能創新。歷史上曾經有些創新就是由普通人做出的，未來，這種概率會大大加大。

一切業務數據化，導致了記錄的大規模沉澱，網絡中的記錄不僅在不斷積累，還在不斷疊加和優化。這是人類其他記錄手段，例如印刷機，難以解決的一個問題，書籍一經出版，它的知識雖然被保存了下來，但也隨着印刷機固化了，而網絡中的記錄永遠在更新，一個知識點，可以在網上不斷地被閱讀、編輯、討論，例如維基百科、大眾點評的條目。在這個不斷循環的過程中，網絡上的記錄是動態的、跳躍的、有機勾連的，因此查閱書籍和查閱網絡的效果完全不同。

在這些記錄中，有相當一部分是個人的行為和觀點。這些數據疊加在一起，產生了一種新的經驗和權威：眾智。

在歷史上，“不能記錄、不能量化”曾經是一種力量，藉助這種力量，組織、公司和專家可以蒙蔽、控制和操縱普通的個體，但因為普適記錄，這種力量被消解和轉移

了，專家正在被大眾共識和眾智取代，這種共識的呈現形式就是數據。例如，我們在買一件商品的時候，可以清楚地看到有多少人購買了這件商品，有多少個評論、多少次“點讚”，這對我們的消費行為直接產生影響。

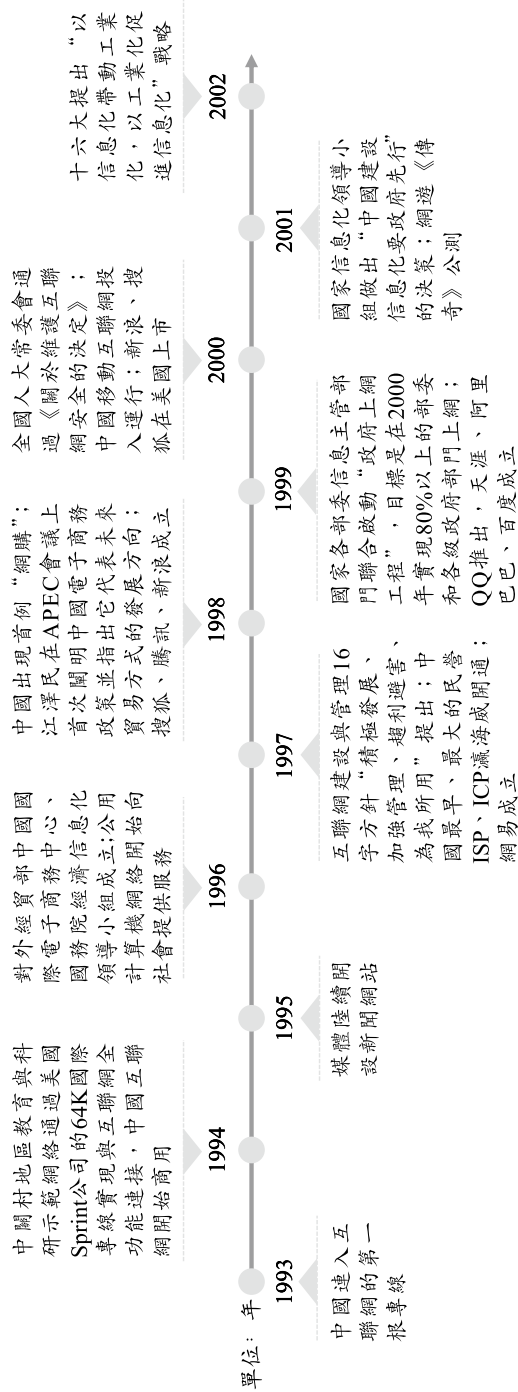
數據在給天下帶來新的光明。所有的數據都可能發光，它類似於自然之光、文字之光。我們在夜空中看到的星星，大部分都閃爍着微弱的光芒，但事實上它們像太陽一樣，每時每刻都在散發着灼熱的光，具有巨大的能量，只是距離我們太過遙遠。距離太陽系最近的恆星是“比鄰星”，它和地球相距 4 光年，也就是說，以每秒 30 萬公里的速度行走的光，需要 4 年多才能抵達，它和地球的距離比太陽和地球的距離要遠約 27 萬倍。我們之所以感覺它的光很弱，是因為它遠。只要我們一靠近，就會感受到它釋放出的巨大能量。數據也一樣，它距離我們似乎很遠，我們觸摸不到，但數據一旦為我們所用，就會釋放出大片的光明。

從長遠看，我對中國文明的發展抱以樂觀的態度，原因在於，近 30 年來，我們抓住了互聯網作為整個社會基礎設施所潛藏的機遇。歐洲是活字印刷機的故鄉，也是工業革命的策源地，但正像我們錯過了印刷機一樣，歐洲沒有抓住新一輪互聯網革命的機遇。迄今為止，歐洲還沒有出現一家世界級的互聯網公司，它落後了，工業革命時代的榮光，如今已經星光漸暗。

互聯網興起於 20 世紀 90 年代初。1991 年，美國政府取消了對互聯網使用的限制。此後，互聯網從一個用於軍事、研究的專門網絡變成了一個沒有限制的、可公開訪問的商業網絡。之後，世界各地無數的企業及個人紛紛湧入，帶來了互聯網發展史上的一個巨大飛躍。中國一直緊鑼密鼓地籌備，不僅參與了、跟上了，現在甚至在局部領域開始引領這個潮流。

2005 年，中國的網民數量首次突破了 1 億。2017 年末，中國的網民數量已經達到 7.72 億（其中手機網民 7.53 億）。¹ 全球有五大互聯網公司，中國已經佔據兩席。中國搭上了互聯網的快車，成為一個不折不扣的互聯網大國。

互聯網興起之初，更多的是作為一種信息傳播工具。回顧歷史，其實也是驚心動魄，中國曾經存在和互聯網失之交臂的可能，如果與互聯網失之交臂，幾乎可以肯定，



圖結 一-1 1993—2002 年：中國互聯網的黃金 10 年

中國不會出現今天的繁榮，所有互聯網帶來的經濟變化，電子商務、雲計算、社交網絡、移動支付、人工智能，其進程都將大大減緩，抑或根本不會出現。

從 20 世紀 80 年代開始，中國的領導人就對計算機和網絡持歡迎的、肯定的態度，持開放的心態，採取“先發展，後管理”的方式。在相當長的一段時間裡，中國的互聯網產業沒有明確的主管部門，這給互聯網在中國的發展提供了巨大的空間。

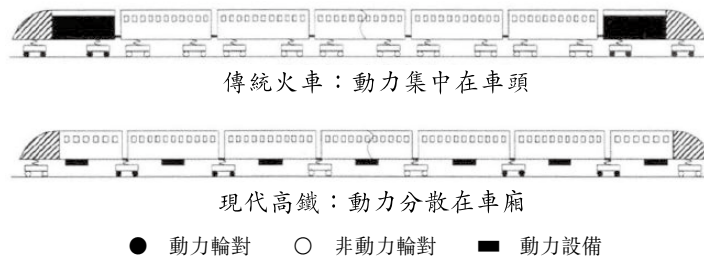
“先發展，後管理”的決定，帶來了巨大的紅利，這個紅利目前還在釋放當中，追根溯源，就是因為我們採用了開放的心態面對互聯網。今天，中國的領導層也充分體認到了這個決策的正確性，當新的大數據、人工智能的浪潮剛剛湧起的時候，面對新的機遇，中國用更積極的態勢張開了雙臂。

2015 年 8 月，國務院印發《促進大數據發展行動綱要》，倡導“用數據說話，用數據管理，用數據決策，用數據創新”。2017 年 7 月，國務院出台了《新一代人工智能發展規劃》（後文簡稱《規劃》），《規劃》中指出，中國必須主動求變應變，牢牢把握重大歷史機遇，主動謀劃、搶佔先機，引領世界人工智能發展新潮流。《規劃》確定了中國人工智能發展的“三步走”目標：到 2020 年，總體技術和應用與世界先進水平同步；到 2025 年，基礎理論實現重大突破，部分技術與應用達到世界領先水平；到 2030 年，理論、技術與應用總體達到世界領先水平，成為世界主要人工智能創新中心。

這份《規劃》中有很多超前的、醒目的、出人意料的表述，例如人機協同、群智開放、智能經濟、智能政府、跨媒體協同處理、人機協同增強智能等，從這些詞語中，我看到了中國政府的雄心，也體會到最高領導層抓住了時代機遇的欣喜之情。200 多年前，中國錯過了工業革命，這一直是我們的切膚之痛，但到目前為止，我們抓住了信息革命的機遇。

天下在變。今天出門，我常常搭乘高鐵，高鐵已經成為中國的一張名片。相較於傳統的火車，它最大的特點是“快”，之所以快，是因為它新的動力結構，每次登上高鐵，我都會從它的動力結構聯想到整個社會的變化。

在火車漫長的發展歷程中，“火車頭”一直是牽引整個車廂前行的唯一動力，所謂“火車跑得快，全靠車頭帶”，這種傳統的動力是集中的、單一的，但“一拖多”限制了



圖結一2 傳統火車與現代高鐵的動力分佈區別

火車前進的速度。高鐵的動力是分佈式的，它的每一節車廂單元都有動力，車廂是自己跑，不是被車頭拖着的，多節動力車廂構成的整體車身處於一個高度協同的狀態。這種協同，不僅是機械協同、能量協同，還是信息協同、數據協同。這樣一來，高鐵當然跑得快。

高鐵的這種動力配置和雲計算、大數據的架構何其相似。我們原來的計算體系是，數據集中存儲在一個地方，需要計算的時候，再把它們拉出去，常常是一部小型機拖着一個存儲機櫃，存的存，算的算；而分佈式的雲計算，它的每一個節點都既有存儲又有計算，因而提供了充沛的動力和功能模塊。一台普通的計算機只要和雲的體系相連，就不僅可以承擔存儲任務，也可以承擔計算任務。這意味着一個節點既提供原材料，又有加工的能力，而不僅僅是被動地執行。

再看區塊鏈的架構，它也是一種分佈式的記錄體系。一個交易發生後，區塊鏈會在所有的節點記錄備案，即每一個節點都可以存儲記錄，如果把區塊鏈比作一個公司，就相當於人人都是會計，人人手中都有一個賬本，所有的節點在鏈條中一律平等、互相監督、實時對賬，每一個節點都可以在必要的時候充當傳統結構中的主節點。

高鐵、雲計算、區塊鏈，當它們的架構被融合進整個社會的時候，我們社會發展的動力結構就會發生改變。

這種變化，是我對中國社會進步的信心之源。

我認為，中國的領導人已經敏銳地感受到了這種時代的變化。2015年，“大眾創

業、萬眾創新”被寫入《政府工作報告》，“雙創”要激活的是億萬人的創造力，其最終目的，是要調動每一個人的能力來推動國家和社會的發展。換個角度看，國家的發展需要每一個人的能量，這也是國家對國民主體價值的尊重和認同。

中國的改革開放已經進入了第 41 個年頭，過去 40 年，更多是靠國家驅動、精英驅動，到了今天，我們需要全體國民來參與、驅動。一個國家的發展要有強大的、更持續的動力，這種動力，必然根植於這個國家的每一個國民。

技術正在賦能個體，文明正在翻開新的篇章，從國家驅動到國民驅動，這個正在轉變中的動力機制才是中國繼續蓬勃發展、再次領先世界最為根本的保證。果能如此，我們就能開創一個新時代，作為人口大國、數據大國，我們未來可期。

注釋

- 1 《中華人民共和國 2017 年國民經濟和社會發展統計公報》，國家統計局，2018.2.28。

後記

POST-MEMORY

野路無人自還

這是繼《大數據》《數據之巔》之後，我關於數據的第三本書。和前兩本書相比，我嘗試引入一些變化。

一是視角變了，前兩本書都是在留美期間寫的，更多的是從世界看中國，本書提筆之時，我已經回國工作三年，做過了企業高管，也開啟了自主創業，已接上國內地氣，這本書是我立足中國的變化，對當下發展趨勢和改革方向的思考。

二是在讀了 8 年多的英文書之後，我開始重新翻閱中文書，特別是歷史經典，我希望從傳統的中國智慧中獲得啟發，希望不僅用中國的語言，還要用中國的智慧闡述現代的問題，並提出解決方案。

例如寫到量子力學時，陶淵明這兩句詩跳入了我的眼簾：“人生無根蒂，飄如陌上塵。分散逐風轉，此已非常身。”這是《雜詩十二首》第一首的開頭。靖節先生的大意是：人生在世就像漂泊在路上的塵粒一樣，四處分散、隨風飄轉，此身歷盡艱難，已非原來的我了。這啟發我想到，現代人就是城市中的粒子，他們像粒子一樣互相作用，人類社會更適用量子力學，人也像微觀世界的粒子一樣“測不準”。

又如寫到人臉識別時，我追溯了中國古代畫工的歷史。西晉文學家傅咸（239—294）曾經專門作《畫像賦》，他概括說，貴族之所以要畫像，是“惟年命之遒短，速流光之有經，疾沒世而不稱，貴立身而揚名”，即其根本目的是對抗遺忘、存名後世。也

就是說，最早的繪畫也源於和文字幾乎一樣的目的：記錄。這在東方和西方是一樣的。

說記錄是人性的需要，其實也不為過。

本書的第三個變化，是需要面向未來，回答諸多前沿的問題。例如，由於深度學習的提出，最近 5 年人工智能取得了巨大進步，但人工智能是不是還會出現新的方法？深度學習是向數據學習，智能是用數據餵出來的，如果有一天，機器能夠向書本學習，它會不會更接近於人？又如，人類的駕駛員會出交通事故，醫生可以有、也肯定有誤診，為甚麼我們就不能允許自動駕駛出事故，不能接受人工智能有誤診呢？再如，人工智能的發展是不是也有邊界？人類擁有一些無法言說的隱性知識，它們很難被規則化、語言化，用文字、數據都無法有效表達，這也是為甚麼醫生年紀越大越值錢，因為隱性的知識只跟時間、年齡和經驗的積累有關，它們難以言表，因此也難以傳承。那問題便是，在數文明之外，未來又會不會出現新的文明形態？

在算法、機器智能大幅躍進，向人類逼近的同時，人類的生活卻越來越程式化，在向機器靠攏。20 世紀 30 年代，喜劇大師卓別林曾經創作《摩登時代》，塑造了一個因為在流水線上整天重複擰螺絲釘而異化的工人形象。大部分人沒有意識到的是，手機也是一台設備，它也有流水線，軟件就是它的流水線，但這條流水線在雲端，是不可見的。如果說工業社會的異化只影響了流水線上的工人，那數據社會的異化正在波及幾乎每一個人。今天，無論是在機場地鐵、街頭路口，還是辦公室、客廳、臥室，我們無時無刻不在看手機，而且也在重複一兩個動作：滑屏和點擊。

就人的本性而言，我贊同人工智能先驅明斯基的看法，大部分人在大部分時候就是一台機器。因此，人工智能只要達到人類的“平均智能”，就可以在很多場合代替很多人，而人工智能超越“平均人”僅僅是時間的問題。

但人工智能是否能為普通人所用呢？例如，把芯片植入我們的大腦，讓我們可以快速調取各種各樣的信息，人類就可以像機器人“沃森”一樣回答各種各樣的問題，那“平均人”的水平是不是會大幅提升，變成“增強人”？這種方法會奏效嗎？人類的創造力會不會因此增強？

此外，算法正在把很多事情都變成數據進行匹配，算法的普及減少了整個社會的

隨機性，生活更經濟了，但是不是更更好了，我們也不得而知。再如，整個社會因為數據而形成了一個強關聯的系統，但這個系統又存在大量的漏洞，智能社會是更脆弱還是更穩健？

這些都是在人類現有的知識邊界凸現的新問題，本書就是在嘗試回答這些問題。

很多個深夜，我一人在書房徘徊，或獨坐在電腦面前，一遍遍校正自己的邏輯和基點，這時候，我的腦海中會反覆回響起南唐李煜的一首詞：“心事數莖白髮，生涯一片青山。空山有雪相待，野路無人自還。”

這首詞描摹的是一個在雪山和野外徘徊的獨行人，字詞之間透着一股冷寂和無奈。我寫作期間，突然意識到這個古代的獨行人就是今天的一個創新者，這種冷寂無奈就是人類的創新之苦。創新就是要向邊界和邊緣突進，那裡大雪滿山，也可能是空山，你徒手攀登，可能空手而回；那裡是無人區，你必須在無人區思考、徘徊、拓進。這種冷寂和孤獨是常人難以忍受的，只要是人，在無人區就待不久，因為寂寥和寒冷，我們只能半路折返回來，所謂“野路無人自還”。

但這不要緊，一個人只要向前突破一點點，就是對人類文明的貢獻。文明的進步從來都是這樣產生的。

回答新的問題，就是創新，這是我撰寫本書最深的感受。

我的啟發是，現實中的創新者，一定要善於用語言和行動不斷突破周邊物理世界和人際關係的邊界。

我的另一個感受是，求人學、求文學可以借鑑中國古代的智慧，但若是求科學、求邏輯、求批判、求精確，那我認為，寄望於大部分中國的書就是緣木求魚。換句話說，從中國過去的智慧裡找不到現代社會的答案。科學技術正在重塑天下，人還是那個人，但世界已經不是那個世界了。今天的現代化，要着力於“數治”，中國的傳統智慧提供不了答案，這也正是我寫作本書的原因。

相較於前兩本書，本書的寫作還有一個困難。我從阿里巴巴離職之後，萌生了一個想法，希望創建一家優秀的數據科技諮詢公司，本書的寫作，是在諸多商業項目中見縫插針完成的。人和機器之間的另外一個共同點，就是人腦需要“預熱”，有很多次

我無奈地發現，預熱的時間竟比寫作的時間還要長，在商業項目和寫作之間穿梭，我常感愧疚，因為兩件事可能都沒有做好。

本書最後能夠完成，有賴家人和朋友的支持。

我首先要感謝太太的理解、兒女的信任。其次要感謝雄安新區的陳剛書記、廣州市人大的陳建華主任以及時任浙江省人大常委會副主任的毛光烈先生，他們關注大數據在中國的發展，對我的思考研究鼓勵有加。我還要感謝南京市發改委的沈劍榮主任、何軍副主任，蘇州工業園區的許文清局長、段晴毅局長。2017年，我承擔了為南京市新型智慧城市做規劃、為蘇州工業園區設計城市大腦兩個項目，這兩個項目幫助我了解了中國個別發達城市的智慧城市建設實況，本書中不少案例都來自南京和蘇州的調查和實踐。

我要特別感謝涂新輝先生，從2012年我們因書相識，結緣已經6年，他在事業發展上給了我諸多提點，幫助我應對人情世故中的“不測風雲”，可謂用“心”陪伴。在這個過程中，我發現了人生的新方向、新滋味和新追求，得一摯友，何其幸哉。

我要感謝公司的各位合夥人、同事以及業界專家。高路通讀過初稿，提出了不少修改補充意見，葛育民、胡曉萌協助我收集、整理了大量的素材。我還要感謝蔣永軍和張英潔，他們為本書的面世做了許多細緻的準備工作。

我最後要感謝錦天城律師事務所、大成律師事務所，它們優秀的律師團隊為我處理了大量的法律事務，我才能專心致志、心無旁騖。

手機和微信的普及，已經極大地改變了人們的閱讀習慣。在本書的寫作中，蔣永軍先生曾經多次提醒我，千萬不要大段地議論，這既費力又不討好。於是，我希望這本書，是一場對話。雖然是我在講，你在聽，但你可以隨時合上書本，來我的微信公眾號上提問，不管你在哪裡，只要有網絡，你我的距離，就是一個二維碼這麼近。我們的交流，也將成為數據，數據會記錄這溫馨一刻，這也是數文明發展的真實記錄。

是的，數據和人工智能將會重塑天下，人性的溫馨卻不會改變，這就是世界的變與不變。

涂子沛

2018年7月6日



大事記

EVENT MEMORY

- B.C.3500 蘇美爾人發明一種象形文字，說明文字從圖形和符號演變而來。商朝出現甲骨文，表明中國是世界上較早發明文字的國家
- 105 蔡倫改進造紙術，開始大規模造紙，當時中國的造紙術領先於世界
- 450 北魏皇帝拓跋燾因不滿司徒崔浩在《國史》中的記述而誅其全族，史稱“國史之獄”
- 626 李世民發動“玄武門之變”並繼承帝位，開創中國歷史上少有的盛世
- 819 韓愈因得罪皇帝被貶，從長安歷經 100 多天到達廣東潮州
- 960 趙匡胤發動“陳橋兵變”推翻後周稱帝，建立北宋
- 1041 畢昇發明活字印刷，但字模因用膠泥製作而易碎，無法重複使用
- 1101 蘇軾獲赦並從海南出發北歸，行走數月，途中病逝於江蘇常州
- 1275 元軍進攻南宋，文天祥召集三萬窯工勤王，吉州窯因此戰亂而衰落
- 1454 德國人古騰堡在歐洲發明活字印刷術，促進西方信息、知識的大規模傳播
- 1494 意大利數學家帕喬利提出複式記賬法，開創現代會計學
- 1519 西班牙殖民者來到墨西哥，開始長達 400 年的殖民掠奪
- 1605 德國天文學家開普勒確立三大行星運動定律，為萬有引力的發現奠定基礎
- 1687 牛頓系統總結物體運動的三大定律，首次將宏觀物體的運動概括於一個理論
- 1700 德國人伯特格爾通過三萬多次燒瓷實驗燒製出瓷器，打破中國技術壟斷
- 1727 牛頓去世，英國為他舉行國葬並留下經典的墓誌銘
- 1759 英國人韋奇伍德歷經 5000 多次實驗並最終推陳出新，發明骨瓷
- 1773 拉普拉斯把萬有引力定律應用到整個太陽系，回答木星軌道收縮、土星軌道膨脹難題
- 1774 歌德出版了代表作《少年維特的煩惱》，帶來時代性的轟動並引發“維特效應”

- 1776 潘恩的《常識》出版，從思想上對美國獨立產生巨大影響，一本書催生了一個國家
- 1781 “清代第一大貪污案”甘肅冒賑案案發，66 人被列入死刑名單
- 1791 《紅樓夢》迎“大轉折時代”，首個印刷版《紅樓夢》完成，被稱為“程甲本”
- 1808 拿破崙要求猶太人必須取姓，西班牙殖民菲律賓時也用同樣做法，以便於社會治理
- 1814 拉普拉斯提出“拉普拉斯之妖”理論，決定論登峰造極
- 1819 莫里森首次用活字印刷術在中國印成《聖經》，該技術由此在中國被推廣和應用
- 1823 倫敦使用煤氣路燈照明，在預防犯罪方面起到巨大作用，引來其他城市效仿
- 1837 美國成立首家信用公司，在信用評估領域走在世界前列
- 1839 達蓋爾發明攝影法，奠定現代攝影基本工藝的基礎
- 1841 美國國會在議會廳專設“記者席”，但議員指責記者斷章取義，雙方關係交惡
- 1860 英國維多利亞女王首次拍攝個人肖像照，並創下全國銷售 50 多萬張的紀錄
- 1877 愛迪生設計出人類第一個錄音機模型，次年成立愛迪生留聲機公司
- 1882 愛迪生在美國紐約建立了世界上第一個供電系統
霍爾瑞斯在打孔卡實驗的基礎上發明打孔製表機，促進信息記錄規範化
- 1884 NCR 成立並在全球售出 2200 多萬台收銀機，帶來商業變革
- 1886 梅特卡夫大力倡導工時記錄，他認為所有的管理記錄都應該被保留
- 1897 留聲機進入西藏，喇嘛用它錄製了禱告語
- 1899 嚴復翻譯《國富論》，譯書院購買該書的版權，中國版稅制度由此發軔
- 1911 IBM 的前身 CTR 成立，主打產品出勤記錄鐘
- 1935 美國頒佈《社會保障法案》和《工資工時法案》，催生打孔製表機需求，IBM 由此壟斷打孔製表機市場
- 1937 美國金門大橋建成，成為人類征服自然的標誌和美國的一個象徵
- 1938 德國、美國發明了基於磁帶的錄音機
- 1939 羅斯福公開批評新聞界歪曲真相，而後他繞開媒體，藉助廣播進行“爐邊談話”

- 1940 剛被發明的磁帶錄音機進入白宮，日後影響了美國的歷史進程和政治生態
- 1946 人類發明的第一台計算機被用於處理美國人口普查的數據
- 1948 IBM 生產了自己的第一台大型計算機，每秒可運算數千次
- 1952 首款商用電子計算機 IBM 701 誕生。之後，塞繆爾在它上面開發了跳棋程序“checker”，讓世人首次認識到計算機具備一定智能
- 1956 美國達特茅斯會議上，“人工智能”成為新的領域，該會議被認為是人工智能的開端
- 1957 美國康奈爾大學計算機教授羅森布拉特提出“感知器”概念
- 1959 塞繆爾在跳棋程序的基礎上，正式提出“機器學習”的概念
- 1960 人工智能專家布萊索在美國成立全景實驗室，受國防部資助而研究人臉識別
- 1961 雅各布斯出版名著《美國大城市的死與生》並發展了“街道眼”概念
- 1962 有約 7000 多年歷史的賈湖刻符在河南出土，它被認為可能是漢字的源頭
- 1964 恩格爾巴特設計出世界上第一個鼠標，互聯網訪問者的行為通過鼠標被轉化成數據
“雅各貝利斯訴俄亥俄州”案中，電影《情人》的“淫穢”認定影響案件判決
IBM 打造“半自動商業體系”並應用於票務系統，極大地變革了航空業
- 1965 IBM 發明新的條碼技術，把收銀自動化推向新高度
- 1968 中國人自行設計和建造的南京長江大橋建成，改變長江南北交通格局
- 1969 三位日本科學家在人臉識別領域取得突破進展，實現在複雜的圖形中識別出人臉
明斯基在名著《感知器》中否定多層神經網絡，使仿生派陷入 10 多年頹廢期
尼克遜當選美國總統，後因“水門事件”辭職，成為美國首位任期內辭職的總統
- 1973 金出武雄創出當時人臉識別正確率最好成績，成為人臉識別幾何時代的代表人物
- 1978 埃克曼創造一種科學可靠的方法分析人類面部表情，並組成“面部表情編碼系統”
美國出台《總統記錄法案》，認定總統錄音的所有權屬於國家

- 1979 美國開始通過公共事務有線電視網直播國會的會議和辯論實況
- 1981 第一個商業化鼠標誕生
習慣面對鏡頭的列根就職總統，在使用錄音時也使用錄像，把白宮記錄推上新高度
- 1982 香港一名出租車司機先後殺害 4 名女乘客，因缺乏記錄遲遲無法破案
- 1985 Windows 系統面世，此後每年都不斷地升級、打補丁
- 1986 美國“挑戰者號”航天飛機因一個小橡膠圈變形，而在空中爆炸
辛頓和魯梅哈特提出反向傳播算法，推動神經網絡在學術界的復甦
- 1989 加拿大阿爾伯塔大學年開發跳棋程序“Chinook”並在 1994 年戰勝跳棋冠軍汀斯雷
- 1991 美國取消對互聯網使用的限制，互聯網成為可公開訪問的商業網絡
- 1992 為方便機器識別車牌，中國出台新的車輛號牌標準
美國國防部高級研究計劃局等聯合發起 FERET 行動
- 1993 中國連入互聯網的第一根專線
- 1994 中關村地區教育與科研示范網絡實現互聯網全功能連接，中國互聯網開始商用
“專家系統之父”費根鮑姆因推動人工智能進入“邏輯推理 + 專家知識”新階段而獲圖靈獎
- 1995 中國媒體陸續開設新聞網站
- 1996 對外經貿部中國國際電子商務中心、國務院經濟信息化領導小組成立
廣東南海市成為中國第一個接入互聯網的縣級市
- 1996 Viisage Technology 公司等成為美國人臉識別商業化的早期公司
- 1997 中國提出互聯網建設與管理 16 字方針
美國新墨西哥等州的駕照管理部門利用人臉識別解決“一人多證”難題
- 1998 江澤民在 APEC 會議上首次闡明中國電子商務政策並指出它代表未來貿易方式的發展方向
中國出現首例“網購”

- 倫敦紐漢區在全球率先把人臉識別整合進天網監控系統
- 1999 國家各部委信息主管部門聯合啟動“政府上網工程”
- 2000 全國人大常委會通過《關於維護互聯網安全的決定》
中國移動互聯網投入運行
- 2001 國家信息化領導小組做出“中國建設信息化要政府先行”的決策
坦帕市成為美國第一個應用動態人臉識別系統的城市，因效果不佳被警方停用
“9·11”事件中攝像頭拍到兩名劫機者卻未報警，促使人們重新關注人臉識別技術
維基百科成立，現已成為全球最大的知識分享網站、最重要的百科全書
- 2002 十六大提出“以信息化帶動工業化，以工業化促進信息化”戰略，同時還首提
“政治文明”，開啟何為政治文明的討論
- 2005 中國的網民數量首破 1 億
倫敦發生公交系統爆炸，警方運用眼力戰 5 天後鎖定嫌疑人
- 2006 辛頓提出“深度網絡”概念
- 2007 北京在全國率先規定“在公共場所設置公共安全圖像信息系統，應當設置標識”
蘋果推出第一代智能手機 iPhone，智能手機開始普及
- 2008 奧巴馬建立個人競選網站，史上首次嘗試通過數據對選民分類，以推送定制
消息
- 2009 李飛飛等華裔學者發起建立圖像數據庫 ImageNet，次年依此舉行圖像識別
競賽
- 2010 美國警方用“綜合自動指紋識別系統”破獲 1991 年的聖貝納迪諾血案，獲 FBI
生物識別案件年度大獎
中國開始試行火車票實名制，到 2012 年所有客車都實行車票實名制
大學生馬躍在北京地鐵墜軌身亡，家屬調閱監控視頻的過程不暢
加州大學聖地牙哥分校研發 CERT，實現視頻流中的人臉自動檢測
- 2011 因定價算法出錯，亞馬遜的 *The Making of a Fly* 一書創 2369 萬美元“天價”
思科全球研發中心前總裁博諾米提出“霧計算”的框架和概念

- 2012 奧巴馬競選連任，通過個人競選網站和 Facebook 聯動
美國國會宣佈公眾可以通過網絡收看立法會議實況並下載會議的歷史錄像
陝西安監局原局長楊達才的照片遭“人肉”並牽出受賄等問題，被稱為“錶叔”
辛頓率隊參加 ImageNet 圖像識別大賽，深度學習引起人工智能領域關注
武漢警方成立全國首支視頻偵查支隊，視頻成為重要的破案線索
周克華在南京作案，當地調集大量警力肉眼查看監控視頻，引發硬盤和眼藥水同時脫銷，堪稱中國治安史上的傳奇
- 2013 十八屆三中全會提出“全面深化改革的總目標是完善和發展中國特色社會主義制度，推進國家治理體系和治理能力現代化”
最高法院開始建立“失信被執行人名單庫”，進行有效的社會信用治理
中國發起“精準扶貧”行動
波士頓馬拉松發生爆炸案，警方成立人眼小組調閱現場視頻，三天確認嫌犯
美國加州聖地牙哥警方開始配備便攜式人臉識別設備，隨時驗證行人身份
- 2014 杭州 8200 輛公交車安裝攝像頭，可自動拍攝違規車輛並把照片提交給交警部門
手機淘寶上線“拍立淘”功能，用戶可通過“以圖搜物”檢索商品
黑客“劫持”10 多萬台物聯網設備，發起 70 萬份垃圾郵件的攻擊，凸顯智能社會的軟肋
- 2015 國務院印發《促進大數據發展行動綱要》，倡導用數據說話、管理、決策和創新
“大眾創業、萬眾創新”被寫入《政府工作報告》，成為國家發展的重大推動力
美國首宗算法合謀案宣判，被告因“夥同他人合謀控制商品的網上銷售價格”而被罰
國家發改委提出，到 2020 年實現公共安全視頻監控“全域覆蓋、全網共享、全時可用、全程可控”
馬雲首次提出，大數據讓計劃和預判成可能，到 2030 年計劃經濟將成更優越的系統
杭州警方運用“物證雲”系統，破獲 13 年前的之江花園滅門血案並抓獲兇手

南京試水視頻開放，在“我的南京”App 推出“主幹道監控”和“實時路況”服務

- 2016 鐵路部門規定暫停向動車上抽煙的乘客售票，破解乘客無懼罰款而抽煙的難題
AlphaGo 以 4：1 戰勝世界圍棋冠軍李世石，成為人工智能崛起的標誌性事件
美團網 CEO 王興先生提出“互聯網已經進入了下半場”論斷

中央政法委推出以視頻監控聯網應用為重點的“雪亮工程”

浙江啟用高速公路不繫安全帶違法檢測系統，它能自動識別沒繫安全帶的圖片並推至交警執法系統

浙江啟動“最多跑一次”改革，因辦事效率提升顯著，全國各地紛紛效仿

美國女星卡戴珊因頻繁在社交網絡分享生活動態，在巴黎遭持槍搶劫

在總統大選最後角逐時期，《華盛頓郵報》披露特朗普勾引已婚女性的錄音

滴滴上線“分享行程”功能，通過分享行程數據保障乘車安全，被大量用戶使用
公安部規定，民警執法時要自覺接受群眾圍觀拍攝監督，習慣在“鏡頭”下執法
公安部等出台新規，規定電子數據可作為刑事案件證據

警方通過“Y-DNA 染色體檢驗”技術，破獲造成 11 名女性遇害的甘肅“白銀案”

南京玄武區在全國首創二維碼門牌，優化戶籍管理，該做法被其他省市效仿

美國西南航空因數據路由器的部分故障導致業務癱瘓，成千上萬旅客滯留

美國 Dyn 公司遭大規模“拒絕訪問服務”攻擊，導致多地斷網，半天損失上億美元

德國紐倫堡大學發佈“Face2Face”的應用，實現視頻中的“表情移植”

- 2017 中國網民數量達 7.72 億，其中手機網民 7.53 億，中國成為互聯網大國

中國《新一代人工智能發展規劃》指出，要把握機遇，引領世界人工智能發展新潮流

機器人索菲亞被沙特授予公民身份，成為史上首個獲公民身份的機器人

年度全球最大“黑天鵝事件”英國脫歐公投中，脫歐派以 52% 對 48% 的微弱優勢勝出

中國女留學生章瑩穎失蹤案中，美國警方通過監控視頻確認涉案車輛並鎖定嫌犯
 ImageNet 圖像識別最小錯誤率為 2.25%，深度學習算法已完全超越人類“眼力”
 Facebook 推出人臉識別應用“魔術照片”，準確率達 98%

蘋果手機配置人臉識別工具，用戶可用人臉解鎖屏幕和支付，可靠性高於 99%

中國烏鎮圍棋峰會上，AlphaGo 以 3：0 戰勝世界排名第一的中國冠軍柯潔

南京市玄武區推出“玄微警”，實現旅客快速認證、入住酒店等功能

江蘇宿遷啟用“行人闖紅燈人臉識別系統”，識別準確率超 90%

中國公安開始全面普及帶有人臉比對功能的移動警務終端

支付寶推出“車牌付”，通過車牌自動識別實現賬戶自動扣除過路費

馬化騰透露，每天數億張人臉照片被上傳至騰訊平台，這些照片數據被用於人臉識別

加州警方通過基因分析匹配技術，確認涉嫌 12 起兇殺案、45 起強姦案的“金州殺手”

準大學生徐玉玉被“精準詐騙”後死亡，引起全國對隱私泄露風險的重視

中國多地幼兒園被曝虐童，引發對公共視頻調取查閱權限的討論

華盛頓大學研究人員通過機器學習製作出仿真奧巴馬講話的視頻

中國網友在微博首次曝出被大數據“殺熟”的經歷，其後相似案例湧現

2018 “最多跑一次”被寫進李克強總理的政府工作報告

Facebook 爆數據泄露，8700 萬用戶數據被劍橋分析公司用於多國選舉中的選民分析

愛奇藝在美國納斯達克上市，市值超 150 億美元，視頻版權受到認可

支付寶發生“年度個人賬單”事件，渾水摸魚“竊取數據”做法受質疑

鄭州鐵警在全國鐵路系統中率先使用人臉比對警務眼鏡，為篩查不法分子帶來便利

浙江杭州第十一中學引進“智慧課堂行為管理系統”，輔助教師進行課堂管理

科大訊飛宣佈，其語音合成技術實現讓機器開口說話，且機器模仿語音可以假亂真