

大数据战争

人工智能时代不能不说的事

BIG DATA, AI
and PEOPLE


何渊——等著

数据是——商业竞争中先发制人的利器
数据是——政治时局中影响判断的工具
数据是——复刻完整的你的隐私追踪者

信息泄露、大数据“杀熟”、数据垄断、数据安全
谁在掌控财富，又是谁的游戏



北京大学出版社
PEKING UNIVERSITY PRESS

 天册合规研究丛书 01

大数据战争

人工智能时代不能不说的事

何渊——等著



北京大学出版社
PEKING UNIVERSITY PRESS

图书在版编目(CIP)数据

大数据战争:人工智能时代不能不说的事/何渊等著. —北京:
北京大学出版社, 2019. 6

ISBN 978-7-301-30367-2

I. ①大… II. ①何… III. ①数据处理 ②人工智能
IV. ①TP274 ②TP18

中国版本图书馆 CIP 数据核字(2019)第 033434 号

- 书 名** 大数据战争:人工智能时代不能不说的事
DASHUJU ZHANZHENG; RENONG ZHINENG SHIDAI
BUNENGBUSHUO DE SHI
- 著作责任者** 何 渊 等著
- 责任编辑** 朱梅全
- 标准书号** ISBN 978-7-301-30367-2
- 出版发行** 北京大学出版社
- 地 址** 北京市海淀区成府路 205 号 100871
- 网 址** <http://www.pup.cn> 新浪微博: @北京大学出版社
- 电子信箱** sdyy_2005@126.com
- 电 话** 邮购部 010-62752015 发行部 010-62750672
编辑部 021-62071998
- 印刷者** 三河市博文印刷有限公司
- 经 销 者** 新华书店
- 880 毫米×1230 毫米 A5 7.875 印张 151 千字
2019 年 6 月第 1 版 2019 年 6 月第 1 次印刷
- 定 价** 39.00 元

未经许可,不得以任何方式复制或抄袭本书之部分或全部内容。

版权所有,侵权必究

举报电话:010-62752024 电子信箱:fd@pup.pku.edu.cn

图书如有印装质量问题,请与出版部联系,电话:010-62756370



BIG DATA, AI
and PEOPLE >>> >>

名家推荐

大数据正在成为新经济的土壤。毫无疑问，数据是一种新的重要资源，很多公司甚至公开承认数据就是它们的资产。数据是资源又是资产，这就是个人与个人、公司与公司、国家与国家之间将会频发数据战争的根本原因。这是时代的话题，需要大众的关注、表达和行动！本书梳理了数据战争的多维细分战场，对如何达成数据空间的和平治理进行了思考，值得参考和阅读。

涂子沛

著名大数据专家

《大数据》《数据之巅》《数文明》作者

从古至今信息量随着文明程度的提高而快速增加，近年来，大数据让我们对世界事物的发展变化产生了新的认知。在信息量大到一定程度后，配合上前所未有的强大算力及机器学习算法，使从前根据低信息量推理出来的逻辑及思想模型出现破溃。世界充斥着海量数据分析、神经网络算法、自

主化智能等科技产品，人类在迫不得已下与机器进行互动或博弈，伦理问题逐步开始浮现。例如，当你知道自己的 DNA 的排列会更容易得血癌时，你会为此多买一些保险吗？如果保险公司知道了，保险公司有权拒绝受保吗？

下面我再分享一个更典型的例子，在 2018 年 3 月 18 日晚 10 点左右，一名市民在美国亚利桑那州坦佩市骑车横穿马路，被一辆自动驾驶汽车撞倒，不幸身亡。虽然车上有驾驶员，但当时汽车完全由自动驾驶系统（人工智能）控制。和其他涉及人与人工智能技术交互的事故一样，这起事故引发了一系列的道德和责任问题：开发该公司的公司在防止该系统出问题方面负有怎样的责任？坐在驾驶位置上的“司机”有责任吗？为自动驾驶而作的安全测试足够吗？现今的道路设计已经足够应付自动驾驶了吗？更有人提出，在意外发生的一刹那间，如果自动驾驶系统要作出一个决定，在驾驶员与横过马路的受害者之间选择减少伤害，抉择权应该在谁手上呢？一下子好像难以找到好的答案。

我知道国内有很多有心人，一直在为科技安全落地而努力。本书来得正合时，以何渊先生为首的一班有志者在科技伦理这个空白里填入更多的国内外相关案例，很好地让大家对未来的科技有更深认知。在下有幸作为大数据实践先行

者之一，为这本新书写推荐语是我的荣幸，也是为社会作些许贡献。

车品觉

香港科技园公司董事

阿里巴巴集团原数据委员会会长

我第一次见到何渊老师是在2017年秋上海交通大学凯原法学院，那天中午我专程过去与几位法学院老师讨论合规问题，我当时特别想了解法学家们是从哪些角度关注合规问题，合规在法学院教学课程设置中有哪些教学价值。

何老师在讨论结束后很谦虚地问了一些问题，随后就邀请我加入他主持的关于数据合规的微信群。我进去后吃了一惊：几百人的一个群，汇集了学者、执业律师、政府官员、法官等，他们都和数据工作相关。在群里，也有很多国企、民企及外企的公司法务同行。很快，我被这个群里的文章、信息和资料吸引住了，大家关注问题的深入性及宽广的国际视野让我大受启发。我自2016年兼任我就职的公司的数据隐私负责人以来，第一次大规模地直接接触到这么多高质量的数据法相关讨论和案例内容，受益良多。在随后与何老师多次定期接触和讨论中，我了解到他非常广泛、系统地研究

“数据合规”“数据治理”的立法及司法实践问题，他对中外相关法律及案例的钻研极有敏锐度和前瞻性。

《大数据战争》一书深刻、全面地阐述了人工智能时代最强有力的资产——数据对人类社会的影响，包括经济、政治、国家安全及道德伦理。本书中呈现的法律思想给我提供了独特的思考角度，帮助我重新认识大数据，同时教我紧密关注有影响力的案例和法律实践。感谢何老师为我和其他读者提供了许多有益的启示，希望本书能让更多人关注科技和科技法，并让人们进一步思考科技应该如何推动人类和社会文明的进步。

张曦

拜耳大中华区副总裁、总法律顾问

人工智能是人类社会未来 30 年发展的引擎。未来国与国之间的战争，主要争夺的是资产，尤其是数据资产，而不再是土地与石油。得数据者得天下。可以说，大数据资产是涉及国家安全的战略性资产。

为了推动大数据产业的健康发展，国家出台了一系列政策法规，以规范支持大数据产业的合法合规发展。大数据安全是“1”，大数据技术与各类大数据场景应用，是“1”后面

的无数个“0”，如果“1”没有了，那么一切也都归零了。所以说，大数据安全是整个大数据产业的基石。

何渊教授团队是业内屈指可数的既对法律法规有深刻理论造诣，又对产业发展有实战经验的理论与实践完美结合的团队。何教授团队倾力奉献的《大数据战争》一书，内容翔实丰富，案例剖析深刻生动，是难得一见的对读者朋友有巨大帮助的好书。

在20年IT从业、15年大数据从业的过程中，我目睹了行业中一些人因法律意识淡薄，利欲熏心地去非法缓存个人数据做“死库”，不按法律法规要求去做脱敏脱密加工数据产品，更有甚者直接倒卖个人信息数据，进而被定罪判刑。特此，强烈向所有大数据从业人员或对大数据感兴趣的朋友推荐本书，相信大家会从本书案例分析与政策解读中受益匪浅。

汤寒林

江苏省大数据交易和流通工程实验室主任

华东江苏大数据交易中心董事、总经理

世界正在悄然进行一场无硝烟的 大数据战争（代序）

在小说《流浪地球》中，刘慈欣提出了一个有趣的话题：未来，如果太阳的灾变将炸毁和吞没地球，人类的唯一选择只能是向外太空移民，那么应当如何移民呢？刘慈欣给出了两个方案。

第一个是飞船方案。即通过建造像上海或纽约那么大的飞船，带着人、植物种子及动物胚胎，像挪亚方舟一样移民到其他行星。飞船方案的好处是技术成熟、有灵活性及风险可控。但其最大的问题是生态系统的规模。离地球最近的有行星的恒星在 850 光年以外，需 17 万年时间到达，而飞船规模的生态系统根本维持不了。“人类在宇宙间离开了地球，就像婴儿在沙漠里离开了母亲。”

第二个是地球方案。即通过在地球上建造 1 万多座行星发动机，带上地球一起流浪，前往 4.3 光年以外的比邻星，整个移民过程将延续 2500 年，100 代人。地球方案相对靠谱，其最大优势是地球所具有的复杂和多样的生态系统，这是人类已知和未知的所有生存方案的唯一来源。但地球方案的最大问题是让地球脱离太阳系并漂移到比邻星，从能量的

角度是很难做到的，因为光速飞行会使得能量的消耗变得近乎无穷大。

于是有了第三个方案，即大数据方案。吴军在《科技史纲 60 讲》中提出，按照质量高低次序排列，分别是自然人、人类胚胎、人类基因以及人类基因信息等，而与质量相关的能量是决定移民成功与否的关键因素。从这个意义上讲，大数据方案也许最为可靠，即用超光速将纳米机器人送到 850 光年以外的行星，让它们为人类建设“殖民地”，然后再直接将人类基因信息而不是把人的肉身送过去。在新行星上，通过 3D 打印出人的 DNA 并复制出人的肉身。至于人的意识，由于其本质上是一种算法系统，在未来也可以数据的形式传输到新的行星，并通过脑机接口输入到复制人的肉身，这样一个完整的人就复原成功了。于是，人类可以逃离即将爆炸的太阳，在距离地球近千光年外的行星上继续繁衍生息。

这场景看起来相当科幻，但并非不可能成为现实。在不远的未来，特别是随着以 5G 和物联网为基础的人工智能时代的到来，数据的作用再怎么强调也不为过。英国著名杂志《经济学人》曾在一篇文章中指出，世界上最重要的资源不再是石油而是数据，但如同石油贸易的冲突几十年来让世界伤痕累累，数据经济的发展也将使得世界面临“数据战争”风险。这也正是本书以《大数据战争》为书名的缘由。可以说，一场没有硝烟的争夺大数据控制权的战争已经悄然在国家、企业、个人之间展开了。

2018年是中国数据合规的元年，这一年数据隐私和数据安全问题频发。年初支付宝年度账单默认勾选《芝麻服务协议》被质疑侵犯隐私权，百度涉嫌侵害消费者个人信息安全被江苏省消保委提起公益诉讼；下半年的大规模数据泄露事件引发了全民的安全担忧，包括华住酒店集团旗下连锁酒店近5亿条用户信息被泄露，12306网站470余万条用户数据在网络上被贩卖。据2018年8月中国消费者协会发布的《APP个人信息泄露情况调查报告》显示，遇到过个人信息泄露情况的人数占比为85.2%，没有遇到过的仅占14.8%。

与个人信息相关的犯罪也屡见不鲜，上市公司“数据堂”涉嫌侵犯公民个人信息罪被查，简历大数据公司“巧达科技”非法交易个人信息达数亿条，引起社会的广泛关注，也被有关部门查封。另外，算法涉及的伦理问题也开始显现，“同房不同价”等大数据“杀熟”让携程等企业备受争议，大数据让“价格歧视”具有了现实可能性；“精准营销”更是令人胆战心惊，明明只是在电商平台上搜索过某商品，新闻资讯类APP上却出现了相同商品的广告。

与此同时，这个世界也不安宁，欧盟和美国就数据保护的国家立法管辖权展开了“殊死争夺”，其本质是数据本地化和数据全球化的对抗。王志安在《交大法学》上发表的文章中指出，欧盟《通用数据保护条例》（GDPR）试图建立以个人数据权利保护为基础的法律规则高地，宽泛的规则域外效力和严格的数据跨境移转制度等体现了鲜明的数据本地化特

征。而美国则在通过“避风港规则”和“隐私盾协议”与欧盟加强合作的同时，又积极推动数据全球化规则，通过《澄清境外数据的合法使用法案》（即 CLOUD 法案）强调美国企业存储和披露全球数据的法律义务，以实现美国在全球的数据霸权主义。

另外一个趋势也特别值得注意，美国、欧盟、日本三方正在加速数据保护规则的融合，试图在数字经济领域用双方贸易规则取代 WTO 框架下的多边贸易规则，欧盟和日本之间已实现包括数据流通在内的自由贸易，欧盟和美国之间也有“隐私盾协议”，美国和日本之间就数字经济多次发布了部长级的联合声明。面对这种极其不利的国际形势，中国如何在数据规则方面迅速融入世界，甚至主导国际标准的制定，而不被美国、欧盟、日本排除在数据标准话语权之外，已成为一个不可回避的紧急任务。

正是在这种数据规则大变局的国际背景下，我们开始逐渐形成了要撰写一本《大数据战争》的强烈愿望，希望通过对国内外数据法典型案（事）例的整理、对国际数据治理经验的总结，为中国的数字立法提供理论储备，更为中国企业的数字合规提供实践指导。基于此，我们精心选取了最为重要的数字法案（事）例，其中既有经典案例，也有最新案例；我们还邀请了一线实务专家来撰写本书，他们中既有亲历案件的公司法务，也有深谙实务的专业律师，还有精通外语和把握趋势的学术专家。具体而言，本书从以下几个角度来设计框架：

一是个人与企业的关系角度。本书从朱烨诉百度案谈 cookie 隐私权，认为网络用户披着以技术为名的“皇帝的新衣”，一举一动都被企业注视着；本书从任甲玉诉百度案讨论“被遗忘权”或“删除权”，认为中国法律并没有规定“被遗忘权”，要想保护这类一般人格利益，必须证明其具有正当性和应予保护的必要性；本书还从庞理鹏诉东航、趣拿公司案讨论个人信息的安全问题，认为合理界定企业责任边界对数字经济的发展至关重要。

二是企业与企业的关系角度。首先是企业间的竞争关系。本书从新浪微博诉脉脉案谈数据竞争规则，认为第三方通过 Open API 获取用户信息时应遵循“用户授权”+“平台授权”+“用户授权”的三重授权原则；本书从大众点评网诉百度案看数据的爬虫规则，认为目前司法实践大都从反不正当竞争角度出发，考虑的更多的是数据抓取、使用行为是否正当，而数据的使用规则和权属划分有待进一步探索；本书从淘宝诉美景案讨论大数据产品的权益边界，认为淘宝“生意参谋”属于竞争法意义上的财产权益，构成了淘宝的竞争优势；本书还讨论了最近发生的“头腾大战”，认为微信/QQ 与抖音/多闪争夺的不仅仅是用户昵称及图片等数据，而且还包括这些数据后面的“关系链”。

其次是企业间的合作关系。本书复盘了 Facebook 的“数据门”事件，讨论了开放平台的失败和重生。本书认为，该事件改变了平台未来的商业模式和监管方式，也改变了我们

的情感、利益和观念。

三是个人与国家关系的角度。本书聚焦于美国法中的“合理隐私期待”原则，重点复盘了凯洛诉美国案（*Kyllo v. United States*），详细论述了用热像仪观察屋内是否种植大麻侵犯了公民隐私。美国联邦最高法院认为，如果没有对宪法保护区域的实际入侵，通过使用感官增强技术是无法获得任何关于公民住宅内部的信息的，这种获取信息的行为构成了法律上的“搜查”。

四是国家与国家关系的角度。本书从微软诉美国司法部案解读美国 CLOUD 法案，论述了美国在数据法领域的长臂管辖原则，并对中国企业提升海外合规风险能力提出了建议；本书还讨论了欧盟 GDPR 首个执法通知 AIQ 案，该案是依据 GDPR 的长臂管辖原则作出处罚的第一案，还涉及剑桥分析公司涉嫌通过提供选民数据操纵英国脱欧公投一事。

五是数据黑产的角度。本书从数据堂案件出发，强调应对数据黑产是看不见的战争，更是看得见的威胁。目前数据黑产有蔓延的趋势，在数据中间商的撮合之下，上达数据源头，下接数据需求者，形成了一个日益庞大的地下黑色数据产业链，有待于政府各机关进一步加强执法强度和形成执法合力。

可惜的是，中国个人信息保护的相关法律并不完备。虽然《网络安全法》《民法总则》等法律规定了侵权者的民事责任，《刑法》及相关司法解释规定了侵犯公民个人信息罪，强调了侵权者的刑事责任，作为推荐性国标的《信息安全技术

《个人信息安全规范》也对个人数据保护提供了执法的标准，但由于中国没有针对个人信息违法行为的行政处罚等前置程序，直接从民事责任过渡到刑事责任，难以形成执法合力，“要么没事，要么坐牢”的现状让很多人铤而走险。

值得庆幸的是，中国立法机关开始重视个人信息保护法律体系的建设，《个人信息保护法》和《数据安全法》已纳入十三届全国人大常委会立法规划的第一类项目；执法机关也开始频频出手，采取约谈、责令整改及罚款等方式打击各类与个人信息相关的违法行为。对企业而言，未来的数据合规能力涉及“生死存亡”问题，监管机关的“达摩克利斯之剑”随时都有可能落下，正所谓“数据合规，合则生，不合则亡”！

这里，我要感谢本书的所有作者，没有大家一起的努力，不可能有这部独一无二的原创作品。他们分别是张瑶、宁宣凤、吴涵、冯坚坚、袁立志、麻策、王磊、许可、尹云霞、周梦媛、黄琰童。谢谢他们！

另外，我要感谢天册律师事务所合规团队的同事们，数据合规的执业实践对我构思本书的框架起到至关重要的作用。他们分别是谢涛律师、傅林涌律师、刘森律师以及实习生胡慧雯同学等。谢谢他们！

我还要特别感谢“数据法盟”（DataLaws）这一平台上的所有参与者，正是通过“DataLaw 数据合规沙龙”“DataLaw 读书会”以及“DataLaw 公益翻译小组”等不同方式的深度参与和无私奉献，使得我们共同铸就了一个有关数据隐私和

网络安全的高端知识共同体，后续我们将通过公开出版系列丛书的形式总结世界各国的监管法律及政策、分享跨国公司和“独角兽”科技公司的数据合规经验，最终实现“以全球视野，为中国数据立法建言；以中国智慧，为国际数据规则献策”的目标。

最后，我要感谢北京大学出版社朱梅全编辑，没有他的催促和策划，这部作品很难有这么高的品质。

人工智能是人类自工业革命以来最大的变革，必将对人类社会产生结构性的根本影响，而大数据正是其中最重要的变量。身处这个剧烈变革时代的前夜，我想以狄更斯在《双城记》中的一段诗句作为结尾，与诸君共勉：

这是一个最好的时代，这是一个最坏的时代；
这是一个智慧的年代，这是一个愚蠢的年代；
这是一个信仰的时期，这是一个怀疑的时期；
这是一个光明的季节，这是一个黑暗的季节；
这是希望之春，这是失望之冬；
人们面前应有尽有，人们面前一无所有；
人们正踏上天堂之路，人们正走向地狱之门。

是为序。

何 渊

2019年3月28日

于上海交通大学法学院