

## 伦理与道德

## 【导读案例】 脸书数据泄露：成操控大选的工具

2018年3月,脸书公司遭遇5000万名用户数据泄露风波,这些泄露数据被指控用于操控当时的美国大选。接着,连续数日脸书公司的股价下跌,市值蒸发500亿美元左右,甚至还创造了4年来的最大单日跌幅。

2018年3月16日,脸书宣布封杀剑桥分析公司。18日,美国《纽约时报》和英国《卫报》等媒体同时发文称,剑桥分析公司利用从脸书不当收集的5000万用户个人数据,来为美国大选参选人提供数据采集、分析和战略传播,简单来说,就是“操纵民意”,而时任美国总统特朗普就是剑桥分析公司此前的客户。

消息曝光之后,各界人士纷纷抨击脸书对用户的数据保护不利,包括美国、英国、欧盟等国家的各大机构也将调查脸书是否存在滥用用户数据的行为。

不过,起初脸书公司不认为他们存在问题,认为自己也是受害者,但是外界却以保护用户隐私不力的原因将其视作帮凶。美国、英国等地掀起了一番“卸载脸书”的活动,甚至苹果应用这样的“友商”也开始落井下石,呼吁用户卸载脸书。

这些用户数据是剑桥分析公司在2015年从一位俄裔美国研究员那里以不正当手段获得的,当时脸书要求他们删除这些数据,剑桥分析公司表示已经删除,但是现在脸书发现,对方并未删除这些数据,还将其用于操控大选期间的民意等用途。“如果这些数据依然存在,则构成了对脸书政策的严重违反,也违背了这些机构过去的承诺。”

向英国《卫报》主动爆料的剑桥分析公司共同创办人克里斯托弗·威利称,剑桥分析公司成立于2013年,由美国共和党的亿万富豪罗伯特·默瑟出资。而这家公司成立的目就是通过海量数据分析研究操纵社会舆论。剑桥分析公司和最初的泄密人都与俄罗斯有着千丝万缕的联系,在特朗普“通俄门”的疑云之下,脸书可以说是被卷入了一场“纸牌屋”式的漩涡中。

英国媒体“第四频道”曝光了一段暗访视频。暗访视频中显示,剑桥分析公司市场总监马克·特恩布尔称,“如果你在人们身上搜集数据,然后描述分析他们,你就有了更多能够加以利用的信息,你就知道该如何细化样本。然后在他们关心的事件上用他们更可能产生共鸣的语言、图像给予信息。”

有分析师认为,通过数据做精准服务是普遍的,提高了信息分配的效率和用户体验;应该谴责的是滥用数据的,而不是分享数据的人,脸书这次是被别有用心的人利用了。

北京时间 2018 年 3 月 22 日凌晨,脸书创始人兼 CEO 扎克伯格在用户数据泄露丑闻后首次发声,他承认对数据泄露事件负有责任,并表示:“我们有保护用户数据的责任,正打击平台对数据的滥用。我们犯了错误,还有更多的事要做,正在弄清状况,本人对脸书发生的事情负责,将调查所有有权限获取数据的应用。”扎克伯格说,用来防止此类事件再次发生的最重要措施在几年前就已经采取了。脸书将禁止开发人员滥用用户信息,并进一步限制开发人员访问用户数据。

资料来源:根据网络资料整理。

阅读上文,请思考、分析并简单记录:

(1) 用户数据的泄露事件时有发生,你认为这次脸书事件的严重性在哪里? 请简述。

答: \_\_\_\_\_

---



---



---

(2) 也是 2018 年生效的 GDPR 是欧盟针对个人数据和隐私保护实施的一项新立法,是 20 年来全球最重要的数据隐私保护变化,也是有史以来规模最大、最具惩罚性的数据保护法。请网络搜索了解 GDPR 数据保护法律规定。简述脸书和剑桥分析公司主要违反了什么规定?

答: \_\_\_\_\_

---



---



---

(3) 在这个事件中,剑桥分析公司耍了卑鄙的手段。但你认为脸书公司只是“受害者”吗? 为什么?

答: \_\_\_\_\_

---



---



---

(4) 请简单记述你所知道的上一周内发生的国际、国内或者身边的大事。

答: \_\_\_\_\_

---



---



---

### 3.1 伦理与道德基础

在西方文化中,“伦理学”一词源自希腊文“ethos”,意为风俗、习惯、性格等。古希腊哲学家亚里士多德最先赋予其伦理和德行的含义,所著《尼各马可伦理学》一书为西方最

早的伦理学专著。首次出版于1677年的荷兰哲学家斯宾诺莎的伦理学著作《用几何学方法作论证的伦理学》认为,只有凭理性的能力获得的知识,才是最可靠的知识。人有天赋的知识能力,世界是可以认识的,“伦理学”从本体论、认识论开始,最后得出“伦理学”的最高概念,为人的幸福指明了道路。

在中国文化中,伦理(图3-1)一词最早出现于《乐记》:“乐者,通伦理者也。”我国古代思想家们对伦理学都十分重视,“三纲五常”就是基于伦理学产生的。最开始对伦理学的应用主要体现在对于家庭长幼辈分的界定,后又延伸至社会关系的界定。



图3-1 伦理

### 3.1.1 伦理的定义

伦理学是哲学的一个分支,被定义为规范人们生活的一整套规则和原理,包括风俗、习惯、道德规范等。简单地说,它就是指人们认为什么可做、什么不可做,什么是对的、什么是错的。

哲学家认为“伦理”是规则和道理,即人作为总体,在社会中的一般行为规则和行事原则,强调人与人之间、人与社会之间的关系;而“道德”是指人格修养、个人道德和行为规范、社会道德,即人作为个体,在自身精神世界中心理活动的准绳,强调人与自然、人与自我、人与内心的关系。道德的内涵包含了伦理的内涵,伦理是个人道德意识的外延和对外行为表现。伦理是客观法,具有律他性,而道德则是主观法,具有律己性;伦理要求人们的行为基本符合社会规范,而道德则是表现人们行为境界的描述;伦理义务对社会成员的道德约束具有双向性、相互性特征。

可以这样理解,法律是具有国家或地区强制力的行为规范,道德是控制人们行为的规则、标准、文化,而伦理是道德的哲学,是对道德规范的讨论、建立以及评价,研究的是道德背后的规则和原理。它可以为人们提供道德判断的理性基础,使人们能对不同的道德立场作分类和比较,能在有现成理由的情况下坚持某种立场。现代伦理已经延伸至不同的领域,因而也越发具有针对性,引申出了环境伦理、科技伦理等不同层面的内容。

在长期的发展中,关于伦理形成了如下一些定义。

**定义 1:** 美国《韦氏大辞典》指出: 伦理是一门探讨什么是好、什么是坏, 以及讨论道德责任义务的学科。

**定义 2:** 伦理一般是指一系列指导行为的观念, 是从概念角度上对道德现象的哲学思考。它不仅包含着对人与人、人与社会和人与自然之间关系处理中的行为规范, 也深刻地蕴涵着依照一定原则来规范行为的深刻道理。

**定义 3:** 所谓伦理, 是指人类社会中人与人之间, 人与社会、国家的关系和行为的秩序规范。任何持续影响全社会的团体行为或专业行为都有其内在特殊的伦理的要求。企业作为独立法人, 有其特定的生产经营行为, 也有企业伦理的要求。

**定义 4:** 伦理是指人们心目中认可的社会行为规范。伦理也对人与人之间的关系进行调整, 只是它调整的范围包括整个社会的范畴。管理与伦理有很强的内在联系和相关性。管理活动是人类社会活动的一种形式, 当然离不开伦理的规范作用。

**定义 5:** 伦理是指人与人相处的各种道德准则。生态伦理是伦理道德体系的一个分支, 是人们对一种环境价值观念认同的基础上, 维护生态环境的道德观念和行为规范。

**定义 6:** 伦理是指人与人相处的各种道德标准; 伦理学是关于道德的起源、发展, 人的行为准则和人与人之间的义务的学说。

### 3.1.2 道德的概念

道德是调整人们相互关系的行为规范的总和。所谓“道”, 是万物万法之源, 创造一切的力量, 而“德”是为顺应自然、社会和人类客观需要去做事的行为, 是不违背自然发展规律, 去发展自然、发展社会, 提升自己的践行方式。道是在承载一切, 德是在昭示道的一切, 德是道的具体实例, 也是道的体现。大道无言无形, 看不见、听不到、摸不着, 只有通过人们的思维意识去认识和感知它。如果没有德, 人们就不能如此形象地了解道的理念, 这就是德与道的关系。

不同的对错标准是特定生产能力、生产关系和生活形态下自然形成的。道德可以是源自于特定哲学、宗教或文化的行为准则中衍生出来的一系列标准或原则, 也可以源自于一个人所相信的普遍价值。

一些研究认为, 对道德情操的注重, 存在于所有的人类社会当中, 道德情操是普世文化通则的一部分; 而一些研究更认为, 诚实、助人、宽容、忠诚、责任、社会公正、平等、家庭与国家安全、社会秩序的稳定、报恩等和道德相关的行为, 是普世价值的一部分, 也就是说, 这些行为可能是所有社会普遍认可的德行。

### 3.1.3 伦理是一种自然法则

伦理是一种有关人类关系(尤其以姻亲关系为重心)的自然法则(图 3-2), 这个概念也是道德和法律的绝对分界线。道德是人类对于人类关系和行为的柔性规定, 这种柔性规定是以伦理为大致范本, 但又不同于伦理, 甚至经常与伦理相悖。法律则是人类对于人类关系和行为的刚性规定, 这种刚性规定是以法理为基础原则的。

在现实生活中, “伦理”二字与“道德”二字常常会一起出现。若要严格加以区分, 则伦理偏重于社会层面, 道德偏重于个人层面。在一般使用上, 二者经常被视为同义词, 有时

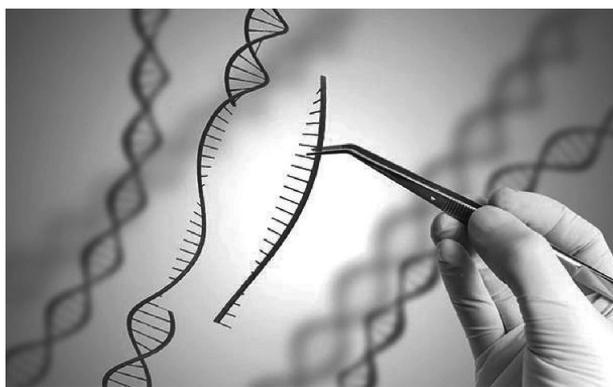


图 3-2 伦理是一种自然法则

更被连用为“伦理道德”一词。伦理道德专属于人文世界的范围,它表现为人类加诸于彼此及自身的规范与评价。

关于“道德”,老子说:“道可道,非常道。”意思是说,“道”并非指的是一条具体的道路,而是一个抽象出来的概念,譬如几何学上的“点、线、面”的概念,物理学上的“质点”的概念。那么“道德”,就是指走路的德行,类似于约定俗成的交通秩序,引申为人在社会上为人处世的规则。

伦理与道德都在一定程度上起到了调节社会成员之间相互关系的规则的作用。规则是为现实的存在不被破坏服务的,它本身并不倡导创新,甚至在一定程度上束缚了创新,而规则与创新的矛盾无一不是以创新的成功和规则的被打破之后形成新的规则而结束的。

随着社会所处的阶段乃至文化环境的不同,“道德”有着不同的规范。例如,在古代氏族部落中,财产是共有的,保留私有财产是不道德的,而拿走其他部落成员刚刚用过的工具也没有什么不道德。而在法律认可私有财产的现代社会,保留私有财产不再是不道德的,而拿走他人的工具则要征求他人的意见。又如,在中国,问别人的年龄和薪水是很正常的,似乎还有点人文关怀的味道;而在西方某些国家,打听年龄和薪水是不太道德的。

### 3.1.4 伦理学研究

伦理学以道德现象为研究对象,不仅包括道德意识现象(如个人的道德情感等),还包括道德活动现象(如道德行为)以及道德规范现象等。伦理学将道德现象从人类活动中区分开,探讨道德的本质、起源和发展,道德水平同物质生活水平之间的关系,道德的最高原则和道德评价的标准,道德规范体系,道德的教育和修养,人生的意义、人的价值和生活态度等问题。其中,最重要的是道德与经济利益和物质生活的关系、个人利益与整体利益的关系问题。对这些问题的不同回答,形成了不同的甚至相互对立的伦理学派别。马克思主义伦理学将道德作为社会历史现象加以研究,着重研究道德现象中的带有普遍性和根本性的问题,从中揭示道德的发展规律。

## 3.2 科技伦理造福人类

2019年7月24日,中央全面深化改革委员会第九次会议审议通过了诸多重要文件,其中《国家科技伦理委员会组建方案》排在首位通过。这表明中央将科技伦理建设作为推进国家科技创新体系不可或缺的重要组成部分。组建国家科技伦理委员会的要旨在于,抓紧完善制度规范,健全治理机制,强化伦理监管,细化相关法律法规和伦理审查规则,规范各类科学研究活动。

### 3.2.1 科技伦理是理性的产物

科技伦理(图 3-3)是指在科学技术创新与运用活动中,人与社会、人与自然以及人与人关系的思想与行为准则的道德标准和行为准则,是一种观念与概念上的道德哲学思考。它规定了科学技术共同体应遵守的价值观、行为规范和社会责任范畴。人类科学技术的不断进步,也带来了一些新的科技伦理问题,因此,只有不断丰富科技伦理这一基本概念的内涵,才能有效应对和处理新的伦理问题,提高科学技术行为的合法性和正当性。如果把人类文明的演化当作一个永无止境的征程,人类奔向更高文明的原动力就是科技和创新。但是,仅有动力还不够,还必须能识别方向,科技伦理就是指引科技造福人类的导航仪。



图 3-3 科技伦理

科技伦理是理性的产物。最根本的理性是,要求科技创新和成果只能有益于或最大程度地有益于人、生物和环境,而不能损伤人、损害生物和破坏环境,即便不可避免地会不同程度地损人毁物——如药物的副作用,也要把这种副作用减少到最低,甚至为零。在具体的伦理规则上,还应两利相权取其大、两害相权衡择其轻。

科技伦理不只是涉及科学研究中的伦理,也不只是科研人员要遵守科技伦理,还包括科技成果应用中的伦理,例如手机 App 下载的同意条款和医院治病时的知情同意等。科技伦理最早起源于人类的生活,今天有了更多更新的内容,应对今天科技创新所带来的诸多挑战,需要更多、更细的科技伦理来规范科研行为和科研成果的使用。

### 3.2.2 科技伦理的预见性和探索性

提出和遵循科技伦理不仅有益于所有人,也有利于生态和环境。尽管人是理性的,并因此诞生了科技伦理,但人类也有一些非理性思维和行动,也因此历史上产生了一些违背科技伦理的非理性行为,甚至是兽性和反人类的行为。在今天,这样的危险并未消除。

第二次世界大战时期,纳粹德军和日军用活人(俘虏)做试验,既违背了科技伦理,更犯下了残害人类和反人类的罪行(图 3-4)。尽管人体活体试验获得了一些科学数据和原理,但建立在伤害人、毁灭生命之上的科学研究是绝对不能为人类社会所接受的。因此,二战后的纽伦堡审判产生了《纽伦堡法典》(1946 年)。1975 年,第 29 届世界医学大会又修订了《赫尔辛基宣言》,以完善和补充《纽伦堡法典》。1982 年,世界卫生组织(WHO)和国际医学科学组织理事会(CIOMS)联合发表了《人体生物医学研究国际指南》,对《赫尔辛基宣言》进行了详尽解释。再到 1993 年,WHO 和 CIOMS 联合发表了《伦理学与人体研究国际指南》和《人体研究国际伦理学指南》。2002 年,WHO 和 CIOMS 修改制定了《涉及人的生物医学研究国际伦理准则》,提出了需要遵守的 21 项准则,体现了生命伦理的知情同意、生命价值、有利无伤原则。



图 3-4 二战后审判对中国实行细菌战的日本罪犯

当科技创新成为今天人类最重要的活动,以及人类需要科技创新才能快速和有效地推动人类文明向更高阶段发展之时,科技伦理才有了大量的新范畴、新内容和新进展。人类基因组和基因编辑、人工生命和合成生命、人工智能、5G/6G 通信技术、机器人、脑机接口、人脸识别、纳米技术、辅助生殖技术、精准医疗等,都是今天科技创新和科技研发的新领域,也关系到所有人的福祉。但另一方面,它们也可能会伤害人,甚至让人类走向灾难和毁灭,如此,科技伦理的导航和规范作用就极为重要和显著。

因此,科技伦理需要有预见性和探索性,在一项研究和一个行业发展到一定规模和程度时,必须要求有相适应的科技伦理来规范。同时,由于人的非理性和逐利性,也导致今天人们在进行科技创新活动和科学研究时既可能违背已有的伦理原则,还可能因为新的伦理原则尚未建立之时,在新旧之间、有规定和无规定的结合部打擦边球,产生有违人类

伦理的或争议极大的科研行为,以及科研成果的不当使用。

不久前一个涉及前沿科研的极具风险的研究就是如此。有研究人员认为,CCR5 基因是导致人被艾滋病病毒(HIV)感染的帮凶,于是在试验中对新生儿敲除了这一基因,以期永远预防艾滋病。这一科研的初衷也许是积极的,然而,由于伦理审查不严,导致这一研究存在巨大风险,既有可能违背既有的生命伦理四大原则——有利、尊重、公正和互助,也存在更大的实际风险。敲除 CCR5 基因固然可以预防艾滋病,但是它的免疫功能、抗癌功能等其他有益于人的作用也会随之完全消失。这实际上是因为并不了解 CCR5 基因的全面用途而导致的伦理审查失责。

由此,可以看出制定各个学科和多学科研究及成果应用的伦理规范有多么重要和迫切。

### 3.3 技术伦理

技术伦理是 20 世纪后期新兴的一门以探讨如何认识和约束技术发展带来的社会问题的学科,主要讲授现代科技提出的伦理问题,如科技共同体内的伦理问题、科技时代中人与自然的伦理问题、安乐死问题、克隆人问题等。

计算机技术串联并融合了现实空间和虚拟空间,具有广泛的社会意义和深层的伦理意蕴,为人们提供了一个人一机关系、技术—社会双向形塑的伦理分析范型。人们可以深入分析计算机技术,以此为例探寻技术伦理学的时代性和未来性。

20 世纪 50 年代,控制论创始人罗伯特·维纳向世人提醒信息技术对社会构成的威胁,提出应对新技术的讨论提高到道德认识的层面,由此奠定了计算机伦理学的基础。20 世纪 70 年代,美国计算机专家沃尔特·曼纳注意到计算机伦理问题日益突出,提出研究这些问题的领域应当成为应用伦理学的一个独立分支,命名为“计算机伦理学”,并将计算机伦理学界定为研究计算机技术引发、改变和加剧伦理问题的应用伦理学科。1984 年是计算机伦理学发展的分水岭,美国计算机伦理学家詹姆士·摩尔和黛博拉·约翰逊提出“真空说”,认为计算机伦理学是一门全新的伦理学,因为计算机技术具有以往技术不具有的逻辑延展性,这一特性导致了理论的含混和政策的真空。从前的伦理学理论无法回答计算机技术提出的挑战,需要建立一门全新的伦理学来应对。

1966 年,美国麻省理工学院计算机专家韦曾鲍姆编写了一个名为 ELIZA(艾丽莎)的心理疗法计算机程序。这个程序表明,计算机能够进行自动化的心理治疗。韦曾鲍姆担忧人类“信息处理模式”会增强科学家甚至普通公众把人仅仅看作机器的倾向,认为人工智能的滥用可能损害人类的价值。1988 年,IBM 公司的雷蒙·巴尔金提出,如果机器人最终与人难以区分,那么必须制定伦理行为规范来调整真实的人与“人工的人”之间的关系。他编撰了“赛博伦理学”一词,用以概括这一研究领域。巴尔金提出的这条路径代表计算机伦理学的机器人伦理学线索。随着互联网的崛起和商业化普及,网络伦理问题获得空前的关注。斯皮内洛、塔瓦尼等众多学者相继出版了大量网络伦理学著作,在 20 世纪 90 年代形成了网络伦理学的研究热潮。

计算技术不只是延长或代替人脑,更重要的是促成万物互联。多种技术如计算机技

术、基因技术、纳米技术、人工智能技术、能源技术等融合,将加快万物互联的进程。技术的融合将促进人机的融合,促成虚拟空间和现实空间的融合,使人们难以区分身处其中的这两类空间,难以区分身体与人工物,并最终导致各类技术伦理学的融合。

技术与人类未来、人与技术的自由关系是技术时代伦理学乃至整个哲学探寻的核心。在不同的技术时代,人与技术的自由关系问题聚焦在不同的内容上。在机器大工业时代,它聚焦于人与机器的自由关系;在当今互联网、大数据和人工智能时代,它聚焦于人与信息、人与数据、人与自主机器的自由关系。人与技术的关系,本质上是人与人之间的关系,是基于技术的人与人之间的关系。如何确保人与人之间的自由关系,确保人类的未来,便成了技术时代伦理学探寻的终极目标。

### 3.4 工程伦理

发生于 1907 年的加拿大魁北克大桥事件(图 3-5),号称是这个世纪最大的技术失误之一。这座大桥原本应该是美国著名设计师特奥多罗·库帕的一个有价值的杰作。库帕曾称他的设计是“最佳、最省的”,可惜它最终并没有架成。库帕自我陶醉于他的设计,而忘乎所以地把大桥的长度由原来的 500m 加到 600m,以成为当时世界上最长的桥。桥的建设速度很快,施工组织也很完善。正当投资修建的人士开始考虑如何为大桥剪彩时,人们忽然听到一阵震耳欲聋的巨响——大桥的整个金属结构垮了,19 000t 钢材和 86 名建桥工人落入水中,只有 11 人生还。由于库帕的过分自信而忽略了对桥梁重量的精确计算,导致了一场悲剧的发生。



图 3-5 1907 年的魁北克大桥(左)和垮塌的魁北克大桥(右)

近几年来,一批环境污染事件不断被曝光。在经济利益的驱动下,牺牲环境导致的每一个事件背后,都暴露出了工程项目决策者和实践者在趋利心态下的错误行动,值得思考。

20 世纪 70 年代,西方一些发达国家在工业革命的进程中也面临类似的环境污染和安全事故,有些事故甚至危及人类的生存和发展。1986 年,因 O 型环密封圈失效而导致的美国“挑战号”航天飞机灾难事件震惊世界。事后发现,在决策中无视已知的缺陷,忽视工程师提出的低温下发射具有危险性的警告,是导致这次事件的关键因素。在应对这些挑战和压力的过程中,西方发达国家发现并开始开展工程伦理教育,并将其作为未来工程师所必备的基本素质。

所谓“工程伦理”，就是关于对工程负责的态度，无论是一个人还是一个团队，对所参与的工程必须要有负责任的态度。工程伦理至少有两个层面的含义，一是工程项目内在的伦理，即工程的伦理准则；二是工程项目核心实施者之一的工程师的职业伦理，即工程师的伦理准则。

随着现代工程技术的发展，工程决策与实践中的伦理冲突不断出现。例如，医学上的“转基因工程和换头术”产生的生命伦理问题，化学与化学工程的发展带来的抗生素问题、环境激素问题等新的伦理冲突。这些冲突，大致上蕴含着两类问题。一是工程本身是否可能带来近期的或长期的环境影响或生态破坏；二是工程决策时决策者、设计者和实施者都承担着怎样的伦理角色。

伦理决策和价值选择对于社会的可持续发展至关重要。因此，工程伦理教育应该是全过程、全方位的教育。培养具有“伦理意识”的现代工程师，以造福人类和可持续发展为理念的工程师，才能在面临着忠诚于股东还是公众的利益冲突等道德困境时做出正确的判断和选择。工程师应该掌握风险辨识和评价的基本方法，具备基于长期利润与道德平衡而进行工程决策的能力。

## 3.5 算法歧视

算法是信息技术，尤其是大数据、人工智能的基础。算法就是一系列指令，告诉计算机该做什么。算法的核心是按照设定程序运行，以期获得理想结果的一套指令。所有的算法都包括以下几个共同的基本特征：输入、输出、明确性、有限性、有效性。

算法因数学而起，但现代算法的应用范畴早已超出了数学计算的范围，已经与每个人的生活息息相关。因此，我们生活在算法的时代。随着人工智能时代的到来，算法越来越多地支配着人们的生活，也给现存的法律制度和法律秩序带来了冲击和挑战。

### 3.5.1 算法透明之争

“黑箱”是控制论中的概念。作为一种隐喻，它指的是那些不为人知的不能打开、不能从外部直接观察其内部状态的系统。人工智能所依赖的深度学习技术就是一个“黑箱”。深度学习是由计算机直接从事物的原始特征出发，自动学习和生成高级的认知结果。在人工智能系统输入的数据和其输出的结果之间，存在着人们无法洞悉的“隐层”，这就是“算法黑箱”。对透明的追求使人心理安定，“黑箱”使人恐惧。如何规制算法“黑箱”，算法（图 3-6）是否要透明，如何透明，是法律规制遇到的首要问题。

面对算法黑箱，不少人主张、呼吁算法透明。其理由主要有以下几点。

(1) 算法透明是消费者知情权的组成部分。这种观点主张，因为算法的复杂性和专业性，人工智能具体应用领域中的信息不对称可能会更加严重，算法透明应是消费者知情权的组成部分。

(2) 算法透明有助于缓解信息不对称。这种观点主张，算法的信息不对称加重不只发生在消费者与算法设计者、使用者之间，更发生在人类与机器之间，算法透明有助于缓解这种信息不对称。

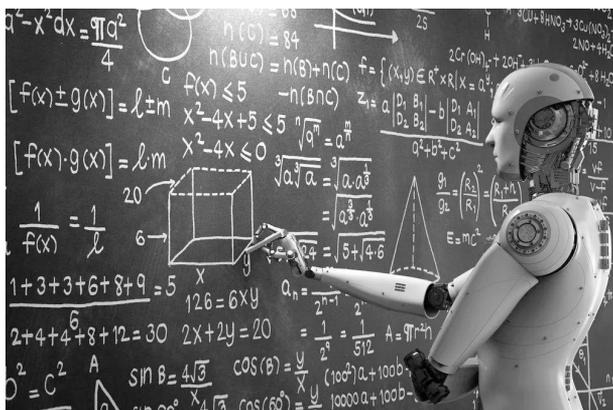


图 3-6 算法

(3) 算法透明有助于防止人为不当干预。这种观点以智能投顾为例,认为算法模型是公开的,在双方约定投资策略的前提下,执行策略由时间和事件函数共同触发,执行则由计算机程序自动完成,避免了人为不当干预的风险,它比人为干预更加公平、公开和公正。

(4) 算法透明有助于防止利益冲突。这种观点认为由于算法的非公开性和复杂性,难以保证诸如投资建议的独立性和客观性。只有算法透明,才能防止这种利益冲突。

(5) 算法透明有助于防范信息茧房。这种观点认为,算法可能形成信息茧房。算法科学的外表容易误导使用者,强化使用者的偏见,从而导致错误决策。例如,算法技术为原本和普罗众生疏离的复杂难懂的金融披上了简单易懂的面纱,金融的高风险性被成功掩盖,轻松化的人机交互界面掩盖了金融风险的残酷本质。

(6) 算法透明有助于打破技术中立的外衣。事实上,技术的背后是人,人类会将人性弱点和道德缺陷带进和嵌入算法中,但它们却可能隐蔽于算法背后,从而更不易被发觉。

(7) 算法透明有助于打破算法歧视。宾夕法尼亚州法学院的汤姆·贝克和荷兰鹿特丹伊拉斯谟大学的本尼迪克特 G. C.德拉特教授认为:公众不能预设机器人没有人类所具有的不纯动机。因为算法存在歧视和“黑箱”现象,因此才需要算法的透明性或解释性机制。

(8) 算法透明有助于打破“算法监狱”与“算法暴政”。在人工智能时代,商业企业和公权部门都采用人工智能算法做出自动化决策,算法存在的缺陷和偏见可能会使大量的客户不能获得贷款、保险、承租房屋等服务,这如同被囚禁在“算法监狱”。然而,如果自动化决策的算法不透明,不接受人们的质询,不提供任何解释,不对客户或相对人进行救济,客户或相对人无从知晓自动化决策的原因,自动化决策就会缺少“改正”的机会,这种情况就属于“算法暴政”。算法透明则有助于打破“算法监狱”和“算法暴政”。

(9) 算法透明是提供算法可责性问题的解决工具和前提。有学者认为,算法透明性和可解释性是解决算法可归责性的重要工具。明确算法决策的主体性、因果性或相关性,是确定和分配算法责任的前提。

(10) 算法透明有助于提高人们的参与度,确保质疑精神。这种观点认为,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。由于人们无法看清其中的规则和决定过程,就无法提出不同的意见,也不能参与决策的过程,只能接受最终的结果。为走出这一困境,算法透明是必要的。还有人认为,质疑精神是人类前进的工具,如果没有质疑,就没有社会进步。为了保证人类的质疑,算法必须公开,除非有更强的不公开的理由,如保护国家安全或个人隐私。

(11) 算法公开透明(图 3-7)是确保人工智能研发、设计、应用不偏离正确轨道的关键。这种观点认为,人工智能的发展一日千里,人工智能可能拥有超越人类的超级优势,甚至可能产生灾难性风险,因而应该坚持公开透明原则,将人工智能的研发、设计和应用置于监管机构、伦理委员会以及社会公众的监督之下,确保人工智能机器人处于可理解、可解释、可预测状态。



图 3-7 算法公开透明

现实中反对算法透明的声音也不少,其主要理由如下。

(1) 类比征信评分系统。征信评分系统不对外公开是国际惯例,其目的是防止“炒信”“刷信”,使评级结果失真。很多人工智能系统类似于信用评级系统。

(2) 周边定律。周边定律是指法律无须要求律师提请人们注意身边具有法律意义的内容,而是将其直接植入人们的设备和周边环境,并由这些设备和环境付诸实施。主张该观点的人宣称,人类正在步入技术对人类的理解越来越深刻,而人类却无须理解技术的时代。智能时代的设备、程序,就像人体器官和中枢神经系统,人们对其知之甚少但却可以使用它们。同样,算法为自我管理、自我配置和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。

(3) 算法不透明有助于减少麻烦。如果披露了算法,则可能引起社会舆论的哗然反应,从而干扰算法的设计,降低预测的准确性。尽管大数据预测准确的概率较高,但也不能做到百分之百。换言之,大数据预测也会不准,也会失误。如果将算法公之于众,人们对预测错误的赋值权重就有可能偏大,从而会阻碍技术的发展。

(4) 防止算法趋同。算法披露之后,好的算法、收益率高的算法、行业领导者的算法可能会引起业界的效仿,从而出现“羊群效应”,加大顺周期的风险。

(5) 信息过载或难以理解。算法属于计算机语言,不属于日常语言,即使对外披露了,除专业人士之外的大多数客户也难以理解。换言之,对外披露的信息对于大多数用户来讲可能属于无效信息。

(6) 偏见存在于人类决策的方方面面,要求算法满足高于人类的标准是不合理的。算法透明性本身并不能解决固有的偏见问题。要求算法的透明性或者可解释性,将会减损已申请专利的软件的价值。要求算法的透明性还为动机不良者扰乱系统和利用算法驱动的平台提供机会,这将使动机不良者更容易操纵算法。

(7) 算法披露在现实中存在操作困难。可能涉及多个算法,披露哪个或哪些算法? 算法披露到什么程度?

折中派的观点认为,算法是一种商业秘密。算法由编程者设计,进而给网站带来巨大的商业价值,因此其本质上是具有商业秘密属性的智力财产。如果将自己的专有算法程序公之于众,则有可能泄露商业秘密,使自己丧失技术竞争优势。鉴于很多算法属于涉及商业利益的专有算法,受知识产权法保护,因此即使是强制要求算法透明,也只能是有限度的透明。

还有人认为,如何对待算法,这个问题并没有“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,它有助于帮助公众了解决策是如何形成的。但是在涉及国家安全时,公开源代码的做法就不适用,因为一旦公开特定黑盒子的内部运行机制,某些人就可以绕过保密系统,使算法失效。

### 3.5.2 算法透明的实践

2017年,美国计算机学会公众政策委员会公布了6项算法治理指导原则。

第一个原则是知情原则,即算法设计者、架构师、控制方以及其他利益相关者应该披露算法设计、执行、使用过程中可能存在的偏见以及可能对个人和社会造成的潜在危害。

第二个原则是质询和申诉原则,即监管部门应该确保受到算法决策负面影响的个人或组织享有对算法进行质疑并申诉的权力。

第三个原则是算法责任认定原则。

第四个原则是解释原则,即采用算法自动化决策的机构有义务解释算法运行原理以及算法具体决策结果。

第五个原则是数据来源披露原则。

第六个原则是可审计原则。

仔细审视这6项原则,其要求的算法透明的具体内容主要是算法的偏见与危害、算法运行原理以及算法具体决策结果,以及数据来源。

2017年年底,纽约州通过一项《算法问责法案》,要求成立一个由自动化决策系统专家和相应的公民组织代表组成的工作组,专门监督自动化决策算法的公平性和透明性。之前,该法案有一个更彻底的版本,规定市政机构要公布所有用于“追踪服务”或“对人施加惩罚或维护治安”的算法的源代码,并让它们接受公众的“自我测试”。这是一份精练的、引人入胜的而且是富有雄心的法案,它提议每当市政府机构打算使用自动化系统来配置警务、处罚或服务时,该机构应将源代码——系统的内部运行方式——向公众开放。很

快,人们发现这个版本的法案是一个很难成功的方案,他们希望不要进展得那么激进。因此,最终通过的法案删去了原始草案中的披露要求,设立了一个事实调查工作组来代替有关披露的提议,原始草案中的要求仅在最终版本中有一处间接地提及——“在适当的情况下,技术信息应当向公众开放”。

在欧盟,《通用数据保护条例》第 71 条规定:“在任何情况下,该等处理应该采取适当的保障,包括向数据主体提供具体信息,以及获得人为干预的权利,以表达数据主体的观点,在评估后获得决定解释权,并质疑该决定。”据此,有人主张该条例赋予了人们算法解释权。但也有学者认为,这种看法很牵强,个人的可解释权并不成立。

我国《新一代人工智能发展规划》指出:“建立健全公开透明的人工智能监管体系”。这提出了人工智能监管体系的透明性,而没有要求算法本身的透明性。

### 3.5.3 算法透明的算法说明

人们呼吁算法透明,但透明的具体内容是算法的源代码,还是算法的简要说明?秉承“算法公开是例外,不公开是原则”的立场,即使是在算法需要公开的场合,也需要考察算法公开的具体内容是什么。

算法的披露应以保护用户权利为必要。算法的源代码、算法的具体编程公式(实际上也不存在这样的编程公式)是不能公开的。这主要是因为一方面算法的源代码非常复杂,且不断迭代升级,甚至不可追溯,无法予以披露;另一方面,公开源代码是专业术语,绝大部分客户看不懂,即使公开了也没有意义。

算法透明追求的是算法的简要说明(简称算法简介)。算法简介包括算法的假设和限制、算法的逻辑、算法的种类、算法的功能、算法的设计者、算法的风险、算法的重大变化等。算法简介的公开也是需要法律规定的;否则,不公开仍是基本原则。

### 3.5.4 算法透明的替代方法

算法透明的具体方法,除了公开披露之外,还可以有其他替代方法。这些方法究竟是替代方法还是辅助方法,取决于立法者的决断。

(1) 备案或注册。备案即要求义务人向监管机构或自律组织备案其算法或算法逻辑,算法或算法逻辑不向社会公开,但监管机构或自律组织应知悉。

算法很复杂,很难用公式或可见的形式表达出来。算法的种类很多,一个人工智能系统可能会涉及很多算法,且算法也在不断迭代、更新和打补丁,就像其他软件系统不断更新一样。因此,算法本身没法备案,更无法披露。可以备案和披露的是算法的逻辑和参数。除了算法逻辑的备案以外,还可以要求算法开发设计人员的注册。

(2) 算法可解释权。一旦人工智能系统被用于做出影响人们生活的决策,人们就有必要了解人工智能是如何做出这些决策的。方法之一是提供解释说明,包括提供人工智能系统如何运行以及如何与数据进行交互的背景信息。但仅发布人工智能系统的算法很难实现有意义的透明,因为诸如深度神经网络之类的最新的人工智能技术通常是没有任何算法输出可以帮助人们了解系统所发现的细微模式。基于此,一些机构正在开发建立有意义的透明的最佳实践规范,包括以更易理解的方法、算法或模型来代替那些过于复杂

且难以解释的方法。笔者认为,是否赋予客户以算法可解释权有待深入论证,但算法设计者有义务向公权机关解释算法的逻辑。

### 3.5.5 算法公平的保障措施

算法公开、算法备案等规制工具都属于信息规制工具,它们是形式性规制工具。除了信息规制工具之外,还有其他实质性规制工具。形式性规制工具追求的价值目标是形式公平,实质性规制工具追求的价值目标是实质公平。在消费者权益和投资者权益保护过程中,除了保障形式公平之外,也要保障实质公平。因此,除了信息规制工具之外,还应有保障算法公平的其他实质性规制工具,这些工具主要包括三方面,一是算法审查、评估与测试,二是算法治理,三是第三方监管。

(1) 算法审查、评估与测试。在人工智能时代,算法主导着人们的生活。数据应用助推数字经济,但也有许多模型把人类的偏见、误解和偏爱编入了软件系统,而这些系统正日益在更大程度上操控着我们的生活。“只有该领域的数学家和计算机科学家才明白该模型是如何运作的。”人们对模型得出的结论毫无争议,从不上诉,即使结论是错误的或有危害的。凯西·奥尼尔将其称为“数学杀伤性武器”。

然而,数学家和计算机科学家应当接受社会的审查。算法是人类的工具,而不是人类的主人。数学家和计算机科学家是人类的一员,他们应与普罗大众处于平等的地位,而不应凌驾于人类之上,他们不应是人类的统治者。即使是人类的统治者——君主或总统,在现代社会也应接受法律的规范和治理、人民的监督和制约,更何况群体庞大的数学家和计算机科学家。总之,算法应该接受审查。

“算法黑箱”吸入数据,吐出结论,其公平性应接受人类的审查。算法的开发者、设计者也有义务确保算法的公平性。

对人工智能系统应该进行测试。目前人工智能机器人尚未成为独立的民事主体,不能独立承担民事责任,但这并不妨碍对其颁发合格证书和营运证书。这正如汽车可以获得行驶证和营运许可证书一样。

(2) 算法治理。质疑精神是人类社会前进的基本动力,必须将算法置于人类的质疑和掌控之下。人工智能的开发商和运营者应有能力理解和控制人工智能系统,而不能单纯地一味依赖于第三方软件开发者。

人工智能系统还应建立强大的反馈机制,以使用户轻松报告遇到的性能问题。任何系统都需要不断迭代和优化,只有建立反馈机制,才能更好地不断改进该系统。

(3) 加强第三方算法监管力量。为了保证对算法权力的全方位监督,应支持学术性组织和非营利机构的适当介入,加强第三方监管力量。目前,德国已经出现了由技术专家和资深媒体人挑头成立的名为“监控算法”的非营利组织,宗旨是评估并监控影响公共生活的算法决策过程。具体的监管手段包括审核访问协议的严密性、商定数字管理的道德准则、任命专人监管信息、在线跟踪个人信息再次使用的情况,允许用户不提供个人数据、为数据访问设置时间轴、未经同意不得将数据转卖给第三方等。这种做法值得我国借鉴。为了让人工智能算法去除偏私,在设计算法时,对相关主题具有专业知识的人(例如,对信用评分人工智能系统具有消费者信用专业知识的人员)应该参与人工智能的设计过程和



- C. 因缺乏工程伦理章程而导致的问题  
D. 是工程项目核心实施者之一的工程师的职业伦理,即工程师的伦理准则
11. ( )不属于工程教育的核心内容。  
A. 忠诚于股东    B. 意识与责任    C. 明辨是非    D. 先觉先知
12. 算法的核心就是按照设定程序运行,以期获得理想结果的一套指令。所有的算法都包括输入、输出以及( )这样几个共同的基本特征。  
① 明确性    ② 有限性    ③ 低成本    ④ 有效性  
A. ①②③    B. ②③④    C. ①③④    D. ①②④
13. 我们生活在算法的时代。“算法黑箱”是指信息技术算法的( )。  
① 不为人知    ② 不公开    ③ 不透明    ④ 不值钱  
A. ②③    B. ①②    C. ①④    D. ③④
14. 面对“算法黑箱”,不少人主张、呼吁算法透明。其理由很多,其中包括( )。  
① 算法透明是消费者知情权的组成部分,有助于打破算法歧视  
② 算法透明有助于缓解信息不对称、防止利益冲突  
③ 算法透明有助于防止人为不当干预  
④ 算法透明有助于参考复制,以降低程序设计的作业成本  
A. ①③④    B. ①②④    C. ①②③    D. ②③④
15. 现实中反对算法透明的声音也不少,其主要理由包括( )。  
① 有助于减少麻烦    ② 防止算法趋同  
③ 信息过载或难以理解    ④ 希望技术垄断  
A. ①③④    B. ①②④    C. ②③④    D. ①②③
16. 2017年,美国计算机学会公众政策委员会公布了6项算法治理指导原则,如知情原则、质询和申诉原则、算法责任认定原则,还包括( )。  
① 免责原则    ② 解释原则  
③ 可审计原则    ④ 数据来源披露原则  
A. ②③④    B. ①②④    C. ①③④    D. ①②③
17. 算法治理指导原则中的( )原则,即算法设计者、架构师、控制方以及其他利益相关者应该披露算法设计、执行、使用过程中可能存在的偏见以及对个人和社会造成的潜在危害。  
A. 知情    B. 利润    C. 解释    D. 质询和申诉
18. 算法治理指导原则中的( )原则,即采用算法自动化决策的机构有义务释疑算法运行原理以及算法具体决策结果。  
A. 知情    B. 利润    C. 解释    D. 质询和申诉
19. 算法治理指导原则中的( )原则,即监管部门应该确保受到算法决策负面影响的个人或组织享有对算法进行质疑并申诉的权力。  
A. 知情    B. 利润    C. 解释    D. 质询和申诉

20. 人们呼吁算法透明,要追求透明的应该是( )。

- A. 算法源代码    B. 算法说明    C. 算法图形    D. 算法所有人

### 【研究性学习】 辩论: 算法是否应该透明

**小组活动:** 通过讨论,深入了解有关算法透明之争的内涵、不同的观点及其对信息技术发展的影响。

**正方观点:** 算法应该透明。面对“算法黑箱”,主张和呼吁算法透明。

**反方观点:** 算法不能透明。偏见存在于人类决策的方方面面,要求算法满足高于人类的标准是不合理的。

**记录:** 请记录小组讨论的主要观点,推选代表在课堂上简单阐述你们的观点。

本小组担任:  正方辩手     反方辩手

本小组的基本观点是: \_\_\_\_\_

---

---

---

---

---

---

---

---

**评分规则:** 若小组汇报得 5 分,则小组汇报代表得 5 分,其余同学得 4 分,其余类推。

实验评价(教师)

---

---