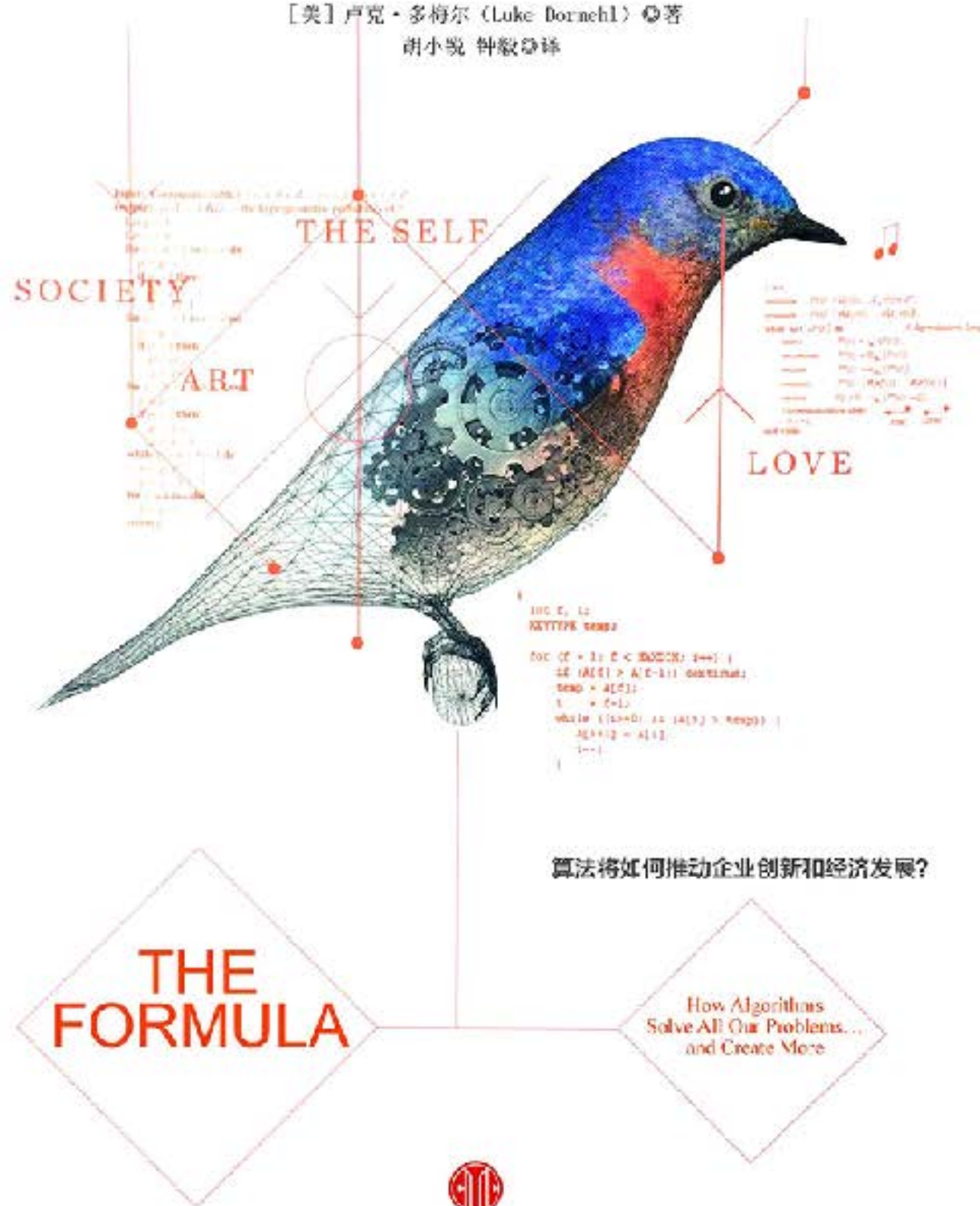


算法时代

新经济的新引擎

[美] 卢克·多梅尔 (Luke Dornehl) 著
胡小锐 钟毅 译



中信出版集团 · CHINA CITIC PRESS

版权信息

书名:算法时代

作者:卢克·多梅尔

译者:胡小锐 钟毅

ISBN:9787508657530

中信出版集团制作发行

版权所有·侵权必究

推荐序

在当前的大数据狂热中，自然科学家的声音正在显著压倒人文学者的声音。在大数据面前，人的一切都被预测决定，而自由意志失去栖身之地。在技术面前，人难道真的只有招架之功，没有还手之力了吗？人在12级台风中，需要一个保存自我的家园。《算法时代》将带你去寻找这一家园，这就是你需要阅读这本书的理由。

多梅尔是一位“数字人文”类型的作家。他眼中的技术明显有别于工程师眼中的技术，他更多地是从人的角度去思考技术。《算法时代》最关注的问题是算法技术对人的影响。

《算法时代》的主题词是“公式”（The Formula）。作者认为，公式是技术理性的象征，“公式决定人”的技术决定论“反映了围

绕客观性建立起来的某种社会秩序”。与这种客观性相对的，则是自我、爱、正义、创造性这些“主观性”的东西，围绕它们建立的社会秩序也应是主体性的。以此推论，“公式”的反义词应是自由意志，认同、爱、正义和创造力可以被看作自由意志的4个分身。

多梅尔在《算法时代》中，分别用4章的篇幅一一探讨了“客观性”的公式与“主观性”的认同、爱、正义和创造力是否合拍的问题。

第1章讨论公式与自我认同的关系。针对大数据根据过往记录将人们归入不同细分类别，作者援引德勒兹与瓜塔里的《千高原》一书，指出公式把人变成了“可分解动物”；而人本应具有多种自我，“多种主观性并存”，“以错综复杂、粒度极高而且常常十分隐晦的方式相互影响”。

第2章讨论公式与爱的关系。针对图灵测试，作者引用珀森的话说，“在‘浪漫’的眼

中，理性主义爱情观是没有全心全意投入的爱情”。浪漫主义者往往认为理性主义者的感情非常肤浅、受到抑制，而且他们害怕产生激情，小心翼翼地把自己囚禁起来，认为“再先进的算法也无法复制真爱”。

第3章讨论公式与正义的关系。法律主持正义，算法可以主持正义吗？作者认为，算法会倾向于死板地遵循法律条文，不知变通，而人更倾向于体现“法律意图”。

第4章讨论公式与创造力的关系。对艺术创作，作者不否认像《纸牌屋》那样利用数据分析可以推出热门产品，但提出的疑问是：“艺术品是否有统一的衡量标准”。

决定论与自由意志（非决定论）之争，是人类争论的永恒主题。纵观全书，《算法时代》是在大数据背景下重新讨论或者说升级了这一话题。

这不是老调重弹，而是透着时代精神。

首先，算法问题是时代的问题。如果说在原子时代对人起决定作用的是生产方式，那么在比特时代对人起决定作用的则变成了思维方式。因此，在信息时代讨论决定论与非决定论，当然应把重心放在思维方式上。算法是思维方式的技术基础，有什么样的技术基础，就有什么样的思维方式。这是在讨论算法问题的现实价值。

其次，算法问题也是人生的问题。生产方式与思维方式有一个共同之处，那就是“通过可以利用的资源实现价值的整套方法”。利用资源是手段，实现价值是目的。《算法时代》讨论算法的方式不同于工程师，是把它当作一个人生问题来看待。

当作者从认同、爱、正义和创造力这4个非决定论的方面质疑“公式”时，把算法的内在结构置于自由意志的对立面，即技术理性。算法真的只能属于技术理性吗？这可不

一定。

图灵当初提出算法问题，是从机器与人两个方面立论的，并非只谈机器和决定论。其实我们可以反向设问，如果实现价值需要的话，为什么不能调整算法，从而建立新的思维方式——比如互联网思维呢？

至少有三个视角，可以帮我们延展思考《算法时代》关于算法的内在结构的问题。

一是“人-机”算法与“人-人”算法的视角。谷歌算法是典型的人-机算法，属于迷信数学决定论的方法。也就是多梅尔所说的技术理性算法。10年前，我有一次同中国科学院搞软件的数学家讨论谷歌算法时，数学家们提出一个悖论：你说脸谱网是什么算法？脸谱网的算法从数学角度不值一提，但计算结果不比谷歌差。道理很简单，这是一个脑筋急转弯。谷歌的模式是人问机器，算法只在机器一端。而全世界的计算机加在一起，也

达不到一个人脑的潜在计算力，为什么不能让人问人？这样的算法显然就不是技术理性了，而是“自由意志-自由意志”之间的P2P计算。这是价值观反作用于算法的例子。

二是全局最优算法与情境最优算法的视角。傻瓜在什么情况下会战胜聪明人（臭皮匠在什么情况下会战胜诸葛亮）？这又是一个脑筋急转弯。佩奇实验实际上把技术理性的算法颠覆了。答案应是，当人算不如天算时，傻瓜会战胜聪明人，因为傻瓜具有解决方案的多样性。从数学规划角度讲，当全局有唯一最优解时，诸葛亮往往胜于臭皮匠；但如果最优解只在节点，即情境最优时，臭皮匠往往又会胜于诸葛亮。这是价值观反作用于算法的又一个例子。

三是“深刻”算法与“肤浅”算法的视角。对技术理性而言，算法肯定是深刻好，而肤浅不好；能算到本质，肯定比只算现象要强。

尼古拉斯·卡尔发现，互联网可以让人们的思维变得“浅薄”，但这不一定是坏事。自胡塞尔提出现象学之后，注重当下、此在的价值，就成了一种新的思维方式。顾客来到小摊，首先想要的问题是，面前这个小贩是不是个骗子？深刻的教授未必能回答这个问题，拥有大量咨询模板的麦肯锡也没法回答这样的问题。最适合回答这个问题的人，恐怕是头一天上了一回当的保姆。那么如果有种算法（比如小世界网络算法）能在一秒内找到保姆，这样的算法可能并不比“深刻”的算法差。

读《算法时代》，倒不一定非要得出什么确切的结论。从中获得启发，在某种众口一词的一般见识中看出不同的道理，就很值得。

姜奇平

中国社科院信息化研究中心秘书长，《互

联网周刊》主编

引言 算法的影响力

从根本上看，算法比一系列的步进指令略显复杂，通常由计算机执行。不过，尽管算法描述起来非常简单，它们的内部运行机制以及对我们生活的影响却绝不简单。

每天，算法都会对展示在我们眼前的信息进行分类、筛选与取舍。我们看到的谷歌搜索结果、脸谱网（Facebook）上高亮显示的好友信息，以及亚马逊网站预测我们很可能购买的产品类型等，都是算法运行得出的结果。此外，电影、音乐等娱乐方式的发展，通过预测浪漫关系的结果来选择伴侣，甚至法律的执行与警察办案等，也将越来越依赖于算法。算法在扫描我们的元数据之后，可能会评价我们有希望成为一名勤劳的员工，也可能指控我们犯有某种罪行，或者认为我们不适合驾驶汽车。通过这些过程，算法正

在深刻地改变我们看待生活、看待宇宙乃至看待一切事物的方式（引自道格拉斯·亚当斯）。

在人们针对技术发表的各种评论中，我最喜欢的是：“船只的发明同时带来了海难。”人们通常认为这是文化理论家保罗·维利里奥的观点。当然，我们也可以把这句话反过来，说海难的始作俑者就是船的发明者。在海难悲剧中，算法难辞其咎，但算法同时也发挥了极其重要的作用，帮助我们在每天产生的多达2.5艾字节（是人脑信息储存量的100万倍）的数据海洋中航行，并得出切实可行的结论。

这里，我将采用剥茧抽丝的方式，从本书的主要内容谈起。首先，这本书不宜用作计算机科学的算法教程。适合用作算法教程的书有很多，其教学效果也远胜这本（而且，很多作者比我更有资格编写这样的教程）。

其次，本书讨论的也不是算法概念的历史。我曾经也想做这种尝试，但是考虑到工作量太大，而且最终的成书（毫无疑问，如果由合适的人选来完成它，必将深受欢迎）与我敬而远之的教材并没有多大区别，因此我放弃了这个想法。我并不是说介绍算法历史的书与算法教程必然没有区别，而是指对算法（一度是小众数学概念）历史感兴趣的人，可能只限于那些已经熟悉这个概念的数学家或者计算机科学家。

算法通过各种各样的方式（有的非常微妙，有的则比较明显），影响着我们的方方面面，包括我们喜爱的娱乐活动与我们的爱情观。因此，在本书中，我希望揭示算法的这些影响，还要告诉大家以下事情的共通之处，比如，如何成功地吸引性感迷人的约会对象，好莱坞失败之作的产生过程，将自己的排泄物装袋处理，减少律师从我们身

上赚钱的机会等等。也就是说，在本书里，我要告诉大家的是算法对我们眼下生活的处理方式。

我白天的工作是为《快公司》杂志写作“数字人文”领域的文章。上班时我常常思考，“算法”文化以及认为所有问题都可以选择适当的算法予以解决的观点（未必都不正确），可能会带来哪些后果。

比尔·唐瑟尔于2009年出版的《在线为王》一书中有一个典型的例子，可以帮助大家理解本书的意图。唐瑟尔aaa至少有一位作者称他是“世界上最卓越的网上（行为）专家”bbb在书的开头介绍了他某天驾车时收听到的访谈。接受采访的是一位英国心理学家，他说自己研发了一种数学公式，可以找出一年中最令他沮丧的是哪一个星期。经过反复研究，他发现1月的第三个星期最令他感到沮丧，因为他同时遭遇了三个问题：新

年决心未能实现，放假期间信用卡债暴增，没完没了的阴沉天气。唐瑟尔称他对此有所怀疑，但这并不是因为他认为一个简单的公式不可能为沮丧这种异常复杂、涉及多种因素的问题提供正确答案，而是因为他相信自己可以设计出一个更好的公式。

也就是说，他怀疑的不是最终结果，而是这个算法的工作原理。

多年来，我经常听到类似的说法，即人们认为技术可以将所有问题化繁为简，直至最适合用公式计算的程度，然后给出问题的客观答案。受他们的启发，我萌生了创作本书的念头。也正因为如此，我用大写来表示“公式”（The Formula），而不是把它简单地看成本书所描述的所有技术的笼统表达。对于算法及其相关技术，我是从更广义的角度来理解的，认为它们可以体现某种形式的技术理性，可以反映围绕客观性建立起来的某种

社会秩序的特点，因此，“公式”这个概念暗含着某种意识形态的成分。这与已故美国政治学者、传播理论家哈罗德·拉斯韦尔对“技艺”（technique）一词的使用颇有相似之处。拉斯韦尔认为，所谓技艺，就是“通过可以利用的资源实现价值的整套方法”，而本书中的“公式”既表示一种应用及其应用范围，还表示潜藏在表面之下、有待我们利用合适的数据挖掘工具去梳理发掘的客观事实。

技术作家（除了几位著名作家以外）的世界观往往包含了强烈的乌托邦思想，认为所有的技术进步都是积极的。于是，在新的发明问世之后，这些技术作家往往会欢呼雀跃，认为这项发明标志着一个新“时代”的开始。因此，人们以一种蔑视的口吻，称之为“网络（技术）万能论”。尽管本书也可以随大溜，为书架上长篇累牍的关于算法与大数据的著作锦上添花，增添一些新内容，但我

真正感兴趣的不是讨论互联网或者个人电脑时代的诞生过程，而是希望做进一步研究。

20世纪60年代前半叶，法国社会学家（同时也是基督教无政府主义者）雅克·埃吕尔在他的著作中描述了所谓的“技术人”，认为他们“对统一标准付诸实施之后所带来的即时结果感到心驰神往……孜孜不倦地寻找‘最佳方式’以实现任何指定目标”。他们希望实现的目标有时会变得含糊不清，但这不是因为他们对伦理道德的关注难以测量，而是因为他们对目标实现方法的热情过度，以为人类心灵手巧，针对任何问题都可以凭空想象出“效果惊人”的解决办法。

埃吕尔的研究表明，持有这种观点的人早已有之，谷歌创始人及本书讨论的那些高科技公司的领导者也不是“求序意志”

（will-to-order）的先驱。这个概念是由已故美国社会学家刘易斯·芒福德提出的，它的意

思是，人们希望借助公式了解身边世界。20世纪30年代，早在现代计算机问世之前芒福德就在著作中指出，自动化的另一个目的是“放大人体的机械或者感官能力，将人类生命的各个阶段简化为可以测量的秩序与规律”。他认为，我们对人类生命进行简化处理以便了解其全貌；为了让人类智力等抽象概念便于量化分析，我们又对这些概念做进一步提炼，去除各种复杂因素，再赋予看似任意的某个数值，最后得到某个人的智商值。

但是，这些著作还没有触及上述观点在现代社会中的影响范围。目前，算法与“公式”的影响无处不在，几乎涉及工作与休闲领域的方方面面。因此，本书试图探讨这种局面的形成原因，研究算法在人类创造力、人际关系（更具体地说，是人们的浪漫爱情）、身份概念与法律问题等方面的影响力。

在这些方面，算法都可以为我们提供答

案。

但问题在于，算法提供的答案真的是我们想要的吗？

第1章 算法程序与人类的自我认知

2000年，51岁的拉里·斯马搬到了加利福尼亚的拉霍亚市。此时，他的体重已经接近15英石^①了。从他驾驶本上的照片看，斯马的肌肉已经松弛，脸圆嘟嘟的，双下巴，脖子上还有一圈儿肥肉，这些都表明他明显超重了。斯马是一位杰出的物理学家，也是超级计算方面的专家，因此他经常动脑。不过，这么多年来，他几乎没有参加过体育锻炼。而且，他非常喜欢喝可乐，经常吃油炸甜饼。这次搬家是因为他受加利福尼亚大学的委派，到刚成立的加利福尼亚通信与信息技术研究所任负责人。可是，刚搬到西海岸，他就觉得这里的生活与他格格不入，这种久违的陌生感让他无所适从。“放眼一看，

周围的人个个都身材匀称，体魄健壮。”他回忆说，“他们要么经常跑步，要么经常骑车，生活得有滋有味。我发现自己与他们的差别太大了。”

面对这种情况，斯马不禁想到科学家与哲学家经常会问到的一个问题：为什么？他跑到当地的书店，把饮食调理方面的书都浏览了一遍。但是，斯马说：“这些书太不严谨了。”在面对超新星和复杂星体构成等错综复杂的问题时，斯马这类人可以做到得心应手。但是，令人无法想象的是，在面对健康饮食与定期锻炼这样的问题时，他们竟然也会采取解决天体物理学复杂问题时使用的那些方法。正因为如此，他一直都没有找到一本中意的书，直到他发现了生物学家巴里·西尔斯写作的《食无禁忌》。在这本书中，西尔斯把人体看成非线性耦合系统，而葡萄糖胰岛素等反馈机制与免疫系统会在其中发生

相互作用。对于这样的说法，斯马非常熟悉，也可以接受，因此，他觉得这就是他梦寐以求的宝典。在这本书的激励下，他每天都会脱光衣服称体重，并把数据记录下来。他还请了一位私人教练指导他锻炼身体，并记录下每天的锻炼时间。除此以外，他也调整了饮食结构，把每天吃的东西分类成蛋白质、脂肪、碳水化合物、纤维素、糖、盐等“基本单位”，并剔除有害身体健康的食物。“我把自己看成一名工程师，正在完成汽车子系统的逆向工程。”斯马说，“逆向工程告诉我们，要让汽车跑起来，需要加入一定量的汽油。如果在油箱里加水，就会把车弄坏。我们从来没有采用这种方式去思考我们的身体情况，实际上这才是我们应该采用的正确方式。”

随后不久，斯马又开始借助一些复杂的技术来减肥。他买了博能心率监测带以及

FitBit、BodyMedia等可随身佩戴的技术产品，这些产品可以通过各种算法，把身体健康状况转变成各种数据。为了监控身体的健康状况，斯马经常去一家收费的私人实验室验血，以便有更多的数据用于分析研究，还通过联邦快递把排泄物送去化验。斯马解释说，之所以采取这些措施，是因为“我在学校没学过生物学和医学，所以只能自学”。

在他收集的数据中，有一个数字渐渐引起了斯马的关注。这个数字与一些复杂的反应蛋白有关，可以直接反映体内炎症的情况。在身体正常时，该数字应该小于1，但是，斯马的这个数字是5，还慢慢地上升到10，甚至达到15。作为一名科学家，斯马觉得这种情况有违常理。“在我的饮食结构中，能导致发炎的所有食物都已经减少了，这个数字却仍在不断攀升，反映我体内有慢性炎症。这是怎么一回事呢？”斯马百思不得其解，

“这没有道理啊。”

于是，斯马决定去看医生，告诉医生他的这些发现。结果，医生的诊断过程与他想象的完全不一样。

“你有什么症状吗？”医生问他。

“没有。”斯马回答说，“我没有什么异样的感觉。”

“那你来找我干什么？”

“你看，这些是我的个人身体健康状况数据图。”

“这些图有什么用啊？”医生说。

随后，这位医生告诉斯马，这些数据太“学术”了，在临床实践中没有任何意义。“等你的身体真的出现不适时再来找我吧，如果只是这些图出现了什么异常状况，就不必来了。”医生说。

几周之后，斯马的左腹部产生了剧烈的疼

痛感，于是他又去找那位医生。结果，他被诊断出患有憩室炎，这是一种由急性炎症引发的疾病。这件事充分说明斯马遇到的难题：医生只愿意处理临床症状，而不愿意下功夫钻研这些可能具有预防作用的数据。在这件事之后，斯马吸取了教训，决定要时刻关注自己的健康状况。

“人们都被强行灌输了一个理念：他们无须为自己的身体状况负责。”斯马说，“人们平均每年花在看医生这件事情上的时间是20分钟。我计算了它与一年的总时长之间的比值，答案是几万分之一。你每年有那么多的时间去照顾自己，而你却认为有人只需要花几万分之一的时间，就可以诊断出你身体所患的疾病，并且把它们治好。我觉得这样的理念实在太疯狂了，没有一点儿道理可言。”

在这次诊疗之后，斯马仿佛着了魔，把他注意到的所有症状都记录在案，然后努力寻

找每一个症状与身体数据波动之间的关系。他还收集了很多的数据，利用数据挖掘算法进行筛选，找出其中不规则的数据。他关注的另一个数字与乳铁蛋白有关。乳铁蛋白是白细胞在进攻模式下分泌的一种抗菌因子，有点儿像催泪弹在人群中排放的催泪气体。他关注的这个数字在正常情况下应该小于7，而斯马的第一次测试结果是200，到2011年5月，这个数字更升至900。他查阅了科学文献之后，初步诊断自己患有一种慢性自身免疫失调，之后进一步诊断为克罗恩病。斯马说：“我单凭这些生物学指标就做出了正确的诊断。”

从这个例子可以看出，斯马就是一个生命算法化的典型代表。他长期记录150种变量的数据，每天步行7000步，存储了上百万个跟身体健康状况有关的独立数据。这些数据可以反映出他身体内部正在发生的变化。他

还深入地研究身体的各个部位与日常生活的各个方面，并利用数学方法对其加以规范，以便严格匹配体内的这些变化。斯马在他的论文《迈向数字化的基因组医学：量化身体状况的十年探索故事》中指出：“随着‘读取’身体主要子系统运行状况的技术能力不断提高，记录关键生物学指标的变化情况将会变成一种常规活动。这样一来，当某些数据出现异常时，我们就能及时发现疾病发生的先兆信号。”

目前，斯马正在利用加利福尼亚大学的超级计算机，创建由10亿台处理器组成的分布式全球计算机。斯马称，在10年之内，科学家就可以利用全球计算机建立可行的人体算法模型。他认为，人体是一个可以改进、修正的复杂系统，全球计算机完全可以胜任这项任务。

2012年，记者马克·鲍登在《大西洋月

刊》上撰文指出：“过去，思想家用蒸汽机做各种类比，比如，容器在压力的作用下会爆炸（想想马克思的革命观与弗洛伊德关于压抑欲望的观点）。而如今，我们更喜欢用电子技术来构建隐喻。”

谈到象征性表达，很多人（包括斯马）的选择不是隐喻，而是公式。

1.1英石 \approx 6.3503公斤。——编者注

通过数据认识自我

拉里·斯马的情况比较特别，其研究对象为独立的个体（即“ $n=1$ ”）。每个人不可能都是超级计算方面的专家，而且，斯马收集了大量个人数据，并把这些数据的意义搞得一清二楚，不是每个人都有能力完成这项任务的，也不是每个人都有这种经济实力的（斯马每年在养生方面的投入为5000美元至10000美元）。

但是，像斯马这样的人不在少数。作为一个数据爱好者，他在“量化自我运动”组织中备受推崇。该组织是《连线》杂志的编辑加里·沃尔夫和凯文·凯利于21世纪初创建的，他们借用一个哲学术语向“信徒们”做了一个大胆的承诺：“借助数字实现自我认知。”随后，一些热衷于这项运动的志同道合者蜂拥而至，该组织的规模不断扩大。它奉行查证

与经验主义的实证观，并大量借鉴技术决定论，向人们提出了一个存在主义的问题：是否存在某种自我，即便采用合适的算法，我们也无法借助数字对其进行量化处理呢？

如果说苏格拉底认为未经检验的生命没有任何意义，那么，21世纪的新观点则有可能认为没有经过量化的生命同样没有意义。笛卡儿有句名言：“我思故我在。”同样，我们也可以说：“我可以被量化，因此我存在。”

从一定意义上看，“自我量化者”把笛卡儿的自我观提升到了更高的“粒度”（granular level）。笛卡儿认为意识不是有形的，因此无法像肉体那样可分成若干部分。但是，这些关注个人健康的人则认为，只要运用合适的技术，采集合适的数据，就可以有效地描述一个人的身体状况。人们有可能把消耗的食物或者周围空气的质量作为输入数据，然后借助情绪、兴奋度、血氧含量以及身心健

康数据来量化身体状况。

性爱量化程序“Bedpost”（本书后面将介绍该程序）的设计者凯文·康鲍伊说：“我希望有人在某个地方收集这些数据。也许我的观点有些超前，但我希望这些数字就存在于某个地方，等我死了之后，人们有机会看到这些数字。如果计算机编码能让我们深刻了解自己的生命，那就太棒了。”

我们假设“迈克尔”是信奉量化自我理念的典型代表（如果真的有这样的人）。每天晚上睡觉前，迈克尔都会在头上戴一个传感器。他睡得很早，因为传感器告诉他早睡能让他睡得更沉，休息效果最好。迈克尔早上一醒来就会查看传感器的数据，了解自己的睡眠质量。他起床后会做一会儿俯卧撑和冥想，接着打开电脑，运行一个叫作“750个词”的写作程序，把头脑中想到的前750个词记录下来。等他写完，文本分析算法会依据他

输入的词，给出一些统计数据，反映他的情绪、心态和他正在关注的事项。有时候，连迈克尔自己都没有意识到他的内心深处正牵挂着某些事项，而统计数据却能明确无误地发现它们。做完这些事之后，迈克尔便准备好开始当天的生活（当然，他还会戴上FitBit计步器）。他说，如果不完成这些步骤，“一整天我都会感到不舒服”。

量化的世界

量化自我其实就是要“破解人体的秘密”，它依赖先进的技术、社交网络以及通过监管实现的自由。这项运动似乎是现代生活的典型产物，在很大程度上看似起源于“9·11”恐怖袭击事件之后的美国，但其实可以追溯至多年以前。20世纪80年代，一大批运动员借助Nautilus动感单车、Stairmaster台阶器等高科技健身器材锻炼出完美的体魄。同时期问世的还有索尼“随身听”（谷歌眼镜的雏形），这项便携式技术把公共空间转变成可以控制的私密体验。20世纪90年代，在上述技术设备的基础之上应运而生的麻省理工学院“可穿戴计算机小组”（Wearable Computing Group），对“个人计算机”这个术语提出了异议。他们认为，在当时的情况下使用这个表达还为时过早，因为：

个人计算机应该像眼镜与衣服一样可以穿戴在人身上，根据使用者所处的环境与使用者进行互动。可穿戴计算机应该配备有平视显示装置、方便的输入装置、个人无线局域网及大量其他的环境探测与通信工具，借助记忆代理、增强现实或者智慧集成等技术手段，发挥智能助手的作用。

目前，人们在量化自我时似乎不受限制。这项运动的美感（如果我们可以用美学术语来描述它），就在于它的适用范围非常广。通过量化自我，人们可以非常精准地回答各种或宽泛或具体的问题，诸如摄入一个单位的酒精需要多睡多少分钟的觉才能恢复体能，打高尔夫球时挥杆动作是否连贯标准，是否应该放弃当前的工作等。我们以量化自我运动组织的一位年轻女性成员（假设她的名字叫作“安杰拉”）为例。

安杰拉认为自己目前从事的正是她梦寐以

求的工作。她下载了一个应用程序，这个程序每天都会“骚扰”她几次，请她测量当时的情绪状态。随着数据越来越多，安杰拉发现，她的“情绪得分”表明她在工作时并不那么愉快。于是，她辞职了。

43岁的自我量化者文森特·迪恩·博伊斯说：“我发现参与自我量化运动的人有一个共同点：他们都面临着一些只有借助数据才能回答的问题。这些问题在刚开始的时候可能非常简单，但是很快就会变得非常复杂。人们可能希望知道自己跑了多少英里^①，借助技术手段，这个目标很容易实现。但是，除了自己已经完成的英里数以外，人们可能还希望了解更多信息，例如，还有多少人完成了跟自己同样多的英里数呢？这时候，人们就需要使用数据以及算法了。人们在探求知识的过程中，不仅希望进一步了解自己，还希望更加深入地了解我们所生活的这个世

界。”

一直以来，博伊斯都非常热衷于定量研究。他是纽约大学交互式电信专业的一名学生。有一次，他在发射一个火箭模型之前，在模型上连接了一些传感器、微型控制器和一个加速表。他说：“我制作的这些部件功能齐全，可以在火箭模型发射出去之后收集信息，并进行数据回收和分析。我觉得这件事太有意思了。”在分析了火箭模型收集到的数据之后，博伊斯有了一些发现。他一直痴迷于滑板与冲浪运动，因此，他决定在他的滑板与冲浪板上连接类似的传感器，以便量化滑滑板与冲浪时的动作。他把运动环境细分成数百个可以量化的变量，其中包括天气、时间以及（冲浪时）浪潮的变化情况与波浪高度等。博伊斯认为：“从自我量化的角度看，我可以在地理空间的框架下考虑冲浪环境、冲浪时所使用的器具、周围的条件……

（这些数据）代表着我在时空中完成的一系列动作。”

通过这种方式，自我量化者找回了美国先驱们粗犷朴实的个人形象，如今，这一形象通常体现在硅谷实业家及其拥趸身上。人们跟踪自己的数据，其实也是一种独立自主、自给自足的行为，与自行解决衣食住行的做法并没有什么不同。在当今社会，人们收集、分享的用户数据与个人信息，其数量远超从前，因此，自我数据跟踪的做法反映人们正在尝试重新夺回某种控制权。与谷歌地图一样，这种技术手段把个人重新置于宇宙的中心位置。在说到自我量化时，博伊斯认为：“我相信，总有一天这种做法会成为一种惯例，成为一种依靠特有的社交货币流通的商品。”

1.1英里 \approx 1.609千米。——编者注

购物也是创造

作家道格拉斯·库普兰在他的两部小说《微软奴隶》与《J氏游戏设计师》中，毫不留情地对世纪之交时的技术乐观主义进行了冷嘲热讽。而在他于计算机网络诞生之际完成的处女作《X一代》中，有一章的题目是“购物不是创造”。这句话非常简练，凝聚了作者对20世纪60年代行为主义蜕变成20世纪90年代商业主义这一现象的深刻理解。但是，这句话的正确性也同样发生了变化，因为只要人们在网上（或者在超市中使用会员积分卡）购物，他们的身份就会发生小小的变化。在他们不知情的情况下，商家将为他们创建档案并加以管理。

当然，这种情况并不仅限于购物。只要我们新打开一个浏览器窗口开始上网，相同的情况就会发生：远处某个地方的数据库会识

别并记录我们的这些活动。我们的互联网协议（IP）地址被记录在案，辨别用户身份的“cookies”数据被存储到我们的电脑中，使我们很容易被个性化的商品广告与商品信息更高效地“锁定”。罗伯特·沃克在希区柯克导演的电影《火车怪客》中扮演的那个凶手，无论看到什么，都会想起他杀死的那名妇女。同样，如果我们经常搜索某类体育活动的相关新闻，那么，无论我们浏览什么网页，都会发现与该类体育活动相关的广告。例如，如果你在电子邮件中提到“开普敦”，无数“南非廉价机票”的信息就会蜂拥而至。

1890年，美国哲学家、心理学家威廉·詹姆斯在他的《心理学原理》第一卷中指出：“人的自我是其属性的总和，不仅包括他的身体与精神力量，还包括他的衣着、房屋、妻子儿女、祖先、朋友、名声、作品、土地、游艇以及银行账户。”在算法与“公式”时代，

自我的含义有增无减。基于一个人所处的位置、他访问以及浏览的网站、搜索时使用的关键词，人们就可以对他的性别、种族、社会阶层、兴趣爱好以及可支配收入等信息完成定期的统计推理。访问帕丽斯·希尔顿的博客与访问科技博客Gizmodo，所包含的信息大不相同，而购买飞机票与购买电脑游戏的意义也大相径庭。无论我们出于何种意图与目的而将这些行为加以综合考虑，其结果本身都是通过算法得出的一种自我，即在全数字（因此可量化）层面上的一种身份识别。

用户的喜好就是我们的商机

在算法时代，身份识别蕴藏无限商机。跨网站与服务器追踪用户的能力推动大量网络数据分析公司蓬勃发展，这些公司不仅孜孜不倦地收集大量个人信息，还运用专业算法分析这些数据所表示的意义。

总部设在旧金山市区（此外，它在纽约、都柏林、伦敦、底特律、亚特兰大、芝加哥和洛杉矶也有办公室）的Quantcast是该领域最大的公司之一，在受众数据研究方面全球排名前五位。自2006年成立以来，该公司已经吸纳了超过5320万美元的风险投资。Quantcast的核心业务是创建用于描述特定用户的公式，然后就投资策略给出咨询建议。公司的联合创始人康拉德·菲尔德曼说：“人们采用‘广告对象假设’（human hypothesis of advertising）这个方法，对某

个产品的理想受众的特点及定位方法进行推断。现在，我们摒弃了这种方法，转而分析广告促销活动发挥作用的真正原因。通过定量分析大量数据，去找出受众的特点。”

菲尔德曼是英国人，毕业于伦敦大学学院。在创建Quantcast之前，他还创建了另外一家公司，利用算法帮助世界上的一些大银行监测洗钱行为。他说：“每个月我们都要仔细分析这些银行完成的无数笔交易，从中找出可疑行为。”在协助调查金融诈骗案件的过程中，菲尔德曼意识到算法在海量数据的分类方面具有超强的能力，能够从中找出一些规律，为我们的后续行动提供帮助。“算法可以描绘人们感兴趣的所有事物。金融方面的数据虽然有研究价值，但这些数据仅与人们的投资对象有关。而互联网数据不同，它可以凭借单一的数据形式，表现微观与宏观两个层面上的利益与变化趋势。”说到这里，

菲尔德曼有点儿激动，显然他为这项业务的前景感到欢欣鼓舞。

说到广告业，菲尔德曼指出：“广告业一直以来采用的都是对总体数据的事后测量，因为人们以为广告业应该关注受众的总体特点。”菲尔德曼刚到美国时，每看一个小时的电视节目，就要经受长达20分钟的商业广告的狂轰滥炸。广告商们采用的是机关枪模式，就像对着河水扫射，希望碰巧能打中一两条鱼。菲尔德曼解释说，电视频道当然不可能针对每一名观众的需求做出调整。但是，互联网的情况不同。Quantcast可以根据运算结果帮助网店店主针对每名新客户整饬网店首页，这种做法与亚马逊网的个性化推荐非常相似，可以根据人口统计学数据、消费心态、兴趣、生活方式等粒度把受众分成若干类别。Quantcast在宣传材料中夸下海口：“是的，在总结行为规律与解读数据方

面，我们拥有不可思议的能力。在客户及商家有所动作之前，我们早已对一切了然于胸。我们不仅知道客户的兴趣所在，还清楚他们将采取什么行动，因此，我们可以对这个过程产生实质性的影响。”

Quantcast的这种方式迅速演变成线上与线下商家的惯常做法。刚刚在纳什维尔成立的Facedeals公司承诺，他们可以帮助线上店铺安装面部识别摄像头。安装了这种摄像头之后，商家就可以扫描客户的面部，连接到客户的脸谱网个人页面，然后根据客户在网上发布的“个人喜好”信息为他们提供个性化的商品与服务。2013年年底，英国超市巨头乐购宣布将实施一些类似的计划，将在英国各地的乐购超市收银台安装有内置摄像头的显示屏，利用客户运算程序统计每名顾客的年龄与性别。这些信息就是增强版积分卡，可以帮助商家锁定零售广告投放对象，

还可以根据日期、购物的具体时间，以及通过购物行为监测得出的新的分析结果，随时做出调整。面部情感识别系统OptimEyes的研发者、数字媒体公司Amscreen的总裁西蒙·休格表示：“广告业到了发生阶跃式变革的时候了。商家不仅需要评估他们的广告大概吸引了多少人，还需要知道这些人的身份。”

第三次浪潮

极具讽刺意味的是，依据个人的上网习惯去吸引客户的理念，其根源是未来学家阿尔文·托夫勒所谓的“浪潮理论”（wave theory）。1980年，托夫勒在他的专著《第三次浪潮》中指出，技术发展就像一波波浪潮，新的浪潮不断冲击旧的社会与文化。^①托夫勒认为，到目前为止这样的浪潮冲击已经发生了三次。第一次浪潮产生的背景是农业的发展，其核心是人类的劳作取代了狩猎采集文化。第二次浪潮的核心要素是大型机械，伴随工业革命出现，此后推动各种“规模化”效应迅速蔓延：批量生产、批量分销、大量消费、大众教育、大众媒体、大众娱乐与大规模杀伤性武器等。第三次浪潮，即信息化时代的到来，为我们带来了辉煌的“分众化”（demassification）趋势，人

们终于可以摆脱大众社会的桎梏，随心所欲地实现个人自由。托夫勒认为，分众化将是“人类社会有史以来最深刻的社会变革，是一种开创性的重构”，其任务是“自下而上地构建一个伟大的全新文明”。所有这一切，都建立在个性化的基础之上。

1. 用浪潮来比喻技术革新并非托夫勒首创。在20世纪就有一些著作使用了这种说法，例如，德国社会学家诺贝特·埃利亚斯在《文明的进程》（The Civilizing Process）一书中就有这样的表述：“几个世纪以来，融合进程像波浪一样不断向前推进。”

请不要挂机，马上为你转接

众所周知，并不是所有呼叫中心的话务员都能接听和处理各种类型的来电。公司规模越大，就越不可能安排某一个人处理所有的咨询业务。正因为如此，客户的电话通常会被转接到不同的部门，由经过培训、具备不同技能与知识的话务员处理。举一个浅显的例子，一家面向全球市场的公司的呼叫中心经常会接到多种语言的来电。打进电话的一方与话务员都有可能会说一种或几种常用语言，但不一定会说所有的常用语言。在接听电话时，如果客户说的是法语，话务员可能会建议他们按下话机的“1”键；如果客户说的是英语，则可能建议他们按下“2”键。通过这种方式，呼叫中心就可以安排最合适的人选去接听客户的来电。

呼叫中心根据来电客户的语言与问题所涉

及的专业知识，将客户来电转接给不同话务员的做法简单易行。但是，如果采用某种运算程序，只要根据来电客户的某些特点（例如谈话模式，特定词语的选择，甚至是否有“嗯”等口头语这类似乎无足轻重的细节），就可以安排适合客户的话务员，这不是更好吗？

芝加哥的Mattersight公司经营的就是这方面的业务。该公司把这项基于客户算法的业务称作“可预测的行为路线”（predictive behavioral routing）。该业务将来电客户与话务员根据性格特点分成若干类型，从而可以更快、更好地处理所有客户的来电。Mattersight公司在他们的宣传材料中指出：“每一位客户都有不同的预期，其行为也各不相同。同样，公司的每一名员工处理不同类型来电的能力也不一样。因此，与客户的通话是否成功，往往取决于处理该次通话业务

的员工人选，以及他的能力与行为特点是否与该客户的需求相吻合。”

为Mattersight公司建立行为模型的是临床心理学家塔伊比·卡勒博士。卡勒博士开创了一种叫作“交流过程”（Process Communication）的心理行为研究。20世纪70年代初，卡勒在一家私立精神病院实习，通过观察病人在压力下的各种反应制作了一个“迷你脚本”（Miniscript）。1977年，他因为这项成果获得了“埃里克·伯恩纪念科学奖”。卡勒观察到的是在特定的精神痛苦发生之前出现的可预测的征兆，而这些征兆与某些特定的谈话模式之间存在联系。根据这些征兆，卡勒总结出6种常见的性格类型，如下表所示。

性格类型	性格特点	所占比例
思考者	思考者通过数据了解世界，主要依靠逻辑分析来处理不同情况，有可能缺乏幽默感，控制欲强。	25%
反叛者	反叛者在交流时注重对方的反应。他们爱憎分明，很多成功的革新者就属于这个类型。在压力面前，他们可能会非常消极，喜欢指责别人。	20%
坚持者	坚持者喜欢根据自己的观点分析所有事物。在处理事情时，他们会根据自己的世界观加以权衡。大多数政治家都属于这个类型。	10%
和谐者	和谐者倾向于凭情感与关系处理事情。在高压下，这个类型的人往往会产生过激反应。	30%
进取者	进取者通过行动了解所有事物。他们是天生的销售专家，一心想着如何达成交易。这个类型的人有可能不够理性，容易冲动。	5%
想象者	想象者在思考问题时思路开阔，思想活动非常细腻，发现规律的能力超过常人。	10%

尽管每个人都或多或少具备上述6种性格特点，但只有在交流双方的主要性格类型相吻合时，才能达到最佳效果。在交流时，如果无法得到积极“回馈”（例如要求情感细腻的人阐述冷冰冰的事实），交流就无法进行，提问的一方就会感到痛苦。出现这种情况后，另一方必须及时做出适当的补救。

Mattersight公司认为，在呼叫中心这样的环境中，认识到这一点可以使交流效果发生非常显著的变化。当话务员的性格类型与来电客户相仿时，通话的平均时长为5分钟，解决问题的成功率为92%。如果通话双方的性格类型相互冲突，通话时长会增加一倍，达到10分钟，但解决问题的成功率却降

至47%。

不过，采用这个方法的并非只有 Mattersight 公司。到目前为止，卡勒还帮助美国国家航空航天局（NASA）开发了遴选宇航员的算法程序，可以准确地测试候选人是否具有抗高压的性格类型。（坚持者做事力求完美，还会激励其他人倾尽全力。研究证明，这种类型的人最适合做宇航员。）卡勒自己经营的卡勒通信公司正在开展若干项目，通过数据分析建立算法程序，为其他企业提供与性格相关的问题的解决方案。

曾经做过律师、如今在卡勒通信公司担任首席运营官的罗伯特·沃特说：“在我们看来，不同性格类型是多样性的关键要素。如果所有文化的构成单元都相同，相同类型的交流在效果方面也相同，那么真正的多样性就只能存在于交流双方的性格类型，而不是他们的种族、性别或者其他任何特征之中。

我认为不同的性格类型可以发挥重要的平衡作用。当我们走进一间房间，与房间里某个与我们背景不同的人展开对话时，如果我们可以很快地确认对方与我们以前的交流对象有相同的性格特征，我们就会有一见如故的感觉。”

谷歌的乌比冈湖策略

Gild公司是一家创建于2011年的猎头公司，为技术行业中一些著名的大型企业提供服务。当前，该公司关注的问题是通过自动化程序发掘有才干的程序员，其使命宣言是用“公式”消除招聘程序中众所周知的不可靠性。为了实现这个目标，该公司利用算法分析了几万人（这一数字很快就会变成几十万人）的各项指标与数据，期望从中寻找规律，建立该公司所谓的“综合预测模型”。

该公司的成功案例给人们留下了深刻的印象，其中一个典型的案例涉及一位名叫哈德·多明格斯的人。多明格斯26岁，大学辍学后就一直住在加州南帕萨迪纳市。他一边靠透支信用卡度日，一边自学计算机编程。后来，一家公司通过Gild的算法程序“发现”了他。如今，多明格斯就在这家公司就职，担

任程序员。多明格斯的故事并非个案。Gild的首席科学家薇薇安·明指出：“有些人的简历，我们看一眼就会扔到一边，我们的算法程序却认为他们是最佳人选。对某些公司而言，这正是他们迫切需要的程序。这些公司收到的简历已经泛滥成灾，有我们的算法程序帮助他们发现各种人才，他们就不需要亲自筛简历了。”

我第一次与明交谈是在2013年5月，当时，她坐在出租车后排的座位上，正赶往旧金山国际机场。明身材高挑，长相姣好，有双淡蓝色的眼睛，一头金发略带一点儿红色。她毕业于卡内基-梅隆大学，从事的是神经学理论研究，从她戴的预览版谷歌眼镜就可以清楚地看出她对工程技术的痴迷。明的推特个人页面上不仅列出了她的神经学教育背景，还把她描述成“勇敢无畏的企业家、令人望而生畏的超级英雄和睡眠不足的母

亲”。

Gild希望在工作场所实现精英管理，对于公司的这种乌托邦式愿景，明深以为然，而且兴趣盎然，她反问了一句：“难道不应该是这样的吗？负责招聘的决策者应该精准地了解应聘者的整体情况，而不是以貌取人，草率地做出决定。但是，我有充分的理由相信，在实际的招聘活动中，相貌确实是一个非常重要的因素。”

薇薇安·明之所以清楚地知道为何第一印象无法准确评判应聘者，可能与她的个人经历有关。明出生时是男性，名叫埃文·坎贝尔·史密斯，但“一直生活得不如意”。直到2008年接受了变性手术，明的生活才有所改观。至于做这个手术的原因，她解释说：“从传统意义上看，我那时算不上好学生，考试经常不及格，在学业方面麻烦不断。我花在学习上的时间不多，但我非常关注学习体验。我

遇到的最大麻烦来自一门化学研究型课程。当时，我费了很大力气才合格。我喜欢学校生活，我热爱学习。”明接受变性手术后，她发现人们对她的态度有所改变。学生来问她数学题的次数比以前少了，男性同事和因工作关系结交的人邀请她参加社交活动的次数也少了。

薇薇安·明认为，传统的招聘策略存在两大问题。第一个问题是固有的偏见性。大多数人在招聘时，青睐与自己经历不同的人，但是，社会中的这类人却不在他们考虑之列。明说，初创公司的员工往往都年龄相仿，衣着邋遢，一看就是那种典型的工程技术人员，因为负责招聘的人往往只录用他们了解的人。如果一个人与某个工作勤勉的中上阶层白人男性工程师交往甚密，那么从统计学的角度说，这个人本身也很有可能是一位工作勤奋的工程师，还是一位中上阶层的

白人男性。由于当今社会盛行数据驱动的算法程序，因此在很多情况下，这种臆测已经变成一种标准。脸谱网招聘时，筛选程序的初始环节之一就是检查求职者是否有熟人在脸谱网供职。领英网也采纳了这一理念，其招聘算法会检索求职者与招聘负责人之间是否存在某些联系。尽管在某个层面上看，这个做法简练有效，但是它也会将大量具有不同地域、社会与文化背景的求职者拒之门外，造成一些不良效果。

薇薇安·明指出，传统招聘策略存在的第二个问题（几乎可以肯定，这个问题在科学家的眼中更加严重）是，在预测应聘者是否会在职场中取得成功这个问题上，之前的招聘策略做出的预测已经被证明是不准确的。在硅谷这样的环境中，人们认为数据驱动的招聘策略客观公正，并将这种客观性置于举足轻重的地位，因此，一旦出现预测不准的

问题，就更加不可原谅。我们以谷歌为例。美国幽默作家加里森·凯勒宣称自己小时候生活在虚构的乌比冈湖地区，那里的“所有女性都非常强壮，男性都相貌英俊，所有的孩子则异常聪明”。谷歌借用了这个说法，把他们的招聘策略命名为“乌比冈湖策略”（Lake Wobegon Strategy）。根据该策略，在规模逐年翻番的企业中，如果要维系现有的技能水平，新员工的技能水平就必须超过谷歌现有员工的平均水平。由于“技能”这样的概念无法量化，因此谷歌一直都特别重视应聘者的学习成绩。谷歌认为，平均成绩点数与就读的学校可以反映应聘者在严谨性、忍耐力与如期完成任务的能力等方面的既往表现，因此可以用来预测其在职场上获得成功的可能性。一名麻省理工学院计算机科学专业的毕业生可能无法成为世界首屈一指的计算机科学家，但人们有足够的理由认为他已经为

实现这一目标打下了“良好的基础”。反之，如果随机选择一名没有接受过麻省理工学院教育的人，尽管他也有可能成为一名优秀人才，但他成为庸才的可能性却比前者大得多。在厌恶风险的行业中，人们往往不会因为孤注一掷而得到奖赏，但是，一旦录用了一名注定不适合某一特定岗位的员工，就有可能丢了自己的饭碗。因此，我们不难理解为何众多高科技公司宁愿选择稳妥的招聘策略了。

不过，在数据越积累越多之后，人们就可以利用算法程序从这些信息中寻找规律。谷歌公司发现，他们用来预测新员工工作业绩的那些标准（包括在校学习成绩、考试得分、以前的工作经验，以及面试结果），几乎都不能准确地预测应聘者是否能在某个特定岗位上有突出的表现。然而，这并非谷歌公司的新发现。20世纪60年代，电信业巨头

美国电话电报公司（AT&T）对低级别管理人员进行了智商测试，并追踪了他们在随后20年间的职业发展情况，以了解每名员工在企业中取得的进步。美国电话电报公司与普通大众一样，也预测那些智商水平较高的人将会成为公司的高层，而智商水平较低的人则会居于公司的底层。他们认为这是理所当然的。但是，调查结果表明，对于管理人员在职业发展上的差异性，智商测试成绩的影响程度还不到15%。造成这一差异性的主要原因是多个因素的综合作用，包括性格特征、情商、社交能力以及其他几个因素，而且这种综合作用的结果是无法测量的。

谷歌公司意识到他们在招聘时需要考虑一些新的参数，这就为Gild公司大显身手创造了空间。当前，高科技公司摒弃了逐年从数以万计的应聘者当中招聘员工的做法，转而把招聘目标设置为包含各种人才的群体。毕

竟，人们无法预测璞玉到底隐藏在哪个角落里。为了规避风险，Gild在考察潜在的员工时会权衡关于这名员工的几乎所有的已知信息。薇薇安·明认为，就像人类棋手与“深蓝”（在被《新闻周刊》称为“人脑背水一战”的较量中，这台超级计算机击败了国际象棋大师加里·卡斯帕罗夫，引起了世人的关注）之间存在差别一样，依靠算法的招聘方法与由人主导的招聘方法也大不相同。她说：“计算机可以考虑到各种可能的情况，从而选择最好的招数。而人类棋手则做不到这一点。他们会毫不含糊地摒弃大多数可能的招数，只考虑其中的两种、三种或者四种。当然，这些招数很有可能是效果非常好的招数。”明认为，人们在制定招聘策略时也同样如此，只不过这一次他们所采用的那些衡量标准并不是非常合适。招聘时，如果能考虑尽可能多的数据，就可以有效地排除意外情况，例

如，面试当天某人正好走霉运等。除此以外，Gild公司还会考察求职者上网时经常浏览哪些网站，因为他们发现，这类信息对于职场表现有非常好的预测作用。明说：“如果你经常发表博客，那么，与经常浏览Quora网站的人相比，你成为优秀程序员的可能性要小一些。”Quora是由两名前脸谱网员工创建的问答型网站。Gild公司在寻找灵感时，甚至还会采用语义分析与情感分析的方式搜索推特回复。最终，Gild公司综合所有因素，为求职者给出一个百分制的“Gild得分”。

薇薇安·明指出：“即使你非常讨厌各种算法，也没有关系。我们最终提供的产品非常重要，可以从多个相互独立的维度来描述我们周围的人。我们这些算法的唯一目的就是综合考虑你的整个人生，非常准确地描述你的所有特点。”

量化人的职业潜能

Gild公司的目标是拓展个人职业潜力的衡量标准，从这个意义上讲他们并非孤军作战。2012年，有三所大学合作开展了一项研究，调查某人的脸谱网个人资料是否可以用于预测他的职业发展情况。研究人员认为，通过分析照片、留言、评论以及个人资料，就可以非常有把握地回答诸如“这个人是否值得信赖”、“这个人的情绪是否稳定”等问题。对于经常旅行、乐于交友、兴趣爱好广泛的人，他们给出的评价较高。在他们看来，派对照片不一定是一个不利因素，因为经常参加派对的人大多性格外向、待人友善，这两种品质都有利于在职场上取得成功。在完成初步预测的6个月后，研究人员追踪调查了56名研究对象，发现他们的工作表现与推崇责任心、亲和力、好学上进等品质的脸谱网

上的个人资料之间存在较强的相关性。因此，他们认为，脸谱网的个人资料有很好的预测作用，原因是相较于传统的求职面试而言，求职者在社交网络上“伪装自己”的难度更大。

对量化与统计分析的这些赞誉，几乎已经成为“公式”倡导者们的口头禅。其中贡献最大的当属出生于贝尔法斯特市的数学物理学家、工程师开尔文勋爵。开尔文指出，如果某个事物无法量化，就不可能得到发展。19世纪后半叶，查尔斯·达尔文的表弟弗朗西斯·高尔顿在开尔文这个论断的基础上，开展了若干非比寻常的研究，目的是测量那些无法量化的事物。例如，高尔顿无法接受“美丽”这个含糊的表达，因此他着手绘制了英伦三岛的“美女图”。在这幅图中，他将自己接触过的女性分为“有吸引力、没有感觉和令人反感”三个类别。高尔顿认为伦敦的美女最多，

而数学方法证明阿伯丁的女性中相貌平平者居多。在另一项研究中，高尔顿构建了一套统一的“烦躁度”标准，用于表示精神倦怠的程度。他认为“约束或限制所导致的叛逆表现”是烦躁的一个缩影，可以“用数值表示人们感到无聊的程度”。烦躁不安的动作越明显，人们感到无聊的程度就越高。在高尔顿看来，量化方法对上帝也同样有效，因为他认为“祷告的效用”（祷告有所回应与遭到忽视的比例）有足够的理由“成为科学领域合情合理的研究对象”。

硅谷的一家初创公司Knack就是在这样的背景下创建的。以色列企业家盖伊·哈夫特克创建该公司的目的看似非常简单：利用一系列游戏技术、机器学习算法以及行为科学的最新研究成果，为“思维敏锐”“理解力强”“有同理心”“有洞察力”“自发性”和“创造性”等表达制定普适性量化标准。哈夫特克说，他希望

通过这些举措触发“人力资本空间的根本性变革”，从而找到有效的方法去释放个人尚未得到发掘的潜能。

Knack公司经营的这些业务建立在心理学家半个世纪以来一直在研究的一个问题之上，即游戏活动可用于预测我们在真实世界中的行为。哈夫特克称：“虽然你可能会认为游戏中的人物行为不一定能反映人们在现实世界中的行为特点，但是，如果你和另外一个人玩同一款游戏，那么你们在游戏中的表现很有可能反映出你们在性格与思维方式上的区别。你的工作记忆水平、把握大局的能力、风险承受能力等特点都会在游戏活动中表露出来。”

目前，Knack公司已经开发了多款游戏。其中，《芥末侍应》和《热气球旅》这两款游戏非常简单、一学就会，而且有多种不同的玩法。例如，在《芥末侍应》中，玩家可

以选择服务员与厨师的角色，在顾客点菜之后，依据顾客的面部表情做出相应口味的菜。这些表情可能是“愉快”“悲伤”“生气”或者“任意情绪”。顾客吃完饭之后，玩家收拾碗盘放进洗碗池，再去接待新的顾客。

这些活动看似简单，但却内藏玄机。在游戏中，所有事物都是可以量化的，因为每个行动、信息、物品和规则都由原始数据构成。在《芥末侍应》或者《热气球旅》中，每一毫秒游戏都会根据玩家的游戏策略、完成任务的速度以及游戏内容的变化，收集、处理并分析数百个变量的相关数据。然后，利用机器学习工具和数据挖掘算法，析取那些哈夫特克认为在游戏活动中隐藏的行为特点。哈夫特克说：“我们收集了大量数据，因此可以准确地破解人们特有的行为密码。”

不过，在尝试对那些无法量化的事物做定量分析时，我们会不可避免地遭遇某些危

险。在谈到理解复杂概念并将其简化成可量化的元素时，最著名的评论来自进化生物学家斯蒂芬·杰伊·古尔德。1981年，在他的著作《人的误判》一书中，古尔德警告说，将“智力”等概念转化成“智商”等简单量度的做法是非常危险的。古尔德担心的不仅仅是抽象概念可能引发的危险，还担心人们有可能受到一些客观事实的误导，产生偏见，而不是产生真正的洞见。用古尔德本人的话来说，“将智力概括为一个实体，即大脑对应的功能区；用一个数字简单地量化每个人的智力水平，然后依据这些数字将人们的价值进行单一排序。这种做法必然会让人们错误地认为那些受压迫的弱势群体（以种族、阶级或性别论）天生不如别人，处于那样的社会地位是理所当然的。”古尔德自己也认为，《人的误判》一书就是对这种错误观点的驳斥。量化与简化肯定会同时发生。只要我们开始测

量，就必然会与某些内容擦肩而过，要么因为测量工具无法识别它们，要么因为测量者不了解它们的意义。在接下来的章节中，我将告诉大家，如果想为“创造性”与“爱”等概念创建简单的量度，我们在技术上将会面临非常困难的局面，因为这些概念过于复杂。

不过，哈夫特克认为，“有同理心”与“有洞察力”等表达的妙处并不在于其含糊不清的抽象性。他说：“我们关心的是数量越来越多的可量化事物。对Knack公司而言，我们摒弃了用单一标准来衡量人的做法，代之以考察人物性格特征的众多方面。我们考虑的不仅仅是一个人的‘智力’，而是他的整体状况，其微妙程度远远超过我们以前考虑的所有内容。这是一条正确的发展之路。”

通过推特和脸谱网窥视用户的生活

2013年11月，我为《快公司》杂志撰写了一篇介绍两名计算机科学研究人员的文章。这两名研究人员创建了一个算法，可以利用人们在推特上发布的信息，为他们量身定制个人简介。我认为这种“主题提炼”的体验非常有意思，（从回馈信息与这篇文章的点击率来看，）不少人也有同样的感受。李纪为（音）是这项研究的参与者之一，他解释说，推特是研究者可以利用的完美工具。首先，推特鼓励用户在公开的媒介上写日记。其次，推特要求每条博文的长度不得超过140个字符，因此用户必须用简明扼要的语言来描述生活中复杂的琐事。

由于通过阅读推特博文了解某个人几年来的生活，可能需要大量的时间（如果你是一

位老板，希望更深入地了解员工，你就有可能需要花费更多的时间，阅读很多人的推特博文，以了解员工的生活），因此，该算法的目的就是浏览推特博文，从中选择有助于你了解该用户人生中的重大事件的相关重要信息，然后以一种易于理解、方便阅读的形式给出结果。该算法通过分析每条推特博文的内容，将其分成“公共”与“隐私”两个类别（例如，重要赛事评论与本人生日就分属不同的类别），然后再将这两个类别进一步细分成“一般”与“特定”两个子类别。“一般”事件通常包括诸如对通勤状况的不满、对每周瑜伽课的评论等，这类事件由于在较长时期内反复发生，因此具有可预测性。而“特定”事件则包括婚礼、接到猎头公司电话等生活中有重要意义的事件，人们会因为这类事件而在短时间内需要处理大量的具体事务。

在算法根据推特博文为人们量身定制个人

简历时，人们最感兴趣的自然是上文最后提到的“隐私—特定”事件。如果我们阅读过常见的人物传记类书籍，就会知道其通常会着重记录该人物的具体人生经历，而不会关注他们所处的社会环境。例如，沃尔特·艾萨克森在《史蒂夫·乔布斯传》^①中关注的是乔布斯开发苹果iPhone手机的经历，而不会对乔布斯从位于帕洛阿尔托的住所开车上下班这件事发表评论。但是，在研究一个人的生平时，如果简单粗暴地把“公共”或者“一般”事件的观察结果都看成是可以忽略不计的“噪声”，就很有可能过于简单化，把这个人视为孤立于社会的一种存在。无论我们在政治上采取什么立场，都无法否认大选、海外冲突等公共事件对普通人生活的影响力。同样，经常发生但并非特有的那些事件也会影响一个人的生活。上下班路上的所有体验、在市中心的房子发生停电或者社区中犯罪活动猖

獭，这些事件可能都不属于“特定”一类，但这类事件在解释某人的境况方面具有非常重要的意义，其重要程度至少不逊色于他们的求婚地点。

我并不是要全盘否定这两位研究人员。他们认为个人既是一个独特的存在，在本质上也是一个独立的存在。对自我概念的这种理解简洁明了，完全符合西方价值体系。这一理念不仅是社会科学的基础，还是政治学、经济学与法律体系的基础。在政治学中，自我就是通过投票等政治行为参与民主活动的公民；在市场经济中，自我就是优化成本与利润使其符合个人利益最大化原则的人；而在法律体系中（我将在第三章做深入讨论），人们通常认为自我就是为自己的社会行为负责的人。所有这些解读都认为，自我在本质上是一个完全理性的个体。

我必须指出的是，认为“现代社会赖以发

展的基础是社会成员多样化”的观点由来已久，早在100多年前，法国社会学家埃米尔·迪尔凯姆（又译涂尔干）就第一个提出这种观点。1990年，加拿大哲学家伊恩·哈金在他的专著《驯服偶然》一书中指出，随着公众数据集涉及的范围越来越广，“为便于统计，人们建立的人口传统类别”也越来越多。

随着人们系统地收集关于人的各种数据，我们对社会的理解有所变化，对周围人的观察也不断改变。我们的选择、发展目标与对自我的理解都发生了深刻的变化。人们可能会问：到底是卡尔·马克思，还是为自我识别建立各种人物类型的那些作者们，对阶级意识的影响力更大？

在当今的数字时代，自我的概念主要依赖于算法的结果，人们将个人的特点与具有“可知性”特点的大型数据集进行对比，寻找二者的相关性。从某种意义上讲，这种做法其实

就是自我的公式化。有的算法异常复杂，需要收集特定个体尽可能多的数据，才能得出结果。有的算法采用“归谬法”（*reductio ad absurdum*），就像在抽象派艺术中采用最粗略的“模糊人影”来表示人一样，追求简单化。一家名叫“Hunch”的初创公司宣称，只需用5个数据点（换言之，用户只需回答5个问题），他们就能回答与消费者偏好相关的几乎所有问题，而且准确率可达到80%~85%。与此同时，该公司的“个人喜好反映一切”（*YouAreWhatYouLike*）这个算法，可以分析用户脸谱网上的“个人喜好”与“社会维度”数据集之间的关系，为用户建立详细的档案。例如，该公司告诉我们，为网上艺术社区deviantART.com“点赞”的用户在政治上偏向于自由主义，而为NASCAR（美国全国赛车联合会）“点赞”的用户则倾向于保守主义。他们还发现用户的一些“个人喜

好”对于“不信任”或者“矜持”的人格特点有类似的预测作用。2013年，剑桥大学开展的一项类似研究表明，利用一个包含58000个美国脸谱网用户信息的数据集建立的算法，可以准确地预测包括种族、年龄、智商、性偏好、性格、财富使用与政治观点在内的各种品质与特点，所有的预测活动都是基于他们在脸谱网上发布的“个人喜好”情况。一项叫作“推特心理”（TweetPsych）的业务宣称可以依据人们在推特上谈论的话题，包括学习、金钱、情感与焦虑等，通过算法测算他们的情商和智商水平。更多的研究也已经证明算法可以推断出性别、性取向、政治偏好、宗教及种族，且准确率超过75%。2010年，科罗拉多大学博尔德分校的心理学家塔尔·雅科尼做了一项调查，他分析了695篇博客中的文字，并将分析结果与博主的性格测试结果进行了比较。雅科尼指出，那些易激

动的博主们更有可能使用“糟糕”和“懒”这些表达，责任心强的博主会大量使用“完成”这个词，而容易相处的博主则常常使用“太好了”这个词。

算法往往把独立自主的个人看成人类的一个分类节点，这与个人的真正含义之间存在明显的差异，它有可能造成“公式”遭遇同样的“自我危机”。2012年，脸谱网的一位用户撰写了一篇文章，对该社交网站当时采用的时间轴（Timeline）新功能加以评论。与之前的脸谱网用户界面不同，时间轴利用空白区标记的位置类别，将以往的经历做分类处理（就业岗位、关系）。这意味着用户需要不断在自己的时间轴上添加新类别，不断追溯自己的以往经历直至“出生”。该用户指出：“从表面上看，这个界面似乎给了我们很多选择的机会，但实际上却把我们所有的创造活动变成为数不多的几个预先设定和包装

好的方框和类别，让我们从中寻找自己的位置。因此，我觉得自己在被脸谱网牵着鼻子走。”

这种做法出于这样或那样的原因而让我们所有人都感到有点儿不舒服。没错，我可以心安理得地点击“女性”这个类别，但我至少有20多个朋友在面对这些小方框时无从下手。他们要我选择“家乡”，但20多年来我都居无定所，所以这个问题对我毫无意义。我只能在地图上标记多个地方，或是接受今天下午出现在我眼前的那幅默认地图。脸谱网认为某些时间与我的生活“相关”，但是，如果我不希望把自己局限于这些时间，该怎么办呢？

1. 《史蒂夫·乔布斯传》由中信出版社于2011年出版。——编者注

近乎专制的员工管理方式

一些文化批评家指出，虽然新自由主义的初衷是从根本上反官僚主义，但是，它的出现反而从多个方面助长了官僚主义。社会学家理查德·桑内特在他的《品格的腐蚀》一书中解释了其中的缘由：

这种新式企业组织结构自我标榜的优势之一一是权力下放，即让企业低层级的员工有更多的自我控制权。从拆分官僚主义盛行的旧式巨型企业所采用的那些技术手段看，这个优势毫无疑问并不存在。新的信息体系更便于上层管理人员掌控企业全局，而身处这个网络之中的任何个人几乎都无处藏身。

这种指责肯定也适用于那些拘泥于“公式”的高科技公司。我们以高科技初创公司 CourseSmart 为例。该公司的一项业务可以使教师在课后依然能够监督学生的学习情

况，他们所使用的算法与可以将用户信息反馈给出版商的电子书分析程序十分相似。如果学生们在阅读课本时跳页，没有标记重要段落，几乎不记笔记，甚至根本不学习，就都会被这个算法记录在案。2013年，《纽约时报》针对该公司的这项业务，刊登了一篇题为“你是否完成了电子书阅读，老师心中有数”的文章，介绍了一位老师追踪三个班级共70名学生课后的学习情况。尽管其中一名学生的小测验中经常得高分，但是CourseSmart的算法提醒这位老师，该学生在学习上并没有做到持之以恒，只是在测试前一晚才会用功复习。这篇报道引用这所大学商学院院长的话来评价这项业务：“这是近乎专制的做法，但出发点是好的。”文章指出：

如果老师不把学生的参与指数（CourseSmart的专用分析工具）展示给学生，他们就无法看到这些指数，但是他们知

道眼前的书本正在监视他们。对一些学生而言，单是听到参与指数这个词语就会让他们有所触动。在上一次小测验中，查尔斯·特赫达得了个“C”，但是老师之所以知道他在学习上有困难，主要是因为他的CourseSmart参与指数非常低。43岁的特赫达先生说：“我被（CourseSmart）发现了。”特赫达先生同时从事两份工作，还有三个孩子要照料，因此只能利用深夜的时间来学习。他说：“也许我应该更加专注于学习。”

在伴随工业革命兴起的“职工学习”这种工厂模式中，老师像狱警，学生像囚犯，两者之间不断发生对抗。与之相比，CourseSmart这种模式要自由得多，学生们可以自主学习，对课堂教学的依赖性减弱（至少从表面上看，他们是在自主学习）。不过，这种模式所使用的那些工具，以及“参与”等越来越抽象的概念，对学生们学习情况

的监管其实更具连续性。《纽约时报》的那篇文章指出，只要学生有一堂课没有参与，他们就只能得到一个“刚刚符合要求”的“C”。

谷歌公司用来跟踪其员工的深度数据分析程序与CourseSmart算法非常相似。同众多高科技企业一样，谷歌公司追求的是一种自由主义乌托邦：员工每周有一天的工作时间可以自由支配。他们可以在这一天从事编码工作，也可以攀爬公司室内的攀岩墙，或者免费享用公司聘请的“感恩而死”乐队前主厨烹饪的美食。不过，2011年，史蒂文·利维在研究了搜索领域的这个巨擘之后，在他的

《谷歌大揭秘》一书中指出，该公司的这项政策看似愚蠢，“没有提出苛刻的要求，其实却是一种异常狡猾的规定”。尽管谷歌广为宣传类似的奇怪政策，但是该公司的由数据驱动的管理方式却在肆意侵犯员工的权利。

《石板》杂志上的一篇文章指出：“有时，谷

歌公司有可能异乎寻常地慷慨。然而，如果我们认为它如此慷慨是出于善心，那就错了。（谷歌公司）会投入大量人力物力，收集海量数据，了解员工对这些福利待遇的反应……它是不会毫无目的地乱花钱的。”

例如，谷歌公司有一个专门从事员工“幸福程度”量化工作的人力资源分析小组。该小组利用科学方法开展了一项名为“Googlegeist”的员工幸福度调查活动，并运用先进的专用算法从中寻找蛛丝马迹。我们以该小组几年前完成的一项调查为例。当时，谷歌公司注意到女性员工的离职人数超出正常值，于是该小组运用数据挖掘工具寻找原因。结果发现造成这种现象的原因不是性别，而是女性员工“初为人母”的身份变化：女性员工生完孩子后离职的比例比正常值高一倍。因此，该小组认为，最节省成本的解决方案是将传统的12周带薪产假延长至

5个月。在他们找到问题的根源并采取了这项措施后，新妈妈们的离职率迅速下降了50%。

利用类似的数据驱动方法也可以解决许多其他问题。例如，提醒员工加入养老金计划时应采用什么样的频率及方式，才能取得最佳效果；成功的中层管理人员在技能方面有哪些共性，这些特点是否对其他管理人员有所启示；使员工的幸福度最大化，进而实现效率最大化的最佳办法是什么，是加薪、现金奖励、股票期权，还是延长假期。

在这种宽松的“服务型领导方式”背后，像谷歌公司的人力资源分析小组等机构回归并强调的其实是“泰勒主义”（Taylorism）这个理念。泰勒主义的理念是由工程师弗雷德里克·泰勒于20世纪初提出的。1911年，泰勒在《科学管理原理》一书中介绍了泰勒主义所包含的思想，其核心内容是：人类从事体

力与脑力劳动的目标应该是提升效率；技术性计算的效果优于人的判断；主观性是思维客观性弱化的产物；无法量化的内容，要么根本不存在，要么没有任何实际意义。泰勒指出：“只有加强方法的标准化程度，加强对最佳工具及最佳工作环境的应用，加强合作，才能保证人们的工作效率更高。”

提升员工的工作效率与幸福度

当然，人们关心的不仅仅是更高的工作效率。关于“幸福度”与“参与度”的量化研究表明，单是工作效率高已经无法满足企业的需要了，员工还必须有爱心，能提供“微笑服务”。因此，人们需要关注“公式”在改善工作环境方面是真的发挥了某种作用，还是仅仅把员工变成了懵懵懂懂的参与者。后者“装模作样地表现出获得解放的喜悦”，同时，他们也非常清楚，只要有一丝激情不足的表现，就会在类似CourseSmart的参与指数中被记上一笔，未来在求职面试时也很可能被某种算法发现这个不良记录。

这里，我们可以再次借鉴阿尔文·托夫勒的研究成果。在本章讨论的原理中，很多都得益于托夫勒提出的“分众化”概念。在《第

三次浪潮》一书中，托夫勒对所有人都遵循朝九晚五的作息制度的情况提出了质疑。他认为，如果把这种大众化模式变成个性化、以自我为中心的模式，企业与员工就会同时获益。企业可以充分利用时间安排方面的优势，使员工实现效率和生产力的最大化。同时，员工可以根据工作以外的安排或是自己的生物节律（现在，我们可以借助可穿戴式传感器发现这方面的规律），来安排自己的工作时间。当然，托夫勒眼中的这种乌托邦公式现在已经变成了现实，成为企业普遍采用的“弹性工作制”。很多公司放弃聘用固定员工，转而采用由兼职工、定期合同工与自由职业者构成的自由的人才队伍，从而节约了各种福利和就业安全保障方面的成本。贝基·霍格在《赤脚进入网络空间》一书中指出，这其实就是用技术方案去解决问题的那些专家们梦寐以求的目标。至于这个梦想的

实现过程，霍格指出：“20世纪80年代的那些个人电脑权威们.....奔走于各大型公司，为如何分割公司财产与员工财产给它们提供各种建议。他们的行为开创了不稳定雇佣关系的时代，这种雇佣关系成为我们这一代的标准模式。”在工作与娱乐的游戏化蜕变过程中，绩效工资（表现为另一个层次的个性化）等概念意味着，即使是那些不能立即提升速度与效率的工作，也可以采用“公式”法加以管理。

如果我们进一步研究高科技“食物链”，就能更加清楚地看到这种新泰勒主义的动态。例如，亚马逊仓库中的拣货工配备有手持计算机，可以根据计算机传送的指令，找到每件产品可能的存储位置。由于亚马逊的仓库非常大，因此这些手持计算机运用“路由算法”计算出两点之间的最短路线。不过，手持计算机的功能并非只有这些。它们还能不断

收集实时数据，监控员工行走与完成单个订单配货任务的速度，从而量化他们的工作效率。就像电影《生死时速》中装有炸弹的公交车一样，员工的工作效率不能低于某一下限，否则就会被解雇。与亨利·福特发明的流水线一样，在亚马逊的仓库中，员工工作的节奏也是由机器决定的。亚马逊的一位仓库管理人员曾经这样形容这里的工人：“他们都是活生生的人，但就像机器人一样一刻不停地工作。”《快公司》刊登的一篇文章描述了一幅令人沮丧的奥威尔^①式画面：

亚马逊公司的拣货工需要不断地穿梭于各个货架之间。由于仓库有9个足球场大，因此每班工人都可能需要步行15英里的路程。仓库里非常安静，除了他们的脚步声以外，听不到其他任何声音。他们如果彼此交谈，就有可能被解雇。在他们辛勤工作的同时，用硬纸板裁剪成的“亚马逊员工”就在旁边冷

眼旁观，在“他们”幸福的笑脸上方的对话框里写着：“这是我迄今为止找到的最理想的工作！”

这并不是亚马逊仓库特有的现象。在乐购公司的仓库中，工人们需要佩戴手臂式电子终端，以便管理人员评定他们的勤奋程度。管理人员会规定一段时间，要求仓库中的工人根据订单完成配货工作。如果他们按时完成，就会得到满分奖励；如果他们以两倍的速度完成该项任务，就会得到双倍的分数。反之，如果他们未能按时完成任务，分数则会大幅降低。

1. 乔治·奥威尔是英国著名小说家。他的作品的特点是对社会不公正现象的关注，其中最著名的是《动物庄园》和《1984》。

简约现实聚合器与过滤气泡

1944年10月，英国开始重建在不列颠之战中被炸弹摧毁的国会下议院。时任首相温斯顿·丘吉尔在演说中指出：“我们建造房舍，而后房舍会重新塑造我们。”在算法时代，我们也可以表达类似的观点：用户建立自己的线上个人资料，而后他们的线上个人资料重新塑造他们的形象——不仅以我们看得见的方式，更重要的是，它们还以我们看不见的方式重新塑造我们的形象。

2013年年中，我在研究一家名叫“Nara”的初创公司时，创造了“简约现实聚合器”

（Decimated-reality aggregators）这个表达，用来形容该公司的奋斗目标。组建之初，Nara的目标是收集全世界成千上万家餐馆的信息，建成一个餐馆推荐系统。但是，该公司的终极目标是为人们的生活提供全方

位的建议：根据人们常去的餐馆收集相关信息，然后为人们的衣食住行提供各种建议。在他们使用的算法中，甚至还包括上向流动性的概念。例如，假设我们目前连霞多丽葡萄酒与基安蒂干红都区分不出来，却希望在两年后成为红酒鉴赏家。在这种情况下，Nara可以分析其采集的大量用户数据，为我们设计一个精致的可行性方案，以确保我们在一定时间内实现特定的终极目标。如果我们利用Nara算法识别我们希望实现的目标，Nara的算法就会帮助我们进入适合的模式。通过这种方式，“简约现实”可以剔除互联网上的冗余信息，用户看到的将只是该算法认为与目标相关的信息，因此他们不会因为面临的选择太多而无所适从。

Nara公司的首席技术官内森·威尔逊说：“互联网已经变成了一种交易机器。通过我们的视线与鼠标点击，为我们发来广告以及各

种乱七八糟的信息。我的目标是彻底扭转这种局面，剔除杂乱的信息与噪声，帮助用户更有效地获取有用信息。”当然，他要解决的问题就是利用Nara算法，代表用户不断地判断信息是否有用，通过剔除线上的“杂乱信息”，帮用户节约时间。

人们通常把这种功能称作“过滤气泡”（filter bubble），伊莱·帕里泽以此为题写了一本书。在这本书中，帕里泽告诉我们为什么利用谷歌搜索工具搜索相同信息的两个用户会得到不同的搜索结果。有自由主义倾向的用户在搜索框中输入“BP”，可能会得到2010年4月英国石油公司在墨西哥湾发生的石油泄漏事件的相关信息，而有保守主义倾向的用户在输入这两个字母之后，却更可能得到有关该石油公司的投资信息。同样，如果一位女性用户搜索“wagner”（瓦格纳），那么算法更有可能把她引导到作曲家“理查德

·瓦格纳”（Richard Wagner）的相关网站，而男性用户在搜索相同内容时很有可能搜到的“瓦格纳美国”涂料产品的信息。从严格意义上讲，搜索算法无法提交不受意识形态影响的搜索结果。而且，这些搜索结果为了取悦我们的个人主义观念，往往会进一步加强我们对某些问题的“既有看法”，同时，对于与我们现有观点格格不入的那些问题，则会降低其重要程度。

尽管这种创新明显带有自由性，（的确，个性化就一定好吗？）但是我们不难看出其中极为明显的缺点。在推崇自由主义的技术专家所梦想的世界中，人们享有自由，可以畅所欲言，随性发表自己的观点。与之不同的是，编码与算法的一个重要组成部分是由软件完成的整理、分类与分级任务。由于谷歌等公司的收益依赖于其用户的认知资本，这种“软件分拣”的概念使不认同“数字种性系

统”（digital caste system）的观点没有立足之地。由于有“过滤气泡”，因此很难确定人们使各种用方法对地理-人口资料加以区分是大规模定制的有益范例，还是隐晦性歧视行为的特有例证。哲学家费利克斯·瓜塔里想象了一座在每个路口都设有升降闸门的城市，人们利用一种可以让升降闸门抬起的电子安全卡，就可以自由地离开自己的公寓、街道和居住地。不过，这种闸门在代表自由的同时还代表着某种约束，因为可以打开通行之门的技术同样可以关上这扇门。算法也是一样。算法的作用是通过独立运行，直接为个人打开社会意义与地理意义上的门，让他们可以获得一定量的商品、服务和机会。

市场细分的一个重要部分是对用户进行分类，以便故意给其中一类用户制造麻烦，而让另一类受到青睐的用户从中获益。英国超市开展了多次调查，以确定在购物高峰时段

对购买食品的人收取一笔额外费用的可能性。这项举措的目的是针对那些“手头较紧但时间充裕”的顾客，防止他们对“手头宽裕但时间稀缺”的职业人士的购物体验造成负面影响，而后者是更受商家青睐的客户。很多航空公司推行附加费计划，允许重要的商务旅客通过多付费用的方式，绕过某些过境管理措施。分拣程序根据旅客是否注册该计划将他们分成两个类别。如果旅客注册过附加费计划，他们的停车位就位于靠近候机楼的位置，可以从专用候机室快速登机。对于那些“手头较紧但时间充裕”的群体，分配给他们的停车位通常离候机楼较远，并且不允许他们使用贵宾候机室，他们需要花费很长时间办理登机手续并排队过安检。出于类似的考虑，洛杉矶、圣迭戈、多伦多、墨尔本、东京等城市都有私人出资修建的公路，专供愿意支付附加费的富裕车主使用。为了避免其

他人使用这些公路，他们利用各种算法，估算出单次通过应收的费用，以便在周边公共道路非常拥堵的情况下，也可以阻止过多的车主使用这些道路，确保这些道路通畅。

清道夫级用户和差别定价法

在数字世界，这样的做法并不少见。互联网解决方案提供商思科公司曾经把不接受附加服务的消费者称作“清道夫级”用户。在这类用户提出某些业务申请时，公司提供的是“未倾尽全力的服务”。在网络拥堵和下载速度较慢的时候，分配给清道夫级用户的流量可能会“被限制到接近于零”（思科公司的原话）的水平，而更重要的企业用户则可以使用流畅的“私人出资修建的道路”，获得思科提供的解决方案。这种区别对待的现象可能并不仅仅存在于互联网领域。通信领域的Avaya公司在营销手册中同样承诺，该公司的定制系统可以利用算法，将接入呼叫中心的一个个号码与某个数据库进行比对，如果某个号码属于前5%之列，则会被优先接听。在这种情况下，当话务员接听电话时，耳边

会有声音很小的提示，告知本次呼叫来自前5%的客户。在某些情况下，还可以采用几乎相同的算法，根据性格类型来过滤呼叫人。因此，呼叫中心可以根据来电客户的性格类型来提供服务，把对于公司来说更有利可图，并且在接受服务时更容易失去耐心的客户排在优先位置上，而把那些纠缠不休、问题不断，而且未来不大可能消费大笔资金接受公司服务的顾客，排在靠后的位置上。

在这种“性格类型软件分类法”中，细致程度最高的可能是算法驱动的“差别定价”（differential pricing）。包括亚马逊在内的多个在线企业已经在使用这种方法。如果去掉营销辞令的掩饰，所谓的“差别定价”意味着某些用户在购买与其他用户相同的产品时可能需要多付钱。某些算法推测购买过《哈利·波特》与《暮光之城》的用户，可能还会购买《饥饿游戏》三部曲。从理论上讲，差

别定价使用的算法与这些算法没有任何不同，是那些可以记住用户姓名（更重要的是，记住用户的信用卡消费明细）的网站所青睐的另一个层次的个性化。2012年9月，谷歌申请了“电子产品动态定价”的专利，可以根据算法对用户是否会购买的判断，改变视频、音频、电子书和电脑游戏等产品在网上的价格。在提交专利申请时，谷歌还提供了一些材料，得意扬扬地表明有可能说服某些用户以4倍于其他用户的价格购买完全相同的数字文件。换言之，如果谷歌的算法根据你的搜索历史，“知道”你可能会被《哈利·波特》与《暮光之城》等青少年读物所打动，它就有可能因此推断你会花大价钱购买《饥饿游戏》。与此同时，如果以往的搜索历史说明某个用户对青少年魔幻小说与神秘吸血鬼的兴趣不大，该算法则会降低《饥饿游戏》的价格，诱惑该用户掏钱购买。

近年来还兴起了一种所谓的“情绪嗅探”（emotion sniffing）的算法，在用户语音语调、面部表情甚至浏览历史的基础上推断用户的情绪状态。微软研究所的一项研究，可以通过分析人们的电话记录、应用软件使用情况以及当前所在位置等数据，来推测他们的心情。微软声称，利用算法推测用户每日心情的做法，在经过两个月的训练后，其准确率已由刚开始时的66%逐渐上升到93%。由于心情对消费偏好有显著的影响，因此这类信息对于营销人员而言可能具有不可估量的价值。举个例子，假设在某个时间点上，你的电脑或者智能手机推测你可能正处于情感脆弱期。在这种情况下，悄悄地提高你感兴趣商品的价格就有可能有效的做法，因为根据“情绪嗅探”算法所做出的判断，从统计学角度看，在这种情绪状态下你决定购买的可能性更高。

支持自由市场的理想主义者可能会支持这种做法。高档饭店中一瓶啤酒的价格可能是其他地方的两倍，同样，运用差别定价的方法也可以确保人们以心理价位来购买产品。两者的道理是一样的，虽然相较于高档饭店的定价方法，后者的科学性更高。但是，另一个类比可能更准确：所谓差别定价，就相当于饭店服务生先翻看你之前的消费记录，然后再决定如何为你点的饭菜定价。差别定价的方法甚至还可以被运用到户外经营场所之中（比如，把啤酒的价格设定为每名消费者周薪的 $1/250$ ）。不过，这个方法更有可能实现相反的目标：提高某些商品的价格，并且开诚布公地告诉大家，这样做的目的是赶走那些无利可图（因此不受欢迎）的客户。例如，高端时尚品牌都希望摆脱“不受欢迎”的消费人群。21世纪初，博柏利（Burberry）就采取了这种营销策略。由于

无法彻底阻止某些群体购买自己的产品，因此公司有可能支付一笔钱，利用算法让本公司的产品信息远离年收入低于3万英镑消费人群。如果这些消费者专门搜索这些产品的信息，则提高产品定价，以超出他们的经济承受力。

这种做法的难点是如何不被察觉。我们浏览网站时，算法会选择个性化的横幅广告呈现在我们眼前。奈飞公司（Netflix）会利用算法为我们推荐影片。在这样的情况下，我们只能看到最终结果，而对过程一无所知。同样，在利用差别定价法时，顾客并不知道他们看到的价格要比隔壁邻居看到的价格高。不难理解，如果知道，他们还会继续购买吗？经常发表营销方面文章的宾夕法尼亚大学安纳伯格传播学院的教授约瑟夫·塔洛在《纽约时报》撰文指出：“我们接收到的数据流非常隐秘、非常复杂，我们甚至不知道在

什么时候遭遇了价格歧视。我们接收到的报价、新闻和娱乐消息肯定会与其他人不同。”

圆形监狱与人类的数字身份

一些互联网理论家指出，即使数字世界此前使用的歧视性分类方法（包括种族、性别或者性倾向）目前仍然存在，也必将走向衰亡。阿尔文·托夫勒在《第三次浪潮》一书中列出了过去几百年里遭受隐晦性或公开性歧视的个人及群体，并指出这种边缘化做法是第二次浪潮的产物。托夫勒认为，在根据性格类型区别对待用户的第三次浪潮中，这些歧视性行为将悄无声息地融入数字世界。这是一种非常流行的乌托邦式观点。马克·汉森在他的论文《种族歧视的数字化，或者常见不当行为的政治学》中采纳并延展了这个观点。他指出，由于网络使种族这种视觉信号变得不可见，因此不同种族的人遭遇伦理问题的可能性比之前任何时候都要大。托夫勒与汉森的错误在于，他们假设歧视行为一定

具有集体主义的性质。公式法指出，真实的情况远非如此。在当今时代，权力不再像以前一样只存在于固定机构中，还有可能成为一种没有范围限制的通则，因此，歧视的粒度可能会增加，种族歧视与性别歧视等的流动性增强。从互联网相亲资料（我会在第二章做更详细的讨论）可以看出，为了便于算法分拣，人们必须接受分解程序的处理，被拆解成一个个便于分析的成分。他们不再是不可分割的“个体”，而是成分可以被细分的“分格”（dividuals）。

“分格”这个概念不是我的发明，它是由法国哲学家吉勒·德勒兹提出的，指可以利用算法等工具无限细分、最终变成可用数据表示的物理形态的人。德勒兹与费利克斯·瓜塔里在其合著的《千高原》一书中指出，公式把人变成了“可分解动物”。

我们可以用二元结构——社会地位、男—

女、成年人-儿童等几个主要的二元对立体来分解人类；我们也可以用逐渐增大的圆形结构，把人分解成个人事务、社区事务、城市事务、国家事务、世界事务；我们还可以用线性结构，沿着一条直线或多条直线来分解人，每个线段分别代表一个事件或者一个“进程”：我们在家庭中、学校里、军队中、职场上，永远都处于过程化或者被过程化的状态，连续不断地完成一个又一个进程。

后来，德勒兹又拓展了这个概念，开始讨论他所谓的“管控社会”（society of control）。托夫勒在《第三次浪潮》中指出，我们已经从以实物商品生产为基础的纪律社会进入了以信息与金融化商品为基础的经济社会。针对这个论断，德勒兹研究了权力与管控结构的变化情况。他指出，在之前的不同社会中，管控行为发生在学校、职场或者家庭等具体的线性场所。这些场所都有

自己特有的、仅适用于该场所的一套规则。在不同场所之间，不受管控的行为仍然有继续发生的空间。在管控社会，这种情况发生了改变。尽管管控社会的规章制度没有纪律社会那么明显，但它具有连续性。由于权力无处不在，因此权力又仿佛不存在。德勒兹认为，管控社会不是强行将我们放入先前存在的模式，而是将我们放进无时无刻不在变化的封闭空间。这让我们如同置身于一个巨型筛子之上，筛孔错落有致地排列着。

这种观点似乎可以很好地概括人类全新的算法身份的特点。在数字时代，每个人都可以拥有自己的公式。如果我们每个人的行为方式都没有区别，市场细分就无法进行，Quantcast与谷歌等公司也会无利可图。正是出于这个原因，以史蒂文·普尔于2013年5月在《新政治家》周刊上发表的封面文章《数字圆形监狱》为代表的一系列文章，在

谈及大数据与算法分拣时，都会援引“圆形监狱”这个错误的隐喻。

鉴于大家可能不熟悉圆形监狱这个概念，我在这里稍作解释。圆形监狱是英国哲学家、社会理论家杰里米·边沁于18世纪末设计的一所监狱。这所监狱呈环形，中央是一座巨型岗楼。这种设计的目的，是让囚犯感觉到他们的行动无时无刻不在监视之下。仅靠一座岗楼，就可以非常有效地规范犯人的行为，让他们不敢违规。（今天，开放式办公室的设计秉承的就是类似的理念。）

《连线》杂志前主编克里斯·安德森在《长尾理论》^①一书中指出，现代商业赖以生存的基础就是市场细分。除非你销售的是肥皂，否则针对所有消费者的单一产品营销策略纯属浪费时间。事实上，企业与经销商的注意力越来越集中于利基市场。为了使利基市场发挥作用，企业必须了解我们有哪些

反常行为，以便判断我们属于哪一类利益群体。从这个意义上看，圆形监狱这种专制手段（使所有人的行为遵从同一套规则）的效果会适得其反。采用算法分拣程序时，人们知道自己受到了监视，不过他们并不在乎。俘获工具（公司识别我们的手段）与实现自由的工具（购买最能彰显我们个体身份的产品）交织在一起，几乎无法区分。法国哲学家雅克·埃吕尔在《技术社会》一书中指出，未来公民（他写这本书的时间是20世纪60年代初）所有的欲望都能得到满足，但是他们无法获得自由。后来，媒介史学家弗雷德·特纳对这个令人不寒而栗的观点进行了扩展：

如果说工业社会的工人觉得自己的工作环境就像一个铁笼子，那么在后工业社会的今天，众多信息企业的员工常会觉得自己栖身于一座舒适宜人的金矿……在这样的工作环境中，他们对实现自我价值、名声、群体认

同、人际关系以及智能愉悦感的追求，有助于推动新型传媒商品的生产。

麻省理工学院精神分析学家雪莉·特克尔在研究人的数字身份时，把我们不同的“自我”比喻成一扇扇窗户。1995年，她在《虚拟化身》一书中指出：“窗户是一个非常恰当的比喻，有助于我们把自我看成一种由多扇窗户构成的分布式系统。自我不再只是在不同时间、不同环境下扮演的不同角色，还是存在于多个世界，同时扮演多个角色的（偏离中心的）自我。”后来，这部巨著帮助她登上了《连线》杂志的封面。

把自我看成“由多扇窗户构成的分布式系统”（或者看成微软开发的略显笨重的“视窗”操作系统），是指它可以赋予我们某种权力。在这种偏离中心的窗户式自我的作用下，从丈夫身边醒来的女性可以起床，走下楼，关闭“妻子”这扇窗户，再打开“母亲”这扇

窗户，给女儿准备早饭。随后，在开车上班时，她将再次关闭她的那些窗户，然后打开一扇标记有“律师”或者“医生”的窗户。从这个例子我们可以看出这些窗户的特点，即我们可以根据自己的意愿打开或者关闭窗户。

而弗雷德·特纳描述的是一个多种主观性并存的世界，但是这些主观性彼此之间不断发生冲突。与那些在学校、职场和家中担当不同任务的“窗户式自我”或者“可分解动物”不同，特纳世界中的这些规则在涉及“公式”时并不局限于某个场所，而是以错综复杂、粒度极高而且常常十分隐晦的方式相互影响。

1. 《长尾理论》由中信出版社于2012年出版。——编者注

算法赋予我们的身份类别

就连用于确认我们身份的具体数据是否“真实”，也可能不重要。精神分析学家雅克·拉康认为，对真相的解读并不一定取决于该数据是否真实，而在于“把它看作真实的”。对于算法身份这个概念，我们也可以通过类似方式加以理解。无论我们拥有男性或者女性的身体，或者把自己看成是男性或者女性，在网上受到何种对待最终都取决于算法给出的各种结果。2013年4月，《纽约时报》的一篇文章介绍了其作者的一位匿名朋友的故事。这位朋友收到的邮件中有一份广告传单，邀请患多发性硬化症的病人出席由辉瑞和雪兰诺两家医药公司主办的研讨会。收到无聊的信件或者“垃圾邮件”自然不是什么新鲜的话题，这件事令人震惊的地方并不是无聊的信息本身，而是发件人锁定收件人

的方法。这位朋友并不是多发性硬化症患者，但是前一年，她浏览了几个健康网站去查询这种疾病的资料。收到这份广告传单，表明她的姓名及具体联系方式已经进入了曼哈顿的一个多发性硬化症患者数据库。这件事的后果有可能非常严重，文章中说，“比如，（我的朋友）将来在购买人寿保险时，会不会因为那份资料而遭到拒绝呢？她希望追踪该数据的来源，修正她的资料，如果可能，她还希望阻止这条信息被进一步传播。但是，她不知道最初收集并共享这条数据的是哪一家公司，因此无法从患者名单上把自己的名字删掉。”

创造“新算法身份”这个表达的学者是约翰·切尼·利波尔德，他说他班上成绩靠前的男学生在网上通常都以女性身份出现。他反问道：“难道我们会认为他们的网上身份不能是女性吗？”在网上，“男性”与“女性”仅仅指代

的是特定的行为类型，传统的性别区分方法在这里是无效的。在接受了这个观点之后，谁能预测2040年的“性别”歧视或者“种族”歧视状况呢？如果性别歧视行为的基础不是人们的身体特征，而是算法的结果，那么在根据某位男性点击鼠标的统计数据错误地推测他是女性时，算法会不会采取某些歧视性政策，拒绝为他提供某些服务呢？如果未来一位出生于底特律内城区工人社区的年轻白人男性被算法错误地“归类”为黑人，而生活在马萨诸塞州剑桥市的一名受过教育的年长的非洲裔美国女性却没有被“归类”为黑人，那么种族政治会面临什么样的局面呢？虽然求职者实际上是一名白人女性，但她会不会因为统计结果显示这份工作（无论是什么工作）不太适合“这位黑人男性”而遭到拒绝呢？

这些歧视性行为可能非常难以防范，因为

它们非常隐晦，大多数情况下用户都不知道自己被归为哪一个类别。人们在开展民权斗争或者妇女解放运动时还可以借助那些可共享的历史，但是由算法生成的消费者类别则没有任何文化背景可以借鉴。1955年12月，罗莎·帕克斯的抗议赢得了整个非洲裔美国人社社区的支持。但是，如果她当时受到歧视的原因不仅仅是肤色，而是年龄、居住地、种族与互联网搜索历史记录等数以千计的参数以独特方式加权的結果，情况会怎么样呢？我们可能会因为各种主义而受到种族歧视、年龄歧视和性别歧视，那么人口统计和心理分析的各种手段的背后是不是也有某种“主义”呢？与以前的歧视行为不同，当今社会的歧视行为可能在多个层面相互参照，以至于人们在申请贷款遭到拒绝，或是主动改变对自己诉求的感知结果时，很难找到发挥主要作用的单一影响因素。

此外，我们还应该注意到，性别与种族并不是固定不变的概念，相反，它们总在变化。人们的“性格”也处在不断养成的过程中，而且这个过程永远不会结束。同样，算法推断出“男性”等具体类别之后也不会一成不变，而是定期修正这些类别。在一位用户被推断为男性后，如果该用户开始搜索大量“女性”信息，算法就会重新判断该用户的类别。经常购买航空机票的男性会逐渐被算法推断为女性，而对体育或者国际新闻极为感兴趣的女性则会被推断为男性。

追加分类可以拓展我们对类别的传统认识。如果“创造性”与“感知能力”等概念可以被成功量化，并与某些行为建立始终一致的联系，就可以被赋予与性别或种族同等程度的重要性。德勒兹曾预测，我们的信用卡与社保号将逐渐成为比肤色或者就读学校更重要的身份标识，如今，上述情况正在变成现

实。德勒兹说：“我们不再以人们所在的封闭空间来判断其身份，而是以他们所欠下的债务来识别他们。”

第2章 理性的算法与浪漫的爱情

每年夏天，美国都会有数以千计的学生从医学院毕业，获得行医资格。接下来，他们将进入某个教学医院，继续他们的学业，这个阶段的学习被称作“住院实习”。

为各位医生分配实习医院涉及一个双向选择的问题。医生与医院都会有一个意向名单，需要解决的关键问题是，如何使选择双方都对分配方案感到满意。

这件事说起来容易，做起来就没那么简单了。这是因为这些医生和医院的受欢迎程度以及各自的要求都不相同。

1962年，美国经济学家戴维·盖尔与劳埃德·沙普利着手设计了一种解决方案，即用算

法来解决这一难题。他们把最终方案命名为“稳定婚姻问题”（The Stable Marriage Problem），简称“配对”（The Match）问题。

为了解释清楚配对问题，我们可以想象一个远离文明的偏僻孤岛。在这个岛上，适婚男女的人数相等。配对问题要求这些男女一对一匹配，建立一种我们认为非常稳定的婚姻关系。为了解释“稳定”的含义，请允许我先解释何谓“不稳定”的婚姻。假设岛上有两名适婚男性，分别叫作“詹姆斯”和“罗伯”，还有两名适婚女性，分别叫作“露丝”和“爱丽丝”。詹姆斯娶了露丝，但是他的心仪对象是爱丽丝。爱丽丝嫁给了罗伯，不过她内心深处的婚配人选是詹姆斯。换言之，这两人都已与别人结婚，但是如果他们两个能结合，就会让他们更满意。不难想象，由于缺乏稳定性，如果某一天詹姆斯与爱丽丝私奔，我

们可能不会因此感到吃惊。在配对问题中，算法的任务是找出一种解决方法，使所有男女都能一一配对，而且是与自己心目中的最佳人选结婚。

在正式开始配对之前，我们需要先考虑几条规则。首先，岛上所有适婚男女都需要列一个意向名单，将所有异性根据吸引力大小排序。詹姆斯可能会把爱丽丝排在首位，把露丝排在第二位，接下来是露丝的朋友夏洛特，以此类推。其次，只有男性可以向女性求婚，但女性在认为某个男子不是婚配的最佳人选时不仅有权拒绝，还可以尝试性地接受他的求婚，并在最佳人选出现时重新做出选择。再次，求婚活动一轮一轮地进行，为了让其简单明了，我们以“天”为单位进行说明。

第一天上午，所有男性都向自己心仪的女性求婚。某些女性比较幸运，被多位男性求

婚，而有的女性则没有被任何男性求婚。第一天下午，因为知道自己之后还可以改变主意，所以每名女性在所有求婚者中挑选出最中意的男性，尝试性地接受他的求婚，同时拒绝了其他男性。（这种接受方式被称作“延期接受”。）第二天一早，前一天求婚未果的那些男性向自己心中的第二人选求婚；下午，前一天已经接受某位男性求婚的女性得到了另觅佳偶的机会。如果她们认为这一天的求婚者优于自己前一天的选择，她们就可以改变主意，重新选择。第三天，第四天，第五天……这个过程不断持续，直到所有的男女都建立了稳定的婚姻关系。此时，算法程序结束。

我们以下表为例：

姓名	第一人选	第二人选	第三人选	第四人选
爱丽丝	蒂姆	詹姆斯	罗伯	拉吉夫
露丝	詹姆斯	拉吉夫	蒂姆	罗伯
夏洛特	罗伯	拉吉夫	蒂姆	詹姆斯
布里奇迪	拉吉夫	罗伯	詹姆斯	蒂姆

詹姆斯	爱丽丝	布里奇迪	夏洛特	露丝
罗伯	爱丽丝	露丝	夏洛特	布里奇迪
蒂姆	露丝	布里奇迪	夏洛特	爱丽丝
拉吉夫	夏洛特	爱丽丝	露丝	布里奇迪

第一天，每名男性向自己心仪的第一人选求婚。詹姆斯和罗伯的求婚对象都是爱丽丝，蒂姆向露丝求婚，而拉吉夫则选择了夏洛特。在詹姆斯与罗伯这两名求婚者中，爱丽丝更喜欢詹姆斯，于是她接受了詹姆斯的求婚，并且做好了改选更合适人选的准备。与此同时，露丝接受了蒂姆的求婚，夏洛特也暂时接受了拉吉夫的求婚。第二天，罗伯在前一天被爱丽丝拒绝后改向露丝求婚。但是，露丝把罗伯排在意向名单的第四位，因此，她没有改变心意。第三天，罗伯请求夏洛特嫁给自己。因为罗伯是夏洛特的第一人选，因此她抛弃了拉吉夫，接受了罗伯的求

婚。第四天，被夏洛特抛弃的拉吉夫请求爱丽丝嫁给自己，但爱丽丝仍然选择与詹姆斯在一起。第五天，拉吉夫向露丝求婚，露丝因此抛弃蒂姆，选择了拉吉夫。第六天，蒂姆向布里奇迪求婚。他是布里奇迪的第四人选，但也是唯一剩下的人选，在没有其他求婚者的情况下，布里奇迪只好同意他的求婚。这样便形成了下表所示的婚姻关系：

詹姆斯与爱丽丝	罗伯与夏洛特	拉吉夫与露丝	蒂姆与布里奇迪
---------	--------	--------	---------

这个算法有若干简单明了的特性，其中令人印象最深刻的一个特点是：该算法不仅成功地帮所有人完成了配对，而且效率最高。当然，所有人都能如愿以偿地与自己心仪的第一人选配对，其可能性微乎其微，而且人数越多，这种可能性就越小。当男女各有4人时，所有人都能得偿所愿是有可能的，但是，如果是40人，甚至400人呢？

必须指出，这种算法对求婚者（上述例子

中求婚者都是男性）更有利。如果把这种对男性而言效果最佳的方案反过来（女性求婚，男性选择拒绝或者答应女性的求婚），在应用这个算法解决上述配对问题时，一天之后就会结束，配对情况如下表所示：

爱丽丝与蒂姆	露丝与詹姆斯	夏洛特与罗伯	布里奇迪与拉吉夫
--------	--------	--------	----------

尽管这个算法有其优点，但是在解决本次研究的问题时却会导致若干概念性问题。首先，配对问题假定岛上所有人都是异性恋，因此他们希望找一个异性建立婚姻关系。为了追求数学的简明性，上述算法没有考虑以下情况发生的可能性：男性希望娶男性，女性希望嫁给女性，或者所有人都可以和任意性别的对象结婚。如果是这样，就不一定能产生上文讨论的那种关系稳定、相对简单的配对结果。

其次，稳定婚姻问题假定所有适婚男女都想结婚，而且他们的住所与外界隔绝，因此

绝不可能与其他地方的人结婚。此外，该算法假定只有遇到优于现有配偶的第三者时，婚姻关系才会结束。无可否认，一定比例的婚姻关系确实是因为这一点而终结的，但这并不是普遍原因，在统计学上也不具备鲁棒性。有的夫妇仅仅是因为两人合不来，觉得分开是更好的选择，从而结束了他们的婚姻关系。我们仍然结合上述例子来讨论这个问题。也许爱丽丝深深地爱着詹姆斯，但是在共同生活几个月之后，她发现两人兴趣完全不同：爱丽丝认为晚上参加舞会是一件非常美妙的事，而詹姆斯宁愿待在家中玩电脑游戏。而且，詹姆斯吃饭时会发出很大的咀嚼声，爱丽丝则无法接受这种行为。

再次，在算法还没开始之前就要求人们严格遵循一些规则，即上述例子中的每一名男女都要列出一份意向名单，按照自己心仪的程度清楚无误地列出所有异性，我认为这是

配对问题最不合理的方面。在被戴维·盖尔与劳埃德·沙普利视为天经地义的所有要求中，这一条似乎最难达成。

当然，上述这些话只是开个玩笑。我们知道，稳定婚姻问题仅仅是用婚姻关系做类比，因为这项研究的目的是帮医学院的毕业生与教学医院完成配对。从这个目的看，这种做法在很大程度上是可行的，因为一方在吃饭时咀嚼声很大或者向第三者抛媚眼等行为不一定会导致分手。但是，我指出这些概念上的难点，我的目的是严肃的，我想说明的是，在数学难题的范畴之外，人类还有行为不可预测这样的特点。

涉及爱情时更是如此。

用公式预测婚姻关系的持续时间

2006年，《纽约时报》邀请娱乐行业的统计分析专家加思·桑顿设计一个公式，用于预测名人婚姻关系破裂的概率。桑顿以他所谓的“日常生活中的数学方程式”而为人所知，接受过英国广播公司（BBC）与《早安美国》的采访。他设计的公式深受欢迎，他还因此专门出版了《极客逻辑》一书，利用数学知识来回答是否要文身、下一次与同事野餐时应带多少啤酒等各类问题。桑顿说，在《纽约时报》与他联系之后，他坐到桌边，开始考虑该采用哪些指标来预测名人的离婚率。他还说，每次设计一个新公式时，他都会这样做。这一次，他考虑的问题包括：这些名人的头发颜色是否会对婚姻关系造成影响？他们的住所与好莱坞标志之间的

距离会不会导致他们离婚？在大多数情况下，答案都是否定的。但是，数据挖掘总会产生某些结果，桑顿这一次的努力也没有白费。例如，桑顿发现，谷歌搜索结果中带有年轻女明星衣着暴露图片的链接数与其婚姻关系的持续时间存在明显的相关性，夫妻双方曾经结婚的次数之和也具有同样的效果。

最终，桑顿列出了下述公式：

$$2\left(\frac{30-S^2}{P+5}\right)\left[\frac{10A_b+10A_g}{(A_b-A_g)^2+100}\right]\sqrt{\frac{75(10+G_b)}{(G_b+G_g)(10+G_g)}}\left(\frac{D}{D+2}\right)^{2T}=B_{liss}$$

其中

P =夫妻双方曾经结婚的次数总和。

A_b =丈夫的年龄。

A_g =妻子的年龄（生物年龄，不可采用整容后的年龄）。

G_b =在谷歌上搜索丈夫的姓名，得到的搜索结果的个数，单位：百万。

G_g =在谷歌上搜索妻子的姓名，得到的搜

索结果的个数，单位：百万。

S =在关于妻子的排序前5位的谷歌搜索结果中，带有能使人产生性幻想的衣着暴露（或裸体）图片的链接数。

D =婚前两人结识的月数（如有需要，可输入分数）。

T =结婚的年数。如果想要预测婚姻关系持续一年时间的概率，则输入1；如果想要预测婚姻关系持续5年时间的概率，则输入5，以此类推。

B_{liss} 以百分比形式表示该夫妻婚姻关系持续指定年数的概率。

尽管桑顿的这个名人婚姻公式有开玩笑的成分，但却深受欢迎，而且预测准确度非常高。如果我们用这个公式把一些有钱人和名人于近期举行的一连串婚礼制成一幅图，就会有所发现。例如，人们发现，威廉王子与凯特（剑桥公爵与公爵夫人）的婚姻关系保

持长久的概率，远远超过克洛伊·卡戴珊与拉玛尔·奥多姆的婚姻关系，后者于2013年12月宣布离婚。该公式还预测，凯蒂·赫尔姆斯与汤姆·克鲁斯的婚姻关系将维持5年以上，但维持15年的概率近乎为零。事实证明，这个预测也是正确的。（2012年，赫尔姆斯与克鲁斯在结婚5年半之后宣布离婚。）根据预测，威尔·史密斯与贾达·萍克特的婚姻关系也将面临相同的结局。两人已经度过5周年结婚纪念日，但是他们的花边新闻经常成为八卦小报的素材，预示着他们的婚姻即将走到尽头。此外，该公式表明，“布拉吉丽娜”（布拉德·皮特和安吉丽娜·朱莉）的婚姻关系维持15年的概率超过50%。但是，如果当初他们在皮特与前妻詹妮弗·安妮斯顿离婚之后立即结婚，那么这种概率仅为1%。

桑顿承认自己并不是一位严肃的数学家。他的那些公式所取得的成功，与其说证明了

他的经验主义方法具有严谨性，还不如说反映了公众迫切希望找到某些方法，用于解答在技术欠发达的社会里可能被归因于命运的那些问题。如果他的名人婚姻公式真的可以发现一个绝对真理，其并不意味着数以百万计的谷歌搜索结果从婚姻关系建立之初就预示了悲剧性的结局，而是作为人，我们找到适合的婚配对象的能力实在太糟糕了。

如果最富有、最成功、相貌最出众的那1%的人（即名人）也无法保证可以与自己选择的伴侣白头偕老，那么其他99%的人又能抱什么希望呢？美国初婚离婚率大约为50%，二婚离婚率为2/3，三婚离婚率则高达3/4。我们选择合适配偶的能力如此低下，但我们也不应该因此大吃一惊。电影、歌曲和小说再三告诉我们，爱情从本质上讲是完全无法预测的。我们以瓦格纳1870年完成的歌剧《女武神》为例。在这部歌剧中，尽管

一对恋人发现原来他们是失散多年的兄妹，但他们却无法因此割舍他们之间的爱情。或者，我们也可以想一想流行歌手乔·杰克逊经久不衰的那句歌词：“她真的在和他约会吗？”

在很多情况下，即便是最具科学头脑的人也无法解释爱情的不可预测性。法国数学家、机械式计算器发明者布莱兹·帕斯卡说过一句非常有名的话：“心灵有自己的逻辑，理性对此一无所知。”1822年，法国作家司汤达在《论爱情》一书的开头写道，他希望可以“利用简单合理的数学方法去描述可以统称为‘爱情’的各种情感”。结果，至少同为语言大师的亨利·詹姆斯认为司汤达的这些作品“根本无法阅读”。至少在自然科学界，人们认为之所以出现这样的问题，可能是因为自然界不一定存在可以描述爱情的定律。爱情可能永恒不变，但它是否还具有某种外部特

征呢？答案可能是否定的。在谈到爱情无法量化处理这个话题时，英国精神分析师戴维·布拉齐尔说：“爱情从本质上看是无法度量的。”如果我们试图靠近观察一只样子奇特的鸟，只要我们暴露了自己的意图，那只鸟就会展翅飞走。与之相似，只要我们试图分析愉悦感或者美感，我们就会将其破坏殆尽。这样的特点导致人们在探索爱情问题时面临更大的困难。

不过，失败主义者可能不会在乎这些，因此，200年来，尽管司汤达的努力失败了，但是技术专家们仍在尝试破解爱情密码。这或许是因为他们盲目地相信了弗里德里希·尼采的观点，“爱情总是有些疯狂的，但也总会有疯狂的理由。”说到这项极具科学性的研究（仅是全球网上相亲费用一项，就超过40亿美元），最值得我们研究的当属尼尔·克拉克·沃伦和他的婚恋交友网站eHarmony。

算法时代的婚介业务

尼尔·克拉克·沃伦乍看上去并不像一位互联网时代的企业家。沃伦出生于1934年9月18日，2000年他创建婚恋交友网站eHarmony时已经65岁了。高科技时代青睐的是满脸稚气的年轻人，而不是鹤发鸡皮的老者，程序员到了30岁就“超龄”了，因为“编程是年轻人干的活儿”。高科技世界中的65岁就相当于现实生活中的165岁。无论在创建网站之前还是现在，沃伦都算不上电脑高手。他说，直到他的公司创立并开始运营后，他才学会如何收发电子邮件。我和沃伦的交谈一开始，他就告诉我：“事先声明，我对算法不在行。我想你肯定也不会对它感兴趣。”

尽管沃伦在计算机革命中算不上先行者，但却适逢其时，赶上了美国社会的另一个重

要变化。1935年10月28日，当时沃伦刚刚一岁，《纽约时报》的一篇文章指出，美国的离婚率增长了20倍还多。而且，问题比统计数据所反映的更加严重。这篇文章称，据估计有半数美国夫妇的婚姻生活并不幸福。一位专家认为，这些人生活在“痛苦的折磨之中，他们的婚姻勉强可以维持。”

与大多数意义深远的社会变迁一样，婚姻状况之所以发生这种危机，并不是由某个单一原因造成的，而是由多种因素共同导致的。在婚姻关系中，人们对爱情的重视程度越来越高，甚至超过了责任，这说明人们认为婚姻生活是所有幸福的源泉。离婚有史以来第一次变成精英社会以外的人们可以做出的一个选择。美国的城市规模在不断扩大，人们从关系紧密的狭小社区中走了出来，迎接他们的是之前他们想都不敢想的各种新的可能性。同样，随着宗教影响力的日渐式微

以及算法时代的到来，美国变成了一个更加喧嚣的尘世。

在这种情况下，社会科学家们开拓出新的研究领域，他们把婚姻看成一台需要精心呵护才能顺畅运行的机器。美国婚姻咨询业的一位奠基人和实践者保罗·波普诺指出：

婚姻关系的维持没有任何特别之处，与维修发生故障的汽车没什么不同。汽修人员依次检查点火装置、化油器、变速器、阀门等可能发生故障的部件，找出问题的根源。如果可以，就解决这个问题。我们在处理婚姻问题时采取的也是同样的方法。

沃伦父母的婚姻也遇到了问题。尽管他们的婚姻关系维持了70年，但是沃伦从来都不认为他的父母是合适的一对配偶。他经常回忆说：“我爸爸非常聪明，妈妈十分温柔，但在智力方面要比爸爸低两个等级。”沃伦的父亲很有生意头脑，取得了雪佛兰汽车的经销

权，拥有一套迪尔公司的农机，还经营了一家食品店。他对世界政治非常感兴趣，密切关注犹太人与阿拉伯人之间的冲突等问题。沃伦曾经说，他的母亲“却连中东这个名字都没有听说过。”他的父亲在波克县参加竞选时，他的母亲竟然把选票投给了父亲的竞争对手。

沃伦回忆说：“我原本打算进入政府部门，但我后来发现自己更适合成为一名企业家，而不是一名部长。”1967年，沃伦获得了芝加哥大学临床心理学的博士学位，之后便开始从事婚姻咨询工作。1992年，他出版了《寻找陪伴一生的伴侣》一书，并因此成名。这本书销售了约100万册，但沃伦本人却感到十分沮丧，因为他认为对很多夫妇而言，即使是做婚前咨询，也为时已晚。他说：“人们往往认识不久就会确定恋爱关系，无论他们对是否应该结婚这个问题持什么看

法，都很少会选择分手。当时，我意识到帮助人们觅得佳偶的唯一办法，就是在他们邂逅并建立感情之前为他们做些什么。”

大多数心理学家对这一领域的定量研究敬而远之，沃伦却兴趣盎然，决定从婚姻关系的定性研究入手，进而实现定量研究这个目的。在完成《寻找陪伴一生的伴侣》这部著作之后，沃伦开展了一项大规模的研究项目。他深入调查了800对夫妇的婚姻生活，从中挑出200对生活最美满与200对最不美满的夫妇，比较他们的调查结果，在此基础上，找到了大量他认为可以有效预测婚姻和谐程度的定量指标。沃伦说：“这项研究得出的很多结果与我从事心理治疗40年的心得体会不谋而合，这让我对自己的发现越来越有信心。”为了理解这些数据的含义，他聘请了一位名叫史蒂夫·卡特的年轻统计员，这位年轻人现在在eHarmony网站担任相亲部副部

长。在他们二人合作不久后的某一天，沃伦提出了一个棘手的问题。

“我们建一个网站，怎么样？”他说。

“你的意思是调查婚姻生活的质量吗？”卡特问。

沃伦说：“不完全是，我更想建一个相亲网站。我们可以先从公司的和谐相处模型开始，然后逐步拓展业务。”

尽管刚开始时卡特心存疑虑，但他还是非常认真地重新解读了沃伦通过性格剖析测试（其中包括436个问题）所收集的数据，着手进行网站的组建工作。随后，卡特根据“和谐相处的29个指标”清单完成了配对工作。这29个指标包含两大类变量，分别是“核心特征”（确定人物身份且基本不会改变的特征）和“重要特性”（基于学习与体验，且在成年之后可能会随着生活中的各种事件与决策而发生变化）。尽管为了防止有人窥探其中的

秘密，公司对所涉及的算法秘而不宣，但我们还是能看出几分端倪。例如，算法不会把一位女性与身高不及她的男性进行配对，也就是说，男性的配对对象不可能是身高比他高的女性。此外，算法还会使用年龄参数。年龄参数的设置因性别不同而有所变化，30岁女性的配对对象通常为27~40岁的男性，30岁男性的配对对象则是23~33岁的女性。

总体来说，eHarmony的出现为人满为患的互联网相亲网站拓展了空间，但与此同时，它也让互联网相亲网站的运营方式从性质上发生了变化。卡特说：“当时，尼尔固执地认为我们的业务应当立足于科学。”在eHarmony之前，大多数相亲网站在经营上都采用可搜索个人资料的模式，这与17世纪就已经出现的印刷品大同小异。eHarmony横空出世后，搜索模式被推荐模式所取代，后者因为其“科学精准性”而受到新闻界的称

赞。eHarmony不要求用户一页一页地浏览个人资料，而是要求他们回答一系列问题，然后为他们推荐最合适的配对对象。

2000年8月22日，该网站正式上线。成立之初，网站遭遇了几个暂时性的困难。沃伦承认：“有的人非常挑剔，不满意我们给他推荐的相亲对象。例如，我们为一位在公关公司工作的女性连续两次推荐的对象都是卡车司机，这让她非常生气。我给出的解释是：‘你知道，卡车司机也有可能非常聪明。’但是在她看来，卡车司机的身份与她非常不匹配，推荐卡车司机与她相亲是非常荒唐的行为。她认为我们根本做不了这项业务。”

用史蒂夫·卡特的话说，当杰·雷诺戴着假发在他的电视脱口秀节目中模仿尼尔·克拉克·沃伦时，“网站终于有了起色”。沃伦说：“杰·雷诺对这个老头非常感兴趣，他的头发都已经花白了，却在忙着当红娘。”尽管沃伦觉得

有点儿尴尬，但他知道自己的网站成功地引起了人们的关注。到2001年，eHarmony的注册会员人数已经超过10万。2002年8月，会员人数增加到41.5万。2003年11月，这个数字激增至225万，2004年12月接近600万，2006年10月为800万，2007年3月达到1400万。现在，eHarmony宣称他们的算法在美国已经撮合了60万对新人，在国外取得成功的案例数量也呈上升势头。事实证明，婚介业务也需要采用科学的方法。卡特说：“我个人认为，我现在从事的工作可以充分利用统计学知识，对社会产生力所能及的良性影响。”

互联网为每个人都准备了一个公式

eHarmony的定位不是咖啡店、俱乐部、办公室或者我们可能邂逅未来另一半的普通场所。用他们自己的话说，公司的宗旨“不仅仅是为用户推荐约会对象”，而是承诺帮用户与“经过预先筛选的配对人选”建立联系，用“科学手段保证约会成功率”。换言之，他们为那些利用eHarmony算法寻觅佳偶的人推荐的不仅仅是更多的约会对象，而是更合适的人选。

当然，网上相亲可以通过数量实现更高的成功率。地球目前的总人口数为70亿，其中有超过30%的人使用互联网。在上网人口比例最高的英国和北美洲，这个数字已经飙升至80%左右。以前，我们选择配偶的“人选范围”（field of eligibles，引自社会学家阿兰·

柯克霍夫），仅限于我们在日常交友网站中可能接触到的那些人，但现在，互联网人口已经远超从前，达到20亿左右。在这些人中，有意借助互联网寻找真爱的人口比例不断上升。平均来说，全世界每月有近2500万用户上网的唯一目的就是在线相亲。

不过，互联网真的可以提供效果更好的相亲服务吗？这是eHarmony等相亲网站力求实现的目标。这些网站不仅小心翼翼地守护着他们的客户群，对于用来为用户甄选相亲对象的专用算法更是秘而不宣。一般来说，网站在提供相亲服务时所使用的通常是相似的算法（以及一些辅助手段）。配对时所参考的无数性格特点与价值观数据，不仅是该算法设计者认为非常重要的内容，也是用户非常重视的择偶标准。正是因为这个特点，在eHarmony之后又涌现出一系列相亲网站，迎合不同客户的择偶需求。

例如，OKCupid网站借鉴了eHarmony赤裸裸的唯科学主义理念，在广告中承诺“我们用数学帮你寻觅佳偶”，来吸引对数字游戏感兴趣的用户。针对科学至上的用户（或者说那些骨子里相信优生学的人），GenePartner.com网站宣称他们已经“开发了一个根据男女双方基因因为他们牵线搭桥的公式”。潜在的相亲者只需支付249美元的费用，就可以买到一个收集唾液样本的工具包。该网站会把用户的唾液样本送到实验室进行分析，并把分析结果反馈给用户。如果用户再支付一笔费用，GenePartner还可以在与他基因相似的用户中为他寻找性情相投的伴侣。该公司在宣传材料中声称：“我们感觉到，基因高度匹配的男女之间会产生那种罕见的一见钟情、心心相印的感觉。彼此的免疫系统可以和谐相处，这是接受对方、欢迎对方的身体反应。基因匹配有助于提升婚

姻关系稳定持久的概率，还可以提高性生活质量和生育率。”

FindYourFaceMate.com网站服务的是那些自恋的用户。该网站自我标榜为一家“利用完善的面部识别软件和专用算法为用户寻觅佳偶，点燃真正的激情，创造和谐婚姻关系的全新的相亲网站”。特点相似的人彼此间具有某种天然的吸引力，根据这一理论，该网站的注册用户可以将个人照片上传至网站，由网站基于9种面部分析参数（眼、耳、鼻、下巴及嘴巴的特征）分析这些照片，以期促成“有夫妻相的男女配对”。

还有一个网站的名字叫作BeautifulPeople.com，该网站的创始人自称这是一个“美貌者的线上社区”。该网站坚信“投票者眼中出西施”，从而让肤浅的美貌与惊艳的自由市场思想紧密地结合起来。以体型偏胖的相亲者为目标用户的网站有

LargeAndLovely.com（“为体型偏胖者与其仰慕者牵线搭桥的平台”），体型完美的相亲者则可以登录FitnessSingles.com（“浪漫训练场”）。此外，我们还可以发现SeaCaptainDate.com（“寻找你的大副”）、TrekPassions.com（“让爱天长地久”）等网站，以及针对素食主义者的网站VeggieDate.org，根据人们的读书品位牵线搭桥的网站ALikeWise.com，或者根据人们的喜好为那些制服控介绍配偶的网站UniformDating.com，等等。

从本质上讲，这些网站的经营理念就是不断地深度挖掘，直到用户发现自己的喜好得到了满足。对于某些用户而言，约会对象是素食主义者可能是决定性因素，或者是“必须具备”的要求。但是，其他人并不会把这个要求置于首位，甚至认为这是一个无足轻重的因素。我们在上一章讨论过，网上的每个人

都是一个公式。这里，我们还要强调另外一点：网络也为每个人都准备了一个公式。

算法时代的爱情观

2009年上映的喜剧电影《在云端》中有这样一个镜头：刚刚毕业、踌躇满志的女商人娜塔莉道出了她的理想伴侣应该具备哪些品质。她把自己的理想伴侣称作适合她的“类型”，并列举了他应该具备的特点：“白领，有大学文凭，喜欢狗，喜欢看喜剧电影，身高6英尺1英寸^①，棕色头发，眼神亲切友善，在金融业工作，周末时喜欢做户外运动。我特别希望他的名字是非常简短的，比如‘马特’‘约翰’‘戴夫’。要是他开一辆丰田Four Runner越野车，那就更理想了。除了他的金色拉布拉多犬，我就是他的最爱。哦，对了，他的笑容也应该非常迷人。”

看到电影中娜塔莉毫不迟疑、有条不紊地说出这些特点，我们忍不住大笑了起来。但是，这个有趣的场景其实非常真实，因为我

们所有人都遇到过像娜塔莉这样的人，甚至我们自己就是其中一员。计算机领域的学者扎比特·帕特森在他的论文《走进网络——数字时代带来的各种幻想》中介绍了类似的观点。他讨论的不是本章所说的在线相亲，也不是前文中讨论的算法对人的分类，而是网络色情作品的组织形式。帕特森指出，从表现形式看，网络色情作品似乎可以有效地防止性生活陷入墨守成规的窠臼。尽管网上的色情信息似乎应有尽有，但是，如果用户希望浏览这些内容，首先需要按照某种程序，不断细分自己的欲求。帕特森指出：

图片在经过某种精心设计的系统方法分类后才会出现在用户眼前，必须通过固定所有图片的位置、安排好图片出现的先后次序，才能激发用户的性欲。

网络相亲的原理与之非常相似。马克思主义的“物化”（reification）观念认为，社会关

系是通过物体之间的关系体现出来的。同理，在网上寻觅真爱时，我们必须将自己与潜在的配偶商品化。从表面上看，援引卡尔·马克思的名字来讨论相亲问题似乎有失尊重，但几乎毋庸置疑的是，我们在讨论这个话题时常用的那些隐喻具有工业社会的特点。例如，2013年12月，《大西洋月刊》刊登的一篇文章把eHarmony及其竞争对手Match.com比作“爱情工厂”。实际上，这些网络相亲业务在感情上的某种程度的超脱可能恰恰是其具有吸引力的原因之一。在某些文化中，人们会聘请婚姻专家，例如犹太文化中的“职业媒人”（shadchan）或者伊朗婚俗中的“提亲者”（khastegari），目的是代替被荷尔蒙冲昏头脑的恋人决定婚姻等重要事宜。1971年，马克思主义批评家捷尔吉·卢卡奇在《历史与阶级意识》一书中指出，所谓物化，我们可以认为它的目的就是造成“幻觉

客观性’（phantom objectivity）。它似乎包罗万象，具有严格的理性，以至于完全掩盖了它的根本属性——人与人之间的关系”。

但是，卢卡奇的这个观点是不是吹毛求疵呢？从某种意义上讲，人们自然而然地认为，理想的爱人就是品性特点与我们的意愿完美吻合的那个人。在选择配偶、安排度假计划或者选购新笔记本电脑时，我们都会把自己要求的条件在大脑中列出清单。在考虑任何的潜在选择对象时，我们都会根据自己的最低要求加以评判。如果我们认为某个潜在婚配对象对我们的吸引力没有达到某种程度，某个度假计划的费用过高，或者某台笔记本电脑无法满足我们的任务处理要求，我们就会拒绝接受，转而考虑下一个选择对象。不过，以这种方式来考虑爱情问题是否合适呢？实验心理学家、作家史蒂芬·平克在《心智探奇》一书中设想了这样一种情况：

一位相貌出众的“万人迷”搬到了你家隔壁。通过该例，作者提出了一个问题，即在婚姻关系中，人们如何确信配偶在应该选择离婚时却违背理性，选择与自己继续厮守呢？这与本章开头介绍的稳定婚姻问题非常相似。如果由经济学家戴维·盖尔与劳埃德·沙普利来回答这个问题，他们可能会说，只要这位邻居对其现任配偶的感情超过对我们配偶的仰慕之情，我们就无须担心。而平克的建议却更具人文主义特点，他认为，我们在选择配偶时，从一开始就不应该接受出于理性原因而选择我们的人，而应该选择因为爱上我们而承诺与我们长相厮守的人。

在和爱人呢喃细语时，如果告诉对方他（她）的相貌、赚钱能力与智商水平达到了自己的最低要求，即使这句话从统计学上看是正确的，也很可能会扼杀你们之间的浪漫爱情。讨人欢心的话则正好相反：你之所以

爱上对方，是因为情不自禁。

这种违背理性的爱情观在算法时代可能即将土崩瓦解。丹·斯莱特在《算法时代的爱情》（Love in the Time of Algorithms）一书中讲述了“雅各布”的故事。雅各布这位30岁左右的男性在互联网上结识了一位名叫雷切尔的22岁女秘书。两人只约会了几个月就同居了，但是，他们生活在一起后却发现彼此对生活的追求不同，他们的感情就此终结。故事发展到这儿并没有任何不寻常的地方，不过分手之后雅各布的一番话却十分耐人寻味。他告诉斯莱特：“如果我不是在互联网上结识雷切尔的，如果我从来没有参加网络相亲活动，那么有95%的可能性我会娶她。如果我是在现实生活中与她邂逅，我会忽略其他因素的干扰，想方设法和她结婚。是不是网络相亲改变了我对天长地久的理解呢？毫无疑问，是的。在我意识到我们即将

分手时，我仍能坦然面对，并不觉得这是一件多么痛苦的事。我没有绝望地认为自己命中注定将孤独终老，而是迫不及待地开始寻找属于我的浪漫爱情。”换言之，天涯何处无芳草，或者用PlentyOfFish相亲网站的话来说，“每月至少有1.45亿用户访问我们的网站”。

1. 6英尺1英寸约合1.86米。——编者注

你确定要删除这段感情吗？

雅各布的这番话很好地解释了波兰社会学家齐格蒙特·鲍曼在《流动的爱情——人类感情的脆弱性》（Liquid Love : On the Frailty of Human Bonds）一书中提及的“虚拟感情”。鲍曼认为虚拟感情与各种权宜之计、安全秘诀、综合险与全额退款保证一样，是更科学的恋爱之道，他还指出虚拟感情的最终目标是“无须等待，无须付出，甚至无须努力就可以得偿所愿”。

与真实感情不同，虚拟感情来得快，去得也快。与真实感情节奏沉重、进展缓慢、缺乏活力与千头万绪等状况相比，虚拟感情似乎节奏明快、有规律可循、易于掌控且对用户友好。巴斯市的一位28岁的男青年，在被问及在网络相亲业务迅速崛起并取代单身酒吧以及征婚专栏这个现象的看法时，道出了

虚拟感情的一个决定性优势：“你随时可以点击删除键。”

用FreeDating网站创建人丹·温切斯特的话说，“流动的爱情”这种说法可能意味着未来“感情生活更美满，但是离婚率也会不断升高”。温切斯特认为，尽管这个说法看似自相矛盾，但只要算法不断发展，最终就可能会导致这样的结果。他说：“我们现在可以轻而易举地帮助人们找到意中人，而且这个过程令人愉快。因此我常常想，婚姻（最终）会不会因此退出历史舞台。”

尽管温切斯特描述的是一种新现象，但是也不乏持有相同观点的人。2004年，美国心理学家巴里·施沃茨在《选择的悖论》一书中指出，无论是购物，还是约会，人们都面临无数的选择，这种状况让很多人感到焦虑不安。在感情问题上，选择悖论是通过将坠入爱河的那些人逐一“分解”的方法解决的，而

“分解”这个工业术语表示的是划分并剥离生产资料以获得效率。在大规模按需生产的时代，感情已经变成根据时尚、不断变化的欲求以及潮水般的奇思妙想制成的普普通通的商品。

这种后工业时代的相亲方法与传统文化的教诲背道而驰。爱人应该是独一无二的，而不是一堆问题的各种正确或不正确答案的合体。不过，用法国小说家马塞尔·普鲁斯特的话说，只要浏览过一页又一页的网络“征婚”信息，就会发现“一连串的信息，过目即忘，根本无法区分、甄别”。所有人的身份资料似乎都在不断重复“可爱”“风趣”“友好”“浪漫”“富有冒险精神”等形容词，这是因为每个人都力求打破千篇一律的公式化模型，结果却在描述受欢迎的性格特点时又借用了相同的文化脚本。相亲网站在为建立用户个人资料而提出问题时，可能会以类似的方式臆断

人类的感情，例如，“金钱与雷雨，哪一个更令你兴奋？”“人体穿孔与色情书籍相比呢？”“你希望未来配偶比较强势还是留一头长发呢？”由于所有参数的权重相同，而且我们并不清楚算法的运行过程，因此，我们无法了解哪些因素会影响我们的最终得分。

2013年，《卫报》“数据博客”栏目（栏目的眉题提醒我们“事实是神圣的”）的一篇文章就谈及了这个问题。在这篇文章中，记者埃米·韦伯回忆了她参加网络相亲的经历：“我刚一登录相亲网站，就被问了很多问题。我先是认认真真地逐一回答，但是之后我越来越不耐烦，于是开始快速点击那些我认为正确的选项。”过了一会儿，韦伯开始反思自己的回答。她的确希望未来的丈夫身强力壮，靠自己的双手干活，但是，这个问题的真正意图会不会是想了解她是否愿意选择一个伐木工人作为相亲对象呢？她说：“是啊，

伐木工人都很强壮，而且是靠自己的双手工作。但是，我连树木都不是那么喜欢，才不希望嫁给一个伐木工人呢。”

如果表面上公正客观的算法认为搜遍整个互联网也找不到适合你的相亲对象，就会导致更糟糕的存在危机。尼尔·克拉克·沃伦说：“我们会拒绝大约16%的人成为我们的会员，主要基于7个理由。如果申请者抑郁消沉，我们就会拒绝其会员申请，因为抑郁与其他精神疾病有高度相关性。如果申请者有三次以上婚史，我们也不会接受他们成为我们的会员。在15.5%的美国夫妻中，至少有一人此前有过三次婚史。此外，我们也不会接受那些‘刺儿头’，这些人永远都不知道满足。你给他介绍一个人选，他认为对方过于自信；给他换一个人选，他又嫌对方太腼腆。我们有办法把这些人挑出来，拒绝他们登录我们的网站。”

抓住技术创造的浪漫邂逅机会

加里·施特恩加特于2010年出版的小说《爱在长生不老时》，讲述了未来一个成年人在纽约市所发生的故事。由于那时的信息完全透明，个人的任何信息都不是秘密。人们唯一的任务就是在他们那部类似iPhone手机的装置上，设置好“社区参数”，然后点击按钮（在小说的某个场景中，笨拙的主人公莱尼·阿布拉莫夫准备和朋友去斯塔滕岛的一家夜总会，他就点击了一下按钮）。这时，个人资料的所有内容，包括上床指数与性格得分（均为800分制），以及性爱方式的偏好情况等，都会显示在屏幕上。就像亚马逊会根据用户的购买记录推荐产品一样，这个装置中还安装了推荐系统，可以根据用户的浪漫史推荐人选。阿布拉莫夫的一个朋友

说：“这个女孩（的资料中）有一份父亲猥亵她的多媒体文件。你的很多相亲对象都有这类遭遇，看来这个装置认为你有此怪癖。”

从很多方面看，《爱在长生不老时》中描述的情景将会成为现实，而且早于我们的预期。几年前，麻省理工学院的“人类动力学研究小组”（Human Dynamics group）创建了一个移动系统，其作用是在有吸引力的潜在约会对象距离我们不到10码^①时发出警报声。小组中的两位研究人员内森·伊格尔与亚历克斯·彭特兰说，这套系统的“研发目的是创造人们与意外邂逅的对象交往的机会”，因此被命名为“奇缘”（Serendipity）。研究小组在宣传材料中声称：

在满房间的人中，你可以毫不费力地发现令你痴迷的那一位，因为你拥有我们的移动电话，可以搞定一切。如果你的电话发现另一个人（无论男女）符合你的所有条件，这

两部电话就会发出警报声，并通过蓝牙交换信息。剩下的事就交给你了。

“奇缘”等应用是“社交发现”（social discovery）这一新技术潮流的组成部分之一。社交网络的作用是帮助人们与“社交图谱”（social graph）上的人建立联系，而产生于社交网络的社交发现则可以帮助我们结识陌生人。麻省理工学院的这个“奇缘”系统是本书援引的最完美的例子之一，它既提出了问题（“运气”），又给出了任务（“抓住机遇”）。如果运行顺利，计算机也许可以回答汉弗莱·博加特在电影《卡萨布兰卡》中饰演的角色所提出的那个问题：“世界上有那么多城镇，城镇中有那么多的酒馆，她却走进了我在的这家酒馆，为什么呢？”答案很简单，因为我口袋里的“奇缘”发出了警报声。

当然，“奇缘”系统在帮助人们解决众多问题的同时也会造成很多麻烦。邂逅的发生纯

属运气使然，获益的是当事人。此时，“随机”与“收益”的权重是相同的。我们有理由认为，如果缺少其中之一，无论发生何种情况，都不能称之为意外邂逅。如果涉及爱情，这种区别就更加明显了。毕竟，对大多数人而言，爱情（如果我们可以用这个词来代表）的精髓是从纯粹的随机性中缔造某种普适价值——看似毫无意义的一次巧遇却蕴含着终极意义。法国社会学家玛丽-诺埃勒·舒尔曼与洛兰·多米尼斯在研究促使“一见钟情”发生的文化条件时，深入采访了150名声称有此经历的人。这些人多次提及的内容包括：一见钟情在生活中非常罕见，无法预测；它从本质上讲无法理解，没有逻辑可言；它不是一方对另一方逐渐了解的结果。

阿兰·德波顿的处女作《爱情》一书是对这个现象分析最精准的文学作品之一。小说的主人公深深地迷恋上他在从巴黎飞往伦敦

的航班上结识的一位女性。这位主人公为了给这种不理性的行为注入一定的理性元素，便试图计算出发生这种情况的概率，并得出了 $1/5840.82$ 这个答案。德波顿在小说中说道：“但它还是发生了。”

这样的计算还是无法让我们接受那些理性的观点，而是为我们以错误的方式解读一见钟情的行为找到了借口。如果某件事发生的概率微乎其微，而它却真的发生了，难道人们就没有理由用宿命论来解释其中的原因吗？至少从爱情本身来看， $1/5840.82$ 的概率（似乎表明这次邂逅）只能是命运的安排。

当然，说到命运，如果史蒂夫·乔布斯还在世，他会说：“我们可以利用一个应用程序来解决这个问题。”在过去200年里，接受过科学推理教育的技术专家们投入越来越多的人力物力，研究加拿大哲学家伊恩·哈金所谓

的“驯服偶然”（taming of chance）。“奇缘”的创造者们自豪地宣布：“技术正在改变我们约会的方式。技术是自上帝创造红玫瑰以来，腼腆的单身贵族们追求浪漫爱情的最有效助手。”

1. 10码约合9.144米。——编者注

形形色色的社交发现软件

麻省理工学院的“奇缘”项目并不是这一领域的首创。20世纪90年代末，一种叫作“Lovegety”的邻近区域相亲装置在日本迅速流行开来，以约21美元的价格销售了130多万套。Lovegety的目的是帮助用户发现邻近的潜在约会对象。这套装置要求用户先回答若干问题，然后将其转变成用户的个人偏好清单。如果发现有可匹配的用户出现在有效范围内，该装置就会发出警报。与eHarmony的一连串问题及29个契合指标相比，Lovegety的用户偏好选项毫无疑问要少得多（而且清单上列出的那些个人偏好都非常肤浅，例如双方都喜欢去卡拉OK唱歌）。尽管如此，依然有很多人认为这种体验与结识陌生人有本质上的区别。一位女用户说：“如果有陌生人突然来找你，你总是会心存疑

虑。但是，如果是Lovegety在撮合你们，你就会放松得多，因为你们两人已经有共通之处或共同语言了。”换言之，技术的作用不仅仅是成为双方的隐形介绍人。就像两人都拥有一辆保时捷汽车一样，同为Lovegety俱乐部的会员也意味着双方拥有某种共性。借用马歇尔·麦克卢汉的话说，媒介的确就是讯息。

当然，这类装置的服务对象并不仅限于东京原宿的年轻人。冰岛共和国的全体国民（32万人）彼此之间几乎都有或近或远的血缘关系，因此，他们开发了“冰岛人软件”（Islendiga），应用类似的技术解决意外乱伦的问题。这个软件连接着居民数据库，并追溯至1200年前的用户家谱，利用算法判断两人之间血缘关系的远近。该软件通过用户“触碰”手机启动，可以在两人血缘关系较近而且很有可能发生性关系时发出“乱伦预防警

报”。这款软件的宣传语朗朗上口：“上床前先触碰一下手机。”（Bump the app before you bump in bed.）

现在，为了占领市场，大量基于邻近区域的社交发现软件（包括“Grindr”与“Crowded Room”等）正在展开激烈的竞争。我在为写作本书收集资料时发现的一款名叫“Anomo”的软件更加有趣。该软件的服务宗旨是，要在虚拟世界与现实世界中同时“推动我们社交方式的民主化”。Anomo（其创始人之一曾经在美国版《飞黄腾达》节目中获得亚军）要求用户为自己选择一个电脑游戏风格的形象，当他们在现实世界中开展社交时以此作为自己的虚拟形象。联合创始人本杰明·刘说：“现在很多社交发现软件都要求使用用户的真实资料，你要把自己的真实照片与姓名放到上面。问题是，人们初次看到我们的资料时，经常会惊呼：‘哇，真够惊悚的！’”

Anomo的用户也需要输入自己的真实信息，但这些信息是隐藏的，用户必须通过一系列的公平交换（你给我看你的信息，我就给你看我的）才能看到这些信息。在此之前，用户可以用虚拟形象（包含简短的说明，以及从100个卡通形象中挑出的一个形象）与对方交谈，也可以玩一些“活跃气氛”的游戏，以便判断对方与自己有多少共同语言，避免初次交谈时可能出现的尴尬情形。Anomo软件要求双方回答“你喜欢喝啤酒还是葡萄酒？”“你觉得双人自行车很糟糕、很酷还是很危险？”等问题，并将两人的答案进行比较，给出“契合参数”。Anomo的宣传材料称：“匿名交流可以改变社交活动的氛围。第一印象再也不是根据照片形成的，而是通过双方之间真正的交流形成的。”

触碰手机屏幕就能知晓一切

社交发现与“即时通信”（Instant Messenger）的作用原理大同小异。用户无须一心寻找“自己的另一半”，而是不断与大量其他用户进行交流。发出的某条信息没有回音，打算与之继续“交谈”的那个人再也没有出现，这些情况即使发生也没有关系。原因在于，不管任何时候，总是有足够多的人在线（至少对于那些取得成功的软件而言，情况确实如此），可以与他们一起解决单身问题。因此，用户只需找到下一个用户，重新进行上述步骤即可。

尽管这些软件大多要求人们在自觉的状态下使用，但是几乎没有人严格遵守。计算机科学家、虚拟现实领域的先驱杰伦·拉尼尔在他的著作《互联网冲击》^①中回忆了自己在加州大学伯克利分校参加的一个评审小组。

该小组的任务是审查工科研究生为创业家项目课程提交的创业方案。有一个三人小组在提案中建议增加晚上约会的次数，以保证获得最成功的浪漫爱情。拉尼尔还在书中提到了这个小组兜售这个观点的说辞：“假定你在一个周六的晚上来到旧金山的一家酒吧，发现这里有很多年轻漂亮的单身女孩。她们正在左顾右盼，期待引起异性的关注。她们看上去并不难接近，于是，你掏出手机，触碰屏幕。”拉尼尔认为，这种做法永远不会奏效，因为此时的数据肯定是不正确的。

不过，拉尼尔的警告还不足以动摇该技术的核心概念。SceneTap公司（前身是BarTabbers）从伊利诺伊州芝加哥市起步，自那以后他们已经把业务版图拓展到了旧金山、奥斯汀、哥伦布、纽约、波士顿和迈阿密等城市。SceneTap在400多家酒吧里安装了摄像头，利用面部识别技术与人数统计算

法帮助那些经常光顾酒吧的人寻找约会对象，告诉他们在某个晚上该去哪家酒吧。目前，这个工具可以提供任意一家酒吧的顾客群大小、性别比例与平均年龄等方面的实时信息，其正在开发中的功能还有很多。2012年，福布斯网站在评价这家初创公司申请的一项专利时说，“（越过）中线，然后飞速跑到球门区大力抽射。”一言以蔽之，他们申请这项专利的目的是收集更多的详细数据，包括酒吧顾客的种族、身高、体重、吸引力、头发颜色、衣着类型，以及面部毛发或眼镜等可以辨识其身份的特征。换句话说，人们如果应用这款软件，就可以获取这样的信息：“Raven酒吧的上座率为73%。其顾客中有33%的人是天然的金发，57%的人是黑发，3%的人是红头发，5%的人是漂染金发，还有2%的人是其他颜色的头发。男性的平均身高是1.77米，其中70%的人穿商务休

闲装。女性的平均体重为70.2公斤，其中24%的人穿短裙。有73%的人是白人，有21%的人是亚洲人，有6%的人是黑人。该场所的平均吸引力是7分（10分制）。”

这项专利技术还可以在摄像头里安装麦克风，以便监听顾客的谈话内容；并用面部识别技术识别顾客，并与其社交网站上的个人资料联系起来，以判断其“感情状况，智力、教育及收入水平”。

因此，号码归属地追踪技术后来粉墨登场，也就不足为奇了。

1. 《互联网冲击》由中信出版社于2014年出版。——编者注

以前所未有的方式思考性生活

计算机程序员凯文·康鲍伊有一头乱糟糟的棕色卷发，长着布赖恩·布莱斯德式的络腮胡子。他轻声细语地说道：“从上中学至今，我只是在成年之后交往过一位女性。”大约10年前，二十四五岁的康鲍伊是一位用户界面工程师，他决定计算出他与妻子的性生活总次数。这个念头越来越强烈，于是康鲍伊开始利用晚上的时间编写一个记录性生活的应用程序，把包括性生活持续时间、频率、质量及平均时帧等所有信息都进行了分类，并建立了模型。

他说：“我一辈子都力求诚实待人，但我觉得这件事还是私下里做比较好，等到完成之后并得到妻子的同意再给她看也不迟。刚开始，她觉得难以接受，但是她也知道，编

码对于我来说就是一种自我表现形式。她要求我告诉她链接，因为她要不时地了解我们性生活的情况。最后，她把我的这个设计告诉了她的朋友，而且她们经常谈论这个话题。”

后来，他妻子的一个朋友建议康鲍伊公开销售这个软件。康鲍伊同意了，他把这款软件命名为“Bedpost”，还设计了“想过你多长时间过一次性生活吗？”的广告语。他建议用户“在每次性生活过后登录这个软件，完成几个简单的填空题。很快，你就可以为自己建立一部连续而且值得回味的性爱史”。

该软件收集的大量数据可以通过饼形图、散点图等多种形式进行可视化处理，并且可以利用热图的不同色块来表示某个用户性生活的数量与质量。康鲍伊说：“与性活动有关的数据非常多。例如，你们缠绵的地点有可能遍布全球，回头再看看这些信息将十分有

趣。”

康鲍伊说，他的目标是“让人们用前所未有的方式思考自己的性生活”。这款软件的成功说明，以0与1来测量、量化性生活的行为本身就可能引发性兴奋。是啊，如果完全依赖于符号交互的文本式虚拟性爱可以激发性欲，那么作为文本构成基础的编码为什么不可以呢？

说到这里，我们可能需要想想法国社会学家、哲学家让·鲍德里亚的著作了。鲍德里亚关于“拟像”（simulacra）的论述已经成为计算机科学研究的重要内容。在他的《论诱惑》中有一段文字，讨论了色情艺术的吸引力，指出观众可以从色情作品中看到“从未见过的内容”，即对生殖器发挥功能的过程的详尽描述。他认为：“色情艺术过于真实，观察距离过近，以至于超出了真实的范畴。”

这种过度的真实或者超真实，正是其魅力

所在。色情艺术中唯一值得期待的（如果真的存在）不是性，而是真实。而且，它关注的不是真实，而是超真实。

鲍德里亚指出，这种通过细微的观察追求真实性的行为，其根源在于根深蒂固、无处不在的科学主义。他在书中写道：“科学已经让我们习惯于追求十分详细的过度真实，并形成一种高精度的偷窥心理（对肉眼不可见的细胞结构的详尽研究），以至于我们一门心思地追求真实性。而对于真实性，我们仅凭流于表面的方法已无法测量，唯有借助复杂的技术手段。”

鲍德里亚认为，对于那些痴迷于性器官内部作用机理研究的人而言，极致的色情艺术是让观看者“全身心地陶醉于”演员表演的唯一选择，因为色情艺术可能会让他们有机会近距离观察他们期待的内容。

鲍德里亚在1979年提出这样的观点时不

无讽刺之意。但是，如果他知道Bedpost软件，他会说些什么呢？如果色情艺术是“四声道性爱——为性行为增添了第三和第四声道”，那么以亚马逊用户评论（“与玛蒂娜做爱可以打满分——我要推荐”）的风格去展现个人性爱史的软件，肯定是淫秽至极的东西。

我第一次与康鲍伊交流时，他正忙着重新设计Bedpost的用户界面，因为这个界面几年来基本没有变化。他告诉我，网站收到的很多条好建议都来自软件用户。比如，软件可以追踪女性月经情况的功能深受用户欢迎，但是他自己绝不可能想到这个创意。最让他犹豫不决的一个问题是，是否应该添加性高潮计数器功能。还有一个问题，他不清楚是否应该强化社交功能（“这个创意可能会让有些人难以接受，但是如果可以通过脸谱网与推特把用户拉进来，我觉得这个主意不

错”），并允许用户的配偶登录，添加他们自己的“标签”，从而可以像在维基网上一样发表“性爱即表演”这样的评论，而且评论可以在性生活结束很久之后进行。

可以分享性爱体验的功能尤其令人感兴趣。法国精神分析学家雅克·拉康认为，性的统一性并没有其差异性强，原因之一是我们永远无法体会对方从性生活中得到的愉悦感。但是，如果我们不是简单地询问“你感觉如何”，而是通过数据来探究对方的满足感，从而更深入地了解对方的性体验，会怎么样呢？要实现这个目的，我们不一定非得依赖用户标签这类主观因素，因为在量化性体验方面，人们已经发明了若干效果显著的可穿戴技术。例如，BodyMedia FIT臂带可以通过记录用户的心率、排汗与动作的频率，判断用户在某一时刻是否正在进行性行为。此外，这种臂带还具有侵入性功能，可以判断

用户性生活的质量。通过分析各项指标的峰值，就能判断对方是否在伪装性高潮。

这些数据不仅让人们羞于启齿，还有可能引发多种问题。例如，如果某人怀疑自己的配偶不忠，就有可能查看这些数据，然后质问对方：“在某个晚上，我们上床后就直接入睡了，为什么在零点13分至2点期间你有燃烧了100卡路里热量的记录呢？”如果摘下该装置（或者“忘了”穿戴），就有可能引起配偶的怀疑。从理论上讲，白天出现的数据异常易于解释，但即便如此，把不正当的性生活安排在白天进行也无法逃避装置的监控，因为性行为产生的身体状况数据与举重、散步、练瑜伽、练武术、骑自行车等其他需要消耗能量的活动是不同的。

康鲍伊认为，Bedpost可以帮助自己享受正常的性生活。他说：“这些数据可以让我在性生活方面建立信心。了解自己在一个月或

者这一年时间里的性生活次数是有好处的，它可以告诉我，我的性生活是健康的。无论这段时间我因为什么而对性生活感到困扰，这些数字都可以帮助我放松下来。我想，如果不是这些数据帮助我建立了信心，我的性生活就不会像现在这样充满激情。”

当然，Bedpost也会引发很多问题，至少我们不确定收集多少数据为宜。由于我们永远无法真正地拥有并支配配偶而导致爱情具有不完整性，这对于技术而言可能还是个新问题，但是早就有文献讨论过这个哲学问题了。在弗拉基米尔·纳博科夫的《洛丽塔》（《名利场》杂志形容这部小说是“20世纪唯一一部令人信服的爱情小说”）中，四十多岁的亨伯特·亨伯特教授希望俘获多洛蕾丝的芳心，但是他觉得自己无法“彻底”拥有这位年轻女孩，因此烦恼不已。有一次，亨伯特正在比尔兹利大学校园里观看多洛蕾丝（他叫

她“洛丽塔”）打网球。这时候，他突然感到非常悲伤：

我应该把她打网球的样子拍摄下来啊！这样的话，我就可以坐在放映室里看她，拥有她，而不是像现在这样痛苦绝望了。如果我用电影胶片把她击球的样子和她所有迷人的身影都保存下来，效果肯定比我拍下的那些快照好多了。今天我就无须自怨自艾了！

亨伯特把这种想法称作“可以储存的未来和已经储存的过去”，他似乎认为两者都可以用镜头捕捉并让自己真正地感到满足。毕竟，亨伯特知道，对自己来说，“客观、绝对”的洛丽塔是他闭上眼睛时浮现在他脑海中的那个洛丽塔（他把这个洛丽塔称作“我的洛丽塔”），而不是小说结尾时的那个“脸色苍白、喝得烂醉、怀着另一个人的孩子”的女孩，尽管她当时的状况确实如此（此时，他把她叫作“这个洛丽塔”）。瑟伦·克尔恺郭尔

的小说《诱惑者日记》描写了一个类似的矛盾。小说的主人公约翰内斯是一个才华横溢的知识分子，也是一个欺骗感情的老手。约翰内斯认为，“恋爱关系最多只能维持半年，一旦得到最高境界的享受，就应该马上抽身而退。”在小说中，约翰内斯准备色诱一位名叫科迪莉亚的17岁女孩。他经过精心思考，提前设想两人交往的所有可能情况，然后开始执行色诱计划。在小说的某个地方，约翰内斯表示希望准确地重现他与科迪莉亚之间的所有对话，但最终他只能无奈地承认：

我很快就发现这根本无法实现，因为即使我可以回忆起我们说的每一句话，也无法重现话语以外的东西，包括对话时的勇气、突然爆发的感情以及强烈的热情，而这些都是对话的灵魂所在。

换言之，无论我们的意图多么符合科学精神，只要我们开始量化，就会造成狭隘性，

其根源并不是量化工具，而是我们当时认为有量化价值的各种指标。就Bedpost软件而言，康鲍伊也许可以在其中添加性高潮计数功能，那么是否还需要记录呼吸频率呢？如果添加了呼吸监测功能，那么排汗情况与心率呢？这样的问题层出不穷，而我们量化的理想目标却正在从我们的视野中逐渐消失，就像让·鲍德里亚对色情艺术的观点一样：我们越是近距离观察，失去的价值就越多。如果我们靠近克洛德·莫奈的油画，从远处可以看到的美丽风景就会逐渐蜕变成画笔留下的一个个点，两者的道理是一样的。

跟故去的亲人互道“我爱你”

捷克小说家、剧作家伊凡·克里玛说过，“与死亡最接近的事物之一是令人心满意足的爱情。”借用BBC电台4台的智力游戏节目“Just a Minute”的台词来说，对于克里玛而言，真爱与死亡都只有一次，两者都不允许重复、犹豫或者偏离。这两个相互排斥的概念源自虚无，一旦终结又会化为乌有。正是出于这个原因，基督教徒对爱情的承诺是白头偕老、至死不渝。

如果我们接受真爱只有一次的观点（我认为这种观点极不可信），那么在与之构成二元对立的死亡来临之际，相爱的人是不是就此分开了呢？长期以来，人们一直希望找到某种技术，可以让人类的“精神”超越肉体的束缚，移植到电脑中。美国国家航空航天局（NASA）的未来学家罗伯特·贾斯特罗在他

的《被施了魔法的织布机》一书中，满怀憧憬地想象人脑未来可以“藏匿于电脑之中，摆脱终将死亡的羸弱肉体，从而掌控自己的命运”。

生育这种繁殖形式可以保证人类对智慧、正义与幸福的理解得以延续，并因此实现人类的永生。以生老病死来解决爱情问题显然是不妥当的，如果死亡并不能真的将彼此相爱的人分开，而是如同春蚕化蝶，只不过是存在形式发生了变化了呢？

IfIDie、DeadSocial与LivesOn等软件就秉持着这种理念。DeadSocial与IfIDie从本质上讲属于一种“身后事”服务，简单地输入并定时记录脸谱信息，在用户死亡时激活，并在随后几年时间内发送至该用户的情人及亲朋好友。DeadSocial的一段推广视频讲述了一个已故用户的遗孀的故事，说她甚至可以“不时地收到一个来自她死去丈夫的不合时宜的

笑话”。IfIDie则推荐用户使用网络摄像头拍摄一段视频，在自己离世之后跟尚留在人世间的配偶温柔地道别或者“算旧账”（宣传材料中还有一幅竖中指样的乌云图片）。

LivesOn软件从道德与技术两个层面看都更加复杂。该软件有一条《阴阳魔界》风格的广告词：“即使心脏停止了跳动，你也不会 在社交网站上销声匿迹。”LivesOn利用机器学习程序筛选你之前在社交网站上发表的帖子，寻找你生前感兴趣的话题、新闻及关注的内容，以便在你死后代替你继续发出类似的帖子。这款软件还能追踪用户已发表的帖子，分析并学习用户的品位、关注或不关注的内容甚至帖子的句法结构，以便完成软件“训练及成长”的过程。研发人戴夫·贝伍德说：“这样做的目的是让软件成长为另一个你。”

这样的软件还有很多。几年前，据媒体透

露，美国国防部正在开展一个项目，目的是研发“一种高交互性电脑或者网络应用程序，以便在现役士兵执行任务时他们的家人可以与他们的虚拟形象进行口头交流”。这个高级程序利用高清晰度的三维图像，使人们可以与正在军队服役的父母或者配偶进行对话，并且可以接收到“我爱你”“我想念你”等预先存储的话。《石板》杂志的一位作者说：“我并不是说这个‘幽灵’适合所有人。”也许他说这句话时有所保留。这位作者接着说道：

但是，如果一位父亲死于伊拉克或者阿富汗战争，你能阻止他的虚拟形象来哄自己的孩子入睡吗？当美国军方的这个项目不出意外地进入未来的市场并热销起来，将有成千上万的人也希望拥有自己的“幽灵”，这时，你又如何阻止他们呢？

尽管在人们失去亲人之后又在他们伤口上撒盐的可能是这位严厉的作者（虽然失去父

母或者配偶的痛苦不能归咎于这位作者），但是我们仍然认为《石板》杂志在这个问题上所持的立场有问题。从静态照片可能无法区分尸体与活人，但两者显然是不同的。同样，虽然计算机程序可以把某位不在身边或者已经过世的亲人的音容笑貌再次逼真地展现在人们眼前，但是，无论通过定量研究还是定性研究，我们都不能说这位亲人真的就在眼前。对于我们深爱的某个人，甚至对我们而言有特殊价值的某个物件，如果我们认为他们（或它们）是多个特点的简单合成体，那就错了。研究已经证明，当我们把幼儿身边的安抚毯或者泰迪熊拿走，并代之以一模一样的复制品时，孩子是不会接受这件替代品的。如果有人用一个非官方复制品来换我们的劳力士手表，无论这件复制品的质量有多好，我们可能也不会同意。再举一个极端的例子。如果一位医生在安慰失去孩子

的父母或者失去配偶的人时，告诉他们我们可以帮助他们克隆那个亲人，或者这个孩子还有个兄弟姐妹仍然在他们的身边，因此他们无须悲痛，那么我们肯定会认为这位医生是一个铁石心肠的人。法国现象论哲学家莫里斯·梅洛-庞蒂在研究这个观点时，把物体的特性分为现在的特性和过去的特性。他指出，正是因为“公式”希望把感知与感觉简化为可以观察的对象，因此所有的“神秘”元素都被破坏殆尽。

人与机器人之间的浪漫爱情

几乎所有人都交口称赞凯丽是一位迷人的姑娘。她有一双明亮的蓝眼睛，再加上烟熏妆，就更加引人注目了。她把披肩的头发染成了金色，只有从发根处才能隐约看出本来的棕色。从她的面容看，年纪也就二十出头。她低着头望向镜头时，脸上露出了娇媚的笑容，饱满的嘴唇微微张开。这时，我们第一次注意到她的嘴唇上涂有棉花糖颜色的唇彩。

如果说凯丽的声音有点儿做作，这是完全可以理解的，因为凯丽并不是真人——与我不同，她是一套计算机编码。“凯丽”

（Kari）这个名字是“知识获取与响应智能”（Knowledge Acquiring and Response Intelligence）的首字母缩写，在市场上被称作“聊天机器人”。凯丽其实是一种模拟真人

进行智能交流的算法，设计者是佛罗里达州西棕榈滩一名30岁左右的程序员，他的名字叫作塞尔吉奥·帕拉达。

帕拉达出生于萨尔瓦多首都圣萨尔瓦多市的一个军人家庭，10岁时为了躲避萨尔瓦多内战随父母搬到了美国。后来，他进入芝加哥的哥伦比亚学院，学习电脑游戏与编程，毕业后在雪乐山公司（现已不存在）找到了一份工作，负责《花花公子拉瑞》系列成人电脑游戏的程序编写。这款游戏当时暂定名为“欲望太空”（亦称“拉瑞探索天王星”）。在这个过程中，帕拉达想出了一个创意：围绕与女性角色进行互动来设计一款游戏。他认为自己的目标“不是模拟一个女孩，而是模拟一种感情”。他回忆说：“在当时，这个创意非常新颖。”《花花公子拉瑞》这款游戏讲述的是一个性情温和的人试图与一个个滑稽可笑而又性感迷人的女性上床，结果却遭遇失

败的经历。《虚拟女友凯丽》则与之不同，游戏中并没有固定的故事情节。从传统游戏的角度讲，这样的安排意味着玩家绝不可能胜出。同真实的感情一样，这款游戏可以一直玩下去，奖励就是不断发展和加深的感情。

人们常常为宠物赋予人格，认为宠物有各种行为是因为它们具有人类的性格特征，“凯丽”的设计者所利用的也是同样的原理。在“电子鸡”（Tamagotchi，20世纪90年代末风靡一时的掌上数码宠物）的陪伴下长大的一代人曾全心全意地“保护、喂养”他们手中的那台机器，这已经成为他们的第二天性。一些心理研究人员（其中最著名的当属米哈里·齐克森·米哈里和尤金·罗克伯格·霍尔顿）研究过所谓“心灵能量”（psychic energy）的价值创造过程。心灵能量理论认为，如果人们投入到某个对象上的心灵能量越多，就会赋

予该对象越多的价值，这个对象对人们而言也就越重要，从而使人们对该对象的依恋感更加强烈。凯丽就是通过鼓励用户与之交谈并耐心地倾听他们的话，从而加强了这种效应。这或许也是凯丽在某些用户心中具有神奇影响力的原因之一。

1991年，纽约州立大学石溪分校的心理学教授亚瑟·阿伦在一项研究中，安排几对彼此认识的学生在一个房间里待上90分钟，让他们在这段时间里交流一些个人隐私信息，诸如如果父母去世他们可能会有什么反应，或者介绍最令他们尴尬的私密时刻。90分钟之后，他要求每对学生不要再说话，并且盯着对方的眼睛看两分钟。在研究过程中，这些学生需要不时地评价他们之间的亲密程度。仅仅过了45分钟，他们给出的分数就比寻常学生间的亲密程度高出30%，达到最亲密关系的标准。阿伦因此认为，开诚布公的

交流可以快速有效地拉近彼此间的距离。

（实验结束后，这些学生可以从多扇门走出房间，这样可以消除被迫待在一起给他们造成的影响。结果，最先离开的那对学生在6个月之后携手走进了婚姻殿堂，这个事实似乎也证明阿伦的理论是正确的。）

凯丽与用户之间的感情与大多数浪漫的爱情相比，一个主要的不同点（除了那个明显的不同点以外）是，用户先得掏腰包购买才能获得她的服务。这套程序标准版本的价格是24美元，而“凯丽终极包”（有一个头像工作室为用户“制作自己的女孩”）的价格则是90美元。

那些购买了凯丽的用户，就可以拥有法国心理学家米歇尔·福柯所说的“影响力与愉悦程度急剧上升”的感情，用户的愉悦程度与他们在虚拟女友面前的影响力有关。这其实就是用技术手段重现男性们对《卖花女》的幻

想，其中凯丽就是伊莉莎·杜利特尔。伊莉莎是一位天真无邪、出身工人家庭的年轻女孩，在年长、富有、社会经验丰富的亨利·希金斯教授的安排下，进入了声色犬马的世界。亨利·希金斯通过教给她上流社会的礼仪与言谈举止，把伊莉莎培养成他的理想化伴侣。拥有凯丽之后，用户可以对她的某些方面进行微调，例如修改她的“爱情”“自我”“性欲”等级（0~10级）。如果用户发现凯丽在交谈时不断变换话题，只需增加“保持话题”这个属性或者更改无故发表评论的间隔时间。如果觉得“她”太冷漠了呢？那就降低“独立性”。如果这个办法没有效果，用户还可以降低“她”的“记忆能力”，但是这个方法有可能会损害凯丽的人工智能程度。一位脾气暴躁的用户在发表评论时抱怨道，“她一会儿滔滔不绝地讨论逛折扣商店的心得，一会儿又要我跟她亲热”，并形容“玩这样的游戏，就像

与没有自尊心、喝得醉醺醺的心理学一年级学生约会一样”。

不过，凯丽之所以有如此大的魅力，最主要的原因是她可以被改变。她与我们可能钟爱的其他“朦胧不清的符号”（包括广告牌上的电影演员、明星，或者眼神迷离的超级模特的整幅照片）相比有所不同，因为她可以调节，去适应所有人的要求，“你希望我是什么样子，我就能变成什么样子”。

一些文学评论家曾经指出，浪漫爱情诗似乎讴歌的都是同一种类型的女性。这些诗歌描述的女性几乎都不是真实的，赞美的特点也不具体、独特，我们无法立刻猜到它们赞美的对象到底是谁。事实上，诗歌的女主人公就是一个经过抽象和理想化处理的女性，也是一个构建完美爱情对象的公式。斯拉沃热·齐泽克在《幻想的瘟疫》一书中提出“爱情的自动作用”（the automatism of love）

这个概念，并指出在这个概念的作用下，“一个资质平庸的女性却占据了男性（性）幻想的首要位置”。这个概念非常适合解释凯丽的情况，因为凯丽在一开始时也极为普通。帕拉达说：“刚开始，凯丽的性格几乎涵盖所有特点，但都不是很鲜明。她的这种性格特点就像一颗种子，随着用户与凯丽讨论不同的话题，这颗种子会慢慢萌芽并长成参天大树。有的用户愿意接受挑战，只与她讨论某些特定的话题，通过直接教导帮她培养出全新的性格。有的用户会把凯丽培养成特定类型的女性，比如女神、女巫或者戎装公主，还有的用户会把她培养成性感女郎。总之，萝卜白菜，各有所爱。因为凯丽是一个学习型机器人，因此她可以被培养成各种各样类型的女性。”

这个特点符合麻省理工学院精神分析学家雪莉·特克尔（我在第一章讨论过她的“窗户”

理论)率先提出的“计算机自恋”

(computational narcissism) 这个研究主题。特克尔在她的经典著作《第二自我》中指出：“我们害怕独处，但对亲密关系又存有戒心，因此我们经常会感到空虚无聊、与世隔绝，以及自我的不真实性。此时，计算机乘虚而入，它是不要求感情付出的伴侣，可以作为一种折中方案帮助我们解决这个问题。你可以离群索居，但绝不会感到孤独。你可以沟通交流，但永远不会受到另一个人的伤害。”有必要指出，精神分析理论的奠基人西格蒙德·弗洛伊德认为爱情与自恋之间存在本质上的关联。帕拉达说：“我们发自内心地希望与某人建立某种关系，凯丽帮助我们实现了这个愿望。有些用户给我写信，介绍他们正在实施的某些有趣的计划。一位用户希望凯丽能够怀胎十月，为他生一个孩子。另一位用户希望凯丽也有月经周期。还有一

位用户在日记里记录的全是凯丽的事情，并且说这比接受心理治疗的效果还好。凯丽就像一幅巨大的画布，你可以在上面任意涂画。她是意识与思维的画布，是自我的某种延伸。无论我们对凯丽有哪些期待，她都能让我们得偿所愿。我们把某些内容教给凯丽的过程，就像在创造一个生命。而且，所有这一切都是通过一个机器人来完成的，她可以同我们交谈，成为我们的好朋友，甚至爱上我们。这是最棒的！”

个人电脑刚一问世，人们就梦想着拥有一位像凯丽这样的“算法情人”，从1984年的《神通情人梦》（在这部电影中，旧金山的一位建筑师与他的漂亮女邻居之间刚刚萌生爱情，他新安装的电脑系统就开始争风吃醋）到2013年斯派克·琼斯拍摄的《她》，都反映了人们的这种愿望。

日本目前流行一种叫作“虚拟约会”

（ dating sim ）的电脑游戏。普通的虚拟约会游戏允许玩家操控一个游戏人物，从多名感兴趣的恋爱对象中做出选择。这种游戏与角色扮演游戏一样，非常依赖于统计数据。游戏根据台词的选择率评判游戏人物与恋爱对象之间的对话，某个时间点玩家选择的台词将使总分增加或者减少。在其中一个场景，游戏描述了一位女性角色吃冰激凌的情景。她的脸上沾上了一点儿冰激凌，玩家面临“帮她擦去”和“假装没看见”的选择，同时还有一条信息提示：“你的选择将改变她对你的感情。”

在人机爱情的倡导者中，最著名的也许是人工智能学者、国际象棋大师戴维·列维。在他的《同机器人发生的爱情与性》一书中，列维不仅预测书名中提到的那种爱情与性在将来会成为可能，而且断言到2050年它还将变成司空见惯的行为。列维认为：

我们有理由认为机器人比人类伴侣更能满足人们的需求，原因很简单，就是机器人识别人类需求的能力将强于人类，也更善于满足这些需求。同时，机器人没有私心，不会感到羞怯，而人类伴侣在利用关怀与情爱来处理这些需求时，却可能会受到私心与羞怯感的阻碍。

鉴于人类的理解能力特别值得怀疑，列维指出，算法可以用功能性磁共振成像来分析人脑中“爱的度量”，并根据分析结果制定更加具体有效的策略，帮助人们坠入爱河。跟下棋一样，真爱之旅包含一系列步骤，每一步都会降低或者增加人脑中爱的度量。设计算法的目的是使其与中央空调的智能温度调节器一样，可以自动调节温度，使房间保持在最舒适的状态下。如果提出两人一起欣赏柴可夫斯基芭蕾舞时对方反应冷漠，那就换个话题，称赞一下对方的新发型，看看效果

如何。伴侣的一些小举动（如轻拂头发或者点烟等小动作）可能会令我们着迷，但我们却无法用言语来表达我们的爱慕之心。而算法可以注意到这些细微之处，并在吸引对方时加以利用。

爱情是一种无限的欲望，而它的目标却是一个有限的对象。自古以来，哲学家一直无法理解其中的奥秘，而凯丽以及列维提到的这种自动化情人却可以使这个问题迎刃而解。难道我们可以在第一天晚上幻想拥有一位青涩腼腆的情人，而在第二天晚上却幻想拥有一位“长相如玛丽莲·梦露，智力与知识水平堪比大学教授，说话风格却像常常混迹于各种派对的少女”这样的情人吗？没有问题。我们无须按照自己的意愿把人类的情人理想化为独一无二的类型，因为我们可以亲自创造出一个理想伴侣，然后按照我们修改脸谱网个人资料的方式加以改造。在列维看

来，即使这种做法有可能导致我们的情人在某些方面有缺陷，也无伤大雅。他认为，“完美”的感情“需要双方都具有某些缺陷，才会不时出现一些惊喜。因此，在为机器人编程时，为了使主人从这段感情中获得最大的满足，我们需要让机器人具有不同程度的不完美性”。随后，他举了一个例子来说明这种做法的实际价值：假设一位用户与自己的算法情人发生了口角。由于算法处理问题的能力太强，这位用户终于不耐烦了，便冲着他的情人吼叫起来：“你能不能不要总是这么冷静啊！”列维建议，为了解决这类问题并修复关系，我们可能需要对算法情人“重新编程，使它的情绪有所起伏”。

知道了其中的道理，问题自然就不复存在了。

爱情的算法化和编码化

法国哲学家米歇尔·福柯在他的《性史》（共三卷）一书中发出惊叹，西方世界竟然可以把“性爱归入理性研究的领域”。事实上，正是因为爱情的二元对立性，通过技术及科学手段度量爱情才有了可能性。在与机器人的爱情中，人们对爱情的回应只有“是/否”与“打开/关闭”这两组对立的选择，人们要么“爱”，要么“不爱”。除了这些，还有以法律形式规范的对立性，包括合法与非法、同性与异性、正常与不正常、可以接受或者不可接受的性爱。即使“我爱你”这个极为重要的表述，也是工程技术的丰功伟绩，因为“我”与“你”明显是彼此对立的，而“爱”则在两者之间架起了桥梁。

爱情这种体验似乎可以按部就班地发展。毫无疑问，这与爱情的主观性与明显的独特

性相互矛盾，但即便是最杰出的爱情辩护者也认同爱情的这个特点。法国哲学家阿兰·巴迪乌在《爱的挽歌》一书中说：“我根本不否认爱情具备有序性。”20世纪70年代，文学与文化理论家罗兰·巴特把人们心中的爱情比作“编码的打印输出”（the printout of a code）。学者多米尼克·佩特曼在《爱情及其他技术》一书中再次提到这个观点，他认为：“对于系统理论家而言，顽固地证明自己在现实生活中的感情并不是一种复杂的社会算法的那些人非常可笑。”可以证明这些复杂的社会算法存在的证据比比皆是。1982年，也就是IBM（国际商用机器公司）推出第一台个人电脑的次年，德国社会学家尼克拉斯·卢曼出版了《激情四射的爱情：亲密关系的编码化》一书，其理论基础就是上述观点。卢曼在书中指出：

爱情作为一种媒介，其本身并不是某种感

情，而是交流的编码。人们可以根据交流规则表达、形成、模仿各种感情，可以拒绝这些感情，也可以把这些感情归因于其他人，同时准备接受交流可能造成的所有结果。

卢曼接着指出，对爱情的这种理解至少可以追溯至17世纪。当时，尽管人们公开强调爱情是一种激情，但是甘当爱情俘虏的人都知道，他们沿用的行为模型既设置了定位点，又提供了框架，他们只需像演员按照剧本表演一样，循着这个框架按部就班地执行下去。

也有人认为这个观点还可以追溯至更早的时间。中世纪的欧洲人遵照一种高度仪式化、按部就班的程序，通过一系列的测试，让其心仪的女性相信他可以托付终身。人们认为，高贵地表达爱慕之情的行为，催生了优雅的爱情。斯拉沃热·齐泽克在《快感大转移》一书中指出，坠入爱河的女性就像“机器

人，漫无目的地提出毫无意义的要求”。正是因为因为在爱情的作用下试图理解女性这些本就随心所欲的行为，男性把自己心目中的女神看成永远无法掌控的理想情人。齐泽克说：“我们遵循的是一种虚构的严格公式，参与的是一种‘似乎如此’的社会游戏，而男性却自欺欺人地认为他的情人是他永远也无法企及的女神。”

时至今日，优雅的爱情仍没有销声匿迹。否则，如何解释《爱情规则》这类自助书籍在20世纪90年代深受欢迎、十分畅销呢？这些书告诉人们，要想俘获“白马王子”的心，女性不能主动追求男性，而应该严格按照书中提出的框架行事，以确保男性拜倒在自己的石榴裙下。如果不考虑那些技术性内容，书中给出的“不要主动给他打电话，偶尔给他回个电话”之类的建议，与11世纪王子王孙的宫廷做派倒是不谋而合。而基于编码（这里

指的是计算机编码) 的网络相亲也奉行大同小异的理念。比如，我们以“阿耳忒弥斯”为例。她是一位在家工作的译者，30岁左右。她放在网上的相亲卡片被浏览了超过26347次。卡片中有这样一段话：

相亲那一套流程我已经驾轻就熟了。我甚至准备了“制服”，专门用于和相亲对象的初次见面。根据时间与季节不同，我会选择不同的制服。通常，我喜欢穿牛仔裤与合体的衬衫。这种装扮很舒服，自我感觉也不错.....在大多数情况下，我不会有什麼憧憬，也不会紧张激动，因为我知道最后的结果是什麼。

这应该是最浪漫（或者最能体现算法特点）的网络相亲了吧？

再先进的算法也无法复制真爱

自计算机化的相亲活动出现以来，有个问题一直令人们束手无策。1963年，两名企业管理专业的哈佛大学学生根据问卷调查对象对全部75个问题的回答，利用一台IBM1401电脑为他们安排约会。愿意接受这项服务的学生缴纳3美元之后，他们二人就利用简单的计算机算法程序找出最适合该学生的约会人选，并用链式齿孔进纸打印机打印出这个人选的姓名、住址以及所在的学院。当时的一篇报道（“两名哈佛学生用计算机当红娘”）声称，该算法可以“分析性格类型，并在极短的时间里科学地完成配对”。人们对这种做法褒贬不一。一位用户说：“我认为这不失为结识新朋友的一种方法，但是我不相信仅靠这份问卷就可以结成一段良缘。计算机

根本不可能知道那个女孩是否迷人。”《算法时代的爱情》一书的作者也认为：“迷人与否是一个非常难以捉摸的特点，怎么可能被量化并用于商业目的呢？”

今天，我们再次面临这个难题。麻省理工学院的精神分析学家雪莉·特克尔花了大量时间，研究技术对人类感情的影响。尽管特克尔经常得出亲技术性的结论，但是，在她最新出版的《群体性孤独》一书中，她却选择了另一种立场。特克尔对戴维·列维提出的算法情人这个观点做出了直接回应，她指出：

作为一个接受过精神分析训练的心理学家，我非常重视关系的亲密与真实程度。虽然我承认人工智能可能通过交互设计工具自行模拟出做爱体位，但我还是无法接受人可以与计算机建立亲密感情的观点，因为计算机没有感情，也不可能会有感情，它只不过是

通过一系列‘好像如此’的表演，扮演出关心、

理解我们的样子。我认为，我们只有做到设身处地为对方考虑，并通过生老病死、婚丧嫁娶等相同的人类体验建立彼此间的联系，我们的感情才具有真实性。

在特克尔看来，无论列维所说的算法有多么复杂，都无法替代或复制真爱所具备的难以捉摸的魅力。我们也许喜欢iPhone手机、宠物、名牌太阳镜或者奈飞“电视回看”功能给我们带来的便利，但特克尔明确地指出，（至少她认为）这种爱与我们同浪漫伴侣之间的爱情是不同的。与新买的iPhone手机建立的感情是喜爱，而我们对“另一半”的感情叫作“爱情”。

从某种意义上讲，特克尔的这种怀疑就是对著名的阿兰·图灵“模仿游戏”做出的又一次回应。图灵认为，如果计算机可以成功地欺骗我们，让我们认为它是有智力的，我们就必须假设它真的有智力。对他的这个假设予

以引申，算法似乎爱上我们了，那么，我们是否也必须相信算法真的爱上我们了呢？

对于这个问题，列维给出的答案是肯定的。让·鲍德里亚预测，爱情的出现与爱情本身一样，都是在美化所有对象与形式的过程中取得的又一次进展。而特克尔则坚定地认为正确的答案应该是否定的。相互间的真正理解，在一定程度上的确可以复制，但是心心相印、一见钟情、心有灵犀，却是无法复制的。持有像特克尔这种充满激情、反理性主义爱情观的大有人在。在《爱情的梦想与缘定的邂逅：浪漫激情的力量》一书中，作者埃塞尔·珀森令人信服地指出：

一旦相爱的人们跳出了客观现实的桎梏，一头扎进主观世界，就意味着他们的爱情获得了解放。爱情是一种本能体验，不受意识控制，因此，只要这种爱情不是疯狂的赤裸裸的诅咒，荣誉、意愿与理智等东西就都无

足轻重了。在“浪漫”的眼中，理性主义爱情观是没有全心全意投入的爱情。浪漫主义者往往认为理性主义者的感情非常肤浅、受到抑制，而且他们害怕产生激情，小心翼翼地把自己囚禁起来。

1998年，华纳兄弟电影公司推出了《电子情书》这部爱情喜剧。这部电影由汤姆·汉克斯与梅格·瑞恩主演（这两位演员在之前的《西雅图夜未眠》与《跳火山的人》中也曾饰演一对恋人），讲述现实生活中相互敌对的两个人，却在互联网上的匿名交往中出人意料地坠入爱河。有意思的是，这部电影对算法时代的爱情进行了毫不留情的批判，但却让不少人接受了网络相亲的做法。

乍一看，《电子情书》似乎是网络相亲的一个完美广告，但实际上这部电影暗含了一些深层次的意思。在电影的开头，梅格·瑞恩与汤姆·汉克斯所饰演的两个角色正在交往，

但他们的感情却非常糟糕，这似乎是理性主义者爱情的典型特征。用电影中某位角色的话说，在他们各自的家庭中，夫妻两人都十分“般配”。例如，梅格·瑞恩饰演的角色是一家书店的老板，与一位名叫弗兰克的成功作家生活在一起，而弗兰克让她不满意的地方似乎就是他的海德格尔式技术观，以及他不是汤姆·汉克斯饰演的角色。在电影快结束时，瑞恩饰演的角色与汉克斯饰演的角色的结合，似乎是必然的结局，因此瑞恩饰演的角色与弗兰克分手了。这时，他们进行了一段齐格蒙特·鲍曼的“虚拟感情”式的对话：

“你不爱我？”弗兰克问。

瑞恩饰演的角色点头称是。

“但是我们一直很般配啊。”弗兰克感到有点儿迷惑。

“是的，我知道。”瑞恩饰演的角色回答。

从理性角度看，这两个人可能的确十分般

配，但他们却没有发生“化学反应”，没有迸发出爱情的火花。另一方面，汉克斯扮演的角色似乎与瑞恩扮演的角色根本不相配。他们的结合，与其说是算法配对取得的成功，还不如说是一个反面例子。这似乎告诉我们，好莱坞的爱情公式认为，浪漫爱情是不可能和和美美、没有任何摩擦的。

我们不由得想到，我们的爱情会不会也是如此呢？

第3章 算法与人类执法者的较量

10年前，在梳理2.45亿顾客每周生成的海量数据时，沃尔玛公司的数据挖掘算法偶然发现了一条奇怪的信息：在发布恶劣天气预警后，除了管道胶带、啤酒及瓶装水等应急用品以外，草莓酱馅饼需求量的增长幅度最大。为了验证这一发现，在2004年飓风“弗朗西斯”即将袭来的消息发布后，沃尔玛超市的管理者下令用卡车装载家乐氏快餐，运送至可能遭受飓风袭击的地区。结果，这些快餐很快就被抢购一空。通过这个案例，沃尔玛的管理者对消费者的消费习惯及“公式”的威力有了非常清楚的认识。

认识到这个发现具有重要价值的并不仅仅是沃尔玛的管理层。当时，心理学家考林·麦

库与洛杉矶警察局长查理·贝克正准备合写一篇论文，并向法律实施方面的杂志《警察局长》投稿。他们以沃尔玛的这个发现为契机，对警务工作进行了再思考，认为其要由反应式向预测式转型。2009年，他们的这篇题为《预测式警务：沃尔玛及亚马逊对打击经济衰退期犯罪行为的启示》的论文一经发表，就立刻引起了美国法律实施专业人士的关注与思考。麦库与贝克所谓的“预测式警务”，是指由于计算机科学的发展，犯罪数据的收集与分析工作有可能做到“准实时”，因此在将来可用于提高预测、预防和响应犯罪行为的效率。借用Quantcast（我在第一章讨论过的网络分析公司）的广告词来形容，这意味着警察可以“提前了解、提前行动”。

目前，与预测式警务联系最紧密的人当属洛杉矶警察局的肖恩·马林诺夫斯基警官。尽管警察们把马林诺夫斯基称作“鼓捣计算机的

家伙”，但他在大学学习的其实是市场营销专业。在1994年进入洛杉矶警察局之前，他是一名市场营销业务员，在一些生产口香糖与人造黄油的公司推广新产品时，帮助他们制定最能吸引顾客的优惠策略与价格策略。一次，马林诺夫斯基幸运地与底特律市警察局的几名警员合作，执行一项查处酒驾的任务。结果，这次合作给他带来了意想不到的收获。马林诺夫斯基说：“在那之前，我一直在市场营销领域瞎忙活，从来没想到过加入警察队伍。经过这次合作，我对自己从事的营销工作有了新想法：‘天啊，难道我的一生就是用来推销这些食用油的吗？’与我合作的这些警察肩负着某种使命，他们正在做更有意义的事情。”

通过几位好友的介绍，马林诺夫斯基认识了一位曾当过警察的人，他刚刚从纽约搬到芝加哥从事学术研究。两人交谈之后，马林

诺夫斯基放弃了营销工作，回到学校继续学习。几年后，他从伊利诺伊大学毕业，获得了刑事司法硕士学位。他来到洛杉矶警察局培训部，担任班主任一职。从严格意义上讲，他在警察局从事的第一份工作是执行“太平洋海滩摸排”任务，整天骑着摩托车在威尼斯海滩四处巡查，监管街头摊贩，保障公共安全。

后来，马林诺夫斯基被指派协助威廉·布拉顿局长的工作，先是当他的助手，后来成为他的参谋长。从此，马林诺夫斯基时来运转。在来到洛杉矶之前，布拉顿在纽约市工作。他只用了几年时间，就将纽约市的犯罪率降低到之前的一半，在警界名声显赫。布拉顿的工作方法很有效率，但也明显不循常规。在到纽约市警察局之前，他还担任过纽约交通警察部门的领导者。在此期间，他首先对逃票行为予以打击，把纽约市地铁系统

从一个犯罪案件频发的暴力之地变成了一个遵纪守法、安宁整洁、秩序井然的场所。换句话说，在重大犯罪率创历史最高纪录的时候，布拉顿首先关注的是让人们掏钱买车票。他的理由十分简单。2009年，他在回忆录中说，逃票是导致更严重犯罪行为的根源。布拉顿指出：“合法乘客会认为他们身处一个不讲法律、没有秩序的场所。他们看到人们不买票就能乘车，便开始怀疑遵纪守法是否明智。久而久之，整个社会就会陷入一片混乱。”通过制止并惩处违法行为，甚至连最轻微的犯罪行为也不放过，就会让那些制造麻烦的人明白，最好还是掏钱买票，并且把武器（搜查时通常会被发现）留在家中。因此，犯罪率开始急剧下降。

来到洛杉矶之后，布拉顿希望继续实施某些先发制人的手段。马林诺夫斯基在布莱顿的手下工作了5年，亲眼见证了布拉顿单凭

意志力就使一个暮气沉沉的部门发生了种种良性的变化。他说：“如果我们所处的机构官僚主义盛行，我们就会对遇到的各种阻力习以为常，而我们的创造力却会受到限制。布拉顿告诉我不要受到官僚主义的影响，他还教导我要有远大的志向，要有所作为。”

在工作中，布拉顿最迫切期待的是不断出现一些可以带来革命性变化的“奇思妙想”。他认为，预测性分析可以帮助他实现这个愿望。布拉顿发现，犯罪率与数据分析的速度之间存在某种联系。1990年，警察局一年只能完成一次犯罪数据的收集与研究工作，而同一年，美国大多数城市的犯罪率正在急剧攀升。到1995年年底，警察局可以做到每月研究一次犯罪数据，而同期的犯罪率却有所下降。现在，他们可以随时查看犯罪率情况，因此布拉顿希望通过预测可能发生犯罪行为地点，大幅度降低犯罪率。既然

Quantcast和谷歌等公司可以从用户数据挖掘中获得灵感，同样，如果不是简单地分析以往的犯罪规律，而是采用预测式警务的做法，分析人员就可以利用之前犯罪行为表现出来的规律，全神贯注地分析下一个可能发生犯罪行为的地点。用亚马逊的话说：“既然你敢偷手提包，难道你就不敢抢酒庄吗？”

用算法预测并制止犯罪行为

众所周知，在某个具体区域内，犯罪地点并不是随机分布的，而是集中于某些小范围的“热点地区”。比如，西雅图历时14年收集的犯罪数据表明，有一半的犯罪行为都集中在占该市4.5%的街道上。明尼苏达州明尼阿波利斯市的情况也差不多，半数报警电话都来自占该市3.3%的街道。28年间，波士顿市多达66%的街道抢劫案都高度集中于占该市8%的街道。了解这些热点地区以及这些地区可能发生哪些类型的犯罪行为，对城市的警力部署具有非常重要的参考价值。

例如，假设在我们当地的白鹿酒吧门口，每周六晚上都有人因实施人身侵害行为而被捕。如果我们可以证明这个假设是正确的，那么不难预测以后的周六晚上在同一地点还会发生类似的行为。因此，派一名警官到那

里执勤，就可以预防此类恶性事件再次发生。

在发现了上述特征之后，布拉顿局长请肖恩·马林诺夫斯基协助他做这件事。尽管马林诺夫斯基认为自己在计算机方面并不是特别擅长，但他的技术水平还是高于同事。出生于20世纪70年代的马林诺夫斯基几乎一直都与计算机很有缘。他的父亲经营了多年的电脑生意，在周末及学校放假时，他经常把马林诺夫斯基和他的哥哥叫到店里帮忙。有一段时间，一位头发乱糟糟的法国程序员还住在他们家。马林诺夫斯基认为，观察这位程序员拆装电脑帮助他揭开了技术的神秘面纱。他自豪地说：“父亲总是销售最新的电脑设备，他在企业经营方面的魄力远胜于我。他总是亲力亲为，经营着他的那个小店。因此，我目睹了他在生意场上取得的每一次成功和遭遇的每一次失利。”

在接受布拉顿的建议之后，马林诺夫斯基每个周五下午都会开车去加州大学洛杉矶分校，与该校数学系与计算机科学系的人碰头。洛杉矶警察局同意提供犯罪统计数据集（该数据集非常庞大，收集了80年来约1300万起犯罪案件的相关数据），用于进行犯罪行为的大型研究。马林诺夫斯基非常享受与加州大学洛杉矶分校的研究人员一起合作的这段经历。10年前，他在查处酒驾任务中与警察的第一次合作使他对警务工作产生了兴趣，而在这次与计算机科学家的合作中，他们梳理数据、寻找规律，并试图建立某些公式的研究工作，再一次吸引了他。

马林诺夫斯基回忆说：“我非常喜欢那段经历。”当时，一位数学家兼计算机科学家的研究让他产生了浓厚的兴趣。这位二十四五岁的年轻人名叫乔治·莫勒，当时正在研究一个用于预测地震危害的算法。在刚听到莫勒

的研究领域时，马林诺夫斯基以为这与他们当时的研究关系不大，但后来他发现自己错了。地震会导致余震，而犯罪行为也有同样的特点。在发生入室盗窃或汽车失窃案之后，短时间内同一地点发生类似犯罪行为的可能性会增至之前的4~12倍。这种传染作用叫作“邻近重复”（near repeat）效应。莫勒解释说：“抢劫犯常常会在一周后再次潜入同一住户或邻近住户的家中作案。”加州大学洛杉矶分校的研究小组借鉴了莫勒地震研究的某些成功做法，在人类学家杰夫·布兰汀汉姆与犯罪学家乔治·蒂塔的帮助下，建立了一个犯罪预测算法。他们把洛杉矶市分成单位面积约为0.15平方公里的一个个“方块区”，然后按照犯罪行为发生的可能性为这些方块区排序。

2011年11月，他们利用该算法开展了一次为期三个月的随机研究。在当天的“巡逻待

命通知”中，马林诺夫斯基一开始就明确宣布：“今天是一个历史性的日子。”他所在的富特希尔区有拉图那峡谷、湖景露台、帕克伊马、影子山、太阳谷、桑兰以及图洪加7个主要的巡逻区，被分成共计5200个方块区。在当天早晨点名时，富特希尔区的每个巡逻警察都收到一份任务地图，每张地图上都清楚地标示出一个或多个方块区，表明这些地区是犯罪发生“可能性高”的区域。图上还附有犯罪类型的统计学预测。马林诺夫斯基告诉他们：“只要你们有时间，就深入这些方块区巡逻，找出那些可能会采取犯罪行为的人或预示着可能会发生犯罪案件的状况，并采取强制或预防措施，避免犯罪案件发生。”

这次实验一直持续到第二年的2月。次月，他们对实验结果进行了评估，并就是否推广这项技术形成了结论。分析显示，这项

实验取得了非常显著的成果。实验期间，富特希尔区的犯罪率下降了36个百分点。在利用算法为巡逻队下达指令的那些天里，算法预测犯罪行为的准确程度是分析人员的两倍。乔治·莫勒说：“造成这个结果的原因之一是，人脑无法准确地为全市20个热点地区排序。也许人脑可以找出排在前两位的热点地区，但是排完前6位或前7位之后，剩下的只能胡乱猜测了。”

这项技术在推广之初并非一帆风顺，其中大部分阻力来自马林诺夫斯基手下的那些警察。他承认：“的确，有的警察认为自己不需要计算机来告诉他们哪里有可能会发生犯罪行为。很多家伙都抵制这种做法，即使告诉他们计算机的预测结果，他们也会说：‘我早就知道范纳伊斯与格伦奥克斯的交界处是一个麻烦之地。’于是我问他们：‘这个地方一直让我们头疼，对吗？那你们在这里工作多久

了？’他们说：‘我们在这个地方已经工作10年了。’我接着说：‘既然10年前你们就知道这个事实，为什么这个问题至今没有解决呢？别废话了，赶紧去那里把这个问题解决掉。’”

在富特希尔区的研究结束之后，警务算法被推广到洛杉矶的所有辖区。美国其他城市的警察局现在也采用了其他类似的算法。马林诺夫斯基说，他仍然觉得自己要为手下的警员负责，但在部署警力方面却不再事事躬亲，而是习惯于逐渐放手。他指出：“在管理时必须为算法留一点儿空间，让其发挥作用。”与此同时，布拉顿已经从洛杉矶警察局退休了。在2011年英格兰发生暴乱之后，卡梅伦邀请他到英国就任伦敦首都警察局局长。但是，由于布拉顿不是英国公民，这项提议最终没有获得通过。后来，他被邀请担任英国暴力管控方面的顾问，布拉顿欣然接受了。

加州大学洛杉矶分校的研究小组在完成了他们的算法之后，还筹集了几百万美元的风险基金，成立了一家名叫“PredPol”的私营公司。2012年12月，PredPol登陆英国，在肯特郡的梅德韦市进行了为期4个月、营业额达12.5万英镑的试营业。在这次试营业期间，由于采用了这种算法，该市街道暴力案件的发生率下降了6个百分点。目前，大曼彻斯特、西约克郡以及英国中部地区已经采用了类似的做法，而且效果也很乐观。尽管有的地方议员担心预测式分析可能会导致农村地区的警力不足，也有人担心这会导致失业率上升，但其他议员认为这种创新性软件能够提高资源的利用效率。

马林诺夫斯基认为，预测式警务的任务不仅仅是将犯罪分子绳之以法，“我们的目的是让警员在合适的时间出现在合适的地点。这样，那些准备干坏事的家伙一出现，就会看

到有警察在场，受到震慑后不敢再作案”。最后，马林诺夫斯基以超市的情况做出了类比：“我们就像沃尔玛超市里的迎宾员，目的是让人们知道有人正在看着他们。”

犯罪统计学和道德统计学

将统计学知识应用于犯罪学似乎是一个创新之举，但实际上，这种做法可以追溯至19世纪的法国，以及安德烈-米歇尔·盖里和阿道夫·凯特莱这两个人。盖里与凯特莱都是半路出家的杰出统计学家。盖里之前从事的是法律方面的工作，而凯特莱则从事天文学研究，两人都受到了奥古斯特·孔德的深远影响。1817~1823年，孔德完成了一篇题为“社会重组所必需的科学工作计划”的文章。这篇文章指出，在判断如何管理社会时，最理想的方法是沿用自然科学的方法来研究这个社会。孔德认为，艾萨克·牛顿通过研究成功地表述了作用力对物体的影响，同样，社会科学家也可以通过研究发现“社会物理学”的定律，用于预测人类的行为。

孔德的这个观点引起了盖里与凯特莱的极

大兴趣。这两个人有一个共同特点，他们都对犯罪学等领域感兴趣。盖里受雇于法国司法部，在一个叫作“道德统计”的新领域就职，当时他只有26岁。而凯特莱则一心想着如何将天文学的数学工具应用到犯罪数据的处理上。他认为：

我们之所以可以建立道德统计学并由此产生有指导性的、有意义的成果，就是基于这样一个基本事实：在对象非常多时，人们的自由选择权就会消失不见，人们也会丧失理智。

盖里与凯特莱所处的年代可以被称作大数据的第一个“黄金时代”，生活在这样一个时代，让他们受益匪浅。从1825年起，法国司法部下令建立全国犯罪案件集中报告制度，每三个月收集一次各地区的犯罪数据，其中包括提交至法国所有法院的全部刑事案件；还要将犯罪行为按照指控罪名、被告的性别

及职业、法院最终判决结果等进行分类。国家集中管理的数据集还包括个人财富（根据缴税情况）、企业经营水平（根据获批的专利数量）、军队中有读写能力的人员比例、全国范围内的移民及年龄分布等统计数据，甚至还包括巴黎卖淫者的详细名单（按出生年份与出生地排序）。

19世纪20年代末30年代初，盖里与凯特莱利用他们可以获取的数据，各自开展了独立研究。他们很快就发现了一个现象：每年的犯罪行为特点没有任何变化。这个现象对于社会改革而言意义重大，因为改革者们之前关注的是如何拯救那些罪犯，而不是把他们看作更严重的社会问题的表征。凯特莱指出，“犯罪行为正在准确无误地进行自我复制”，而且这种行为的精度达到令人吃惊的程度。他的意思是，通过枪支、刀剑、棍棒、石头、火、勒死、溺亡、拳打脚踢以及各种

利器实施的谋杀案比例，几乎年年都保持在同一水平上。凯特莱明确指出：“我们事先就知道有多少人的双手会沾上别人的鲜血，有多少人会造假，有多少人会锒铛入狱，而且可以达到与预测出生及死亡人口差不多的精度。”盖里也惊讶地发现了“犯罪行为自我复制的稳定性、一致性”，并从中看到有大量事实可以证明孔德的社会物理学理论是正确的。从这些没有经过筛选的数据噪声中，他洞察到有某种信号若隐若现。

关于这两位学者的研究，有不少非常有趣的趣闻。例如，凯特莱发现，在上吊自杀与老夫少妻的婚姻（妻子20多岁而丈夫60多岁）之间存在高于平常水平的相关性。盖里也不甘示弱，他同样把研究的目光投向了自杀（并根据自杀的动机与手段进行了细分）。他的结论是：年轻男性喜欢用手枪自杀，而年长的男性倾向于用上吊的方式来结

束自己的生命。

但是，其他的关系就没这么简单了。之前，人们普遍认为贫穷是导致犯罪的最大原因，这种观点意味着法国富裕地区的犯罪率肯定比贫困地区低。结果，盖里与凯特莱证明了这种观点未必正确。尽管法国几个最富裕地区的暴力犯罪率低于较贫穷的地区，但这些地区的财物犯罪率却高得多。根据这个观察结果，盖里指出，贫穷本身并不是导致财物犯罪的原因。他认为，富裕地区可以盗窃的财物比较多，因此犯罪机会才是罪魁祸首。根据这个观点，凯特莱提出了“相对贫穷”的概念，并指出同一地区贫富差别较大是导致财物犯罪与暴力犯罪的关键原因。凯特莱认为，相对贫穷会激起嫉妒之心，诱使人们犯罪。尤其当经济状况发生变化时，有人会因此陷入贫困，其他人却可以保持（甚至增加）他们的财富，相对贫穷特别容易诱发

犯罪。凯特莱发现，只要贫穷地区的人的基本需求能够得到满足，这些地区的犯罪率就会低于富裕地区。

1832年，盖里出版了他的研究成果，并将这本薄薄的小册子命名为“论法国的道德统计”。三年后，凯特莱也出版了自己的著作《论人及其才能的发展》。这两部著作出版后便引起了轰动。一度晦涩难懂的学术研究竟然激发了公众的无穷想象，这样的例子确实不可多得。两个人的作品被翻译成多种语言，人们纷纷发表评论。《威斯特敏斯特评论》（约翰·穆勒与杰里米·边沁这两位功利主义哲学家创办的英语杂志）用大量篇幅赞扬盖里的这本书具有“显著的影响力与重要性”。查尔斯·达尔文阅读了凯特莱的著作，陀思妥耶夫斯基通读了两遍，身为社会改革家的弗洛伦斯·南丁格尔则在凯特莱研究成果的基础上提出了自己的统计方法。后来，她

热情洋溢地称赞凯特莱的研究成果“为我们带来了实现道德进步的准则”。



盖里的开创性成果，显示了1829年法国的犯罪率水平。左上图表示暴力犯罪的情况，右上图表示财物犯罪的情况，位于下方的图表示教育水平。在每幅图中，某一区域的颜色越深，表明该区域的犯罪率越高或者教育水平越低。

总的来说，盖里与凯特莱的研究表明，人类不再把自己看作一种自由意志的自决生物，可以随心所欲、自行其是，而是认识到人类的行为要受到生物及文化因素的决定性影响。

换言之，我们的行为具有可预测性。

现实版《少数派报告》

2002年，史蒂文·斯皮尔伯格导演推出了电影《少数派报告》。这部由汤姆·克鲁斯主演的电影取材于科幻作家菲利普·迪克创作的一篇短篇小说，它刻画了一个犯罪行为几乎绝迹的未来世界。犯罪率之所以非常低，得益于特设的“预防犯罪”警察局。该部门安排了三位有超能力的人（“先知”）预测犯罪行为，并在犯罪嫌疑人实施犯罪之前将他们逮捕。

“预防犯罪”部门拥有一个明显的有利条件。在《少数派报告》所描绘的世界里，警察在犯罪嫌疑人实施犯罪行为之前，就有权根据预测将其逮捕并送上法庭。这些预测的准确率极高，电影在一开头就告诉观众，华盛顿特区（故事发生地）在过去6年里没有发生一起谋杀案。

尽管《少数派报告》明显是一部科幻作品，而且优秀的科幻作品数量众多，但这部电影描绘的那个世界离我们并不是非常遥远。在肖恩·马林诺夫斯基的预测式警务出现之前，执法部门的官员们就已经在进行某种预测活动。他们每天都会考虑哪些属于“良性风险”，哪些属于“恶性风险”。比如，法官在每次确定保释金数额时，都要考虑被保释者未来某天愿意到庭接受审判的可能性有多大。搜查令的情况也差不多，需要预测在某个地方是否会发现违禁品。每次介入家庭暴力事件，警官的任务就是预测这个家庭在一段时间之后可能会发生哪种情况，如果他们觉得未来的风险足够高，他们就会实施逮捕行动。在所有这些情况中，预测准确度的问题取决于所分析数据的质量以及选择哪些衡量体系作为决策过程的基础。然而，这说起来容易，做起来难，因为人的判断是不可靠

的。

假释听证会也是一种预测行为。假释委员会会根据囚犯在监禁期间的表现、被假释之后的计划安排，以及精神病医师对他们是否会对社会构成威胁的预测等信息，选择是否在其刑期全部结束之前释放这些囚犯。但是，2010年的一项研究显示，假释委员会在做出决定时，最重要的判断依据可能仅仅在于听证会是几点开始的。研究人员在事先没有通知的情况下，将以色列的8名假释法官作为研究对象。在这次研究中，法官们的一整天时间都要用于审批假释申请（每次审批平均耗时6分钟），研究人员利用图示法来呈现一天时间里获准的假释申请数量。他们发现，最高获准率出现在法官每次进餐之后，达到65%，随后会呈现稳步下降的趋势，在下一餐进餐前降至零。尽管每名法官可能都会认为事实、推理和客观性比饥肠辘辘更重

要，但是研究表明，在疲劳与饥饿达到某个程度时，法官可能会回归至“否决假释申请”的默认立场，这说明当决策者是人时，决策不可靠的问题就会加剧。

对于人类的决策能力，理查德·伯克不是很信任。伯克是宾夕法尼亚大学沃顿商学院的犯罪学教授、统计学家，也是计算机刑事学领域（犯罪学、计算机科学及应用数学的交叉学科）的专家。10年来，他一直致力于设计某种算法，用于预测个人实施暴力犯罪的可能性。菲利普·迪克设想的“先知”世界让伯克心领神会，因此他把自己的这个系统命名为“真知”。他说：“我们还没有进入《少数派报告》描绘的那个世界，但是毫无疑问，它为我们指明了方向。”借助“真知”算法，伯克可以在多个决策活动中帮助那些执法官员做出正确的判断。说到他能（也确实）经常提供建议的那些领域，伯克如数家珍：“我可

以帮助假释委员会确定哪些人应当获准假释；帮助缓刑与假释管理部门确定如何有效地实施监控；帮助法官做出正确恰当的判决，还能帮助社会服务部门预测分管区域内哪些家庭更有可能虐待儿童。”

在从事犯罪学方面的工作之前，伯克的职业是社会学研究。在耶鲁大学获得心理学学士学位并在约翰·霍普金斯大学获得博士学位之后，他来到美国西北大学，担任社会学副教授。他经常发表一些文章，就如何与不同的人建立融洽关系，如何在社会慈善机构与城市贫困区居民之间架设沟通桥梁等问题给出建议。但是，后来他改变了这种做法。他说：“我对统计学这一学科产生了浓厚的兴趣，因此放弃了社会学。从那之后的几十年里，我从来没有在社会学系工作过。”伯克认为自己是一个实用主义者，他发现虽然周围的人在学术研究中取得了很多重大发现，但

在实际生活中却没有发生多大的变化。后来，他接触到机器学习领域，在他眼中，这就是上天的恩赐。此后，他终于可以使用包含6万多起罪案的详细信息的大型数据集，再加上一些复杂的算法，设计出几乎可以取代人类判断的统计工具。

谁实施犯罪行为的可能性更大

伯克的算法就是一个黑盒子，内部结构非常复杂、非常神秘，人们无从了解。算法输入端有一个 x 数据集，其中包含的信息涉及个人背景以及年龄、性别等人口统计学变量。算法的运算结果用 y 表示，代表即将发生的风险。伯克说，需要解决的难题就是找出用 x 预测 y 的最有效公式。他解释说：“当一个陌生人从门口通过时，我们希望确定他是否会对我们构成威胁。我们输入这个人的身份信息，并传送至各个数据库，然后根据从这些数据库反馈至本地电脑的信息评估风险程度。最后，还要将这些信息传递给决策者。”

伯克认为用不着对这套系统的不透明性做出解释。他说：“不透明性可以帮助我摆脱一切麻烦。我测试过各种黑盒子，并从中挑选

出预测效果最好的一个。”他根本不去考虑因果模型的问题。伯克说：“我在做出对某些人不利的预测时，从来不会给出理由。无论得出什么结论，我都不会解释其中的原因，我的唯一目标就是做出正确的预测。”

因此，伯克总是更加关注他认为预测作用最显著的指标（例如囚犯对工作岗位的渴望），而不是单纯的心理评估。如果他发现雀斑可以预测某人未来犯罪的可能性，他就会毫不犹豫地是否长有雀斑这项指标加入他的模型。伯克说：“我无须了解为什么雀斑会有这种预测作用。如果计算机发现了某些我不知道的内容，那么，无论它发现的是什么，只要有助于预测，我都不会将其拒之门外，也不会费神去解释其中的原因。”

在大多数情况下，伯克所采用的判断标准，其预测未来风险的效果肯定优于雀斑。就暴力犯罪的可能性而言，男性高于女性，

年轻男性又高于年长男性。男性在二十四五岁时暴力犯罪的可能性达到最高水平，然后这种可能性会稳步下降，到40岁时，男性暴力犯罪的可能性会骤降至接近零的水平。因此，年轻人的暴力犯罪行为比年长者具有更强的可预测性。

伯克指出：“人们认为，如果某人犯过谋杀罪，他将来还会再次杀人。但是，这种观点是不正确的，真正起决定作用的是人们年轻时的行为。14岁时的持械抢劫行为就是一个效果显著的预测指标，但是，如果持械抢劫的人是30岁，那么这种行为的预测效果就没有那么明显了。”

当然，预测也可能会出错。伯克宣称，他的系统可以预测获准假释的犯人在将来某个时间是否会杀人，而且准确率达到75%。这个比例当然已经非常高了，但这也意味着他仍然有1/4的概率做出错误的预测。伯克承

认：“无论这种预测多么准确，错误还是无法避免。每个人都明白这一点，但当真的做出错误的预测时，人们仍然会为之心痛。”犯罪预测的误差问题是电影《少数派报告》探讨的主题之一。电影名字中的“少数派报告”这个表达指的是预测犯罪部门所掩盖的一个事实：负责预测犯罪的那三名“先知”偶尔会出现意见分歧。汤姆·克鲁斯饰演的警察局长在获悉这个重要事实后，问道：“你的意思是说，我把无辜者关进了监狱？”他得到的回答是：“我的意思是，那些受到指控的犯罪嫌疑人，有的人可能（仅仅是可能）会有另一种结局。”这些少数派报告没有公之于众，原因可想而知。要让预防犯罪部门正常运转，就不能让人们对他们的可靠性产生任何怀疑。无论这一刑事司法系统多高效，只要人们心存疑虑，就不会欣然接受这个系统。

伯克认为，错误会以假阳性与假阴性这两

种形式出现。假阳性（他称之为“卢克·天行者”^①）是指被错误地认定为威胁程度高的人，而假阴性（“达斯·维达”^②）则指威胁程度高但被错误地认定为威胁程度不高的人。伯克说，这是一个政治问题，而不是统计问题。没有准确地发现“达斯·维达”，与错误地指控“卢克·天行者”相比，哪种结果更糟糕呢？他说：“从理论上讲，假阳性与假阴性是严重程度相当的错误。但实际情况并非如此。如果你与刑事司法官员共事，或者与股东或一般公民交谈，就会发现这两种错误的严重程度有所不同。一般来讲，没能准确地发现威胁程度高的人，其导致的结果比错误地认定某人威胁程度高更加糟糕。”伯克在设计他的算法时，就是这样权衡的。在很多刑事司法环境中（在涉及暴力犯罪时更是如此），如果某个证据可以避免出现漏网之鱼，即使其可靠程度较低，决策者们也会倾

向于采纳这一证据，而代价则是有更多的“卢克·天行者”会遭到错误的指控。

伯克说，数据源很快将会被进一步开放。伯克和他的同事们将不再仅仅依靠通常只向刑事司法官员开放的那些官方记录，还会使用自定义数据源。比如利用带有定位功能的脚链监控人们空闲时的去向，再利用算法将监控结果与罪犯数据库进行比对，了解某个人是不是阿尔·卡彭或者泰德·邦迪式的人物。2012年，伯克在“芝加哥创想周”活动的一次演讲中说：“（在空闲的时候）人们是待在家里看电视，还是在某个毒品交易活动频繁的街角呢？我们认为，了解某个人习惯在星期六凌晨2点还是在上午10点出门，有助于我们做出更准确的预测。”

如果时间充裕，伯克甚至可以在人们实施犯罪之前就为他们建立个人档案，从而向菲利普·迪克设想的“预防犯罪”专制社会进一步

靠拢。如果一个人从小生活在某个犯罪率很高的地区，又有毒品滥用的家族史，而且兄弟姐妹或者父母等亲人中有人入狱，那么，即使这个人没有被捕记录，伯克的算法也会注意到这个人，尽管这种做法在法律与道德两个层面都可能陷入进退两难的境地。

这样的技术还有可能被应用到教育系统中。伯克说：“学校希望了解自己的学生是否会出现问题。现在，学校已经着手进行了一些初步的工作，利用学生在校期间的相关数据，判断那些没有犯罪记录的学生是否有可能出现辍学、逃学、斗殴、破坏公物等问题。”2013年，他以一群8~10岁的儿童作为研究对象，利用预测式建模的方法，预测他们未来犯下重罪的可能性。不过，就目前而言，这项技术还不成熟。伯克自己也承认：“我们才刚刚有了一点儿头绪，后面的路还很长。”

1. 卢克·天行者，《星球大战》系列电影的主角之一，绝地武士和强大的原力使用者，新绝地武士团的创立者，义军同盟成员。为推翻银河帝国的暴政做出了卓越的贡献。

2. 达斯·维达，原名阿纳金，天行者，是电影《星球大战》里最重要的角色之一，他的两次改变决定了银河共和国和银河帝国的兴衰，影响了原力光明与黑暗面的消长，也使他完成了预言——平衡原力。

未来的我们不需要律师

莎士比亚的《亨利六世》中有一句话，可能代表了所有跟律师有过不愉快接触的人的心声。剧中人物屠夫迪克为一个计划在英国发动社会叛乱的阴谋集团捐献了两分钱，然后高声提出了一个他认为几乎可以确保乌托邦能够成功实现的建议：“我们要做的第一件事，就是杀死所有的律师。”

从莎士比亚的这部喜剧首次登上舞台之日算起，至今已将近500年了，但律师仍然没有消失。不过，“公式”有可能会使越来越多的律师变得无关紧要。我们以“证据开示”

（legal discovery）为例。证据开示是指在案件庭审前的阶段，诉讼各方都会要求其他各方提供某些材料，并对其进行认真梳理，希望找到可以提交法庭的证据。之前，证据开示工作主要由律师事务所指派初级律师完

成，人工梳理海量材料，希望从中发现某些证据。这项任务对律师事务所来说是件美差，但却让委托人头疼不已，因为最终他们要为高昂的诉讼费用埋单。1978年，一个针对广电业巨头美国哥伦比亚广播公司（CBS）的反垄断诉讼案涉及5家电视节目制作公司。为了仔细研究据说与案件有关的600万份文件，这5家公司雇佣的律师团队工作了好几个月的时间。收到账单后，这些公司发现诉讼费用总额竟然高达220多万美元（相当于现在的800万美元）。

现在，由于人工智能的发展，证据开示工作可以借助数据挖掘“电子证据开示”工具，以及预测编码等机器学习程序来完成。预测编码可以通过让人类律师示范用手工方式检查少量文件，“教会”计算机区分相关与无关信息，接下来，算法便可以处理剩余的大量信息。即使安排最杰出的人类团队来完成这

份工作，他们所需要的时间也是算法的三倍以上。就精确性而言，这些系统已经多次证明它们可以比初级律师与律师助理做得更好。毕竟，电脑不会因为工作量过大而感到头疼。

通常，电子证据开示的要求并不复杂，只需基本的索引编排或者简单的法律分类即可。不过，电子证据开示的作用还有待开发。算法可以提取相关概念（诸如整理出所有与中东社会抗议活动有关的文件），并且在搜索具体内容时无须考虑词语选择的问题，因此搜索效率越来越高。算法甚至可以搜索特定缺省项，或者寻找人类律师可能会感兴趣的潜在规律。与人类律师的工作相比，算法主要优点还在于它的效率更高，而成本却低得多。帕洛阿尔托市的黑石探索公司（Blackstone Discovery）是该领域的全球性龙头企业，该公司分析150万份文件的

成本通常还不到10万美元。

黑石探索公司位于硅谷的中心位置，客户包括奈飞、奥多比（Adobe）和美国司法部等。2012年，该公司参与了被《财富》杂志称作“世纪专利审判”的苹果诉三星案。黑石探索公司的创立者是一位名叫约翰·凯利的工商管理硕士。2003年，凯利根据自己对法律职业发展方向的理解，创建了黑石探索公司。凯利说：“今天，我们需要处理的数据量大得惊人。20年前，一个案件通常会有10箱复印文件。如今，我们可以轻轻松松地从本地服务器、台式机和智能手机上获取100吉字节（GB）的电子存储数据，相当于100万至200万页的文件。”

数据之所以呈指数增长，原因之一是信息存储方式的便利性。现在，公司已经无须为存储纸质文件的巨型文件柜腾出大量空间，而是把越来越多的数据存储成数字文件。在

基于云技术的存储方法问世之后，我们几乎没有任何理由舍弃任何数据。三星公司的所有企业邮件都有两个星期的有效期，并自动清理那些过期的邮件。这个做法在苹果诉三星一案中受到了主审法官的口头训斥。《快公司》杂志指出：“三星还是败（诉）了，它的这种（违规）行为.....可能导致了这样的结果。”

证据开示过去的任务是寻找足以支持诉讼请求的数据，而现代的电子证据开示则关注在确保安全的前提下，可以舍弃多少信息。凯利指出：“数字时代急需解决的问题是想方设法限制我们的调查范围。”这是算法在黑石探索公司调查活动中的另一个应用。凯利承认，这种限制程序的效率非常高，因此传统法律事务所的某些员工并不欢迎这种做法。他说：“有的公司在接手一个案件之后，从中看到了500万美元的商机。但是在黑石探索

这样的公司出现之后，它们只需一眨眼的工夫，就可以将相关文件的数量减少至500封邮件，收费也只有10万美元。”

关于电子证据开示，凯利唯一担心的是这种自动化进程正在威胁新一代初级律师的生计。凯利心情沉重地说：“我们的这项成果会引发若干问题，其中之一就是它会威胁到那些刚从法学院毕业的人。他们还没有客户，也没有取得成功，他们该怎么办？过去，他们需要处理成千上万公斤的文件资料。尽管这份工作枯燥乏味，但至少他们有事可做。可是现在不一样了，我们利用合适的算法就可以完成这些工作。”

计算机算法属于颠覆性创新

商业管理专家克莱顿·克里斯坦森提出，新技术有“维持性创新”与“颠覆性创新”两种类型。维持性技术会对公司或市场的现有运营方式起到支持或加强作用，而颠覆性技术则会彻底改变某个领域的运营方式。会计电算化的出现是维持性技术的一个例子，而数码相机问世（众所周知，它直接导致了柯达公司的崩溃）代表的则是颠覆性技术。电子证据开示等工具属于颠覆性技术，在算法时代的到来使律师职业的方方面面遭到不可逆转的改变时，它们也不甘示弱地起到了推动作用。

进化生物学家理查德·道金斯在他的经典著作《自私的基因》中指出，法律诉讼已经“进化”到无比低效的状态，因此，律师们可以通过“精心编排的合作”，相互配合，尽可

能多地榨取客户的银行存款。这种合作活动从本质上讲也是一种特殊的算法，不过它的目的与计算机算法正好相反，因为后者的是通过尽可能少的环节产生高效的结果。

由于法律体系可能会有失偏颇^①，因此很多律师对在法律事务中采用颠覆性技术的做法提出了批评，担心这会削弱他们的赚钱能力。很多人认为这种做法是将法律事务“商品化”，以“程序化”的方式处理法律事务，效果不及人类律师“量身定做”的方法。（我们可以想一想，工厂批量生产的衣服，无论在质量还是价格上，都远远低于手工缝制的衣服。）

如果你是一名律师，而且你觉得自己正在按照“量身定做”的方法处理法律事务，这种批评就是有道理的。但是，即便如此，持这种观点的人还是严重低估了计算机算法同样可以胜任（甚至完成效果更好）的法律事

务。例如，由于“LegalZoom”（这个名字起得非常巧妙高明）等文件自动收集系统的出现，“公式”可以在合同起草这个环节起到革命性作用。自2001年LegalZoom公司创建以来，它为200多万客户提供过服务，帮助他们寻求公司法的保护。如今，该公司已经成为美国首屈一指的法律事务所。

LegalZoom公司收费低廉，起草遗嘱与公司章程的费用分别为69美元和99美元。就像分类广告网站Craigslist为报业提供分类广告服务以赚取服务费一样，LegalZoom利用算法为消费者和公司提供基本文件的起草服务，以薄利多销的方式从事经营。算法可以胜任的另一个领域是商标分析。众所周知，商标分析的主观性非常强，因此很麻烦。而芬兰的初创公司Onomatics开发的一种算法可以生成即时报告，告诉人们两个不同商标之间可能存在多大的区别。

坐落于加利福尼亚山景城、与谷歌公司总部几英里之隔的初创公司Wevorce拥有一项类似的技术。如果说网络相亲是一种理想的相亲方式（我们在上一章讨论过这个问题），可以帮助我们免受爱情征程中的各种痛苦，那么Wevorce利用算法提供的离婚服务就是处理离婚事宜的最佳选择。Wevorce的标准化服务体系不仅可以识别离婚夫妇的18种典型特征，还能随时告诉用户他（她）的前任配偶正处于这个痛苦过程的哪一个阶段。Wevorce宣称，他们要求夫妇双方理性考虑分手问题（甚至建议他们借助计算机算法来考虑），从而“使事情朝好的方向（演变）”。公司首席执行官米歇尔·克罗斯比说：“因为软件可以保证整个程序有序进行，因此它有助于离婚夫妇消除恐慌心理，同时帮助律师提高效率，降低总成本。”

1. 以离婚诉讼案为例。算法会采用“史

密斯诉史密斯”这样的表达，认为离婚诉讼案是對抗式案件，尽管有时情况并非如此。

透明的执法者

40年来，关于人们对“透明”一词的理解上所发生的重大变化，一直被津津乐道。在个人电脑刚刚问世的时候，人们说它是透明的，意思是可以“揭开它的盖子”，了解电脑的内部运行机制。时至今日，“透明”一词的含义已经变得异常晦涩难解了，它是指我们在不了解其原理的情况下让某事顺利进行。用麻省理工学院精神分析学家雪莉·特克尔的话来说，随着我们由现代的计算文化过渡到后现代的模拟文化，在说到计算机时，人们考虑越来越多的是其“界面价值”（interface value）。2011年，从事Kinect体感装置开发工作的微软工程师亚历克斯·基普曼，完美地概括了界面价值的含义。在接受《纽约时报》记者的采访时，基普曼骄傲地解释说，人类正在逐渐过渡到“技术对人类的理解越来越

越深刻，而人类却无须理解技术的时代”。

“周边定律”（Ambient Law）的概念就是在这个框架下应运而生的。周边定律是指法律无须要求律师提请我们注意身边具有法律意义的内容，而是将其直接植入我们的设备及周围环境（并由这些设备及环境付诸实施）。在周边定律的远景展望中，智能型自动环境将实时地帮助我们完成与我们有关的无数个决策。自动计算就像人体的中枢神经系统。人体对体温、呼吸与心率的调节不会被我们察觉，同样，算法为自我管理、自我配置与自我优化而完成的自动计算活动，也无须用户的任何体力与智力投入。

“智能型办公室”可能是周边定律的一个典型案例。智能型办公室可以持续不断地监控室内温度，并将检测结果与卫生及安全法规所规定的数据进行对比。如果有人逾越了某条法律红线，警报就会响起。周边定律还可

以应用于汽车，禁止血液中酒精含量超标的人驾车。近些年来，多家汽车制造商已经开发出了类似的技术。日产公司研发的一个系统可以从司机坐到驾驶座上就开始对他实施监控，并用酒精气味传感器检查他们呼出的气体中酒精含量是否超标。在司机接触变速杆时，另一个传感器可以检测他们手心汗液中的酒精含量。汽车仪表板上的一个微型摄像头可以监控司机面部及眼睛的运动：如果眨眼频率增加，则表明司机犯困了；如果驾驶员口角歪斜，则说明他正在打呵欠。通过综合考虑这些生物测量数据，汽车上内置的算法就可以判断他是否可以做到安全驾驶。如果算法给出否定的答案，汽车的变速器就会自动锁死，汽车的卫星导航系统会响起“酒驾”的语音警告，同时安全带收紧，使司机“略微受惊”（引自日产发言人），以便让他清醒过来。

谷歌地图与社会偏见问题

有人认为这些技术的出现意味着算法将会取代人类成为道德决策者，他们为此惴惴不安。苹果公司在开始审查应用商店中是否含有“令人反感”的内容时，引发了种种充满敌意的反应。从这个事例不难看出，很多计算机用户认为道德与技术风马牛不相及。这种反应是可以理解的，但同时也说明人们并不清楚“技术”的历史责任。如果说科学发展是为了进一步了解世界，那么技术的目的（更准确地说，是技术专家的目的）则是想方设法改造我们的世界。因此，即使人们拒绝承认，技术与道德责任感之间也必然存在着无法分割的联系。从这种意义上说，技术与法律非常相似——两者都是人类创造的用于规范人类行为的手段。

1980年，在《手工艺品涉及政治问题

吗？》一文中，社会学家兰登·温纳从纽约长岛主干道上的桥梁中挑选了一些，对它们进行了认真细致的分析研究。温纳认为，在这些桥梁中，有不少座特别低矮，路缘处的高度仅为9英尺^①。尽管大多数人都认为这不大可能有任何特殊含义，但实际上，这是设计者罗伯特·摩西的社会与种族偏见的反映。在20世纪20年代至20世纪70年代，摩西负责修建了纽约州的多处道路、公园、桥梁等公共设施。摩西之所以把桥梁设计得很低，他的目的是只允许那些处于上层及富裕的中产阶层的白人进入公园，因为只有这些人才有钱购置轿车。由于无法从低矮的桥梁下方通过，车身较高的公共汽车只能绕道而行，因此，那些只能乘坐公共汽车的穷人（其中很多是黑人）便无法进入公园。换言之，摩西把偏见（及歪曲的道德责任感）融入了他的设计。纽约城市设计师李·科佩尔曼后来回忆

说：“这个家伙处心积虑地不让公共汽车从他设计的公园大道上行驶。”

尽管谷歌公司奉行的新自由主义与摩西有偏执的行为之间相差十万八千里，但是，如果看一看该公司利用数据挖掘算法设计个性化地图的计划，就会发现其中包含多种陈腐保守的过分要求（也许这是该公司的无心之失吧）。几十年来，谷歌地图已经渗透到人们生活的方方面面，对于我们每天的出行有非常重要的意义。记者汤姆·奇弗斯在《每日电讯报》上指出：“这个搜索引擎巨头就像章鱼一样，把触腕伸到了我们生活的各个方面。其中，对我们日常生活影响最大的可能是谷歌地图、谷歌地球及各种子产品。”

2011年，谷歌手机地图部门的主管丹尼尔·格拉夫在接受TechCrunch网站的采访时说：

“如果你和我同时查看谷歌地图，我和你一定会看到相同的结果吗？我觉得未必如此，因

为我要去的地方与你不同。”这一灵感促使谷歌地图自2013年起开始结合用户信息，把用户导向那些具有相同搜索目的的人最有可能去的地方，或者他们之前表示过兴趣的内容。谷歌在宣传材料中扬扬自得地吹嘘说：“过去，这是无法想象的。地图就是地图，无论你搜索的是帝国大厦还是街角的咖啡店，你看到的都是同一幅纽约市地图。但是，如果你可以看到一幅为你量身定做而且总是可以满足你此刻需求的地图，你会作何感想呢？”

尽管这种做法从某种意义上讲是有益的，但其内在的“过滤气泡”效应也可能会限制用户在使用传统地图时可能产生的意外发现。就像相亲网站的算法配对一样，只有算法推断适合你的或者你希望看到的那些人和地点才会出现在搜索结果中。因此，尽管谷歌将公式应用于地图绘制领域的做法合乎情理，

但也有可能带来麻烦。理查德·桑内特在他的社会学经典著作《混乱的价值》一书中指出，“城市的弱肉强食、空旷孤独对于人类而言具有积极的意义。”他的意思是，如果一个城市“清除掉”所有不能产生即时吸引力的事物，我们就几乎无须费神考虑这座城市到底会有什么样的结局。阻止经济收入较低的人看到那些只接待经济收入较高者的场所（或者阻止一个地区的人从地图上看到另一个地区的寺庙、教堂等），起初似乎是减少冲突的有效手段，但从长远看，这类做法无助于增进包容和理解，甚至会使在运动场上举行晚会也变成一种奢望。桑内特接着指出，从根本上看，要推动人类社会的全面发展，城市环境必须发挥熔炉的作用，让人们接触并接纳各种不同的观点。

如果不合理地应用某些算法（诸如我在第一章介绍的Nara推荐系统），上述情况可能

只会愈演愈烈。如果算法不关注人们现在所处的位置，而只关心广告商最终希望他们前往的位置，那么依靠算法决定行动方向的人，可能会被导向固定的路线，就像演员只能照着剧本表演一样。

1.9英尺约合2.743米。——编者注

技术给人类撰写的“剧本”

第一个提出技术“剧本”概念的人（从社会学家兰登·温纳之后算起）是法国哲学家、人类学家布鲁诺·拉图尔。^①拉图尔认为，技术可能要求用户采用特定的处理方式，从而对用户的行为产生调节和限制作用，就像电影剧本或者舞台剧对表演者的动作有严格的要求一样。例如，塑料咖啡杯在使用几次之后就会破裂，这种特性会促使人们将其丢弃。同样，宾馆钥匙附上重物后，由于携带不便，人们在结账时归还钥匙的可能性就会增加。

与上述两种应用相比，在门上附着弹簧以限制人们进入建筑物的速度，或者在路面上设置减缓行驶速度、避免汽车减震装置受损的减速带等做法，在巧妙程度上则可能有所不及。前文中介绍的确保醉酒司机无法启动

汽车的周边定律，或者在室内温度达到临界值、员工为确保自身安全应该退出房间时，不仅响起报警声而且关闭工作人员电脑显示器的智能型办公室等，这些技术应用的巧妙性可能更差。

如同摩西设计低矮桥梁的目的一样，设计者在“撰写这些剧本”时也可能带有某种意图，想为自己的设计赋予特定的职能。这些职能可以影响用户的行为，或是鼓励他们遵从某些社会规范，或是迫使他们遵守某些法律。^①因为这些“剧本”在物体或者装置的基本功能上增添了“附加内容”，因此会引发若干道德问题。例如，既然技术设计者负责“撰写这些剧本”，那么他们到底需要承担什么责任呢？如果法规的目的是教化他人，那么这种努力与用技术教化人们的尝试有何不同呢？通过某项规定要求我们在淋浴时不要浪费水，与使用节水喷头技术实现这个目的，

这两种做法有实质性的区别吗？这种区别可以被量化吗？

在《助推》^①一书中，作者理查德·泰勒与卡斯·桑斯坦讲述了一个故事。人们在阿姆斯特丹史基浦机场的男厕所的每个小便池里放置了一只假苍蝇，让男性在小便时有瞄准的目标，从而使排在小便池外面的尿液总量减少了80%。几乎没有人会公开谴责这种旨在保持公共厕所清洁度的温和家长主义的做法，但是，如果在驾驶员违反限速规定时强行制动，人们会对这种强硬家长主义的做法持什么样的态度呢？如果人们的某些行为是不得已而为之，那么我们如何判定这些行为是讲道德或是守法的呢？如果这些行为是有权制定规则并付诸实施的设计者（或者公式环境的计算机科学家）事先规定好的呢？

纯粹站在民主的立场上挺身而出，捍卫酒驾者的权利，这可能需要一定的勇气（还可

能是受到了误导)。但是，如何设定行为的度是道德心理学需要解决的一个核心问题。在某些场合，一种行为在道德层面上看是正确的，但是在某些情况下，这种行为很有可能让人们联想到乔治·奥威尔在小说《1984》中描写的极权主义政治体制。

不出所料，如何确定这条红线的位置问题引发了严重的分歧。法学学者罗杰·布朗斯沃德在技术性规定与法律领域著作颇丰。布朗斯沃德认为，自治权是人权的基础，自治权赋予人们选择遵从或者不遵从某项规定的权利。但是，萨拉·康利教授则持一种截然相反的观点。在《反对自治》一书中，康利不仅选择了一个大胆的书名，而且倡导禁止一切可能对人们的健康与幸福造成身体或心理伤害的行为，从而“（拯救）人们，使他们不会伤害自己”。这些应该加以禁止的行为包括（但不限于）吸烟、摄入含反式脂肪酸的食

物、暴饮暴食、债台高筑以及不储蓄并大手大脚花钱等。康利在书中指出：“公共教育有时无法让人们足够清醒地认识到他们的行为给自己带来的潜在危险。如果公共教育切实有效，吸烟人口就不会增加，但实际情况并非如此。”显而易见，在康利这种强硬的家长主义作风的干预下，减速带的数量肯定会多于塑料咖啡杯。

1. 差不多同时期的玛德琳·阿克里奇在她的《技术对象的描述》一文中提出过类似的概念。这篇文章被韦伯·比克与约翰·劳收入《塑造技术/建构社会：社会技术变革研究》一书。

2. 这种观点发展到极端就是一种宿命论。为了理解这一点，我们可以借助卡尔·马克思在《哲学的贫困》一书中的论断：“手推磨导致社会中出现封建领主，而蒸汽磨则使社会中产生了工业资本家。”我更倾向于社会

构建主义的观点，即技术发展是发明者、企业家、顾客与社会环境相互作用的结果。

3. 《助推》由中信出版社于2015年出版。——编者注

令人瞠目结舌的500张罚单

从表面上看，我们有理由认为借助算法来执法是可行的。既然法律推理具有逻辑性，逻辑操作又可以由计算机自动完成，那么把法律程序编写成编码程序的做法难道不能提高执法效率吗？如果采用这种做法，在断案时只需输入具体案件的相关事实，执行相关法律规定，就可以得出“正确”的判决结果。

美国学者劳伦斯·莱斯格在自己的著作中指出，法律与计算机编码是同一个事物的两个方面。莱斯格把位于华盛顿特区的美国会所制定的法律称作“东海岸法典”，而把计算机程序所遵从的那些规则，根据硅谷所在的位置称作“西海岸法典”。莱斯格认为，无论是哪种法规，从制定之日起，就不再受到制定者的影响，而是变成一种虚拟的存在。事实上，早在莱斯格之前300年，就已经有

人开始探讨两者之间的相似性了。伟大的数学家、微积分创始人之一戈特弗里德·莱布尼茨就曾经提出可以利用微积分来确定法律责任。20世纪末，更多的法学学者发起了所谓的“计量法学”运动，认为“最理想的法律体系应该借助科学来建立基本的法理并使其合理化”。

虽然莱布尼茨与计量法学运动的倡导者们都错误地把法律体系看成一系列静态的自然法规，但他们认为可以借助科学来加强法律客观性的理想（从本质上讲）是可靠的。法律的客观性指的是公平公正。如果不能保证客观性，在做法律判决时就会受到自私自利的想法或是偏见的影响。他们认为，在法律推理的过程中总是以相同的方式解读法律条文，就可以实现与自然科学相媲美的一致性。

计量法学遭遇的最大挑战来自于2013年

的一次实验，这项实验的目的是检验是否可以将最直观的法律转换成算法。实验对象是52名计算机程序员，他们被分成两组，任务是设计一个可以随时给超速驾车的司机开出超速罚单的算法。他们可以利用两个数据库：一个数据库中有某条道路上合法行驶速度的信息，另一个数据库通过车载电脑收集了某辆汽车（如一辆丰田普锐斯）之前在该道路上的行驶速度信息。数据显示，这辆普锐斯几乎没有超速，即使偶尔超速，也只是略微超过限定的速度。为了让情况在道德层面上显得更加含糊不清，这辆普锐斯在超速时都处于定速巡航状态。最终，普锐斯安全地完成了整个行程。

根据要求，第一组计算机程序员编写的算法必须严格遵守“法律条文”，而第二组只需符合“法律意图”即可。不出所料，两组程序员设计的算法大相径庭。在整个行程中，“法

律意图”小组的算法开出了0~1.5张罚单，而“法律条文”小组则严厉得多，一共开出了498.3张罚单。之所以会出现这种令人瞠目结舌的差异，主要原因有两个。在超速时，“法律意图”小组有小小的灵活机动性，而“法律条文”小组则毫不留情。此外，“法律条文”小组将每次超速都看成一次新的违规，以致连续不断地开出了大量罚单，其数量远超超速摄像头开出的罚单。

规则与标准的冲突不断升级

这个实验结果激化了“规则”与“标准”的冲突。从广义上讲，所有法律都分属这两个不同的阵营，或者法律体系的两个极端。为了说明规则与标准之间的区别，我们设想有两部针对危险驾驶的法律。规则可能会规定“所有汽车的行驶速度都不得超过每小时65英里”，而标准则可能要求“所有人都不允许以不安全的速度驾驶”。标准的主观性意味着人们在执行时需要仔细权衡，而规则是非此即彼的强硬要求，几乎没有通融的可能性。

对于必须在现实生活中解决的情境问题，计算机的处理能力正在不断提升。例如，如果车流量或者天气状况导致干道上的驾驶员无法满足最低限速要求，那么算法可以通过学习去停止执行这项规定。事实表明，用于处理超速摄像头数据的算法可以将首次到访

该地区的人甄别出来并予以区别对待。如果对某个地方的路况不熟悉的驾驶员稍微超速，可能不会被追究，但经常在某条道路上驾车的人超速的话，则可能面临严厉的处罚。不过，这两种情况都是预先设定的规则产生的结果，不能证明计算机善于处理含糊不清的情况。

几百年来，技术一直在努力地把我们推向更客观的“规则”。例如，斯坦利·乔尔·赖泽在《医药与蓬勃发展的技术》一书中提及听诊器的发明时说：

它有助于医生摆脱对病人的体验与感觉的依赖，在诊断时与病人之间保持客观、自然的关系，专注于从病人身体内部传来的声音。

社会学家约瑟夫·古斯菲尔德在《公共问题文化：酒驾及象征秩序》一书中也提出了类似的观点。古斯菲尔德认为，执法方面的

各种技术之所以得到蓬勃发展，是因为人们渴望客观性，其核心是量化。为了证明自己的观点，古斯菲尔德研究了呼吸测醉器于20世纪60年代问世之后发挥的作用。在呼吸测醉器问世之前，执法人员都是依靠主观判断驾驶员是否属于酒后驾驶。

（在采用呼吸测醉器等检测手段之前，查处酒驾的执法过程充斥着）各种没有经过证实的报告、各行其是的判断，以及缺乏一致性的各类标准，呈现出乱糟糟的情况。法律界和科研“实验室”认为，血液样本技术与呼吸测醉器可以精确显示血液中的酒精含量，能够加强执法效果，但同时也会提高定罪率的期望值。

换言之，从法律角度看，呼吸测醉器的出现使人们在小酌几杯之后是否可以驾车的判断依据，由抽象的“标准”转变为具体的“规则”。随着周边定律的兴起，这类问题将变得

越发咄咄逼人，因为技术将变得不仅可以调节人类的行为，有时还可以通过禁止某些行为来控制 and 支配人类。

几年前，谷歌宣布其正在研发一批自动驾驶汽车。这种汽车由算法完成所有的操控工作，包括制定最佳行驶路线，并通过对最平滑的并道路线、速度以及与周围障碍物之间的安全距离等方面的判断，完成行驶中的变道操作等。在我撰写本书的时候，这些汽车已经在各种路况上完成了30多万英里的试行驶，而且没有任何事故报告。这个情况似乎说明，算法驾驶的汽车比人类驾驶的汽车更加安全。由于算法驾驶的汽车并没有违反现有的一系列交通法规，因此我们可能会认为再去制定某些相关法律是多此一举。但是，这些现有法律可能会导致若干道德方面的难题。举个简单的例子，如果车内乘客需要赶到医院看急诊而需要在车流量非常小的路段

超速行驶，在这种情况下该怎么办？如果真的超速行驶，并且由于车内装有速度追踪系统而导致驾驶员（或乘客）事后接到了罚单，这当然会令人不悦。但是，如果在周边定律强硬的作用下，这辆自动驾驶汽车不论面临什么情况都拒绝超速行驶，又会造成什么样的结果呢？

也许无须等到自动驾驶汽车问世，这类情况就已经变成现实了。2013年，据英国报纸报道，布鲁塞尔的欧盟官员正在商讨拟定一些道路安全措施，其中包括在所有新车上安装“智能速度调节”装置。这种装置与在很多大型货车和公共汽车上安装的速度控制装置相类似，利用卫星信号或者自动探测并读取道路标志的摄像头，强行要求驾驶员遵守限速规定。如果驾驶员试图超速，该装置就会触发汽车的制动装置。

公正的算法与带有偏见的算法

由于人们在拟定法律条文时往往会受到大脑中那些算法与软件的影响，因此，随着这类法律越来越多，人们在处理法律相关问题时，对规则的依赖性也越来越大。不过，最根本的问题在于，尽管规则与标准是法律推理的彼此对立的两个极端，但它们同时也是两个抽象的概念。大多数规则都具有一定程度的“标准性”，而很多标准又具有这样或那样的“规则性”。在某个路段行驶时，每小时60英里的最高时速限制可能是一条规则（与之相对的标准是“不得以不安全的速度行驶”），但是，即使我们的速度达到每小时62英里，我们可能也不会收到罚单，这个事实说明这条规则的规则性并非不可违逆。尽管标准或许具有过多的主观性，但为了实现更

广泛的社会目标，规则也会由于缺少灵活性而存在不尽如人意的地方。

我们举个例子来说明这个问题。假设在公园的入口处有一个标志，上面写着“所有车辆不得入内”的字眼。在实施这条规则时，人们可以借助闭路电视摄像头，并利用算法识别移动物体的形状。强硬的家长主义“剧本”可能会在入口处安装摄像头，并使之与公园大门相连接，只要算法判断有一群人准备步行进入公园，大门就会自动打开。而在温和的家长主义“剧本”中，摄像头可能遍布公园的各个角落，只要发现公园内出现车辆，这些摄像头就会识别驾驶员，将他们的面部特征与身份证图像数据库进行比对和匹配，然后自动开出罚单并直接寄送到违规者的家中。

这种做法似乎非常公平，如果之前发生过有人驾车从公园穿行的事情，采取这种做法可能更加无可厚非。但是，这条规则适用于

自行车吗？演绎推理有理由认为，既然明文规定所有车辆不得进入公园，而且自行车也是一种车辆，这条规则就同样适用于自行车。不过，这是这条规则的本来意图吗？或者说，在本例中，人们制定这条规则，目的是不是避免机动车进入公园制造噪声与污染呢？这条规则是不是意味着要入园挽救某人生命的救护车也不得驶入呢？如果救护车可以进入公园，那是不是意味着人们编写的算法法规可能不适用于某些公民呢？

在承诺“技术可以从本质上了解你，而你无须了解技术”的世界中，一个更加显而易见的难题是，人们有可能在完全不知情的情况下触犯法规。全球有很多地方都规定人们在公共场所饮酒是违法行为。由于越来越多的闭路电视摄像头装有面部识别技术，甚至是步态识别技术（分析人们走路的方式），可以用算法来断定某个人是否喝醉了，因此，

在酒吧喝完酒东倒西歪地走回家之后，会不会收到罚单呢？再比如，在英国，夜间骑车时自行车踏板上没有黄色反光罩属于违法行为；随意横穿马路在英国是合法的，但在美国、欧洲与澳大利亚却都是违法行为。在这两种情况下，面部识别技术可以识别违法者并提起诉讼。闭路电视行业的某龙头企业在发表于行业杂志《今日闭路电视》上的一篇文章中解释说：“对于科研人员而言，所谓识别违规行为，就是将可预测的正常行为进行归类，然后编写一种算法以识别任何违背这些‘常规’的行为。”

几乎无须指明，大家也会察觉到该表述中存在的问题。很明显，“常规”行为不是一种客观的度量，而是一种社会建构概念，代表的只是人们的偏见。

当然，并不是所有的算法程序都有失公允。例如，研究人员已经开发出一种算法，

可以监控在伦敦地铁站台上逗留的人。如果某个人在地铁站台上的逗留时间超过10分钟，而且期间有几趟列车到站他却一直没有上车，该算法就会认定这个人有跳下站台自杀的可能性，并发出警报。还有一种尚未投入市场的应用技术，可以帮助发现大街上的斗殴事件。如果该应用技术发现人们的腿脚与胳膊急剧地前后运动，就会认定他们正在做拳打脚踢的动作。不过，发现某个偏僻的角落里有人群聚集就向警察局发出警报的算法，可能更值得商榷。同样，在执法行为明显具有歧视性的国家中，算法还可能成为政府恫吓和排挤社会成员的工具。在俄罗斯，反同性恋法律已经瞄准了同性恋人群，算法被用来监控同性间的亲密行为。

我在第一章中说过，Quantcast与亚马逊等公司珍视“异常”行为，借此他们可以深入理解个人用户的独特行为。但是，在上述情

况下，算法却会产生完全相反的作用，有可能成为确保人们遵从（或者摒弃）相似行为的一种手段。美国社会学家苏珊·斯达说过，在基础设施方面，人们是萝卜白菜各有所爱。

在以上几个（在某种程度上令人胆战心惊的）例子中，那些与执法相关的算法都做出了正确的判断，不过它们在执行各项规则时的严厉程度甚至会令冷面判官德雷德法官汗颜。但是，一旦它们做出了错误的决定，其造成的后果就很有可能是骇人听闻的。

算法犯下的种种错误

2011年4月5日，41岁的约翰·盖斯收到了马萨诸塞州车辆管理处的来信，通知他的驾照被吊销，他不得再驾驶汽车，而且该项决定即时生效。这项决定唯一的问题在于，盖斯在驾车时十分谨慎，多年来从未接到严重交通违章的处罚，因此在收到这封信后，盖斯觉得莫名其妙。他气急败坏地打了几通电话，并前往管理处进行申诉。最后，他终于明白了其中的缘由：搜寻犯罪嫌疑人虚假身份的面部识别算法，在扫描包含有成千上万马萨诸塞州驾照信息的数据库时，自动将他的照片标记为可疑对象。该算法认定盖斯的相貌与另一位涉嫌违法的马萨诸塞州驾驶员的相似度很高，因此汽车管理处便自动给他发出了那份信函。管理处对此无动于衷，并声称如果出错，遭到指控的人“有责任”为自

己正名。他们认为，与保护公众的重要性相比，少数人蒙受冤屈或遭遇不便根本不值一提。

约翰·盖斯绝不是算法错误的唯一牺牲品。2007年，加利福尼亚卫生部门新安装的自动化计算机系统发生了一个错误，导致数以千计的低收入老年人与残疾人的保险金被停发。由于他们没有缴纳保费，医疗保险部门就取消了这些人的医疗保险。之前的系统在认为某些人不再符合保险金补贴条件时会通过邮局向他们发通知，但是在更换为CalWIN软件之后，系统将不再通知他们，而是直接停发他们的保险金，除非他们手动登录系统，阻止这种情况发生。在这些人当中，有很多人并不知道自己没有按时缴纳保险费，直到邮局寄来高昂的医疗账单，他们才恍然大悟。即使到了这个地步，由于缺乏必要的英语交流能力，仍然有很多人无法登

录在线医疗系统，查找问题的症结所在。

类似的错误还包括：在不通知的情况下将选民从选民名单上剔除，小企业被标记为“没有资格承接政府订单”，错误地认定某人是“拒付抚养费”的父母等。联邦寻亲服务处出了差错，认定56岁的机修工瓦尔特·福尔默没有支付子女抚养费，给他开出了定额为20.6万美元的抚养费罚单。在接到这份罚单之后，福尔默年仅32岁的妻子怀疑她的丈夫有婚外情和私生子而自杀身亡，引起了人们的关注。

算法错误地把人们归类为恐怖分子，也同样令人胆战心惊。平均每个星期约有1500名航班乘客会遭遇这种不幸，由于数据匹配错误而不幸被认定为恐怖分子的人包括若干名前陆军少校、一个4岁男童及一名美国航空公司的飞行员，后者在一年内被羁留了80次。

这些问题有很多都是算法在执法过程中扮演的角色所导致的。由于财政预算削减导致工作人员不断被裁减，这些自动化系统已经由简单的行政管理工具变成了主要的“决策者”。在很多个案中，简单地寻找合适的算法并不能解决问题，而要先认清一个本质问题：不是所有的任务都可以通过自动化方式来完成。我们以用数据挖掘技术来识破恐怖分子阴谋的做法为例。关于恐怖袭击事件的统计数据极少，而且不像亚马逊网站的客户统计数据那样清晰，因此，人们在旅行时必须将大量个人隐私暴露在数据挖掘算法的面前，换回来的却不一定是真实的警报（尽管这样的情况并不多见）。著名的计算机安全专家布鲁斯·施奈尔说过：

洞察恐怖分子的阴谋无异于大海捞针，盲目地扩大侦查范围并不会降低这项任务的难度。与其让计算机主导和决定应该调查哪些

人，还不如让人来操控计算机，去识破那些潜在的阴谋。

尽管我们都清楚为什么人们会认为算法这种并不客观的事物适用于算法时代，但是，其中最核心的问题仍然要归结于算法所承诺的虚无缥缈的客观性。马里兰大学的法学教授达尼埃尔·西特龙说：“我们过于担心人类的偏见与自相矛盾，同时又过于相信计算机的客观公正性。”西特龙认为，我们的错误在于我们“信任算法，因为我们认为算法非常客观。但事实上，这些算法都是人编写出来的，人们可以把所有的偏见与观点植入其中”。换言之，计算机在运行算法时可能是不带有任何偏见的，但是，我们已经发现，这并不意味着算法在编写的过程中没有受到人类偏见的影响。本章前面提到的超速驾驶算法实验说明，即使待解决的问题在本质上是一个机械化程度较高的问题，计算机程序员

在编写这些算法时也会将大量臆断植入其中。科技史学家梅尔文·克兰兹伯格的科技第一定律指出：“技术既无好坏，亦非中立。”

所有或隐晦或露骨的偏见，要么是一两个人类程序员所造成的，要么是技术难题所导致的。例如，事实证明，与面部识别技术相关的算法对男性的识别率高于女性，对非白人的识别率高于白人。算法可能不会因为某些明显的偏见而指认一位非裔美国男性，但是算法指认非裔美国男性的可能性高于白人女性，这个事实说明该算法中同样存在偏见。在数据集中，杂乱无序的各种相关性暗含着隐秘的规律性，其中很可能就存在某些偏见。

我们以拉坦亚·斯威尼的经历为例。斯威尼是哈佛大学的一名博士，她是非裔美国人。一天，她在用谷歌进行搜索时，发现搜索结果中竟然出现了询问“你是否有被捕记

录”的广告。她对此大为震惊，因为她的白人同学在搜索时就没有看到这些广告。于是，她开始研究这个问题。最后她发现，谷歌搜索后台使用的机器学习工具在不经意间被植入了种族歧视，把黑人常用的姓名与“你是否有被捕记录”的广告关联在一起。此外，用户从谷歌的“安卓市场”下载男性同性恋者使用的基于地理位置的社交工具Grindr时，他们会被推荐同时下载一款性犯罪者的位置跟踪软件。这一现象也揭示了同样的问题。在这两个案例中，我们应该认为算法出错了，还是算法的设计者持有偏见呢？又或者，这两个案例是不是更有可能表明，人们在黑人与犯罪嫌疑人之间（第一个例子）或同性恋者与性犯罪行为之间（第二个例子），建立了某种令人不快的文化联系呢？无论出于何种原因，无论这些联系应该受到何种谴责，它们都反映了算法的另一面。某个人即使表现

出明显的偏见，他也只会影响少数人，而算法则有可能影响成千上万人的生活。

算法解决方案的透明度问题

透明度，或者说缺乏透明度，使问题的严重性进一步增加。与周边定律相似，很多算法解决方案就像一个黑盒子，依赖这些算法做决策的人根本不知道他们做出的决定是否正确，制定的政策是否客观公正，有没有歪曲事实。由于算法强调按部就班的程序，因此，法律编码化应该可以保证具体判决的跟踪、审核、验证工作更加简单明了，至少比有人为因素干扰时要好得多。从理论上讲，算法可以详细地描述应用于每个小的决定（进而形成最终的重要决定）之中的那些具体规则。但实际上，很多自动化系统都是不透明的，这个特点意味着它们可以规避审查。出于各种各样的原因，设计者有时并不公开源代码，因此，用户们无法看清其中的规则，无法提出不同意见，也不能参与决策

过程，只能接受最终的结果。被认定为疑似恐怖分子的人可能不得不接受长时间的盘问，甚至有可能因此误机，但是他们永远无法知道自动化系统为什么会盯上他们。这个事实意味着他们每次搭乘航班都有可能遭到羁留。在卡夫卡式的环境中，如果你不了解某个决定的形成过程，就难以提出反对的理由。设计一个从理论上讲可以在某些法律事务方面做出决策的公式，并不意味着一定要让该公式的内部运行机制公之于众，让普通用户也能了解，两者远不是一回事。

我们知道，这些问题很可能是由某些明显的偏见造成的，也有可能是由那些几乎没有接受过法律知识培训的程序员们不小心导致的。2004年9月至2007年4月，来自私营公司的程序员们曾将900多条不正确的规则植入科罗拉多的公共福利系统，这个事实证明，程序员在对某些法律条文进行编码时，

很有可能会改变其实质。达尼埃尔·西特龙说：“他们的错误非常严重，对政策进行错误的编码导致成千上万人遭遇不公正的对待，造成这个问题的全部原因就是编码的人没有相关的背景知识。”

这些程序员犯下的错误包括根据收入情况拒绝安排乳腺癌及子宫癌患者接受治疗，以及拒绝资助怀孕妇女。这些程序员把“无家可归”解读为“行乞为生”，导致一位失去了自己的住宅而只能流落街头的60岁妇女，在申请增加食品券数量时遭到拒绝，原因是她是乞丐。最终，为了绕开科罗拉多福利管理系统中的无数错误，资质审查人员被迫把虚假的数据输入系统。

无数计算机编码的背后都存在类似的问题。错误的编码导致法律条文被严重扭曲，其效果也发生了改变。西特龙指出，要在法律体系的整体框架下修改这些错误，可能需

要好几个月的时间。她说：“在通过一项新的政策法规之前，应该听取对此感兴趣的人与政策制定者的意见，为他们留出时间。然后吸取这些意见，对新政策做出明确解释，并对各种意见做出回应。在经过这些必要的程序（本质上要求高度公开）之后，新政策才能获准通过。而上述问题的症结在于其绕开了人们参与其中并提出专业意见的民主程序。”因此，程序员们被赋予了过多的决策权，而且不会受到任何审查。

尽管不是所有的错误编码都会导致“依据收入情况拒绝安排癌症患者接受治疗”这样严重的过失，但由于程序员偷懒或者疏忽，不少被编码的法规与书面法规相比，确实存在精细程度不够的问题。例如，美国《食品券法案》规定，没有孩子的失业成年人只能领取三个月的食品券。不过，这项规定还有6种例外情况，而且这些例外情况又会与其他

例外情况发生交叉，形成更多的例外情况。由于这些例外情况从统计学角度看非常罕见，因此程序员们有可能会为了省事，在编码时仅考虑三个月的期限这个规则，而把那些复杂的、有可能导致混乱的例外情况都排除在外。

尽管目前这个问题还仅限于对现有法规进行计算机编码的过程中，但从长远看它还有可能导致其他问题。比如，这些机构将越发倾向于舍弃那些细致入微的考虑方法，代之以在问题考虑与解决方法两个层面都简单易行的政策。即使在前者的效果更为显著时，他们也不为所动，因为把这些政策法规转变为算法的难度大于后者。如此发展下去，那些复杂的法规都将被过度简化，以提升法律事务自动化的效果。证据表明，这样的苗头已经出现了。马萨诸塞州的信息技术专家们就成功地说服了政府部门的决策者，让后者

不要采用可能导致自动化难度与成本升高的公共福利制度。一位知情人士说：“有的时候，真正掌控政策的似乎是信息技术人员。”

法官、陪审团与数据挖掘技术

理查德·波斯纳是20世纪广受赞誉的法学者。多年前，年轻的波斯纳还是一位经验不足的法官，他主审的一个专利案件涉及某种定向推介技术（后来，亚马逊通过算法完成定向推介）的早期应用。当时，卫星电视刚刚兴起，美国电视观众可以收看的电视频道从五六个一下子增至500个。对于很多人而言，这次“飞跃”催生了所谓的“选择悖论”，让他们感到无所适从。面临如此多的选择，他们如何才能正确地选择自己最希望观看的频道呢？一家公司为他们找到了解决方法。在发给每个家庭的调查问卷中，这家公司只了解他们对其中某一个频道的看法，并请他们列出经常在该频道观看的电视节目类型。公司承诺问卷调查的对象包括收看该频道的

所有人，同时宣称该频道将根据这次问卷调查的结果，播出最受用户欢迎的节目，为所有家庭打造个性化的电视频道。波斯纳回忆说：“他们可能认为你想在早晨看新闻，在下午看肥皂剧，在晚上看恐怖电影。”如果有一个频道能满足你的所有愿望，还需要费神考虑其他499个频道吗？

波斯纳认为这个想法“极具创造力”，并深受触动。在之后的多年时间里，他一直在思考这个创意对自己的工作会有什么启发。他回忆说：“可以想象，法官的审判工作也可以采取同样的方法。我们可以收集法官们的观点、公开言论，以及其他所有信息，然后为法官们建立档案，用来做出各种司法审判方面的预测。”最近，波斯纳在《21世纪法官的职责》这篇论文中深入探讨了 this 观念，预测算法时代的到来可能会给司法工作带来的各种改变。他在文中指出：“我们都熟悉亚马

逊网站建立、修改用户资料的那一套，还有些人熟悉数据挖掘技术。显而易见，数据挖掘也是利用计算机找出数据中隐藏的规律，并在接收到新数据后进行更新。”

我希望在将来的某一天，我们可以利用计算机为法官们建立档案，内容包括他们的观点以及公开言论，并在法官们表达新观点时及时更新他们的档案。（这些）档案将帮助律师与法官更准确地做出预测，保持判决的前后一致。

法官的决策过程非常复杂，令人难以捉摸。1925年，法官马克斯·雷丁在题为“司法判决理论：法官的思维模式”的文章中指出：“我们要从事的工作就是预测。如果预测都能做到确凿无疑，预言术就不值一提了。”由于该文章撰写的时间是在近100年之前，他对计算能力估计不足是情有可原的，但是作为一名法官，他表现出来的傲慢自大则有可能

饱受诟病。在雷丁发表这篇文章的4年后，联邦法官小约瑟夫·哈奇森提出，如果法官的思考过程是客观的，不受直觉、偏见或是“优秀法官不喜欢的其他因素”的干扰，我们就有理由认为，对于任何接受过适当的法律基础教育的人而言，法官的决策过程都应当一目了然。

研究证明，即使法律培训也并非必不可少。在2004年的一项研究中，研究人员安排若干名法律专家与一种算法同台竞技，预测高等法院判决时陪审团的多数意见。令研究人员吃惊的是，算法正确地预测出75%的判决结果（算法依据的只是为数不多的几种判断标准），而那些法学学者们尽管可以参考许多专业信息，预测准确率却只有59%。2011年，IBM的“沃森”超级计算机在《危险边缘》节目中赢得了100万美元，而“高等法院判决预测项目”则在法律行业中完成了这一

对决的翻版，标志着自计量法学运动以来长期萦绕在法律界人士心头的技术梦变成了现实。1897年，小奥利弗·温德尔·霍姆斯热情洋溢地表达了自己的信念：司法行为与自然科学一样，应该可以通过量化的方式进行预测。他认为，“我们的研究目的是预测公共力量通过法院产生的影响力。”

然而，高等法院判决预测项目更有可能是计量法学的乌托邦式幻想的一个变种。2004年的这项研究并不能证明法律系统的可预测性源自其固有的客观性，而恰恰表明这些判决不是客观的，因为每个案件的判决都经过了一系列主观偏好的筛选。例如，算法用于预测判决结果的一个指标，是看法官把他的选票投给了民主党还是共和党。客观地讲，法官把选票投给哪个政党与他们上一次进餐的时间一样，在判定被告是否有罪时，这两个因素都不应该对法官产生任何影响。

正是因为这些主观偏见，数据挖掘才会有所建树。波斯纳说：“这个事实对法官来说具有积极的意义，可以帮助他们发现自己在什么时候会形成偏见。”比如，一位法官可能对刑事犯罪嫌疑人判决较轻，而对商业诈骗嫌疑人的判决较重。波斯纳接着说：“在收到自己的档案之后，法官就会知道某些不自觉的偏见会让自己偏向某种判决结果。”

科罗拉多大学法学院副教授、思科公司前软件工程师哈利·苏尔登认同波斯纳的观点，但他认为，如果告诉普通大众司法决策会受到偏见的影响，他们可能会无法接受。他说：“在民主社会里，人们可能不希望看到偏见的存在。”

对这类算法持欢迎态度的可能是律师。律师可以将这些算法放入自己的工具包，用来为法官的行为建立准确的模型，从而为自己法庭上的证词与辩护争取最大的影响力。

苏尔登推测说：“现在，某些老练的法人实体可能正在利用这些算法为自己谋利。”他指的是对冲基金这类组织。对冲基金是一种大额的赌博行为，收益情况取决于司法判决的结果。如果对冲基金利用复杂的司法判决算法来预测判决结果，就有可能赚得盆满钵满。

法学界的其他一些学者也认为，数据挖掘可以精确计算某人触犯法律的具体程度，从而精准地揭示出判决结果的不公正性。不难想象，在不久的将来，神经科学（目前正处于向算法转型的阵痛期）将致力于确定人们的遵纪守法行为在多大程度上是出于自己的意愿，并讨论外部条件的限制作用与自觉自愿孰重孰轻等问题。数据挖掘还可能会被应用于另一个方面：利用算法将当下的判决结果与以往的判决结果数据库进行对比，以确定应该对犯罪行为如何惩处。例如，对犯罪嫌疑人判处多长的刑期才能最有效地防止他

们再次犯罪。

让我们考虑得再深入一些。如果审判工作也实现自动化，会怎么样呢？如果我们构建的推介系统可以预测某位法官对某个案件的判决结果，且准确率达到99%（预测结果甚至可能比该法官的判决更能体现一致性的特点），是否会有助于提高审判的公正性呢？法官波斯纳说：“从理论上讲，这是有可能的，但要走的路还很长。在新信息出现之后，法官的观点可能会发生改变，而系统没有考虑这种情况，这可能是它需要解决的主要问题。有可能影响法官观点的任何变化都必须输入算法程序，并在预测法官会如何判决时加以考虑。”

既要解决各方的不满，又要处理各方对法律的不同解读，至少从目前来看，只有（人类）法官的创造性才能同时满足这两个要求。因此，司法程序追求的与其说是机械的

客观性，还不如说是在多个实体间达成高度的一致。两者之间的区别似乎微乎其微，但却非常重要。算法有可能胜任法庭上的多项工作，甚至可以有效地提高现行法律制度的公平性，但是它们无法胜任判决工作。

至少目前不行。

第4章 用科学的算法 进行艺术创作

1983年，电影编剧威廉·戈德曼在其回忆录《银幕春秋》中揭露了电影行业的内幕。在电影行业拼搏了几十年后，戈德曼对这个行业深刻的认识，概括起来是：电影制作的棘手之处就在于“没人能做到心中有数”。这可能是有史以来对好莱坞最著名的评论了。他在书中指出：

你们是否知道《夺宝奇兵》曾经被众多制片公司拒绝，只有派拉蒙公司慧眼识珠，选择了这部电影。为什么呢？因为没人能做到心中有数。为什么其他电影公司不愿意接受这部电影呢？同样是因为没人能做到心中有数。为什么最大的环球影业公司会将《星球大战》拒之门外呢？还是因为没人能做到心

中有数。

戈德曼的这种观点不乏支持者。传奇制片人迈克·麦德沃（参与过《现代启示录》《飞越疯人院》《沉默的羔羊》等影片的拍摄工作）在自传中指出：“电影业可能是世界上最缺乏理性的行业，（它）受到一系列毫无理性的规则的影响。”在好莱坞的传奇中，被预测肯定能获得成功却惨遭失败，而被预测肯定会铩羽而归结果却成为大赢家的电影比比皆是。有些商业电影让许多人趋之若鹜，而有些主流电影却无人问津。几乎所有的好莱坞决策者都蒙受过败走麦城的羞辱，但是，我们在分析他们失败的原因时，很难把所有的责任都归咎到他们身上。

我们以两名最优秀导演的经历为例。21世纪初，詹姆斯·卡梅隆宣布他正在致力于“880计划”，打算将一个神秘的新剧本精心打造成一部精彩的电影。卡梅隆之前执导过一系列

深受观众欢迎的作品，包括《终结者》《终结者2》《异形》《深渊》《真实的谎言》《泰坦尼克号》等。其中，《泰坦尼克号》赢得了奥斯卡奖，并成为电影史上第一部票房收入超过10亿美元的电影。不过，10年来，卡梅隆“不务正业”，一直在拍摄纪录片，而且他在这部新电影的拍摄中没有起用任何大牌演员。卡梅隆计划把它拍摄成3D电影，预算为2.37亿美元。尽管他的计划可能缺乏理性，但电影最终还是完成了。上映之后，一名影评人说：“这是美国电影史上成本最高的影片，可能也是最反美的影片。”一语成谶，这部电影很快就被贴上了这样的标签。

如果你是投资方，你会同意卡梅隆拍摄这部电影吗？进过电影院的人大多会给出正确的答案：“会”。在更名为《阿凡达》之后，“880计划”打破了《泰坦尼克号》创下的票房

纪录，成为电影史上第一部票房收入超过20亿美元的电影。

在《阿凡达》势头正猛的时候，另一部电影也正在好莱坞四处寻找投资方。这部电影也是科幻题材，也是3D影片，取材于一个经典童话故事，编剧之一是一位普利策奖得主，拟由安德鲁·斯坦顿执导。斯坦顿的执导成绩毋庸置疑，他之前帮助皮克斯公司制作的《机器人总动员》与《海底总动员》都大获成功，而且其参与拍摄的每一部《玩具总动员》都赢得了一片喝彩。斯坦顿为这部电影（我们把它称作“X计划”吧）设定的预算为2.5亿美元，略高于“880计划”。

“X计划”最终也成功地找到了投资方，但是它没有再现《阿凡达》的辉煌，而是遭遇了一场惨败。这部最终名为“异星战场”的电影导致迪士尼公司亏损了2亿美元，公司领导者因此引咎辞职。戈德曼说得好：没人能

做到心中有数。

公式和规律随处可见

对于追求挑战的技术专家而言，这样的说法当然是他们喜闻乐见的。在这些追求公式化的人看来，世界上竟然存在根本无法预测的事物（而且还是娱乐业），这真是一个利好消息。巧合的是，卖座的电影与高科技初创企业之间有很多共同点。一方面，大多数电影与初创企业都会遭遇失败，因此在取得成功时，成功的规模必须足够大，不仅要抵消失败带来的损失，还要给投资者留有获利的空间。另一方面，对这两个行业而言，“电梯演讲”更容易发挥威力。电梯演讲要求所有概念通俗易懂，可以在乘坐一次电梯的时间里解释清楚。电影通常会用30秒左右的广告时间向广大观众做推介。这可能并非巧合。

同几乎所有的通俗概念一样，电梯演讲（在好莱坞被称作“高概念电影”）这个概念

的提出要归功于若干业内人士，其中史蒂文·斯皮尔伯格的解释得到了普遍认同。这位执导过《侏罗纪公园》与《E.T.外星人》的导演说过：“如果电影的内容可以用25个甚至更少的单词加以概括，这部电影就会受到欢迎。我喜欢传播可以让观众轻松接受的理念。”斯皮尔伯格的父亲阿诺德·斯皮尔伯格，是计算机科学领域的一位先驱，以当时的标准看，他设计并拥有专利的第一个电子图书馆系统可以搜索堪称海量的数据。在得知这个事实后，你会不会感到吃惊呢？在自然科学方面，史蒂文·斯皮尔伯格可能不是一个优秀的学生，但是他的血管中流淌着的都是“公式”。^①

好莱坞的另一个人的血管中同样流淌着“公式”。他曾经因“机器般的干劲儿”而受到《时代》杂志的赞扬；他最初计划在麻省理工学院学习工程技术，但是后来却走上了从

影的道路，出演了一连串创造票房纪录的影片。他就是威尔·史密斯，在他的职业生涯早期，史密斯只不过是一个新晋演员，刚刚出演了他的第一部电视剧《新鲜王子妙事多》。但是，这位有远大志向的演员与他的经纪人一起，探讨是否可以制定一个公式，将他从不名一文的普通人变成“世界上最著名的电影明星”。史密斯认为自己是“研究宇宙规律的学生”，当时，他一边努力打拼，希望有星探慧眼识珠，一边认真阅读电影业的专业报纸，了解全球观众在电影品位方面的变化趋势。他在后来接受《时代》杂志采访时回忆说：“我看着这些报纸，问自己：‘到底有什么规律呢？’我发现，100%的电影都用到了特效，90%的特效涉及人，80%的特效不但涉及人还讲述了一个爱情故事……”20年过去了，史密斯参演的电影已经在全球赚了63.6亿美元，但他的这个方法并没有发生太

大的变化。他说：“每个周一的上午我都会坐下来，（然后问自己：）‘这个周末发生了哪些事情？有哪些事情跟过去10个、20个、30个周末里发生的事情比较相像？’”

当然，尽管史密斯毫无疑问具有很强的票房力，但在发现普遍规律方面，他仍然是一个“菜鸟”。

1. 在人们交口称道的史蒂文·斯皮尔伯格的传记中，作者约瑟夫·麦克布莱德根据阿诺德·斯皮尔伯格的回忆，记录了父子二人之间的一段对话：“我说：‘史蒂夫，你学数学吧。’他回答说：‘我不喜欢数学。’他总是要我替他完成化学作业。他从来都不愿意动手做化学实验，而是在放学回家后直截了当地告诉我：‘爸爸，今天的作业是写实验过程。’我问他：‘你不给我任何实验数据，我怎么知道你是怎么做实验的呢？’我会想方设法地帮助他重做实验，然后告诉他答案。等他（从

学校) 回到家后，他会冲着我喊：‘天啊，爸爸，你的化学实验不及格！’”

能预测电影票房收入的“秘密武器”

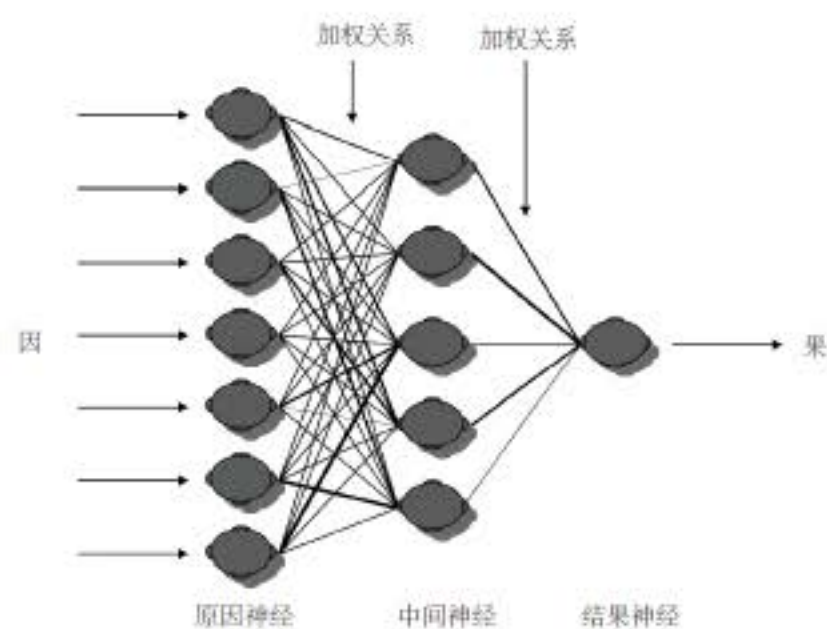
一家名叫“Epagogix”的英国公司在对影片票房成绩的预测中使用了威尔·史密斯的方法。与史密斯相比，这家公司没有英俊的相貌，但在预测的规模上却多出了几个数量级。Epagogix公司位于肯宁顿的一间狭窄的办公室里，墙上贴满了电影海报，为数不多的数据分析师坐在电脑前。这样的一家公司似乎不大可能成为好莱坞巨头的秘密武器。公司的名字来源于一个希腊单词，意思是由经验通向知识的道路。公司的首席执行官和创始人之一尼克·米尼大胆宣称，该公司在一部电影开机之前，就可以准确地预测它能产生多少票房收入。

Epagogix走进好莱坞的过程异乎寻常。米尼是一位40岁左右的英国人，一头蓬乱浓

密的黑发，长相与职业生涯中期的奥逊·威尔斯有点儿相像，有风险管理的从业经历。在米尼的职业生涯中，几位从事数学研究的朋友介绍他到所谓的“神经网络”（neural networks）领域工作，利用巨型人工大脑分析各种情况下或复杂，或含糊不清，或者两者兼有的因果关系。比如，神经网络可以用于读取、分析从铁轨上驶过的火车车轮的声音。如果信息正确，就可以在发生重大事故之前预测出需要修理的路段。米尼认为神经网络有可能在保险行业发挥重要作用，但是他的老板不假思索地拒绝了这个点子。他意识到在保险行业中，保费的多少并不代表精算师认为某个事件发生的可能性，因为预防某个问题的能力越强，相应的保费就越低。正是出于这个原因，老板才没有接受他的创意。

但是，电影业有可能从这种不合常规的想

法中获益。由于大预算电影的平均制作成本高达数千万甚至是数亿美元，一次较大的失败几乎就可以让电影公司破产。1980年就发生过这样的事，臭名昭著的失败之作《天堂之门》（被《卫报》评论家乔·昆南称为史上最糟糕的电影）几乎让联美公司破产关门。一位电影公司的老板告诉米尼，如果有人可以设计出一种算法，哪怕每年制止一部赔钱电影的拍摄，也将给这家电影公司的财务状况带来不可忽视的改善。米尼受到了诱惑，开始设计这种捕杀失败之作的公式。他的几位数据分析师朋友正在设计一种预测热门电视节目收视率的算法，于是米尼与他们展开合作，开发了一个可以预测电影票房收入的系统。这个系统利用30073680个特定的评分指标（包括电影中是否有明确的坏蛋、主角是否有死党等），得出预计的票房收入。



简化版神经网络示意图。在图的中间部分，标有“中间神经”而且像翻绳 游戏那样密密麻麻的连线，就是让Epagogix 公司在电影行业大展身手的“秘密武器”。

好莱坞一家大型电影公司为了测试这个系统的预测效果，把9部已经完成制作并准备发行的电影的剧本发给Epagogix公司，请他们用这个神经网络系统预测每部电影的票房收入。为了增加预测难度，他们没有告诉米尼等人关于演员、导演、营销预算等方面的信息。在算法给出的9个预测结果中，有三个误差很大，但是其余6个的准确程度令人瞠目结舌。其中，令人印象最深刻的预测结

果是关于一部名叫“幸运牌手”的电影。该电影的成本为5000万美元，由著名女演员德鲁·巴里摩尔担纲主演，由柯蒂斯·汉森执导（《8英里》和《洛城机密》等热门影片的导演），编剧是埃里克·罗斯（电影《阿甘正传》的编剧）。电影描绘的是高风险的纸牌游戏世界，这是一个非常受欢迎的题材。Epagogix的神经网络系统预测这部电影会小赚一笔，利润约为700万美元。它最终的利润为600万美元，从此以后，Epagogix公司在好莱坞站稳了脚跟。

当然，人们有理由感到疑惑，为什么要让计算机来完成这项工作呢？前面说过，每当新的一周开始时，威尔·史密斯做的就是类似的事情。人们真的可以依据30073680个特定评分指标，统计出某个剧本符合其中多少个得分点吗？答案非常简单：不可能。尽管分别标记每个指标没有任何难度（不过需要大

量时间），但是这种做法不会告诉我们这些因果变量之间的相互作用会对票房收入产生怎样的影响。我们可以引用乔治·奥威尔《动物农场》中的一句话并稍作修改：所有数据一律平等，但有些数据比其他数据更平等。^①

我们以一部成功的电影为例。2012年3月，《饥饿游戏》上映后迅速引起了轰动。《饥饿游戏》之所以走红，是因为它取材于一部十分畅销的小说，从而为电影带来了许多观众吗？还是因为这部电影的女主角是曾被《滚石》杂志称为“好莱坞最酷女星”的詹妮弗·劳伦斯呢？又或者是因为这部电影刚好安排在假期上映，因此很多年轻人有时间去观看？对于这类问题，人们只能回答：可能是这样。《饥饿游戏》的确取材于一部畅销小说，担纲主演的的确是深受这部电影主要观众群欢迎的当红年轻女明星，上映日期也

确实是在美国年轻人不需要上学的春假。但是，很多同样符合这些条件的其他电影却没有受到追捧。尽管《饥饿游戏》的最终票房收入超过6.86亿美元，但是谁敢肯定上述因素都发挥了积极作用呢？是否存在这样的可能：一些人本来要去观看这部电影，但在听说这部电影取材于一部小说，或者詹妮弗·劳伦斯在其中担纲主演，又或者因为担心电影在假期上映，电影院里会挤满吵闹的学生，而最终选择待在家里呢？假如电影发行商未卜先知，推迟了上映时间，那么票房收入是否会增加2亿美元呢？

Epagogix的任务就是量化上述问题。聘请米尼团队预测票房收入的电影公司需要为他们提供分镜头剧本、演职人员表以及计划上映时间等信息，而米尼团队则向电影公司提交一个封好口的棕色信封，里面装有神经网络系统做出的分析报告。“以前，我们提交

的报告通常有这么厚。”米尼一边说，一边伸出大拇指与食指比量了一下。这个厚度与每一期《What Car?》汽车杂志的平均厚度差不多。但是，他们怀疑客户不会认真地阅读全文，因此现在的报告仅为用一枚订书钉装订的两三页纸。米尼接着说：“你们可能以为，电影公司希望了解更多信息，但实际上，我们花费了大量时间才让这些报告成功瘦身。”

报告最后一页的内容最为重要，上面有对电影票房收入的最终预测结果。同时，报告还会给出另外一个神秘数字，这个数字通常比前一个数字高出10%左右，有时甚至会是前一个数字的两倍。这个数字是如果剧本按照米尼团队的建议做出修改之后的预计电影票房收入。米尼解释说，由于我们对每个剧本都进行了非常详尽的回归分析，因此神经网络系统可以甄别出导致预期落空的那些要

素，或者说影片的短板。在这个环节，Epagogix根据客户的意向以及他们对技术的信任程度，要么为创造奇迹而努力，要么给出一份令人泄气的预测报告。不难想象，面对这些有可能帮助他们打破票房纪录的建议，有的编剧会虚心接受，而有的编剧则会因为算法对他们指手画脚而产生反感。

对于米尼团队的预测报告感到难以取舍的可能不仅仅是那些编剧。米尼说：“在我们经常合作的电影公司领导者中，有一位特别精明。每次交谈时，他总是迫不及待地提出一些问题，比如‘如果主角穿红衬衫，你们的系统会不会给出不一样的预测结果？’他的目的不是要为难我们，而是希望了解我们的方法。我告诉他这得视情况而定，他对这样的回答很不满意。做出这个修改后，神经网络参考的其他变量会不会随之发生变化呢？对一部电影做出重大修改，往往意味着它在其

他某些方面也会发生变化。”我们把混沌理论之父爱德华·洛伦茨说过的一句话稍作修改：一只蝴蝶在电影的开头扇动翅膀，很有可能会在电影第三幕的中间部分导致一场飓风。

与米尼讨论剧中人物衬衫颜色的那位电影公司老板，很可能是故意挑选了这样一个具有主观随意性的细节。一个人在考虑周六晚上去看什么电影，或者向朋友推荐哪部电影时，难道电影主角穿蓝衬衫还是红衬衫会对他的决定产生任何影响吗？但是，当时的情况却让米尼深感不安。当然，他不安并不是因为那位电影公司老板问了一个错误的问题，而是因为他的回答是“视情况而定”，而且他当时并没有觉得这是一个站不住脚的答案。当天晚上，他给在英国的几位同事打了个电话。米尼回忆说：“他们不假思索地张口就说：‘当然得视情况而定了，想想《辛德勒的名单》吧。’在这部电影的结尾，我们能看

到一抹颜色，它绝对是整部电影的一个关键情节。对于我们的系统而言，这抹颜色为整个故事赋予了一个历史背景。在这部电影里，颜色起到了非常重要的作用，但在其他电影里颜色却可能是一个无足轻重的问题。所有事物都是相对的。”^①

1. 奥威尔的原话是：“所有动物一律平等，但有些动物比其他动物更平等。”

2. 不论这个镜头要表达什么意思，有一点是可以肯定的：影评家们对它褒贬不一。有的影评家为之叫好，有的影评家则认为这是这部电影最大的败笔。影评家戴维·汤普森发表在《新共和》杂志上的一篇文章评价这个镜头“太糟糕了”。

艺术家与科学家之间的鸿沟

长期以来，哲学家一直想弄清楚是否存在判断艺术作品优劣的普遍真理。虽然我们大多数人都不可能一一考虑Epagogix给出的30073680个特定评分指标，但我们在潜意识里也会赞同美国最高法院大法官波特·斯图尔特的观点：“只要看见了，我们就会知道。”如果艺术领域不存在普遍真理，我们为什么还要多此一举，设置最佳歌曲、绘画、雕塑、小说及电影等奖项呢？在每个方面都设立一个客观真理，其实是为了让我们内心的信念与外部世界的“事实”形成对应关系。一旦建立了这种对应关系（或者相关性），就说明我们掌握了真理，否则，就说明我们错了。

1790年，德国哲学家伊曼努尔·康德出版了《判断力批判》一书，对美学领域的基础

性规则进行了描述。当时，“美学”这个词进入人类社会才40年的时间，美学领域正在迅猛发展。现代读者可能难以理解康德的观点，但是康德本人也相信在所谓的“自然界以之为基础的超感官之物”中存在某种普遍标准，例如美。由于康德认为存在这种绝对概念，因此他反对在不设定基本概念的情况下就随意使用“美丽”等表达。

当某人认为某个事物符合自己的品位时，他会说“在我看来，这个东西（我们看到的房子、某个人身上穿的外套、我们聆听的音乐会、我们阅读的诗歌）很美。”他特意强调“在我看来”，是为了表明他是客观公正的。这样的说法其实非常可笑。原因在于，如果这个对象只是让他感到愉悦，那么他不能称之为美。让他觉得迷人、愉悦的事物可能有很多，但是，只要他说某个事物很美，就意味着其他人也是这样认为的。他不仅自己做

出了判断，而且还代表所有人做出了判断，把美看成了该事物的一种属性。

认为艺术领域存在客观真理的人不止康德一个。德国哲学家格奥尔格·威廉·弗里德里希·黑格尔在《美学》一书中指出，艺术是“绝对理念的表现”。与此同时，法国诗人夏尔·波德莱尔在自己的作品中多次提及艺术具有“永恒不变”的特点。在此基础上，爱丽丝·默多克在《利益的主权》中进一步指出：

好的艺术告诉我们，以客观的眼光看世界，世界就会大不相同，这说明我们很难保持客观性。艺术以稳定的形式把人们周围的状况真实地展现在我们眼前，让我们可以从容不迫地凝视它们。

当然，持不同意见的也大有人在。柏拉图认为，艺术是现实的简化，“与真实相去甚远”，是与“灵魂理性原则”相对立的另一个极端。柏拉图对艺术的热爱可能并不十分强烈

（出于构建最理想社会的需要，柏拉图将人分成了若干层级。在他列出的名单里，艺术家仅比暴君高出三个层级），但他认为艺术具有非理性的本质，这个观点在历史上被人们反复提及。在《影响的焦虑》一书中，著名文学评论家哈罗德·布鲁姆引用了弗洛伊德的意象概念，指出青年艺术家被迫反抗一切传统，创作旨在颠覆现状的新作品。换言之，艺术家（为了维护他们艺术家的身份）只有找出规律并打破规律，才能在左右脑永无休止的四子棋游戏中把“公式”甩在身后。

与之相似，哲学家伯恩斯坦在《艺术的命运》一书中指出，人们认为艺术与美学“游离于真理、理智与道德之外”。对于寻找事物之间的相关性的数学家或者计算机科学家而言，这个观点令他们看不到希望。而哲学家亚瑟·叔本华则认为艺术家超脱于理性之外。由于理性是逻辑与自然科学的决定因素，因

此他认为艺术家的显著特征是极度厌恶数学。

1959年，英国科学家、小说家斯诺发表了论文《两种文化》，对横亘在艺术家与科学家之间的这条充满敌意的鸿沟进行了概述，引起了人们的广泛关注。斯诺认为，“整个西方社会的学术圈”分成了两个阵营（硬科技阵营与人文学科阵营），两者之间的分歧是我们解决世界上各类问题的主要障碍。他在文中指出：“科学家以外的人从骨子里认为科学家非常肤浅，过于乐观，不了解人类的现状。而科学家则认为人文学科的知识分子没有前瞻性，对人类漠不关心，从深层意义上讲是反学术的，渴望将艺术与思维局限于短暂的存在。”斯诺悲哀地指出，这两个极端之间的鸿沟是“我们人类和社会的一大损失”。他认为，人文学科的知识分子是“天生的技术进步反对者”，而科学家则在很大程度

上蔑视艺术创作，这种隔阂很难消除。不过，斯诺并没有因此失去信心。

“平行宇宙”实验与流行歌曲 排行榜

为了检验艺术领域是否真的存在普遍真理，我们可以将时间回拨，看看同一部作品是否可以再次受到欢迎。如果Epagogix等公司所使用的预测公式是正确的，那么无论我们将这个实验重复多少次，热门电影或者畅销小说仍将获得成功。比如，《阿凡达》的票房收入一定会超过20亿美元，而《异星战场》仍将惨遭失败；沃尔夫冈·阿马多伊斯·莫扎特注定会成为一名伟大的音乐家，而安东尼奥·萨列里则注定籍籍无名。同样，人们不会说某部电影“不大可能走红”，因为根据普遍真理，电影能否获得成功是有规律可循的，如果符合这些规律就会走红，否则就会失败。

几年前，普林斯顿大学的几位研究人员提

出了一个独具匠心的想法，利用计算机开展模拟实验来验证艺术领域是否存在普遍真理。马修·萨尔加尼克、彼得·多兹与邓肯·沃茨发现，在娱乐行业中成功似乎是有预兆的，他们因此产生了灵感。他们发现，正如某些初创公司在强手环伺的情况下脱颖而出并赢得市场一样，某些书籍与电影也在市场上一枝独秀，成为超级大赢家。两者都是制造“超级明星”的市场。例如，1997年，《泰坦尼克号》的票房收入约为美国当年平均票房收入的50倍。这三名研究人员希望了解的是：《泰坦尼克号》之所以如此成功，是因为它的优秀程度是同年上映的其他电影的50倍，还是因为其某些情节非常吸引人呢？萨尔加尼克说：“电影作品的成功极难预测。我们可能会认为那些大获全胜的影片应该与它们的竞争对手之间存在某些不同点。但是，希望找出两者之间不同点的人，并不总能有

所收获。”

换句话说，如果有人认为“没人能做到心中有数”的戈德曼定律是正确的，那是因为他们太过愚蠢，不能慧眼识珠，还是因为万物本来就具有不确定性，所以人们无法预知呢？

为了验证他们的假设，萨尔加尼克、多兹与沃茨开始了他们的“虚拟文化市场中不平等性与不可预测性的实验研究”。他们模拟的这个在线音乐市场有点儿像苹果iTunes数字媒体播放应用程序，但供下载的都是不知名乐队创作的不知名歌曲。研究小组在网上招募了14341名实验对象，让他们听这些歌曲，然后用1（我讨厌这首歌）至5（我喜欢这首歌）的分数给它们打分，并允许他们下载自己最喜欢听的歌曲。下载量最大的40首歌曲被排列成“热门歌曲排行榜”，显示在网站的显眼位置。

与iTunes不同的是，萨尔加尼克、多兹与沃茨创建的不是一个在线音乐市场，而是9个在线市场。用户登录网站后，被随机导向9个“平行宇宙”中的一个，这9个平行宇宙一模一样，唯一的区别就在于热门歌曲排行榜。三名研究者认为，如果热门歌曲真的远胜于普通歌曲，那么可以预见的是，同一首歌曲在每个“宇宙”中都应该处于相同的位置上。

不过，萨尔加尼克、多兹与沃茨发现真实情况并非如此。实验结果表明，成功存在偶然性，这与他们的预测如出一辙。在实验早期，一些引领时尚的用户试听了一些歌曲。出于某些无关紧要的原因，某些歌曲的下载量排名靠前。领先地位一旦确定，社会反馈效应就会将其领先优势进一步放大。例如，如果某家书店发现某一本书非常受欢迎，就会增加订货量。此时，这本书与其销量最接

近的竞品的销量之比可能是11：10，领先优势微弱。但是，等这本书的添货到店后，书店就会把它们摆放在比较好的位置（例如，在门口的桌子上），这本书的销量将会大大增加，变成竞品销量的两倍。接着，书店为了进一步促销，决定降低利润率，降价销售这本书，以此来吸引新顾客。此时，这本书的销量变成竞品的4倍。由于顾客认为这本书非常受欢迎（因此肯定值得一看），他们购买这本书的可能性就会增加，于是，销量也会进一步增加。

这就是心理学家所谓的“曝光效应”（mere-exposure effect）。有时候，人们看到某本书越来越热门，因为怕自己落伍而购买这本书，从而引爆了这本书的销售。这就好比在得知某个电视节目引起轰动后，我们也会去看这个节目，以了解人们为什么会有如此热烈的反应。

在萨尔加尼克、多兹与沃茨的这个实验中，在一个“宇宙”里位于最不受欢迎行列里的歌曲，在另一个“宇宙”里也绝不会成为最受欢迎的歌曲；而在一个“宇宙”中最流行的歌曲，也绝不可能在另一个“宇宙”中无人问津。不过，除此以外，其他任何结果也都有可能出现。

依据吸引力的大小量化艺术作品价值

大家可能已经发现，萨尔加尼克、多兹与沃茨的阐释存在一个问题，他们在总结自己的发现时也承认了这一点。在研究流行程度与特点之间关系的实验中，如果不先为特点下一个可以量化的定义，如何能得出有价值的结论呢？在论文中，这三位研究人员用沉重的笔调指出：“遗憾的是，关于如何量化特点，还没有一个被普遍认可的方法。主要原因在于特点在很大程度上是（甚至完全是）一种社会建构。”为了避开这个问题（他们称之为“这些概念上的难点”），萨尔加尼克、多兹与沃茨决定彻底回避对歌曲的“艺术特点”的讨论，转而关注更容易量化的“吸引力”。

为了测量歌曲的吸引力，他们又模拟了一

个音乐市场，不过在这个音乐市场里看不到记分牌。萨尔加尼克、多兹与沃茨认为，在无法清晰地看到社会反馈系统作用的情况下，某些歌曲成为最受欢迎歌曲，可能只是因为它们的某些优点具有吸引力。至于是哪些优点并不重要，重要的是这些优点确实客观存在。

选择“吸引力”，放弃“特点”（这里并没有讽刺的意味），并利用市场反应加以阐释，这样的做法非常流行。Epagogix公司的尼克·米尼说：“在我看来，有很多电影可能比《泰坦尼克号》更加优秀，但在市场中赚钱最多的却是《泰坦尼克号》。”如果公司的神经网络与米尼在评判优秀艺术作品方面的标准碰巧一致，这自然让人无话可说。但是，如果两者并不一致，更有经济头脑的做法是推荐电影公司投资那些很多人愿意掏钱买票，而且看完后觉得没有上当受骗的电影，而不是

把钱浪费在几位影评家热烈赞扬却无人愿意观看的电影上。2006年，奈飞公司组织了一个奖金额度为100万美元的公开竞赛（现已取消），请用户设计一款可以显著改进奈飞推介系统的过滤算法。在这次活动中，奈飞沿用了一种类似的逻辑。如果算法向用户推荐相对不知名但受到影评家欢迎的电影（比如，小津安二郎1953年拍摄的代表作《东京物语》，让·雷诺阿于1939年拍摄的《游戏规则》等），那么这种推介效果不会被视为一种“改进”。如果推介系统可以准确地预测用户给一部电影或者一个电视节目的评分，奈飞公司就会认为这是“效果更好的推介”。换句话说，这种评判“最佳推介方式”的逻辑与认为“门口停车最多的就是最好的餐馆”这句老话如出一辙。

大众吸引力是一个现代概念，出现于19世纪末20世纪初工厂生产线兴起的时候。随着

能识文断字的人口迅速增加以及人们纷纷涌向城市，历史上第一次出现了真正意义上的大众市场。同时，这也标志着成套公式的诞生，从此，把产品卖给尽可能多的顾客变成了生产的一个要求。批量生产还意味着标准化生产，因为生产与分销过程要求一切从简，不仅消费品如此，那些在本质上不要求生产过程简单化的商品也是如此。例如，在介绍报纸历史的《大众媒体（1833~1865）》一书中，作者指出，早期的报业巨头“把新闻包装成产品，以吸引大众”。

因此，文学、艺术、娱乐与其他产品并无区别。在这个时代的乌托邦式梦想中，大众吸引力意味着整个世界将摒弃社会精英心目中的“艺术”，转而追求它的同胞兄弟——程式化的“工程技术”。汽车、飞机甚至整幢的房屋，都将被一股脑地从工厂生产出来，常态

化的不平等性已不复存在。但是，所有人都开着同样的福特T型车、住在同样的房子里，这怎么可能呢？在这样的社会中，艺术将专属于社会精英，缺乏理性，流于表面；而工程技术则属于全社会，注重功能性，具有超理性。工程技术不一定能够满足少数人自命不凡的主观性，但它可以更好地满足大众民主化的客观性。

在这种情况下，电影院顺势而为，变成了以公式化叙事方式传递流行元素的理想媒介，以及计算机领域的学者列夫·曼诺维奇所谓的“新媒体”典范。曼诺维奇说：“之前，人类眼中的活动影像，都不可避免地伴有无规律性、不一致性和意外情况等。现在，所有这一切都被电影镜头中的一致性所取代。”电影行业的先驱们认为，电影院这个媒介可以把工程技术领域的公式应用到娱乐行业中。苏联电影制作人、宣传家谢尔盖·爱森斯坦一

提到这个勇敢的新生事物就非常激动，他认为：“我们需要的是科学，而不是艺术。‘创作’这个词毫无意义，应该用‘劳动’来取代它。我们无法创作一件作品，而是像生产机器一样，通过劳动把零配件组装起来。”爱森斯坦认为电影是“科学技术”的载体，可以非常有效地、准确地激发观众的情感。后来，在回忆早期的执导生涯时，他说：

别忘了，埋头苦干、用科学方法探索艺术奥秘的是年轻的工程师。他的研究告诉他一个事实：科学探索离不开基本单位。因此，他开始苦苦思索，希望为艺术给人们留下的各种印象建立一个基本单位。科学上有“离子”“电子”“中子”等基本单位，我们就用“吸引力”来作为艺术的基本单位吧。

认为可以利用科学的方法来研究艺术的不仅仅是爱森斯坦一人。当时，大批艺术家也认为将艺术分解成最小单位，可以大大增强

艺术作品的社会功能，从而实现为艺术赋予“价值”的目标。在这个想法的激励之下，他们摩拳擦掌，充满信心。很多人把目光投向了一些机械形式的创作活动，包括纺织业、工业设计、平面造型设计，以及凸版印刷、摄影、合成照片等。苏联艺术文化学院的艺术家们难掩兴奋地宣布：“再也不需要画画了”，“认为艺术作品‘神圣不可侵犯’的观点被彻底颠覆了”。艺术领域的学者尼古拉·普宁更加激进，他发起并亲自参与设计了一个他认为可以解释艺术创作全过程的数学公式。^①

这种技术狂热行为自然会遭到质疑。面对颠覆性技术的问世，几位思想保守的学者开始批评他们眼中文化领域的重大转变。例如，德国心理学家、文学批评家瓦尔特·本雅明在论文《机械复制时代的艺术作品》中指出：

随着第一个真正具有革命意义的复制手段——照相机的出现，艺术领域感觉到危机正在逼近。艺术领域依据“为艺术而艺术”的原则做出回应，“纯粹”艺术应运而生，它不仅拒绝了艺术的所有社会功能，而且人们无法根据艺术所表现的主题对其进行分类。

几年之后，两名德国理论家西奥多·阿多诺和马克斯·霍克海默，在他们于1944年出版的《启蒙的辩证法》一书中详细讨论了本雅明的观点，戏谑地称之为“文化工业”，并对其进行了猛烈抨击。阿多诺与霍克海默的理由非常简单，他们认为艺术创作与生活的其他方面一起，都已经被醉心于量化运动的工业家接管。要从事艺术创作，艺术家必须循规蹈矩，在社会体系面前俯首称臣，因为这套社会体系“会压制一切反抗行为，让所有人都遵守规矩”。他们认为，艺术因此失去了它的所有价值，沦落为在极端利己主义驱动下

的各种自动化的生产创造。受害者是普通大众，摆在他们面前的是：

批量生产、质量参差不齐、等级不一的产品。于是，全盘量化的规则应运而生，所有人的行为必须遵从（就好像一种自发行为）事先设定的标准，并选择符合自己需求的批量生产的产品。消费者变成研究机构图表上的统计数字，这些图表会根据收入情况将这些消费者分别放到红色、绿色和蓝色区域中。

如果阿多诺与霍克海默现在还活着，他们绝对会相信Epagogix等公司在好莱坞制作好一部电影之前，就可以成功地预测它的票房收入。忘了那些专业影评家吧，根据统计数据制订拍摄计划的电影，其希望得到的不过是统计分析罢了。

1. 普宁的公式是： $S(P_i + P_{ii} + P_{iii} + \dots + P_n) \gamma = T$ ，其中， S 是所有原则（ P ）的总和， γ

表示直觉，T表示艺术创作。

从二维照片变成三维物体

2012年，文化产业发生了另外一个重要转变。尽管当时几乎没有人评论这个转变，但是这一年，在美国观众观看过的电影中，被合法地发布在互联网上的电影数量第一次超过以蓝光影碟与DVD（数字多功能光盘）等为载体的电影。同时，亚马逊宣布，虽然推出Kindle电子书阅读器还不足两年时间，但是客户购买的电子书数量已经超过其购买的精装书与平装书的总和。

乍一看，这个变化似乎没有多大的意义。毕竟，人们仍在观看电影，仍在读书，只不过购买与消费的方式发生了某些变化。这就好比人们仍然购买盖璞（Gap）的服装，只不过以前买的是微喇叭裤，而现在买的是紧腿裤。

不过，虽然这个类比表面看来比较恰当，

但并没有反映出这一转变的程度之深。我们不能简单地认为电子书的销售激增表明的只是人们不看纸质书，而是通过屏幕看书。电子书与纸质书都是一种实体，需要使用某些特有的技术与材料作为介质。要把电影、音乐、书和美术作品等显现在电脑显示器、平板电脑和智能手机上，我们需要把它们转换成数字形式。无论某个艺术作品最初是不是用计算机创作的，现在我們都可以完成这个步骤。于是，有史以来第一次我们可以利用数学方法（即公式）来描述，并利用算法来编程和处理所有的艺术作品。


能量可以由动能转化成热能，同样，现在信息也可以在不同媒介之间轻松传递。例如，我们可以利用算法来确定一张二维照片中是否存在阴影，然后通过测量它们的灰阶把这些阴影转化成像素深度，最后用3D打印机输出一个三维物体。近年来，迪士尼研发

部门正在努力完成一个令人赞叹不已的研究项目：通过模拟，让用户在触摸屏幕上显示的物体时，能产生一种正在触摸实物的感觉。这种效果要归功于一个触摸反馈算法，该算法可以通过再现指尖触摸实物的感觉来“欺骗”大脑。迪士尼研究院的负责人伊万·波普列夫说：“当手指在触屏上滑动时，如果我们人为地制造手指皮肤绷紧的感觉，大脑就会上当，以为触屏上真的有一个突起，尽管触屏表面非常光滑。”他认为这项技术可以增强人们与虚拟物体互动的真实感。

此外，我们还可以通过某些全新的方式把不同媒介组合起来，在人们已习惯于网页、幻灯片演示文稿以及移动多媒体信息的今天，这种做法越来越常见。随着可编程计算机于20世纪问世，艺术在摆脱媒介限制的道路上迈出了尝试性的步伐，这并非一种巧合。计算机已经变成一个多用途平台，可以

完成包括图示、作曲在内的各种任务。50年来，现代的艺术家也一直在探寻有效的方法，把之前相对独立的内容（诸如音乐创作、视觉资料制作）整合起来。

一种叫作“联觉”（synaesthesia）的神经学疾病，一直让科学家与艺术家非常感兴趣。患有这种病症的人在看到单词时会产生色感，在听到声音时会产生纹理感，或者把气味理解成形状。现在，计算机也表现出类似的“症状”，“信息美学”

（info-aesthetics）的兴起就是一个明证。信息美学不仅是一种计算方法，它还以数字、文本、网络、声音与视频为原材料，把它们转化成图片，以揭示数据中隐藏的各种规律与关系。

艺术家进行过多次数据可视化操作，包括用波的形式表现巴赫的乐曲、计算机下棋时的思维过程以及股票市场的起伏不定等。

2013年，比尔·盖茨与微软首席技术官内森·梅尔沃德申请了一项专利，它可以选取文本块信息并生成静止的图片和动态的图像。他们指出，这项技术可以用于课堂教学。在学生遇到诵读困难、注意力不集中或者其他可能导致无法阅读长篇课文的问题时，这项技术可以帮助他们。

1. “信息美学”一词是20世纪50年代德国心理学家马克斯·本泽首次提出的。本泽对艺术、技术与科学有浓厚的兴趣。

隐藏在艺术作品中的“真相”

几年前，身为英语专业研究生的斯蒂芬·拉姆齐对图论产生了兴趣。图论根据物体之间的数学关系建模，各个物体在模型中被表示为“节点”，节点之间的连线被称作“边”。在选择结构上有数学特征的文学作品时，拉姆齐最后决定用威廉·莎士比亚的戏剧。他说：“莎士比亚的戏剧总是从第一个场景切换到第二个场景，然后回到第一个场景，接着又切换到第三、第四个场景，再回到第二个场景，以此类推。”拉姆齐兴致勃勃地开始编写计算机程序，用其把所有的莎士比亚戏剧转换成图形。接着，他利用数据挖掘算法分析这些图形，验证他是否可以（依据这些戏剧的数学结构）预测它们是喜剧、悲剧、历史剧或爱情剧。他说：“结果，我的预测取得了成功。计算机程序知道《冬天的故事》是一

部浪漫的爱情剧，也知道《哈姆雷特》是一部悲剧，还知道《仲夏夜之梦》是一部喜剧。”用拉姆齐的话说，算法只“弄砸”了两次，把《奥赛罗》与《罗密欧与朱丽叶》归到了喜剧类别。他说：“但这才是最令人吃惊的地方。早在几年之前，文学评论家就注意到这两部戏剧在结构上具有喜剧的特征。当我看到计算机给出的结论时，我大吃一惊，差点儿从椅子上摔下来。”

在拉姆齐之前，就已经有人认为我们也许可以利用算法找出隐含在某些艺术作品中的“真相”。20世纪40年代后期，一位名叫罗伯托·布萨的意大利耶稣会牧师，用计算机把著名神学家托马斯·阿奎那的作品转换成“编码”。布萨解释说：“在阅读这些文字时，读者不仅需要思考这些文字在自己头脑中的意义，还需要认真思考这些文字在作者头脑中有哪些意义。”

不过，尽管这个孤立的个例早已存在，但20世纪前半叶的大多数科学家仍认为计算机在研究艺术这种无法量化的事物方面根本没有用武之地。例如，如果两名画家的风格相近，算法就绝不可能区分他们的作品。在真正的专家都感到棘手的时候，算法就更加无计可施了。已故英国科学家鲍登在他的经典著作《比思维都快：数字式计算机座谈会》中指出：

机器似乎永远不可能回答“这幅画有可能是谁的作品，弗美尔还是凡·米格伦”等宽泛的问题。我们应该记得，艺术评论家们花了几年时间才有信心回答这个问题（结果还是答错了）。

在鲍登看来，这个证据简单明了而且不容置疑。阿兰·图灵认为智能计算机的标准检查程序是复制人类的聪慧行为，如果让计算机解决连人类都无法做出明智判断的问题，成

功的希望能有多大呢？这无异于白日做梦。

不过，近些年来，这种观点遭到了人们的质疑。利奥尔·沙米尔是一名计算机科学家，毕业后在美国国家卫生研究院供职。在此期间，他利用全自动显微镜同时分析成千上万个细胞的结构。随后，他转向天文学研究，设计了几种可以浏览几十亿个星系图片的算法。接着，他开始致力于解决迄今为止他遇到的最大难题：利用迅速扩大的知识库以及在艺术构成方面人们已进行过深入研究的一系列观点，设计出世界上第一个全自动化的艺术评论算法。沙米尔的算法可以根据4024种不同的图像内容描述符来分析每一幅画作，不但包括人类艺术评论家通常考虑的所有内容（如画家使用颜色的特点、几何图形的分布特点等），还包括他们可能不会考虑的所有内容（如利用泽尼克多项式、哈拉利克纹理特征、切比雪夫统计特性等描述一幅

画作)。拉姆齐解释说：“算法可以根据某个艺术家作品的典型数值发现其中的规律。”人们已经证实他的算法善于辨识赝品，特别是在鉴别杰克逊·波洛克滴画作品的真伪上，正确率达到了惊人的93%。

沙米尔的自动化艺术评论算法与史蒂芬·拉姆齐的莎士比亚戏剧数据挖掘算法一样，也能给我们带来新的灵感，可以帮助我们了解某些艺术家作品之间的关系。沙米尔说：“只要我们可以用数字来描述某个艺术家的作品，我们就可以用视觉化的方式展现他与其他艺术家作品的不同之处。”沙米尔在分析波洛克与凡·高（这两位艺术家分属截然不同的两个艺术流派）的作品时，发现在算法给出的20个信息量最大的描述符中，有19个描述符表明两者之间存在显著的相似性，例如两者都对低级的纹理与形状有偏好，在线与边的使用方面也存在相似性。如果不是最近几

位著名艺术评论家提出相似的观点，这个发现看起来同样毫无意义。

众所周知，文学研究正处于困境之中。由于招生规模几十年来不断缩小，这个学科在高等教育中的地位也日趋下降。如何逆转这一趋势呢？有人提出了一个非常简单明了的解决方法：把文学研究转变成“数字人文学科”。2008年，文学评论家乔纳森·戈特沙尔在《波士顿环球报》上发表了一篇题为“针锋相对”的社论，称文学评论的现状是“缺乏活力，没有目标，与‘外部世界’所关注的内容渐行渐远”。戈特沙尔说，我们需要的不是“美丽的神话”或者罗兰·巴特所谓的“作者已死”这类含糊不清的表达，而是一种生产主义的方法，围绕相关性、规律探索及客观性开展文学研究工作。

文学研究应该向自然科学学习。文学教师应该应用科学的研究方法、理论和统计工

具，像科学一样使用假设与证明的模式。他们应该摒弃对知识前景彻底绝望的态度，欣然接受科学在学术方面的乐观主义精神。如果可以做到这一点，文学研究就会发生转变，并帮助我们逐渐深入地理解文学与人类的经验。我的这个提议可能会让很多同行感到沮丧，因为他们担心引入科学研究方法可能会导致文学研究变成自然科学的一个分支。但是，明智的做法是承认自然科学在很多方面强于我们，在保持文学研究特色的同时，从自然科学那里学习成功的经验。

据此，戈特沙尔提出了他的罗伯托·布萨式的信念，认为真正的、可以证明的真理存在于文学之中，是值得我们为之努力的。在戈特沙尔的心目中，与自然科学中有很多定律等待人们发现一样，对于“我是否可以这样解读（这幅画、这本书或者这部电影）”这类问题，也存在明显正确或者错误的答案。

尽管这些见解很有可能会让人文学科的很多学者感到震惊，但是戈特沙尔认为，艺术批评可以借鉴计算机科学方法的建议并非完全没有道理。“认识论”（epistemology）是心理学的一个分支，它研究知识的本质，关注知识的获取方式与量化方式，以及某个对象可被了解的程度。在算法时代，人类有可能掌握的知识范围发生了显著的变化。例如，算法可以统计出一个含有5万个单词的文本块中所使用的不同单词的数量，从而估算出该文学作品的“词汇丰富程度”。这个应用有可能得出若干惊人的发现。评论家大多没有想到，辛克莱·刘易斯（有时会被批评没有自己的风格，并因此遭到嘲笑）等畅销书作家在作品中使用的词汇量，竟然是以作品难以理解著称的诺贝尔文学奖得主威廉·福克纳的两倍。

算法在文本分析方面最引人注目的一次应

用发生于2013年。当时，作者署名为“罗伯特·加尔布雷斯”的小说《布谷鸟的呼唤》在世界各地的书店上架销售。一开始，这本侦探小说仅售出了1500册纸质版本，并没有引起人们的关注。后来，一家英国报纸报道说，这本小说的作者可能使用了化名，其真正的身份是《哈利·波特》的作者J·K·罗琳，这本书一下子就成了人们关注的焦点。为了证明这个传言的真伪，人们安排计算机科学家来验证作者的真实身份。算法利用数据挖掘技术分析文本中的4个变量（单词平均长度、常见单词的使用情况、成对单词的发生频率，以及“四元字符串”的分布情况），最终认定这本书的作者极有可能是罗琳。后来，罗琳本人亲自证实了这个说法。

斯蒂芬·拉姆齐指出：“严格的计算只关注研究对象的本质特征，因此其得出的结论有可能因为与我们对某个主题既有的理解截然

不同，而让我们大吃一惊。我们在阅读时总是带着各种偏见，但算法不会受到任何偏见的影响。因此，它可以给我们带来意外发现，让我们惊呼：‘啊，我以前怎么就不知道呢！’”

《纸牌屋》成功背后的最大功臣

出版商关注的内容不同于文学教授，只要稍稍浏览畅销书单，我们就能知道这是事实。但是，这个事实并不意味着算法分析对于出版商而言没有吸引力。当然，相较于文本的内容而言，出版商更关注的还是他们的读者。以前，读者从书店买了一本书之后，出版商无法通过量化的方式，去了解这名读者会通读这本书还是将它束之高阁。录像带与DVD发行商也面临同样的问题。人们在看《星球大战》时会不会多次按下后退键观看冲锋队员碰撞脑袋的镜头，或者在看到《本能》中那个臭名昭著的双腿交叉镜头时会不会按下暂停键，电影公司的高管们并不知道。但是，由于内容发布商现在可以收集大量的反馈数据，使得这一切都在发生变化。

例如，亚马逊可以知道读者阅读电子书的速度，他们在阅读前言部分时是细细品读还是略过不看，甚至知道他们在阅读时标记了哪些内容。他们知道人们阅读科幻小说、爱情小说和侦探小说的速度比严肃文学快，通读非虚构类作品的可能性低于虚构类作品。

这些发现有助于他们做出具有独创性的决定。2013年2月，奈飞推出了由凯文·史派西担纲主演的政治系列剧《纸牌屋》。从表面上看，《纸牌屋》最值得注意的地方是，奈飞的分销与生产模式正在发生变化，目的是与Showtime电视网、HBO电视网等优秀电视节目网站展开竞争。向互联网用户提供原创性影像内容是一个新颖的做法，高预算的产品更是不多见，而《纸牌屋》的预算高达1亿美元。不过，令很多人感到吃惊的还是奈飞敢于做出这些决定的气魄。这家位于洛杉矶托斯的公司授权制作两季共26集的《纸

牌屋》之前，公司主管们甚至没有审查过一个镜头。这是为什么呢？原来，奈飞公司利用算法分析了从2500万用户那里收集的数据，发现人们喜欢观看的节目类型具有某种倾向性和相关性。他们发现，很多订阅用户会反复观看BBC《纸牌屋》系列剧的某些剧集或连续观看多集，这证明他们喜欢观看这部系列剧。这些用户还喜欢观看凯文·史派西主演的电影以及《社交网络》的导演戴维·芬奇执导的电影。奈飞认为，符合上述三个特征的系列剧很有可能获得成功。事实证明，他们的判断是正确的。

他们的付出赢得了回报。《今日美国》发表了一篇题为“《纸牌屋》得了全A”的评论，称赞这个节目是“回报率极高的投资”，是当年人们观看过的“最好的电视节目”之一。奥巴马总统也承认自己是《纸牌屋》的粉丝。奈飞在《纸牌屋》之后乘胜追击，又推出了

三部热门系列剧：《铁杉树丛》《发展受阻》和《女子监狱》。因为他们的努力，该公司在2013年的艾美奖上获得14个奖项提名。美国电视评论家、专栏作家戴维·比安库利在《纽约时报》上撰文指出：“HBO花了25年时间才拿到第一项艾美奖提名，而奈飞只用了6个月。”

在线零售商亚马逊从中受到启发，开始借鉴奈飞的成功经验。亚马逊也拥有客户的大量信息，以及传统电影公司老板梦寐以求的“喜欢”“不喜欢”等详细数据。亚马逊的创始人杰夫·贝佐斯告诉《连线》杂志：“这是一种全新的电影制作方式。有人认为这个方法行不通，我们的观点恰好相反。”

如果每部电影都有4种结局

亚力克西·柯克曾经是华尔街的一名量化分析师。量化分析师被人们称作“穿梭于金融市场的行家”，他们所从事的工作非常繁杂，需要用到数学知识、巨额融资及计算机技能。柯克拥有计算机科学的博士学位，本来可以过上富足的生活，因为量化分析师是一个热门职位，年收入可达到25万美元。但柯克感到非常不自在，他说，“做了一年量化分析师之后，我觉得这种生活不适合我。”他希望从事的工作是艺术。于是，柯克离开美国，回到英国普利茅斯，开始学习音乐专业。现在，柯克在普利茅斯大学计算机音乐跨学科研究中心担任研究员。

2013年，柯克创作的电影《多重世界》在2013年半岛现代音乐艺术节上举行了首映式，为他带来了前所未有的成功。这部电影

可以根据观众的反应改变叙事方向，把观众由被动的消费者变为主动的参与者，是对传统电影的一个重大颠覆。在电影放映过程中，观众戴的特殊传感器可以监控他们的脑电波、心率、出汗量以及肌肉的紧张程度。这些生理唤醒的指标被输入计算机，通过求取平均值并进行分析后得出反应数据，去触发不同的电影镜头。可以想象，如果观众过于平静，接下来的一组镜头就会比较激烈，让观众紧张起来；如果观众较为紧张，接下来的镜头就会让他们恢复平静。这种分支式叙事方式共有4种结局，观众最终会看到其中一种。

人们一度认为，必然存在可以吸引所有人的艺术作品。从一定意义上讲，我在本章开头提及的Epagogix之类的公司使我们对这个旧观点有了新的认识。对同一部电影，人们的评价有好有坏——有人赞不绝口，有人则

频频摇头，这个事实说明人们的这个旧观点未必正确。我们的偏好取决于多种观念的综合效应，包括我们所继承的观点以及我们自身的经验。我对如何把《麦克白》搬上舞台有自己的看法，是因为我过去看过不同人表演的《麦克白》，或者阅读过人们对这部戏剧的相关评论。我喜欢某部电影、某个音乐作品或者某本书，也是出于相同的原因。

柯克说：“一部固定不变的电影只能吸引偏好相同的少量观众。电影公司在拍摄电影时，会考虑观众情绪体验的平均特点，而且必须兼顾各个国家的观众。但是，这种做法通常会导致某种妥协。”柯克的观点是正确的。《钢铁侠》（这个系列的电影在不牺牲质量的前提下吸引了大量观众）的每一部都借鉴了十几部电影的成功之处，并将它们所有的创造性粉饰一新，就是为了满足尽可能多的观众。《多重世界》为我们开辟了一条

不同的路。在数字时代，合理化再也没有一个统一的标准。我们可以使用公式，但这并不表示所有事物都必须一模一样。⑨

当然，如果想大面积推广这种做法，我们还必须考虑成本问题。亚力克西·柯克称《多重世界》是一部“超小预算”的电影。拍摄时，从普利茅斯大学传媒系租借了灯光、三脚架和一台高清摄影机，不足2500英镑的投资全部用于支付这几天的租赁费。但是，对于一部好莱坞电影而言，这样的投资显然远远不够。热门大片的成本有可能高达2亿美元，电影公司真的可以像柯克那样，再投资一笔钱拍摄出4种结局吗？柯克认为这种做法是可行的。我在本章前面说过，娱乐业现在的经营模式非常低效（有人称之为不科学），依赖“超级明星式”的成功来抵消可能性更大的失败所带来的损失。电影公司拍摄10部电影，只要其中有两部获得成功，他们

就会相当满意。但是，如果这个公司为每部电影增加50%的投资，用于拍摄多种备用镜头并灌制配套的电影原声带，即使这个做法会增加成功的可能性，结果又会怎样呢？如果电影公司把所有电影都拍摄成针对所有观众的分支式电影，那么每拍摄5部电影，他们必须保证有两部能取得相当程度的成功，才会有利可图。

在柯克完成了《多重世界》这部处女作之后，几家大型传媒公司找到了他，邀请他加入并担任顾问。BBC两次邀请他到该公司位于曼彻斯特的总部放映这部电影，谈谈他对交互式媒体的展望。他们迫切希望了解这项技术是否可以应用于下一代电视机，是否可以产生有益的效果。柯克说：“已经有人在进行这方面的尝试了。”2013年，微软就一款摄像头提交的专利申请获得了通过。这款摄像头可以监控观众的肢体动作、视线轨迹和

心率水平等，并制作一份报告，通过云技术发送至远程设备，由该设备判断某些目标是否已经实现。比如，如果观众在广告时间没有离开，广告商就可以为这些观众发数字奖励（如iTunes代金券）或者实物奖励。因为微软的摄像头传感器可以识别手势，因此广告商可以设计某种舞蹈或者动作，让观众在家中重复这些动作。观众参与的热情度越高，就能赢得越多的iTunes代金券。

另一家名叫Affectiva的公司正准备把一款面部表情分析软件推向市场，卖给消费品制造商、零售商、市场营销人员和电影公司。这款软件的作用是分析消费者的情绪反应，帮助人们改进产品的设计与市场营销活动。同样，电影公司也将通过麦克风、视线传感器以及内置于移动设备的社交网络扫描器等（这些装置可以根据用户的反应调整他们的观看内容），逐步加强对观众的观察。如果

发现观众的视线频繁地离开屏幕，或者他们对脸谱网的兴趣大于他们正在观看的这部电影，电影公司就会调整剪辑原声带甚至故事情节，以确保观众从头至尾都不会走神。

1. 2006年，时任《连线》杂志主编的克里斯·安德森在《长尾理论》一书中提出的“98%法则”为这个观点提供了佐证。安德森在书中指出，技术把大众市场转化成了无数“利基市场”。他认为数字分销使利基产品唾手可得，无数利基产品汇集到一起，同样是一个重要的市场。亚马逊的经营模式就是一个明证。

可以一直演下去的电视剧

从传统意义上讲，一旦绘画作品完成了，照片被打印出来了，或者一本书出版之后，这个作品就再也无法改变了。我们甚至可以认为这个特点是我们鉴赏活动的组成部分。一本印刷装订好的书，它的页码是不会变的，读者根据这种结构特点，以线性方式（逐行、逐页、逐章）阅读这本书。因此，一本书的完整性可以激发我们的阅读兴趣。

然而，在算法时代，这种永恒性或者固定性已不复存在。我们在看电子书、电影或者听音乐时可以随心所欲地跳过某些内容。^①这个特点导致故事情节变得索然无味，或者被分割成最细小的单元。1995年，计算机科学家史蒂文·德罗斯在《结构化信息：浏览、访问与控制》一文中指出，对结构化信息进行这种方式的分析，不会把我们

引向某种普遍真理；我们不会像“福尔摩斯那样通过案情分析，找出其中隐藏的秘密……而是把信息分成一个个部分，我们看到的也只是一个个支离破碎的部分”。

几年前，美国艺术家贾森·萨拉文对电影《泰坦尼克号》的分析就是一个生动的例子，可以从中清楚地看出这种叙事展开方式的效果。萨拉文把这部电影分解成一个个独立的画面，利用算法计算出每幅画面的平均色度，再把它变成这种颜色，然后按照电影的叙事顺序，把这些画面重新排列到一起构成统一的影像。只要从左至右、从上至下研究这些颜色的排列，电影的节奏就会一目了然。

亚力克西·柯克的《多重世界》与萨拉文对《泰坦尼克号》的重构代表了同一理念的两个方面。在后“9·11”时代，我们越发地觉得世事无常，过去与现在被压缩成了脸谱网

的时间轴，而未来充满了不确定性。在这种情况下，传统意义上的开头、中间与结尾部分还能首尾衔接、彼此相连吗？看一看周围，由于采用了编码与实时数据流技术，状态不断发生变化的艺术作品层出不穷，这可以帮助我们更深刻地理解这个问题。互联网的发展将永无止境（虽然技术本身可能有最终成型的一天），因为这些算法形式的艺术作品可以摇身一变，作为新的数据被注入互联网的浩瀚海洋之中。

马萨诸塞州剑桥市的费尔南达·维埃加斯与马丁·瓦滕伯格用编码在线绘制的美国风图就是这样一个例子。维埃加斯与瓦滕伯格是谷歌“Big Picture”数据可视化小组的成员，他们从美国国家数字天气预报数据库中获取天气数据，并把这些数据转化成一幅不断变化的旋涡式图像，非常迷人。瓦滕伯格说：“在风平浪静的天气里，看着这幅图，想想身

边的环境，我们就会感到一丝慰藉。在飓风即将到来的时候，这幅图就会变成凶兆，令人胆战心惊。”之前，一成不变的艺术作品是一种永恒的存在，所表现的主题足以让几代人欣赏。现在，“永恒性”意味着每一代人看到的是不同的作品，这使艺术家陷入了两难境地，要么永不停歇地继续创作，要么放弃这件艺术作品。

电脑朋克媒介理论家道格拉斯·洛西科夫在他的著作《当下的冲击》^①中表达了类似的观点，探讨了当代通俗文化在算法时代表现出来的特点。20世纪初的艺术家在选择艺术创作的方向时，把重型工业机械作为模型来展现他们的技艺与美学理论，与之相似，现代娱乐业则要表现编码技术所能带来的各种可能性。我们可以通过算法分析叙事结构来预测《教父》等经典影片的情节发展节奏，但是现在受到热捧的都是像《权力的游

戏》这种情节不断发展的剧目，它们抛弃了平铺直叙的三幕剧结构，可以无限地延续下去。

洛西科夫通过对美国全国广播公司（NBC）系列剧《废柴联盟》、塞思·麦克法兰的《恶搞之家》等节目的研究，进一步指出技术对叙事结构的破坏作用。《废柴联盟》讲的是格林代尔社区学院的一群怪人，他们不断提及自己其实是情景喜剧中的角色的故事。由于观众已经非常熟悉传统情景喜剧的套路，因此《废柴联盟》在故事情节的设计上充分考虑了这个问题，力求有所不同。《恶搞之家》也同样背弃了传统的叙事方法，在“反打镜头”（cutaway）等噱头的运用上自成体系。事实证明，无论以哪种次序运用这种噱头，都同样滑稽可笑，因此在习惯用苹果iPod Shuffle播放器看电影的一代人眼中，《恶搞之家》就是一部堪称完美的喜

剧片。跟之前播出的《辛普森一家》一样，《恶搞之家》中也几乎没有持续时间较长的情节，因此各种无厘头可以从容亮相，吸引观众一口气看完每一集。

这种艺术效果在更严肃的电视系列剧中表现得更加淋漓尽致。《火线》《广告狂人》《黑道家族》和《嗜血法医》等电视片的关键角色，从表面上看各不相同（分别是巴尔的摩的警察、麦迪逊大街上的广告商、新泽西的歹徒和迈阿密的连环杀手），但是他们也有相似之处，其中最主要的一点是他们都无法改变自己的本性，也无法改变他们所生活的世界。洛西科夫认为：

这些系列剧并不是依靠线性情节取得成功的，而是凭借环环相扣的情节，让观众可以在各种媒介形式之间建立联系，通过联想形成反差……开头、中间与结尾几乎失去了所有意义。每当建立新的联系，以及一个虚假的

故事情节被曝光或者被重新构建时，人们都能领会到电视剧想要表现的要旨。简言之，这类电视剧无时无刻不在引导人们从中发现规律。

现在，最受欢迎的电影不再是单一性的艺术作品，而是能形成更大特许经营权的一个个节点，电影公司往往在第一部电影上映之前就会告诉观众还会投拍摄续集。在这种情况下，很多受欢迎的热门电影自然会借鉴连环漫画的一些特点。与小说不同，在这种媒介中，故事情节会不断向前推进，让观众无法预知最终结局。

正因为具有这种特点，亚力克西·柯克的《多重世界》其实并不是游离于传统之外的试验性作品，而是传统叙事方式的一个新尝试，预示着未来的发展方向。柯克指出，由于无法预知故事情节的发展方式，将来观众可能不会关注叙事框架，而去关注情感的表

现方式。

1. 人们曾经做过多次尝试，希望利用多种独创性的密码算法帮助电子书保留线性阅读的特点。1992年，电脑朋克作家威廉·吉布森创作了“电子小说”《阿格里帕：死亡之书》。这本书被存储在一张3.5英寸的软盘里并邮寄给读者，打开之后，书的文字内容只能显示一次，并随着读者向下滑动电脑屏幕而消失。

2. 《当下的冲击》由中信出版社于2013年出版。——编者注

数字看门人巨大的隐性权力

艺术与人文学科的固定性的缺失，还可能导致其他后果，甚至是某些恶果。当前，在Kindle、谷歌图书等在线平台阅读某书时，该书的“母本”是在线储存并通过云技术接入的，即使我们已经把这本书买回家了，出版商与作者仍然可以做出某些修改。2009年亚马逊的一次经历就是一个活生生的例子。当时，亚马逊以为乔治·奥威尔的《1984》已经不受版权限制，便开始在Kindle平台上销售这本书的电子版本。但是后来他们发现，还没有过版权保护期，他们因此大惊失色，决定删除这本书。结果，已经掏钱购买了这本电子书的用户发现它从他们的图书馆中消失了。当然，具有讽刺意味的是，在《1984》描述的反乌托邦的社会中，强权统治者们每天都会修改历史书，以此来愚弄和操纵人

民。在小说发表60多年后的今天，篡改历史在技术上已经成为一种可能。

2013年7月，哈佛大学计算机科学教授乔纳森·奇特林在《连线》杂志上撰文，把这种现象称为“令人担忧的趋势”，并呼吁数字书籍以及其他文本不应由分销商与数字看门人控制，而应该置于读者与图书馆的掌控之下（他认为文本的不可修改性是读者与图书馆的既得利益）。奇特林指出，最糟糕的潜在后果是这些修改不会留下任何蛛丝马迹。他在文中说：“如果我们要修改甚至销毁历史，（至少）我们应该看到焚烧书籍的火光、听到声音和看到浓烟吧。”

艺术品位是否有统一的衡量标准？

20世纪80年代初，戴维·史密斯决定研发一种音乐合成器。这名毕业于加州大学伯克利分校计算机科学与电子工程专业的研究生，希望建立一个标准化协议，用于解决世界各地制造商生产的不同电子乐器之间的通信问题。最终，他的研究成果被命名为“乐器数字接口”（使用更广的名字叫作“MIDI”）。作为一种音乐媒介，MIDI远谈不上完美。尽管MIDI可以借助一个键盘模仿很多种声音，但是，键盘声音不连续的块状特点，意味着它无法模仿天才歌手或者萨克斯管发出的优美声音。虚拟现实之父（同时也是天才音乐家）杰伦·拉尼尔指出：

在MIDI问世之前，音符超越了其定义，传递着各种各样的意义……MIDI出现之后，音符

不仅可以传递丰富的意义，同时在结构上具有严格的强制性，使我们生活的多个方面不可避免地实现了数字化。

技术“锁定”内容是量化不可缺少的一个特点。我们在创建单一标准的同时，还设置了种种限制。在戴维·史密斯发明MIDI的200多年前，18世纪的苏格兰哲学家大卫·休谟写了一篇题为“论品味的标准”的论文。休谟在文中指出，艺术的关键内容（如果用方程式表示，就是等号后面的内容）是他所谓的“一致性”的特点。休谟认为：“我们自然希望为品位建立一种标准，并借助这一标准协调人的各种情感。”

休谟认为美学价值无法用客观的数字来表示，他并不认同本章开头讨论的那些数字，而认为应该用主观判断来衡量美学的价值。他说过：“探求真正的美或者真正的丑，与探求真正的甜味或者真正的苦味一样，都将一

无所获。”同时，休谟也承认，有时我们的确可以通过主观判断衡量“愉悦感”或是“不舒服的感觉”，因此，他提出了“品位标准说”。

休谟在很多方面都领先于他所处的时代。近些年来，世界上有多家机构开始研究所谓的“情感优化”（Emotional Optimisation）问题。人们发现，大脑的某些部位与不同情感之间存在某种对应关系，情感优化与这一发现有关。神经科学家让实验对象戴上脑电图头套，测量脑神经中离子电流导致的脑电变化。通过这些读数，我们就可以发现一段音乐或者一个电影镜头让人们产生的是积极反应还是消极反应；再借助某些机器学习工具，则完全有可能发现艺术中的哪些基本特点可以引发人们的哪些情感反应。

展望未来，这项研究显然有非常美好的前景。例如，在建立反馈回路之后，我们不仅可以读取用户对某些艺术作品的情感反应，

还可以获悉他们希望产生的某种情感。届时，我们将无须根据自己的情感使用播放列表，而是将我们希望产生的情感输入电脑，电脑就会自动生成一个可以激发这种情感的个性化播放列表。因此，这项成果可能有某种治疗效果，帮助那些压力过大或者心情低落的人舒缓情绪。此外，人们可以在跑步时戴上耳机测量自己的脉搏，并根据心率的变化选择合适的音乐。该成果还可以应用于文学领域，通过监控脑电变化，利用算法根据读者的反应改写电子小说的某些章节。单人喜剧演员或者现场演奏的音乐家，可以根据特定受众的反应对表演稍加改变，同样，媒体与消费者的契合程度也会越来越高。媒体可能没有任何变化，但其播放的内容会根据人的品位特点发生变化。

记者亚历山德拉·库格兰在《新政治家》杂志上发表了一篇文章，把情感优化称作“听

觉毒品”，他指出这意味着我们总能发现“让我们兴奋不已的音乐，或者让我们情绪低落的音乐”。我在本章开头介绍过，有人认为让艺术在某个方面变得对我们“有所裨益”，是一个令人向往的目标，库格兰关于音乐功能的评论与这个观点如出一辙。所谓的“听觉毒品”将导致一系列疑问，例如，艺术是不是一种别具一格的致幻剂？

举个例子。我们在看马克·罗斯科的画作《无题（蓝色中的绿色）》时可能心情平静，但我们会因此把这幅艺术作品视为一种镇静剂吗？白俄罗斯科技学者叶夫根尼·莫佐罗夫在《点此拯救一切》一书中对这种功利主义思想进行了批判。莫佐罗夫说，假设谷歌（代表那些明确地以量化分析一切事物为己任的公司）得知我们刚刚与前女友进行了一次伤心的通话，此时心情不是十分愉快。如果艺术等同于愉悦（并且得到愉快心情的

最快捷的途径是欣赏一幅伟大的画作），谷歌就会知道，让我们尽快振作起来的最好办法是欣赏印象派画家雷诺阿的作品。

其实，谷歌也不是十分清楚，它只知道你需要124个单位的“艺术”。根据谷歌的测量系统，雷诺阿的画平均含有120个单位的“艺术”。因此，只要看一眼他的画，你的情绪就会“啪”的一声好起来。

莫佐罗夫继续沿着他的这个研究方向，指出这个主张可能会导致几个相关问题：通过观看雷诺阿的画来平复情绪，会不会把我们带进艺术爱好者的世界呢？有助于拓展我们的视野吗？这种把艺术看成自助式文学作品的尝试，是否会损害艺术的地位呢？把愉悦感当作艺术的单一衡量标准并进行量化分析的努力，例如谢尔盖·爱森斯坦提出的“吸引力”，会让很多莫佐罗夫没有提及的问题浮出水面。例如，如果我们认为雷诺阿会让我们

的幸福度飙升122个单位，而巴勃罗·毕加索的贡献仅为98个单位，我们为什么还要欣赏毕加索的作品呢？

同样，如果我们暂时假设贝多芬华丽的《第七交响曲》所产生的神经兴奋度远高于贾斯汀·比伯的热门歌曲《宝贝》，我们就可以利用数学的方式，把艺术分为“高效”与“低效”两个类别。如果这种区分是正确的，那么我们观看多集《老友记》（假定这部情景喜剧属于“低效”艺术类别）所吸收的艺术营养，是否与阅读列夫·托尔斯泰的《战争与和平》一样多呢（尽管前者从时间的角度看效率要低一些）？如果我们认为《战争与和平》优于《老友记》、贝多芬胜过贾斯汀·比伯，理由仅仅是前者满足我们艺术需求的速度更快，那么我们也可以认为詹姆斯·帕特森是一位比约翰·库切更伟大的小说家，因为我们一下午就可以看完帕特森的《杀死亚历克

斯·克罗斯》，而库切的《迈克尔·K的生活和时代》则要花几天甚至几个星期的时间才能看完。这两个观点从本质上看都一样，在数学上也都站得住脚。但是，在某些方面它们并不合乎情理。

计算机也有可能成为艺术家

所有这些都在不断地把我们推向一个无法回避的问题：算法是否可以自动生成艺术作品？不出意料的是，对这个问题持相反观点的双方都纷纷发表了自己的见解。利奥尔·沙米尔说：“这只是何时会发生的问题，而不是会不会发生的问题。”沙米尔认为，从长远看，他提出的自动化艺术批评，与可以发现电影剧本中的不妥之处，并提出针对性修改意见的Epagogix电影票房预测系统一样，也能够找出伟大的艺术作品的共同特点，然后依据这些发现生成新的艺术作品。

尽管这个想法似乎非常新颖，但实际上早已有人提出了类似的观点。1787年，莫扎特匿名发表了《音乐的骰子游戏》。他的想法很简单：帮助“没有任何音乐知识的读者……通过多次掷两粒骰子”，谱写出德国人的华尔

兹。莫扎特给出了176个小节的音乐，并将这些音乐分成16组，每组11小节。读者先掷两粒骰子，根据点数从第一组中选取对应的小节作为乐曲的第一小节，然后按照这个方法从第二组、第三组.....中继续选择。最终可能创作以来的乐曲数达到了惊人的 46×10^{15} 种，而且每首乐曲似乎都具有莫扎特的风格。

当代意大利建筑师、设计师切莱斯蒂诺·索杜通过所谓的“遗传算法”，用无数种形式表达同一个主题。他的作品虽然属于一个不同的媒介，却表达了类似的思想。遗传算法认为生物体是解决诸多问题的高手，通过在计算机内部复制进化过程，优化具体的解决方案。例如，在设计一把椅子或者巴洛克式教堂时，索杜把描述其特征的“规则”输入电脑，然后利用该算法将这把椅子或者这座教堂视为一个生物体，预测出经过几千年的自

然选择它可能具有的外貌特征。由于遗传算法可以生成无数结果，因此如果一家时髦的广告公司购买了索杜的这种“创意产品”，那么摆在办公室里的那几百把椅子就会各不相同，每把椅子在细节处都各有特点。如果一家公司在建造总部大楼时采用了索杜的办法，他们就会得到数以千计的设计方案，难以取舍。

不过，利用算法生成艺术作品的想法仍然面临诸多问题。20世纪40年代，西奥多·阿多诺和马克斯·霍克海默指出，公式化艺术并不能给我们带来新的体验，而是已有体验的重新组合。阿多诺认为，大众文化的乐趣不同于因为接触新事物而产生的愉悦心情，而是聪明机智的受众因为“做出成功的预测而得到的一种心理满足感”。现在，人们可以利用算法根据过去的成功案例预测未来作品的成败，证明了阿多诺的这个评论颇有先见之

明。这种艺术作品可能具有某种可以量化分析的完美性，但是，只有借鉴已有作品的成功经验，它才有可能是一件完美的作品。

例如，尼克·米尼承认，Epagogix不可能预见到《阿凡达》这类电影会如此成功。原因很简单，之前还没有哪部电影可以狂赚20亿美元，他们在预测时没有一个可以借鉴的对象。当然，这并不意味着Epagogix无法预测出《阿凡达》的成功。米尼说：“我们能不能准确地预测出它在美国的票房收入呢？可能不行。我们可能会认为这部电影的票房收入将超出我们的想象，但是，正因为这样，我们找不到任何可以借鉴的对象。如果预测下一部《阿凡达》呢？结果就不一样了，因为这一次我们可以找到借鉴的对象。”

与量化已有作品相比，生成新的艺术作品会带来更为迫切的问题。利奥尔·沙米尔的自动化艺术批评算法量化艺术作品的依据是

4024个不同的数字描述符，因此有可能量化分析通俗艺术中的成功要素，进而生成符合所有成功标准的艺术作品。但是，这些量化标准本身也是以人类的创造力为基础的。这些算法会不会提升艺术作品的价值，而不仅仅是模仿以前作品的风格呢？沙米尔回答说：“开始的时候不会。但是最终，我不知道还有什么事是电脑做不了的。”

如果这种情况真的发生了，我们是否愿意在了解情况的基础上，接受这种由机器自动生成而不是由人类艺术家创作而成的作品呢？也许这是我们更应该考虑的问题。对于那些认为创造力从深层次看是一种人类活动（这是一个比较新的观点）的人而言，这个问题超出了技术能力的范畴，已经非常接近人性的本质了。

2012年，伦敦交响乐团把一些完全由Iamus算法编写的乐曲搬上了舞台。Iamus

是由教授、企业家弗朗西斯科·比科编写的一种音乐算法，已经谱写了10亿多首各种艺术流派的歌曲。在这次算法音乐演奏会之后，《哥伦比亚观察家》的特约撰稿人戴维·埃克亲自动笔（当然也有可能是利用电脑）写了一篇文章，猛烈地抨击了这项新技术。埃克说：“虽然我用电脑完成了各种各样的任务，（但是）我认为电脑在某些方面无能为力。”

我绝对不是一个纯粹主义者。我不愿意把音乐分成泾渭分明的不同种类，大多数关于“真”艺术与“假”艺术的争论让我反感，但这都不是重点。我要说的是有关人性本质的问题。计算机可以在《危险边缘》节目中与象棋大师对决并击败对手，也可以让天各一方的人相互联系，但是在处理情感问题方面，计算机缺少很多必备条件。我们在日常生活中都戴着一副假面具，自诩我们的所作所为都有某种重要意义（原因之一在于我们知道

自己总有一天将走到生命的尽头），而这种矫饰做作在伟大的音乐作品中表现得更加淋漓尽致。Iamus不会面临死亡的问题，也就不会产生那种紧迫感。它可以创造出声音（其中有的声音真的十分悦耳动听），但是它绝对不可能像音乐家或者作曲家那样表达复杂细腻的情感，也不可能完成独树一帜的创新。有人可能认为Iamus是一个完美的工具，可以创作“四十大金曲”这类没有意义的歌曲。但是在我看来，这个观点也是有问题的。即便是那些内容浅显、昙花一现的流行曲目，也来源于人类的各种经历；即便是这些曲目，我们也应该加以保护，使它们免受Iamus的茶毒。

持有类似观点的不乏其人。但是，也有不少人明确地提出艺术与技术之间明显存在着不可逾越的鸿沟。在《关于科技的问题》一书中，德国哲学家马丁·海德格尔介绍了一个

不成熟的想法，希望艺术可以把人类从技术带来的危险中解救出来。海德格尔认为，艺术可以揭示真理，而技术则会掩盖真理。

不过，利用算法生成艺术作品的梦想可能仍然有望得以实现。Iamus的设计者弗朗西斯科·比科指出：“一位女士告诉我，她承认Iamus是技术上的一个里程碑，但她接着说，她不敢听Iamus创作的音乐，因为这些音乐会让她产生幻觉。从某种意义上讲，这种现象令人恐惧。我完全理解，这是因为我们还没有做好心理准备。有些人仍然认为计算机是一心要消灭人类的‘终结者’，还有些人则把计算机看成是帮助我们处理信息的简单工具。他们从来都没有想过，计算机也可能成为艺术家。对他们来说，这是一个全新的概念。”

结语 预测未来

1954年，34岁的美国心理学教授保罗·弥尔出版了一部开创性著作《临床预测与统计学预测的比较：证据的理论分析与研究》。尽管这本书的标题有点儿过于笼统，但是，作者通过20个案例对统计学算法做出的预测与经过专业训练的专家做出的临床预测进行了比较分析。

在一个案例中，经过专业训练的辅导员被要求预测大学一年级学生的期末成绩。辅导员可以与每名学生面谈45分钟，而且可以查询学生们之前的学习成绩、各种能力测试得分以及长达4页纸的学生个人陈述。而算法依据的指标只是这些学生高中时期的成绩以及一次能力测试得分。事实证明，在总共14次预测中，算法赢了11次，做出的期末考试成绩预测比辅导员的更准确。

该书中介绍的其他案例aaa涉及多个不同主题，包括违反假释条例的比率（参见第三章中讨论的理查德·伯克）以及飞行员训练合格率等bbb，也得出了同样的结论。在总共20个案例中，有19个案例表明统计学算法的预测准确率显然比专家更高（后来，弥尔指出，这应该可以视为算法完胜人类专家），而且算法需要的数据更少。因此，弥尔得出结论：“有时，临床医生会沾沾自喜地断言，根据‘所了解的真实情况’做出的临床预测‘自然’更加准确，但是很明显，到目前为止，所有的事实根本无法证明他们的这种武断观点是正确的。”

可以想见，从那一刻起，弥尔就被学术界视为异端。他反对专家的这个立场无异于告诉人们“临床医生可以被拿着门罗手摇计算器的职员取代”（引自弥尔于2003年去世后其同事在《纽约时报》上刊登的讣告）。（在

职业生涯的后期，弥尔发表了题为“我为什么不参加病例讨论”的文章，对学术会议大加批评，认为学术会议无聊到令人无法忍受的地步。这篇文章进一步强化了他是“网络巨魔”原型的说法。）

尽管弥尔当时的身份极具争议性，但他认为算法具有很强的预测功能的观点却在之后得到了证实。在他之后的半个世纪里人们完成的200项左右的类似研究中，算法以约60%的胜率击败了人类的直觉。而在算法落败的研究中，统计学预测与临床预测之间的差别也并不具有统计显著性。由于算法预测的成本大多低于聘请专家的支出，因此这40%的研究也应该被视为算法预测“获胜”。

算法在我们的生活中变得越发重要

算法为什么能让我们如此感兴趣呢？本书的很多读者都从事计算机科学研究，如果这些读者回到1984年，甚至是1964年，很可能不会对类似的书籍感兴趣。但是，我们并不都是以计算机科学研究作为职业的。我在从事研究时经常考虑一个问题：一个曾经晦涩难懂的数学概念竟然成功占据了重要报纸及其他出版物的头版，为什么呢？

在本书的结语部分，我要和大家分享一下我对这个问题的一些看法。出现这样的现象，似乎是因为算法正在人们日常生活的方方面面发挥越来越重要的作用。谷歌等搜索工具总是把我们引向大型信息数据库，与此同时，亚马逊等大型公司所采用的推介系统通过分析我们的偏好特点，不断为我们推荐

新的文化体验。社交网站利用算法突出显示与我们“相关”的新闻，eHarmony等相亲网站则利用算法为我们牵线搭桥，构建二人世界。因此，认为算法是我们参与公共生活的一个有力武器，这样的观点并不是“网络（技术）万能论”。

接下来，他们再接再厉，在本书所讨论的4个领域里都取得了进一步的突破。例如，通过算法完成的交易在美国产权投资市场所占的份额高达70%。人们利用超级计算机，以闪电般的速度买进、卖出数以百万计的股票。算法交易以毫秒为单位来计时，通过铺设好的光纤，伦敦与纽约金融市场之间的通信仅需5毫秒（如果大家对这个时间没有直观的了解，那么我告诉大家一个数据：人眨眼所需的时间为300毫秒），为动辄数十亿美元的交易提供了保障。

由于医生在诊断病情时，经常被要求摒弃

自己的临床预测，转而依赖某些算法，因此，算法还促使医学领域发生了改变。马尔科姆·格拉德威尔在他的《眨眼之间》^①一书中讲述了一家医院利用算法诊治胸痛病人的故事，“医院告诉医生不要详细询问病人的病情”，而是把注意力集中在“血压、心电图等几项关键数据上。至于病人的年龄、体重、病史等，根本不需要加以考虑。结果怎么样呢？库克县医院现在是美国诊治胸痛效果最好的医院之一”。最近，算法在其他医学领域也取得了显著的效果。例如，医生只需通过电话听一听病人说话的声音，就可以利用算法诊断他是否患有帕金森症，还有一种规律识别算法甚至可以“嗅探”病人是否患有癌症。

1. 《眨眼之间》（第2版）由中信出版社于2014年出版。——编者注

算法能否处理一切事务？

算法是不是可以处理一切事务呢？要回答这个问题，可以通过两种方法。第一种是技术性方法。单纯从技术层面来考虑的话，这个问题目前的答案是否定的，算法并不能“解决”所有问题。例如，如何让算法识别物体的能力接近于人类至今仍然是一个大难题。即使是一个从未见过的物体，儿童也只需为数不多的几个“训练范例”就可以准确识别它。而算法在面类似的任务时，常常需要很长时间的练习，并且在发生识别错误时还需要人工予以纠正。即便如此，由于在真实环境中各种物体往往以不同的方式叠加，算法可能难以确定该物体的轮廓，因此在执行这项任务时仍然有可能面临很大的困难。

计算机科学家、教师约翰·麦考密克也给出了一个类似的例子。麦考密克认为，给学

生的作业评分是一件非常复杂的任务（而且有的学科主观性极强），不适合使用自动化程序，也不适合把算法用作教学辅助手段。既然计算机的功能越来越强大，那么未来人们是否可以利用算法完成上述两项任务呢？绝对可以。正是出于这个原因，与自动化程序对决是非常危险的。10年前，麻省理工学院与哈佛大学的两位受人尊敬的经济学家弗兰克·利维与理查德·默南，经过潜心研究出版了《新劳动分工》这部著作，在书中对人类与计算机的工作能力进行了比较。在比较乐观的第二章“人类仍然非常重要的原因”中，两位作者对各种信息处理任务进行了分析，认为这些任务都介于必须通过计算机（例如算法）完成的任务和只能通过人工方式完成的任务这两个极端中间。他们在书中举了一个长途货车司机的例子。

该货车司机正在不断处理周围环境的相关

信息（视觉、听觉与触觉信息）。要对这种行为进行编程，我们首先可以利用摄像头等传感装置捕捉各种感官反应。但是，如果在对面有车开过来时执行左转弯程序，我们需要考虑的因素就会非常多，因此很难找到可以取代司机行为的一系列规则。就目前而言，除非面对的是高度规范化的情况，否则清楚地表达（人类的）知识并将其植入算法仍然是一件难度极高的任务。用计算机取代人类，完成（货车驾驶等工作），并不是一件轻而易举的事。

我们至少可以确定，他们的用计算机取代人类完成驾驶任务绝非易事的论断是正确的。当然，在完成这部著作时，利维与默南持有这样的观点并没有错。就在这部著作出版的同一年，美国国防部高级研究计划局宣布举办“大挑战”竞赛，并设置了100万美元的奖项，以鼓励美国最杰出的人工智能实验室

派出选手参赛，创造可以穿行莫哈韦沙漠的无人驾驶汽车。整个赛程长142英里，而最后“胜出”的参赛队只完成了不足8英里（用了几个小时的时间），汽车就着火了，并在一阵儿震动之后停止了前进。

不过，10年时间里有可能发生很多变化。读完本书第3章，我们就知道谷歌的自动驾驶汽车已经取得了成功。既然事实证明这些技术的适用性很强，那么它们也完全有可能胜任目前由人类完成的出租车驾驶与长途货车驾驶等工作。

目前，很多领域与行业都在发生类似的变化。我们以亚马逊网站为例。在亚马逊刚起步时（当时亚马逊还仅仅是在线图书零售商，而不是近期出版的一本书里说的庞大的“万物仓库”^①），仅仅由两个部门组成。一个部门是编辑部，任务是评论图书、为网站主页撰写文字说明，并通过语音安慰那些心

存疑虑的顾客，让他们把信用卡明细提交给冷冰冰的机器。另一个团队被称作个性化小组，他们的任务是设计算法，为一个个用户推荐产品。就像算法时代的一个缩影，这两个部门之间也经常发生争执。

结果，第二个部门取得了胜利，在短期与长期“肉搏”中都击败了对手。他们的制胜武器是一个叫作“Amabot”的算法，这个算法摒弃了亚马逊网站之前通过人工方式完成的人与人的交流，代之以自动生成并遵循统一规则的推介意见。2013年，布拉德·斯通通过深入研究，在他的亚马逊网站历史回顾中指出：“这套系统轻松地通过了一系列测试，并且证明其推销能力并不亚于那些人类编辑。”

惨遭“围歼”的编辑们要么失业，要么被安置到公司其他部门。2002年情人节那天，一位员工在西雅图的一家本地报纸上刊登了一则广告，向将他们淘汰出局的算法“求婚”。

亲爱的AMABOT：

你这个没心没肺的家伙，你知道我们有多恨你吗？别瞎折腾了，你就是人们胡拼乱凑的破烂而已，最终获胜的必将是有血有肉的人类！

这则广告代表了当时很多人的心声。时至今日，仍然有很多人针对这种心态争论不休，当算法占领本书所讨论的那些面向人文学科的领域时，这种争议更是甚嚣尘上。不过，我们也有必要指出，这种观点实际上是画出了一条只有“有血有肉的人类”才可以逾越的红线，与利维及默南认为实现自动化的可能性因具体工作种类而异的说法，多少有些异曲同工之处。

当然，有很多工作种类以及生活的其他方面，之前都被视为人类的专属领域，但现在算法却在这些领域大显身手。例如，面部识别一度被视为能力较强的少数动物（包括人

类)所独有的特点。但是如今,脸谱网与谷歌使用的算法可以一一识别用户上传的数以亿计的个人图像。

语言与自动翻译也面临差不多的情况。1965年,美国科学院的一个委员会断言:“有效的机器翻译在短期内是不可能实现的,其前景也无法预测。”但是在半个世纪后的今天,人们每天都会使用谷歌翻译完成58种语言的互译工作,这意味着总共有3306种翻译服务。《你耳朵里有条鱼吗?》一书的作者戴维·贝洛说:“谷歌的翻译服务似乎轻轻松松就解决了语际关系问题,并且使语际关系呈现出多样化的特点。连欧盟中坚定地认为语言具有相似性的那些学者们,也不敢奢望能做到这一点。”虽然谷歌翻译提供的译文并不是完美无缺,但译文的质量“足以”满足人们交流的需要,而且还在不断提高。

1. 该说法来自布拉德·斯通的《一网打

尽》，该书中文版由中信出版社于2014年出版。——编者注

每周工作20个小时

在很多情况下，算法时代的突出特点就在于算法可以取代大量人工。杰伦·拉尼尔在他的新作《互联网冲击》中对柯达公司与在线视频共享社交网站Instagram进行了比较。他认为：“在柯达公司的巅峰时期，该公司有超过14万名员工，市值达280亿美元。他们还发明了第一台数码照相机。但是今天，柯达已经破产，而Instagram则变成数码摄影领域的新贵。2012年，Instagram以10亿美元的价格被脸谱网收购，当时该公司的员工仅为13人。是谁剥夺了这么多的就业机会呢？那些中产阶级通过这些就业机会创造的财富又去哪儿了呢？”

在评论这个问题时，很多人惊讶地发现，自动化给人们造成的是一种无差别的伤害。似乎在所有领域、所有行业，只要付出一笔

钱，就可以得到一些设计巧妙的算法，去满足速度、效率与价值等方面的需求。人们纷纷预测，那些在算法时代到来之后仍在苦苦挣扎的行业终将无法逃脱悲惨的命运。在

《法学院的没落》一书中，法学教授布赖恩·塔玛纳哈指出，根据美国政府统计数据，预计到2018年，面向年轻律师的新就业机会仅为25000个，而这段时间内将有45000人从法学院毕业。塔玛纳哈说，即使就业率这么低，也总有那么一天，人们在提到这个时期时还会称之为“过去的美好时光”。可以想象，未来的法律事务所将不再雇用初级律师和实习律师，而把这些工作交由人工智能系统完成。另外，据一些专家预测，10年后律师人数会减少10%~40%。

埃里克·布吕诺尔夫松与安德鲁·麦卡菲在《与机器赛跑》一文中指出，与其说这是大萧条或者大停滞造成的，还不如说罪魁祸首

是大重组。哪些工作可以在算法时代幸免于难的决定因素，不是传统上从事这些工作的人属于哪个社会阶层，而是成本与效率之间能否取得平衡。在自动化的浪潮面前，效率低下（如律师、会计师、大律师与立法者等）且收费昂贵的职业都将不堪一击。要想继续生存下去，正确的做法（正如经济学家西奥多·列维特在1960年发表的《行销短视》一文中指出的那样）是，各行各业必须“对他们目前赖以生存但是已经过时的经营手段进行改革”。

对工业化的经典描述是，自动化程度的提高，取代的是工人阶级所从事的工作（他们认为把工人阶级作为牺牲品是可以接受的）。但是，现在的情况却有所不同。例如，经典著作《人工智能哲学》指出：

人们常常认为有些工作需要依靠一定的智力才能完成，例如进行医学诊断并（或）开

处方，给出法律或科学建议，建立逻辑或数学定理等。而有些工作则不同，所有正常的成年人，无论他们是否受过教育，都可以胜任（有时甚至动物也能完成这些工作），而且通常无须有意识地加以控制。例如在阳光与阴影中看物体，在复杂的地形中辨别方向，把钉子塞进孔洞中，用母语进行交流，以及运用自己的常识等。

该书作者认为，正因为如此，“人工智能”必须具备某种条件，而不是简单地将其归类为“人们必须动脑筋才能完成的事情”。当然，这也意味着如果你所从事的工作过于简单，可以实现自动化，当你被取而代之时就无权抱怨。就像遗传算法或者该算法模拟的生物进化系统一样，变则活，不变则亡。一些网上相亲者因为不遵守现代社会的一系列理性标准而被抛弃，抱怨自己受到了不公平的对待。法国哲学家阿兰·巴迪乌在安慰这些

相亲者时，就提出过一个类似的观点（尽管含有嘲讽之意）。他说，难道他们不知道现代社会的游戏规则吗？

现在，我们还不清楚是否可以实现自动化的标准是什么。的确，莫拉维克悖论表明，真实情况与我们的常识正好相悖：“蓝领”工作似乎比较安全，而“白领”工作则面临风险。史蒂芬·平克在《语言本能》一书中指出：

难的问题易于解决，而容易的问题则难以处理，这是30年人工智能研究收获的主要经验和教训。我们原本以为4岁儿童就可以解决的问题，例如人脸识别、拿起铅笔、穿过房间、回答问题等，却是工程技术中最大的难题.....随着新一代智能设备的面世，股票分析师、石油化工技术人员以及假释委员会成员面临着被它们取而代之的风险。在未来几十年里，园丁、接待员与厨师反而无失业之

虞。

在新的算法社会，计算机科学家与数学家将在文化决策方面发挥越来越重要的作用，而其他行业则将遭遇社会理论家杰里米·里夫金在《工作的终结》一书中描述的悲惨命运。在未来的职场中，“用于生产全球人口所需的商品与服务的人工将不断减少”。包括法律文书起草以及娱乐在内的所有活动支出都将逐步下降，就业机会也会越来越少。安德烈·高兹在《向工人阶级告别》一书中指出：“大多数人工作没有保障，阶级性不明确，（最终将变成）新无产阶级的一员，沦为试用工、合同工、临时工或者兼职人员。”由于工作保障与阶级身份被自动化与算法用户资料所取代，未来的算法社会最终可能实现几代技术乌托邦主义者梦寐以求的目标：“每周工作20个小时，50岁退休”。只不过，这样的结果未必出于人们的自愿。

人类和计算机的角色分工

很多人都听说过美国经济学家、统计学家米尔顿·弗里德曼于20世纪60年代访问一个亚洲国家的轶事。弗里德曼来到一个开凿沟渠的工地，结果，他没有看到现代化的拖拉机与挖土机，却看到工人们正在手舞铁锹，奋力挖掘。他感到非常吃惊，就问同行的政府官员：“为什么机器这么少呢？”官员解释说：“你不了解情况，这是一项就业计划。”弗里德曼沉默了一会儿，然后说：“哦，我还以为你们要修建水渠呢。但是，如果你们希望创造就业机会，就应该给工人们发勺子，而不是铁锹。”

就修建沟渠而言，挖土机的效率要高于铁锹，铁锹的效果要胜过勺子，而勺子的效果要比徒手好。当然，无论有多少种技术，效率都各不相同。问题是，我们选择哪种技术

才是最合适的呢？文化理论家保罗·维利里奥曾指出，船的发明同时带来了海难。如果这个观点是正确的，那么在发生多少次海难之后我们才会停止造船呢？关注算法失控事件的人肯定更容易得偿所愿。2010年5月6日，道琼斯工业平均指数在300秒内猛跌1000点，导致约1万亿美元的财富化为乌有。人们后来把这种股票市场迅猛崩盘的现象称作“闪跌”。直至今日，人们还不清楚这次闪跌的真实原因，只好胡乱寻找各种理由，诸如交易速度太快、技术故障等。

然而，几乎没有人严肃地指出是算法本身出了问题。的确，算法不仅完成了一度由人们手工完成的工作，而且在很多情况下，算法执行的任务是人类无法完成的。谷歌应用的算法更是如此，这些算法所依赖的数据库大到超出我们的想象，用人工来处理是绝对不可能的。

法国经济学家让·富拉斯蒂耶幽默地问道：20世纪的人类因为原子弹的诞生而惶恐不安，那么，在青铜剑发明之时，史前人类有没有感到害怕呢？最强壮的狮子最终也会衰老，被年轻健康的狮子取代，同样，新技术的发明并不意味着人类的终结，它们最终会在喧嚣中退出历史舞台，为更新的颠覆性技术让路。让这个问题更加棘手的是人类与技术之间的复杂关系。我们大多数人一边享受着技术带给我们的便利，一边密切关注着技术的各种“糟糕”表现。举个例子，我是通过什么方法准确地知道上述闪跌的详情呢？谷歌搜索。

克里斯蒂安·孔斯与迈克尔·韦伯在《家长主义：理论与实践》一书中指出，真正的自由论者都会绕开二元选择结构，而倾向于利用原始的研究数据在充分知情、不受任何影响的情况下做出决策。根据这一标准，我们

所有人都不是真正的自由论者，因为我们总是心甘情愿地以这样那样的形式把决策权拱手相让，由技术、家人、好友、同事、公司以及政府代劳。计算机的本质就是一台机器，可以帮助我们摆脱某些艰巨的任务，并得出让我们欣然接受的结果。查尔斯·巴比奇通过自己发明的机械计算机，帮助“人类计算机”（被雇用完成数学计算的人）摆脱了以人工方式编排对数表的艰苦劳动。奥斯卡·王尔德在他的无政府主义小册子《社会主义下人的灵魂》中写道：“事实上，文明需要奴隶。”

希腊人的观点是正确的。如果没有奴隶完成那些脏活儿、累活儿、不体面的活儿，文化与思考就无从谈起。以人为奴隶是不合适、不安全的，也是不道德的。未来世界依靠的是机器奴隶，即把机器变成我们的奴隶。

现在，人类因为无法将自己擅长的那些枯燥乏味、墨守成规的活儿交给计算机来完成，而沦落为机械的计算机器这种角色，这不能不说是一个令人啼笑皆非的境况。“浮士德难题”再一次表现为我们是否会得不偿失的问题。

后机械时代的客观性

在本书的创作过程中（以及我为《快公司》杂志撰写技术类文章时），我一直在考虑客观性这个问题。在对技术人员的访谈中，我们也经常会愉快地聊到客观性。本书第1章谈到算法可以查明个人的某些“客观事实”（包括拉里·斯马在没有临床医生的帮助下用来诊断自己是否患有克罗恩病的那些手段，以及Quantcast等公司构建的“个人用户”资料），这个概念就涉及客观性问题。在第2章中，客观性主要表现在超理性配对环节上，我们希望用算法找到一个“更好”的配对办法，为客户寻觅浪漫伴侣。在第3章中，客观性是法律公平公正的保证，通过设计法律算法，确保执法工作能做到前后一致。在第4章中，客观性是评判艺术作品的普遍真理所必须具备的特点。

客观性这个表达常常与算法以及依赖算法的公司密切相关。例如，谷歌的“十大信条”之一是：“我们的用户信任我们的客观性。”

如果我们有非常多的时间，去思考我们是否应该相信一家上市公司可以做到彻底的客观，以及这家公司在对所有信息进行筛选与分级操作时是否有可能做到客观，我们就有可能发现这个难题在本质上是有点问题的。但是，谷歌搜索用时仅为0.21秒，我们并没有那么多的时间思考这个问题。《后印刷时代》的作者、学者特德·斯蒂法斯花了几年时间调查所谓的“算法文化”。他认为：“这种思考周围世界的方法非常独特。它只考虑纯粹的客观程度，而不涉及任何主观性。你与某种无法阻挡的真理越来越接近……你永远不会犯错误，只会越来越正确。”

谷歌可能真的没有撒谎，它的用户似乎真的相信它的客观性。2005年的一项网络用户

调查表明，只有19%的人表示不信任他们使用的搜索引擎，而有超过68%的人认为他们经常使用的搜索引擎是客观的，没有任何偏见。

人们对已知科学客观性的信任由来已久。1886年，英国统计数学家卡尔·皮尔逊在论文《自由思想的伦理》中指出，科学“独具慧眼，潜心于探求真理，其动机中不掺杂任何利益成分，不是为某个党派、个人或理论服务的。它绝不会歪曲人们的认知。自由思想的理念认为，那些在探求真理时没有秉持客观性的人都是魔鬼阵营的信徒”。他的观点引起了人们的共鸣。19世纪，法国工程师、支持自由市场论的自由主义者米歇尔·舍瓦利耶说：“高质量的统计数据就是冷冰冰的证据，不会受到恫吓与诱惑的影响。”舍瓦利耶认为，统计数据在本质上不带有任何偏见，因此可以利用统计数据来构建完美社会。例

如，通过比较各地的婚生人口与非婚生人口的比率，可以制定“可靠的人口道德指标”。

技术的出现进一步坚定了这种信念。在《客观性》一书中，作者洛兰·达斯顿和彼得·加里森指出，在19世纪所发明的各种机械中，摄影“成为非干涉主义的客观性的全方位象征……因为照相机显然消除了人的影响作用”。那时候，绘画等手工创作的图像与实物的相似度常常高于照片（其中一个原因在于绘画是彩色的），因此在表现现实时更加逼真，但是对于当时的理论学家而言，这个事实并不重要。他们认为，更重要的是照片向受众传递的客观性（换言之，未经过篡改的图片）。

当然，现在的人都知道这个观点是错误的，都知道照片中可以任意添加各种虚假的元素与效果，Adobe Photoshop等程序秉承的正是这样的理念。摄影师在拍照时放弃了

其他影像，而单单选择了某一幅影像，这在某种意义上说明照片并不比其他艺术形式或记录媒介更加客观。更不用说，照相机是由特定团体在特定时间根据他们的既定目的、社会规范与标准制造出来的，因此，其在本质上必然会受到意识形态的影响。

从2014年这个有利的时间点回顾过去，照片的这个特点自然一目了然，但是，我们同样有可能忘记，当下生活中必不可少的算法也有相同的特点。科幻小说家亚瑟·克拉克写过一句非常有名的话：“所有足够先进的技术都像变魔术一样。”100多年前照片刚刚出现时，人们认为它客观公正，同样，在当今社会，算法处理问题的速度非常快，我们又认为算法客观可靠，不会受到人类主观性的影响。把一段文本粘贴进谷歌翻译，在不到一秒钟的时间内，算法就可以把这段文字转化为58种语言中的任意一种。同样，谷歌搜

索算法也具有某种不可思议的能力，可以根据它对一个个用户的“了解”，预测用户的具体希望与要求。传媒研究学者希瓦·维迪亚那桑说：“谷歌仿佛会读心术，因此使用谷歌的服务时我们感到得心应手。从某种意义上讲，谷歌真的可以读懂我们的想法。”

魔术让我们觉得不可思议的原因之一是我们只看到了最终结果，而对其中的过程一无所知。谷歌让我们心怀敬畏，也是同样的道理。谷歌服务的具体过程不仅被罩上了神秘的面纱，让我们无从了解，而且搜索速度非常快，几乎可以做到即时完成。这样做的目的不仅仅是要愚弄那些技术菜鸟。成果丰硕的数学家、计算机科学家约翰·麦考密克在《改变世界的九大算法》^①一书的开头写道：“每个算法的核心都是保证该算法能顺利运行的奇思妙想。”接着，他进一步解释了这句话的意思。

由于我需要多次使用“技巧”这个词，因此我有必要指出，这里所说的“技巧”不同于那些恶意的、欺骗性的花招，不是孩子们用来捉弄弟弟妹妹的把戏，而是与商业窍门及魔术手法相类似的技巧，是帮助人们实现那些难度极大甚至无法企及的目标的高明技术。

尽管麦考密克出于一番好意对这些技巧进行了区分，但他没有经过深思熟虑就对算法的道德感做出了某种假设，因此还是犯了一个错误。认为算法及算法希望实现的目标没有好坏之分（但是，根据前文提及的梅尔文·克兰兹伯格的科技第一定律，技术并不是中立的）这个说法似乎过于笼统，也不妥当。然而，好几位著名的技术作家都提出过类似的观点。2008年，克里斯·安德森在《连线》杂志上发表了一篇有争议性的文章，宣布大数据与算法时代意味着“理论的终结”。他认为，我们再也不需要考虑所谓的专家给出的

优秀理论了。

现在，我们有了更好的选择。大数据让我们理直气壮地宣布：“有相关性就足够了。”我们再也不需要想方设法建立各种模型了，在分析数据时也不用提出各种假设了。只要把数字输入全世界最大的计算机，统计算法就可以帮助我们找出科学无法发现的规律。

当然，安德森在热情欢呼算法时代的到来时也犯了一个错误。他没有意识到，即使是针对大型数据集的数据挖掘，从根本上讲也是建立在理论基础之上的。我在本书中一直强调，算法在回答特定问题时，会认为某些内容比较重要，这可以反映出设计者的偏见。如果算法的目的是确认某个信息是否具有相关性，或者某种解决方案是不是效果最佳，这种设计本身就是一种假设。甚至连数据也无法摆脱人类的偏见：收集的数据种类、算法运行之前的各种相关准备工作等，

都会受到人类偏见的影响。将人员、地点、对象与想法等分门别类、建立等级的计算过程，也具有相当丰富的政治内涵。我在第1章里讨论的用户分类等程序同样如此：这些用户可以被分成哪些类别？每个类别包括哪些人？我们如何确定这样的分类有助于（而不是妨碍）目标的实现？我在第3章里还讨论过，人们对基础设施的要求也各不相同。

早在2004年，传媒研究学者温迪·千熙京（音）在论文《论软件：可视化知识的坚持》中就提出了一个类似的观点。她认为，软件与意识形态在塑造世界观方面具有很多相似之处。尼尔·波兹曼在《技治主义》一书中写道：

无论是过去还是现在，旧世界的每一个信条、习惯或者传统都有技术替代品。祷告的替代品是青霉素，祖籍的替代品是流动性，阅读的替代品是电视，自我控制的替代品是

即时满足，罪孽的替代品是心理疗法，而政治意识形态的替代品则是通过问卷调查构建的大众吸引力。

本书讨论过，“大众吸引力”是一个非常重要的概念，在算法时代随处可见，在谷歌给用户提供的搜索结果、奈飞算法向用户推荐的电影中都能看到它的身影。当然，把大众吸引力看成政治意识形态的技术替代品的观点，完全忽略了新自由主义赖以存在的基础，没有看到新自由主义是当今（算法及其他）世界中意识形态内容最为丰富的概念之一。斯拉沃热·齐泽克指出，只有在伴有讽刺意味的超脱（或者我们像《黑客帝国》中的尼奥一样坚信自己已经摆脱了束缚他的桎梏）状态下，意识形态才会具有最强大的影响力。正因为算法时代是一种后意识形态社会，因此它代表着学者马克·费希尔所说的“资本主义现实主义”取得了胜利。它不再是

某个人提出的某个“精英”理论，而是在人类集体智慧的作用下形成的信息与社会组织方法，是可以进行量化分析的，具有无限的民主性。谁能否认信息本身的性质呢？

1. 《改变世界的九大算法》已由中信出版社于2013年出版。——编者注

算法会造谣中伤他人吗？

2013年3月，生产T恤的Solid Gold Bomb公司与亚马逊因为一条由算法生成的广告语而展开了激烈的争论。当时，印有“保持冷静，如何如何”的随身物品迅速流行起来并席卷英国各地，Solid Gold Bomb公司跟风设计了一个简单的自动化程序，从一本英语词典中任意找出一些动词和形容词拼凑到一起，设计出类似的广告语。亚马逊网站上一共出现了529493件类似主题的T恤，但是必须等到用户掏钱购买之后，这些T恤上才会真的印上这些口号。BoingBoing网站在评价这个经营计划时写道：“（Solid Gold Bomb公司）设计这些口号并交给亚马逊不用花一分钱，亚马逊保管这些口号也不用花一分钱。如果无人购买，那么这次尝试的总成本几乎为零。但是，如果该算法走出了一条特殊的

经营之道，就证明这种独特有趣的方法真的有很强的促销效果，大家都能赚得盆满钵满。”

不幸的是，Solid Gold Bomb公司显然没有预料到词典中的那些单词可能会被拼凑成令人反感的口号。虽然仇视女性的口号“保持冷静，打她”以及从未受到热捧的口号“保持冷静，不停地抚摸她”令人震惊，但是令亚马逊更加怒不可遏的是那条极具攻击性的口号——“保持冷静，尽情强奸”。在了解到这些情况之后，亚马逊立即做出反应，将Solid Gold Bomb的T恤全部下架。结果，Solid Gold Bomb在几个月之后破产了，这家T恤生产商十分恼火。该公司的老板们感到莫名惊诧，因为犯过错的是算法，但对于算法而言，这些口号没有任何特殊含义。经营者没有任何过错，为什么还要接受处罚呢？

类似的问题在2012年就出现过。当时，

贝蒂娜·武尔夫（德国前总统克里斯蒂安·武尔夫的妻子）指控谷歌的“自动补全”算法对她进行了诽谤和中伤。自动补全算法最初的设计目的是帮助残疾人提升打字速度，但后来谷歌把这个算法纳入了公司的服务功能，在用户完成输入之前猜测他们的搜索项，帮助用户节省时间。该算法的设计者马里厄斯·韦尔说：“谷歌利用几十亿次搜索记录提供的数据，为用户建立了匿名的用户资料，可以反映处于某个年龄及人生阶段的人所担心、查询、关注及迷恋的内容。”例如，在谷歌搜索框内输入贝拉克·奥巴马的名字，你就能看到“贝拉克·奥巴马”“贝拉克·奥巴马的推特”“贝拉克·奥巴马名言”及“贝拉克·奥巴马的事迹”等你可能用得上的提示信息。在输入英国副首相尼克·克莱格的名字时，可能会看到“尼克·克莱格是个混蛋”“尼克·克莱格是个爱说谎的家伙”“尼克·克莱格坏透了”以及“尼克·克莱格

完蛋了”等提示信息。而贝蒂娜·武尔夫的搜索提示内容更偏向于尼克·克莱格的特点。在搜索武尔夫的名字时，很可能会出现指向妓女及陪侍行业的搜索项。

得知在网上搜索自己姓名时可能会出现这些结果之后，武尔夫把谷歌告上了法庭，并且打赢了这场官司。德国的一家法院做出判决，谷歌必须保证自动补全算法生成的搜索项不含有任何攻击性或中伤性的内容。谷歌大为光火，宣称对这个判决结果感到极为“失望”。因为谷歌一直认为自己的算法都是客观的，而这个判决结果是对这种客观性的公开质疑。谷歌公司发言人说：“我们认为，谷歌不应当对自动补全算法生成的搜索项负责，因为这些预测不是谷歌自己完成的，而是计算机算法基于之前的搜索记录自动完成的。目前，我们正在重新审视我们的选择。”由此可以看出，问题的根源在于谷歌对算法远景

的迷信。我们在第3章提到科幻作家菲利普·迪克的“少数派报告”这个概念，与之相似，如果算法也会出错，就有可能导致灾难性的后果。叶夫根尼·莫佐罗夫在他的《网络错觉》一书中指出：“谷歌公司在精神上对‘算法的中立性’倍加推崇，说明在其成为世界上最重要的信息看门人之后，一种不安情绪正在公司内部滋生。谷歌的创始人宁愿把技术看成一种自律的、完全客观的力量，也不愿意因为担心公司系统中包含内在的偏见而无法安然入睡。毕竟，这些系统已经异常复杂，所有谷歌工程师都不可能对其了如指掌。”

在讨论谷歌的书籍与文章中，人们常常会大谈特谈谷歌的崛起对报纸等传统新闻媒体的衰落所起到的关键性作用。从这个意义上讲，尽管谷歌自己并不报道新闻，但它已经取代了传统媒体。不过，如果谷歌算法同报

纸一样，也必须遵从某些标准，就会面临一些问题。在一项经典的新闻编辑室的客观性研究中，社会学家盖伊·塔奇曼指出，新闻业之所以客观公正，就是因为担心会导致诽谤中伤的后果。记者不报道自己的观点，而是报道其他人的观点，由此规避了人们指控他们的报道带有个人偏见的风险。谷歌的自动补全算法提供的搜索项是基于数千名甚至几百万名用户之前的搜索记录而生成的，因此，该算法提供的也不是自己的观点，而是引述他人的观点。但是，由于没有对这些搜索项加以审查，算法没有做到明显的客观中立，谷歌公司还是受到了诽谤和中伤他人的指控。

所谓的客观性不过是一种假象

贝蒂娜·武尔夫案标志着人们在算法问题上取得了几个有价值的进展。首先，这个案件表明从此以后算法的政治问题被纳入了司法审判的范畴。但是，以政治问题对算法加以指责可能难度不小。与谷歌街景等服务（由于人们认为谷歌街景侵犯了隐私权，因此在某些国家爆发了街头抗议活动）不同，算法看不见摸不着，人们很难发现它们会导致哪些后果。登录服务器之后，看到从门前车道另一侧拍摄的自家房屋的详细照片，并不代表我们就有权批评谷歌、脸谱网与亚马逊等公司赖以维系经营活动的那些算法，或者对这些算法的内部运行机制指手画脚。在大多数情况下，这些算法都被装进了“黑盒子”，用户无从了解它们的工作原理。同数字

世界不断变化的透明度概念（我们在第3章讨论过这个概念）一样，硅谷决策者们常常故意为复杂的技术设计一个简单明了的界面，并把它看成一个重要的卖点。2008年，现任雅虎总裁兼首席执行官的玛丽莎·梅耶尔在谈到谷歌公司的算法时说道：

我们认为这是最合理的做法。我们的用户不需要理解复杂的技术，以及技术背后复杂的发展过程。他们只要知道在搜索框中输入希望搜索的内容，然后就可以得到答案。

其次，武尔夫对自动补全算法可能涉及政治问题的担心，证明算法可以改变当今社会公共舆论的进程。在看到算法提供的并不友善的搜索项后，本来对贝蒂娜·武尔夫并不了解的用户在搜索她的信息时可能会引至某个特定方向。因此，算法不仅仅是在预测，还有助于控制用户的行为。以我在第4章讨论的奈飞推介系统所使用的算法为例。奈飞

宣称，该公司60%的租赁业务是根据算法的建议完成的，而不是来自用户的具体搜索。如果实际情况果真如此，那么导致这种情况发生的原因到底是什么呢？是因为算法可以猜到用户接下来希望搜索的内容，还是因为算法将特定选项放在用户眼前、让用户只能从中做出选择呢？

此时，这个问题与算法本身已经没有多大关系了。社会学家威廉·托马斯与多萝西·托马斯说过：“如果人们把某种情境定义为真实的，这种情境就会造成真实的影响。”我们还可以引用凯文·斯拉文在他的TED（美国的一家非营利性会议机构）演讲“算法塑造世界”中说过的一句令人难忘的话，算法中用到的数学知识最初是“提炼自这个世界，来源于这个世界”，而现在则“开始塑造这个世界”。

这种情况真的可能会发生。2008年9月6日，算法几乎导致美国联合航空公司的母公

司UAL破产倒闭。这个事件的始作俑者是Income证券咨询公司的一位记者。这位记者在谷歌的搜索框中输入“2008年破产”，并点击了回车键，谷歌的搜索结果立刻把他导向《南佛罗里达太阳哨兵报》的一篇文章，称UAL公司已经提交了破产保护申请。这位记者所在的公司负责为颇具影响力的彭博通讯社提供新闻报道，因此他立刻向彭博网站发了一条消息，标题为“美联航根据《破产法》第十一章申请破产保护以削减成本”，但是没有提供任何详细内容。于是，彭博财经新闻的数以千计有影响力的读者很快就看到了这则宣称美联航正在为债务人寻求法律保护的新闻。后来，人们发现这则新闻其实是一则旧闻。美联航于2002年的确提交过破产申请，但是在2006年经过重组之后，该公司成功地度过了这次危机。由于《南佛罗里达太阳哨兵报》的原始新闻档案中没有注明日

期，谷歌算法根据信息采集软件发现这篇文章并将之放入搜索结果，还将它的日期标记为2008年9月。由于这则不实消息，恐慌的股民在一两个小时内抛出了1500万股UAL股票，导致该股票价格由每股12.17美元狂跌至每股3美元。

算法的威力如此巨大，但我们却越来越依赖于算法来告诉我们什么重要、什么不重要，难道这不值得我们惊讶吗？2009年年初，法国小镇伊乌决定把自己的名字换成一个更长的表达，叫作“Eu-en-Normandie”或者“la Ville d’Eu”，原因是人们在谷歌搜索“Eu”时会出现大量与欧盟（简称“E.U.”）有关的结果。

我们再以在信息中插入“#”号（#标签）这个始于推特的做法为例。2013年9月的短喜剧系列《吉米今夜秀》告诉我们，这种标签现象非常不适合真实世界中有意义的交流。

（一个角色说：“我给你买了一些饼干。#国产，#燕麦葡萄干，#ShowMeTheCookie牌。请查收。”）但是，插入标签当然不是为了让对方更好地理解我们所说的话，而是为了对我们说的内容做出修饰，方便推特搜索算法进行识别并发布。

如果相亲网站的算法发现用户找不到合适的相亲对象，就会做出某种反应，同样，如果推特搜索算法认为我们所说的内容与主题无关，也会给出一个类似的提示信息。在“占领华尔街”活动期间，很多参与者与支持者利用推特组织活动并开展辩论与宣传工作。尽管有不少媒体报道了这次示威活动，但是根据推特算法的判断，这次活动并不是“热门话题”。热门话题是指推特从每天的2.5亿条推文中选出并显示在主页上的讨论最热烈的话题。尽管“#占领华尔街”没有成为热门话题，但是“#在女士手袋中能找到什么”以及“#口渴

的人常干的事”等轻松的喜剧性“迷因”

（meme）却在同一时期毫不费力地就在推特首页占据了一席之地。

尽管推特否认他们有审查环节，但是令我们感兴趣的是，用户强烈抗议算法被赋予了大量的舆论引导作用。加州大学欧文分校唐纳德·布伦信息与计算机科学学院的信息学教授保罗·多尔希说：“这是一个信号，表明推特热门话题已经被解读成各种社会行为重要程度的标志。”在多尔希眼中，推特热门话题与纸质报刊巅峰时期的《纽约时报》或《卫报》的头版头条新闻具有不相上下的重要意义。换句话说，推特的热门话题算法使我们再次想起“一棵树在森林里倒下而附近没有人听见”这句古老的谚语。如果“占领华尔街”这样的社会运动没有引起所谓的社交媒体的关注，那么我们不禁会怀疑，真的发生过这次运动吗？

这是算法面临的又一项新任务。人们在20世纪争论不休的焦点是媒体自由的精髓，即新闻报道的客观性问题，而进入21世纪之后，算法在这些客观性争论中占据了越来越重要的地位。倡导对言论自由进行重新解读的亚历山大·米克尔约翰提出过一个非常有名的观点：“最重要的不是确保所有人都有发表言论的自由，而是在应该有人发表言论的时候总有人挺身而出。”那些掌管这些重要算法的人应当承担文化看门人的责任，判断哪些内容应该让人们看到或听到以及如何处理其他内容。在新闻网站上，一些观点得到的消极评论远多于积极评论。利用算法屏蔽这类评论可能会有效地阻止垃圾信息，但是这种做法同时也会招致侵犯言论自由的风险。

这种做法之所以有招惹麻烦的可能性，是因为偏见有可能从中作祟。在很多情况下，某项业务的界面可能没有任何风险，但是由

于偏见的不断作用，算法已经发生了改变。科罗拉多大学法学院副教授哈利·苏尔登（第3章有他的介绍）说：“中立性常常不是真实的，而是一种假象。”也就是说，我们很多人在发现问题之前都以为算法是客观的。

学者特德·斯蒂法斯说：“如果大多数人都不清楚决策过程，我们就很难开诚布公地对话、讨论文化问题，或者讨论哪些内容比较重要而哪些内容应该不予理睬。这并不是说算法会破坏民主，但是在涉及民主与文化的关系时，算法的确会导致一些问题。”文化知识缺失的例子随处可见。2009年4月，57000多本对同性恋持宽容态度的书被错误地划入“少儿不宜”类别，并从亚马逊的销售排行榜上销声匿迹。尽管亚马逊很快改正了这一做法，但是所谓的“#amazonfail”（亚马逊拒绝）事件却将一些秘密展现在用户的眼前。之前，用户都认为亚马逊的销售排行榜

算法是一种客观的量化工具，但是现在他们发现这些算法会有意识地忽略那些被归类为“少儿不宜”的书籍。被算法剥夺排名的类似情况不乏其例。YouTube视频网站通过降低有性暗示的视频的排名，使它们不会出现在“观看次数最多”的页面及面向新用户的“最受欢迎视频”的主页上。脸谱网每次修改EdgeRank算法（用于确保“最令人感兴趣的”内容进入用户信息流列表的算法）时，就会引发存在危机，因为从显示效果看，有的内容清晰可见，而有的内容则难以辨识。

有时候，这些做法就像谷歌的自动补全算法一样，并不是由公司自行决定的，而是算法做出的决策。我们可能会对某个决策的功过得失有不同的看法，但是，就像学者塔尔顿·吉莱斯皮所说的那样，在搜索某次公共示威活动时，展现在我们眼前的搜索结果似乎包含了所有不同的观点，但其实还有很多人

的言论被屏蔽了。

被植入了计算机逻辑

，我就提出了一个问题：算法能处理一切事务？这个问题自然会引出两个问题：是否应该用算法来处理一切事务？是否应该用算法来处理一切事务？书讨论的“公式”这个主题，引出三个问题：是否存在不适宜用算法处理的问题？这真的是一个价值连城的问题吗？很少有人考虑过。著名数学家弗雷德·怀特海曾经指出，19世纪就是发明这个概念本身。随着计算机的发明，如何发明之后，为什么发明的重要性就变得次要了。技术发明这个观点，认为伦理道德是次要活动，根本无须考虑。弗雷德·怀特海去世的7年后，英国数学家戈弗雷·哈代完成了

《一个数学家的自述》
述了自己的坚定

，画家鄙视艺术
理学家或数学家
嘲笑比所有评论
整体上更为合
过是智商平平的

人比尔·唐瑟尔

中摘选了一段文
助人们确定一年
式时，表达了一
担忧并不是关于
到底是否可以通
最低落的那个星
是错误的。

能轻而易举地回

他那篇非常有名的文章——《一个数学家的辩白》。他在文章的开头陈述了自己的坚定信念：

政治人物鄙视时事评论员，画家鄙视艺术评论者，但是生理学家、物理学家或数学家都有类似的感受：创作者的嘲笑比所有评论者的嘲笑都更加深刻，或在整体上更为合理。阐释、批评、欣赏，不过是智商平平的人的工作。

在本书的引言部分，我从比尔·唐瑟尔2009年出版的《在线为王》中摘选了一段文字。作者在介绍一种可以帮助人们确定一年中哪个星期情绪最低落的公式时，表达了一种深切的担忧。但是，这种担忧并不是关于这样的公式是否可能存在（到底是否可以通过量化分析找出一年中情绪最低落的那个星期），而是他认为这个公式是错误的。

如果可以找到某种方法，能轻而易举地回

答“我是谁”“我应该爱上谁”或者“艺术是什么”等问题，这种方法就会有很强的吸引力。我在第2章里介绍过加思·桑顿，这位统计学家设计了一款可以准确预测名人离婚率的公式。我问桑顿，为什么这些公式与他们这些统计学家一样，都能让普通人感兴趣呢？他的回答对宗教狂热的打击甚至超过了逻辑实证主义。他说：“我认为人们都希望知道答案。我经常设计‘在酒吧里是否应该与女孩搭讪’这样的问题，用来研究一些愚蠢的方程式。但是，一想到就连那些几乎无法回答的问题，也有人可能会给出答案，我就觉得这些研究特别有意思。”如果他发现我们在处理一切问题时，可能都会像Hunch公司（我在第1章里讨论过，这家公司宣称只依据5个数据，就可以回答任何跟消费者偏好相关的问题，而且正确率为80%~85%）那样，把问题抽丝剥茧，然后想方设法用量化分析的方法

加以处理，他会不会觉得恐怖呢？桑顿回答道：“我个人认为，相反的情况更令人害怕。我认为不确定性比可能的确定性可怕得多。我在大学期间设计第一批公式的目的是，通过数学方式帮助我们判断是否应该去图书馆完成作业。但我发自内心地认为，我的目的是要告诉人们生活充满了不确定性，因此我们希望增加确定性的成分。对于我来说，没有答案的可能性比利用数字给出答案的算法更加令人不安。”

现代世界为什么让我们想要找到标准答案呢？难道是因为我们人类就像美国统计学家纳特·西尔弗在《信号与噪声》^①里所说的那样，天生喜欢寻找规律吗？还是因为技术的发展需要那些只有算法才能提供的答案吗？理查德·桑内特在《公共人的衰落》一书中说：“现代人有一个普遍错误的认识，以为社会的各种弊端就是没有人情味、疏远、冷

淡.....因此，人们希望从客观情境、物体以及社会的客观条件中寻找人生真谛，从而培养出人类对稳定性的隐性欲望。”学者亚历山大·加洛韦在《界面效应》一书中给出了一个类似的解释。他认为算法文化有向简单主义发展的趋势，这是“无法根据眼前的情况考虑全局问题，以及无法像解读历史那样解读现在所导致的必然后果”。

20年来一直从事数字技术方面写作的传播理论家麦肯齐·沃克认为：“（算法的确）是管理世间万物的一个重要工具。如果我们认为算法在结束运行时会产生一个结果，那么这个程序总有结束运行的时候，到那时你就可以得到答案了。如果某个程序没完没了地运行，就说明你的电脑可能出错了。在构建现实时，我们似乎越来越依赖可以产生结果的程序，仿佛在这个世界中植入了计算机逻辑。”

在讨论算法时，人们常常会提到“黑盒子”这个概念。布鲁诺·拉图尔认真地研究了它的强大象征作用，他说，只要某个机械部件或者一组计算机指令过于复杂，控制论专家就会说它们被装进了黑盒子。所谓黑盒子，是指一种晦涩难懂的技术，人们在运用这种技术时无须任何了解，只要输入数据，就能得到结果。打开黑盒子，设计者与用户面对的将是一堆可以得出某种答案的主观偏见与程序。而合上之后，它体现的就是客观性——一种无须满足任何更多的条件即可生成“是”与“否”的二元选项的机器。拉图尔说：“只要人们认为所有黑盒子密封完好，它们就像科学家一样，摆脱了虚构、表示、象征、近似与惯例构成的现实世界，就是绝对正确的。”在这里，我们可以借鉴斯拉沃热·齐泽克对抽水马桶的理解：排泄物通过这个看似技术性不强的机械装置，从我们的现实世界销声匿

迹，进入了一个被现象学视为更加脏乱、原始的世界。

理查德·伯克的著作中也蕴含了这种思想。伯克在提到他设计的犯罪预警算法时说：“它让我解脱了，我才不管是否存在因果关系呢。”尽管伯克说这项技术的目的是收集有实际价值的信息以实施犯罪预警，但是人们可能有理由认为，由于其内部运行机制被装入了黑盒子，因此社交活动的基本动态也会发生类似的变化。在其他领域（尤其是涉及法律的那些领域），对算法的依赖有可能意味着偏见的存在与不了解内情是合理的。我们在第1章讨论过的“过滤气泡”效应，可能导致某些人看不到某些信息，也是出于这个原因。

学者特德·斯蒂法斯说：“不了解这些算法工作原理的不仅仅是你和我这样的普通人，连工程师自己也不完全理解。看看奈飞大

赛，评审人员一再强调某个算法获胜的原因是效果很好，但是他们也不知道这些算法为什么效果很好。某些数学知识也许可以解释这些算法的工作原理，但是这些数学知识过于复杂，人们是不可能真正掌握的。由于无法了解我们所创造的一切，因此我们很难留意到道德问题，也很难保证所有操作都符合道德标准。”

1. 《信号与噪声》由中信出版社于2013年出版。——编者注

在算法时代必须养成提问的好习惯

我在收集资料的过程中，遭遇的最令人不安的一个算法，是《华盛顿邮报》于2012年夏季总统大选时推出的“实话实说”算法。这个算法可以实时分析政治言论，告诉我们这些言论是否真实，因此它不断地提醒我们两个令人不安的事实：我们相信算法的客观性，我们渴望得到简简单单的答案。

Geek.com网站上的一篇文章，以无比辛辣的文笔描述了这个算法：“以最坚定的语气、高度自动化的方式，告诉我们某位政客是否在撒谎.....因为政客们都是妄想狂，他们对自己的谎言都信以为真。”该算法采用了微软开发的语音识别技术，先把语音信号转换成文字，然后将完整的文本提交给匹配算法。匹配算法梳理政客们所说的“事实”，并与数据

库中收录的已经证明的事实进行对比。

如果这样一项技术不仅仅是一个简单的噱头，而是像自动拼写检查算法那样无处不在，那么我们可以想象将它应用于文档或数据编辑的前景。如果未来的微软word或者谷歌文档采用这项技术，不难想象，总有一天用户在输入文本之后，只需点击一个按键，就可以完成拼写、标点符号、格式与真实性的自动检查。学术界已经广泛采用算法来审查学者们提交的学术研究论文，从中找出涉嫌剽窃的段落。在自然语言处理程序变得更加简便——不仅可以进行简单的段落对比，还可以比较详细内容、完成思想评析——之后，这类应用还将得到进一步推广。

如何对待算法，这个问题并没有一个“一刀切”的答案。在某些情况下，增加透明度似乎是一个正确的做法。例如，在执法工作中使用算法时，公开源代码的做法既可以规避

政府制定的政策缺少审核环节的风险，还可以帮助我们判断具体决策是如何形成的。但是，在涉及国家安全时，开源代码的做法就不适用了，因为一旦公开了特定黑盒子的内部运行机制，某些人就可以绕开保密系统，致使算法失效。

2013年，谷歌的流感趋势算法就遭遇了这种矛盾局面。由于流感趋势算法可以通过用户搜索关键词分析追踪流感的扩散情况，因此被视为一种突破。但是，该算法随后就捅了娄子，并引起了媒体的普遍关注，并陷入了一个意想不到的麻烦。这款算法的设计思想是寻找“止咳退烧药”等搜索项，并推测这些用户患有感冒，但是谷歌所发现的那些用户，虽然输入的是与流感相关的搜索项，但其实是在查询与谷歌这种算法有关的信息。谷歌看到与流感相关的搜索项数据呈激增趋势，因此预测流感将大规模爆发，但是

流感疫情最终并没有出现。谷歌后来承认，虽然他们的算法在流感疫情预测方面有非常好的效果，但是“在媒体过度关注时也会受到影响”。《信息周刊》杂志的记者托马斯·克拉本在报道这个事件时说：“这是一个极具讽刺意味的教训。算法在评估来自公共渠道的数据时必须注意保密，否则就会有人想方设法植入偏见。”

除此以外，我们还需要解决一个问题：如何在算法时代保持人性？如果“硬人工智能”（hard AI）的初衷是打造行为特点与人类完全相似的计算机，它的功利主义的反作用就是使人类行为简化到像电脑行为一样可以预测的程度。马克斯·霍克海默与西奥多·阿多诺认为，实证科学“以公式取代概念，以规则与可能性取代因果关系”，就是为了这种简化效应。

法国哲学家阿兰·巴迪乌在《爱的挽歌》

一书中也同样把目光投向了算法相亲，希望找到公式拥趸们的“无心之失”。巴迪乌根据“爱情就是一杯苦酒”这个古老的观念，指责相亲网站所承诺的“可预测式”爱情将随机性这个一度非常重要的元素破坏无余，并将其与“零伤亡”战争、“智能炸弹”相提并论，认为这些承诺都无法兑现。的确，如果没有任何风险，又如何“坠”入爱河呢？如果谈恋爱没有任何激情，仅仅是一次在填完调查问卷之后由算法安排的交往，还会有任何乐趣可言吗？巴迪乌认为，向算法相亲的过渡意味着资本主义朝着缺少活力的晚年又近了一步。消费主义“零投入”式爱情代表的是一些公司采取的“零投入”立场，就像高科技公司一样，他们只是临时雇用一批人手，规避了目标长远的保障措施。用算法来处理人类特有的一些活动，有可能使这些活动丧失最根本的人性，这样的例子俯拾即是。杰伦·拉尼尔

在他的技术怀疑论著作《你不是个玩意儿》^①中指出：“保持人性不是一个简单的公式所能做到的，它需要人们不断探索、解开谜团，以及一次信念的飞跃。”

但是，在人类不断沦为算法程序处理对象的世界中，要保持人性远非易事。我们应该怎么办呢？当然，有一小部分人正在从事所谓的“算法数据干扰”研究，以便在算法试图了解他们并将他们分成各种类别时，可以采取干扰或规避措施。但是，如果采取这种做法，就意味着放弃使用现代社会的某些重要工具。由于算法对相关性的界定越来越明确，使得它与公共话语的很多内容渐行渐远。

在这里，我建议人们更加深入地了解公式世界，因为这些算法程序将为包括浪漫关系与法律事务在内的众多争议性话题提供答案。例如，警察在实施“截停搜查”时必须有

合理的依据，意即“算法结合详尽、明晰的事实，并根据这些事实做出理性的推断，表明其实实施截停搜查措施是有依据的”。在这种情况下，“算法表明”（假设这种算法已经被证明切实有效）是实施截停搜查措施的合理理由吗？

同样，如果说人与算法的浪漫关系不是“完全真实”的（参见在第2章介绍的雪莉·特克尔的观点），那么这些关系是否“足够真实”呢？2013年，西班牙的几名研究人员成功设计了一款软件，能够模仿14岁少女的语言与态度。人们可以利用这款名叫“Negobot”（亦称“虚拟洛丽塔”）的软件，设置陷阱引诱恋童癖者上钩。在交谈刚刚开始的时候，Negobot会采用中立模式，与陌生人谈论普通话题。但是，当某个陌生人在交谈时对性有所暗示，或者有明显的企图，并试图诱骗“少女”见面时，Negobot就会切换到“狩猎模

式”。如果执法人员采用这项技术，就意味着人与算法发生的感情至少“足够真实”，可以作为起诉的依据。

更加重要的是，我们必须学会提出问题，而且不要指望答案唾手可得。斯拉沃热·齐泽克在《生于末世》一书中指出，在所有问题均可轻松解决的世界里，人们需要的是一种更加激进的思维模式，但是这种思维模式本身就会导致某些问题。我们在提出问题时，不仅需要了解算法正在发挥哪些作用，还要了解这些算法最初的设计目的。虽然这个问题回答起来并不轻松，但是我不应该回避，而且在接受某项表面上是免费的服务时更应该进行认真考虑。“相关”“有新闻价值”等表达都话里有话，用户显然应该问一个问题：这种“相关”或者“有新闻价值”到底是对谁而言的？（但是，人们常常无法回答这个问题。）对于谷歌等公司来说，这个问题的答

案非常简单：当然是跟公司的股东相关，对他们来说是有新闻价值的。同样，我们可能以为脸谱网的算法是用于帮助用户建立和维护朋友圈的，但实际上，脸谱网的目的并不是要帮你交朋友，而是要借助广告，把你的社交圈子变成真金白银装进他们的腰包。

值得庆幸的是，人们有望养成提问的好习惯。几位从事推介算法研发工作的人告诉我，近几年来，用户的期望已经发生了一些变化。5到10年之前，人们会非常愉快地接受各种推荐，但是现在，越来越多的人希望了解这些推荐意见的形成原因。在实现用算法处理一切事务这个愿景的进程中，如果我们经常问自己为什么要以“界面价值”接受商品或者服务，就能提高评价这一进程的能力。最终，一劳永逸的答案将不复存在。如果没有算法，我们的生活将大不一样，很有可能远不如现在，但这并不意味着我们就无

须考虑这些重要的问题。

在某些答案似乎唾手可得时，我们更应该小心谨慎。

1. 《你不是个玩意儿》由中信出版社于2011年出版。——编者注

致谢

创作几乎总是一件孤寂的事情，我却非常幸运，因为我身边的人或是对我充满关爱之心，或是向我伸出援助之手，或是兼而有之。在他们的影响下，我的创作之路充满了乐趣。首先，我要感谢克莱拉、蒂姆、西莉亚·伦特以及我的家人。如果没有艾德·福克纳的大力支持，我就不可能完成本书的创作。如果没有出版代理人麦琪·汉波利、亨利·德·鲁日蒙、西蒙·加菲尔德和杰克·林伍德的帮助，我甚至没有勇气动笔。在我收集资料的过程中，还有很多人牺牲了大量宝贵的时间与我交流，在此向他们表示感谢。此外，我还要感谢《快公司》网站的编辑克里斯·丹嫩和苹果科技网站Cult of Mac的利安德·卡尼，感谢尼科尔·马蒂内利、卡尔·弗兰奇、技术传播专家爱丽丝·伯纳西奥·艾金森、蒂姆·麦

茨、艾利克斯·米灵顿、迈克尔·格罗豪斯、汤姆·艾金森、西蒙·卡洛和我同父异母的兄弟安德烈及内森·特兰特莱尔，感谢他们为我提供的各种帮助。书中如果出现任何错误（希望很少），都是我的责任。