

认真对待个人信息保护的社会结构： AI 科技时代的不对称性挑战

杨 建

内容提要：《个人信息保护法》的出台要有效回应 AI 科技应用带来的个体与共同生活的双重改变。目前学界对人工智能技术冲击下的社会结构的阐述存在不少缺陷。AI 科技的应用具有道德镜像属性与社会工具属性，它构成了社会大挑战。我们共同生活的架构以及诸多价值理念没有被明显消解，但其真实的运作已经被重塑和改变。功利主义和现实主义式论调认为，类似的结构重塑和改变由于在结果是好的，所以并没有什么不对。这种论调既是反事实的，在价值上又无法证立。在隐默且不当的不对称性结构之下，AI 科技构成社会挑战的关键在于：它改变了民主关系中的权力支配状态，进而使得强制力行使的正当性不足。科技公司与政府担负着重新宣誓忠实、重新赢取公众信任的道德与政治义务。AI 科技时代的法律实践需要切实回应社会结构中的不对称性挑战，并被这一回应的成效所检验。

关键词：个人信息 人工智能 不对称性 科技权力 公共法理

杨建，南京师范大学法学院副教授。

《个人信息保护法》以极快的速度在中国落地生效，展现出中国对这一问题的理解和重视程度：它不只关乎个体的权益保护，也关乎主要国家之间的政治竞争，还关乎长久经济利益的占有与分配。这主要是因为信息科技的发展以及经济、社会交往方式的转变，使得个人信息在社会结构中的地位与功能发生了彻底的改变。从法学立基的权利视角来看，这些转变触发了对人格尊严、隐私、意志自治、民主权利、经济平等、社会进步等价值何以保障和平衡的顾虑与吁求。这些价值诉求殊途同归，最终均落脚在个人权益之上，追问着在人工智能普遍应用的时代继续维持、保障和促进个体美好生活的可能性与可行性。对这些追问的成功回答部分取决于我们对上述实践的实践语境——AI 科技全面塑造了和持续塑造着的社会结构的——准确理解与把握。

本文即是力求对该实践语境提供一套全面的、具有竞争力的说明。本文认为，人工智

能技术在目前以及未来很长的一段时间之内,仍然只是主体关系之中的媒介与手段,其带来的社会问题的核心并不像很多人呼吁的那样,是要严肃对待和处理人与智能系统之间的关系。^[1]问题的核心仍在人际主体之间,是要严肃对待和处理由人工智能技术带来的人际主体间关系——特别是其中的权力支配关系——的变化,严肃对待和处理信息新科技更全面地揭露出来的,乃至不对称地强化了主体之间、社会之中的结构性难题与道德弊病。本文将具体说明,为何现有对人工智能之挑战的理解是欠缺现实性的,“人与人工智能的关系”为何不是一个恰当的分析视角。在此基础上,本文以“被冲击下的人际关系”为视角,以“不对称性”为分析概念,呈现当下人工智能最具法理性质的社会挑战:它改变了民主关系中的权力支配状态,进而使得强制力行使的正当性不足,科技巨头与政府需要重新宣誓忠实,重新赢得公众的信任。

一 AI 科技形成挑战了吗?

探讨 AI 科技构成了怎样的挑战有一个不假思索的前提,即这类信息新科技的确构成了挑战。质疑这一点似乎既是反直觉的也是反事实的,仿佛没有太多讨论的必要。但这一想法模糊了不同挑战类型之间的区别,也高估了挑战一词的清晰性。

(一) 构成挑战的条件

某项事物要想构成挑战,逻辑上看,需要具备这样几个条件:第一,是它带来的挑战,倘若这一挑战在它出现之前就存在,它的出现只是落在了这一挑战之中,对原有的问题也无实质的改变,我们便不能将这一挑战归诸于它;第二,它带来的挑战应当具有相当程度的重要性,倘若它确实带来了挑战,但该挑战实在无足轻重,我们便没有充足的理由关注它,又或者它的裨益与它的挑战相比,裨益的重要程度远大于挑战,以致该挑战可以被省略或搁置,那么学理或实践上重要性程度的降低也会消解在认知澄清上的优先次序;第三,由它带来或重塑的重要挑战,要么现有的学理、机制不知如何解决,要么根本无法解决,要么解决的难度非常之高,倘若现有的架构可以有效地或者大体上解决该问题,它也不构成真正的挑战。

AI 技术是否构成挑战便需要接受这三个方面的审视,本文也大致从这三个方面予以证明:第一,通过对 AI 科技应用之社会属性的说明,证明本文所归纳的挑战与 AI 科技之间具备直接关联;第二,通过对 AI 科技塑造的双重不对称性困境的描述,证明它带来的这一挑战是重要的;第三,通过对这一挑战性质的分析,证明这一挑战无法通过对策性的、阶段性的司法或立法应对来解决。在具体展开之前,本文第一部分先明确挑战的类型,以为第二部分和第三部分的理论推进确立基础。

(二) 两种类型的挑战

回到是否构成挑战的问题上来。一旦开始审视,便会遇到第一个困难,即需要明确,

[1] 参见陈景辉:《人工智能的法律挑战:应该从哪里开始》,《比较法研究》2018年第5期;杨庆峰:《从人工智能难题反思 AI 伦理原则》,《哲学分析》2020年第2期。

我们所讨论的是哪种类型的挑战？这样的追问之所以首要，是因为存在着两种指向不同的挑战：概念性的与实践性的，它们的应对方式与解决标准各不相同。

概念性分析本身存在一阶与二阶两个不同层面的划分，本文限定在一阶层面。这一层面的概念性挑战是说，出现了以往的概念无法明确涵盖的对象，或者它的出现动摇了既存概念观的有效性或说明力，它在概念层面挑战了我们的现有认知，并要求我们在学理上重新予以分析、澄清；实践性挑战是说，它在实践场域制造了需要认真面对的道德困境，而我们不知道如何应对，或者就如何行动存在重大分歧，进而需要提出某个竞争性的实践方案并为之辩护。概念性挑战譬如说，主张 AI 科技挑战了人们对人、自由、尊严等基础概念的固有理解；实践性挑战譬如说，认为 AI 科技制造了新的或更严重的社会不公正，侵犯了人们的自由、尊严等重要的个体价值。^[2]

可以看到，概念性挑战与实践性挑战可能关涉同一个对象，譬如说尊严，但即使关涉同一个对象，它们各自仍然构成不同类型的挑战。有关尊严的概念性挑战指向的是尊严的性质，是要重新去界定和回答对于尊严的认识和理解；有关尊严的实践性挑战指向的则是维护或实现尊严的价值这件事，对尊严的维系和保护面临困境，需要思考、提出并证成某个应对的实践方案。

（三）实践性分析具有独立的重要性

当意识到存在两种不同类型的挑战，我们便需要明确，本文选择处理的是 AI 科技的实践性挑战。就概念性分析与实践性分析之间的关系来说，概念性分析具有认知上的优先性，它也是法理学的首要工作。^[3] 从直觉上看，概念性分析与实践性分析之间肯定存在着关联。^[4] 譬如，如何理解尊严通常会对如何维护和实现尊严产生影响：如果我们对尊严的理解发生了改变——不再将尊严与官阶和贵族身份相关联，不再认为勇气是尊严最核心的构成要素，^[5] 我们对尊严的实践也就相应发生改变——不再认可男性在求婚的竞争中诉诸决斗是对尊严的维护。但概念性分析替代不了实践性分析，也无法直接用做

[2] 参见於兴中：《算法社会与人的秉性》，《中国法律评论》2018 年第 2 期，第 58 - 62 页。

[3] 法理学的首要工作是概念性分析，但概念性分析并不是给法学概念下定义。对于这个基础认知，see H. L. A. Hart, *The Concept of Law*, Clarendon Press, 1994.

[4] 根据迈克尔·史密斯(Michael Smith)在《道德问题》中的说明，大多数哲学家认为，二阶问题(元伦理学问题)要比一阶问题(规范伦理学问题)在道德哲学中具有理论上的优先性：只有先在二阶问题上形成妥当的答案，一阶问题才可能得到更为清晰、有效的呈现与理解。这一理论预设虽然并不是不存在挑战，但大体上也适用于当代法哲学的主流讨论。法哲学中的概念分析一般指向法律的性质问题，而如何进行法律解释与推理则属于规范的、实践性的问题。实践性议题与我们应当如何行动联系在一起，要从价值上给出一个具有规范性的结论，因此对应的是一阶问题；与法律的性质相关的概念性议题则与如何认识特定社会实践中的法律现象联系在一起，要在一般法理论层面为实践性议题中的法律给出具有清晰性的认知说明，是关于一阶理论的理论，因此对应的是二阶问题。本文指出，概念性分析本身也有一阶与二阶之分，即是说，在实践性问题的分析中其实也存在着概念性的分析，这一概念分析与元理论层面的概念分析即使指向同一个概念，在理论性质上也分属于不同的理论工作。对于厚概念以及一阶层面的概念性分析，它们与同一层面的实践性分析之间通常具有本文所理解的强关联。这种强关联并不当然能够延展至二阶问题层面的概念性分析，这一层面的概念性分析与一阶实践性分析之间的逻辑空间大到可以包容几乎任何概念性的与实质性的主张。在注意到这一点的基础上，本文想指出的是，无论二阶概念性分析与一阶实践性分析的理论关系如何，都无法推翻本文的核心立场：实践性分析与概念性分析不同，且具有相对的独立性，AI 科技的实践挑战是一个独立存在的重要问题。

[5] See Jeremy Waldron, *Dignity, Rank, & Rights*, Oxford University Press, 2012, pp. 30 - 36.

实践问题的答案,两类分析具有各自的特殊性与独立性。混用两种分析,或者将实践性分析还原为概念性分析,会减损分析的质量,并落入实践性分析最常见也最致命的困境:预设了真假参半因而任人打扮的事实性前提。

在明确文章所处理的挑战类型的基础上,接下来,我们便可以审视 AI 科技是否构成了实践性挑战这个问题。如前所述,本文力图证明,从挑战的来源、挑战的重要性程度、解决挑战的困难性这三个维度来看,AI 科技确实构成了重大的社会挑战。但面对这一重大挑战,学界的诸多认识存在偏颇之处,要么夸大其词,用想象代替现实;要么顾左右而言他,用次要或边缘情节代替核心的问题焦点,这些问题都有必要重视与深入分析。

二 一种错误的分析视角

对于 AI 科技造成的挑战,本文力图批判的“超越论”“取代论”“解构论”等观点均共享了一个错误的分析视角,这一分析视角之所以错误,是因为它的事实性前提是虚假的。尽管建立在虚假事实性前提之上的论证是有效的,但虚假的事实性前提无法排除并非真实的概念空间,进而使得在这些并不真实的概念空间中进行的诸多论证既不是真的可靠,又不是真的具有说服力。

(一) AI 科技的特殊性

具体分析对 AI 科技社会挑战的错误认知之前,需要注意,这些认知共享了一个本文也同意的观点,那就是 AI 科技具有特殊性。AI 科技最突出的不同,在于它具备了强大的学习能力。这种强大的学习能力主要表现在两大方面:远超出人的记忆能力与识别能力,它能够做到海量的记忆,并识别出对象之间的细微差异。^[6] 阅读速度是人类的一亿倍,测算分析数据是人类的一万倍。它不会因为数量过于庞大而像人一样超出记忆与运算的能力,相反,数量越庞大它的运算便越准确有效;它也不会像人一样遗忘或混淆读过的文献,^[7] 它的优点恰恰在于可以永不遗忘,可以迅速在海量数据之间发现关联与区别。

正是因为这样,AI 科技迅速成为了各大研究室和科学家的辅助帮手,并被越来越多地部署在涉及复杂决策、整体决策的工作领域。AI 科技的特殊性不仅体现在它的能力方面,而且体现在它作为工具的有效性方面:作为人类的工具,它比传统的工具好用太多了。它可以在极大程度上增强和拓展我们的能力,帮助我们实现各类认知以及实践目标——这里虽然不加区分地使用“我们”这样的概括性指称,但后文将指出,在 AI 科技的社会应用中,“我们”作为一个同质化的指称并不符合事实,它只是一个惯性的、直觉式的、具有迷惑性的错误用法。

[6] 参见[美]王维嘉著:《暗知识:机器认知如何颠覆商业和社会》,中信出版集团 2019 年版,第 228-235 页。

[7] 这里需要补充交代的是,人类记忆的有限性以及会遗忘的属性是个人生活中的一种保护机制,完美的记忆并不必然是一种“进步”和“优势”,相反,它会衍生出很多负面问题。因此,对于 AI 科技看上去的进步来说,简单地设想与一厢情愿地美化是情绪化的和拍脑袋式的,需要尽力避免。参见[英]麦尔荀伯格著:《大数据·隐私篇:数位时代,“删去”是必要的美德》,林俊宏译,我国台湾地区远见天下文化出版股份有限公司 2015 年版,第 128 页。

AI 科技是一个多好用的新工具呢？试举一例加以说明。根据密歇根大学罗杰·托兰格教授的研究，因为社会期许偏好等原因，人在现实生活中有三分之一的时间在对自己和他人撒谎。在主观信息的调查中，除非人们跟调查者说真话，否则传统的研究根本没有可靠方案来确保研究的真实性，但 AI 科技的出现改变了这一点。^[8] 譬如，大数据的记录、汇聚、梳理和对比，可以非常准确地得知某个人是不是同性恋，是不是种族主义者，任凭他们怎样口头否认或故意误导都无济于事。目前的 AI 科技已经可以帮助政府、商业公司、科研工作者观察、捕捉和记录人们掩藏在内心的部分真实想法和秘密。

（二）人与人工智能的关系视角

正是因为人工智能呈现出如此强大的能力与功用，展示出一副拥有无限可能的远大前景，这让越来越多的人加入不安的队伍，以怀疑的眼光看待它，不断反问人与人工智能之间的关系。从不安审视的角度，在工具性的面相之外，它们似乎已经展现出了“理性化”的可能。尽管还需要借助某些外观的实体（如电脑）存在，人工智能的实质在于它是一种虚拟机器，重要的不是它的形体，而是它能够做的事情。用博登的话来说：它可以像程序员在写程序时那样“思考”问题，它有一套信息处理的机制，仿佛人脑一般。^[9] 这种信息处理机制再辅以算法与大数据，便能展开一种深度学习，进而获得某种程度上近似人类的分析与决策能力。^[10] 倘若如此，那人工智能与人本质上还有什么区别？人工智能的这些特殊性使得认识和把握它带来的挑战变得困难和复杂起来，也促使很多讨论者选择了“人与人工智能的关系”这一维度作为讨论该科技之挑战的视角。

人工智能理性化的潜力呈现出一幅伴有强烈危机感的暗淡图景。现实的世界里，AlphaGo 已经战胜了人类顶尖围棋选手，人们在借助人工智能提升生活的同时，人工智能也在“观察”“记录”和“分析”着它的使用者。我们的生活在不知不觉之间已经逐渐将部分决定权转移给了人工智能，在不少生活场景中，我们似乎仅仅剩下了名义上的而且往往是一次性的同意授权。^[11] 更让人心生恐惧的是，无论同意与否，AI 科技几乎已经无处不在，我们无法真正选择拒绝，无法逃脱人工智能的凝视，我们几乎每天赤身裸体地走在它的注视之下，它似乎看得到我们的底裤，看得穿我们的底色，它突破了我们构筑的阻挡他人窥视我们隐私的物理空间。尽管我们在名义上还保有隐私的权利，但在物理空间和权利实践的维度，隐私权已经在技术上处于有名无实的状态。这让习惯了现代文明观念、重视体面和隐私感的人类忧心忡忡，很难接受。

在 AI 科技的社会应用高度渗透，支配了越来越多的生活场景之后，人们自然会追问，已然在不少方面超越了我们的 AI 会最终学会并全面掌握与人一样的理性能力，进而实现对人类的整体超越吗？^[12] AI 会最终取代人类并实现对我们的统治吗？这样的疑问带着

[8] 参见[美]赛斯·史蒂芬斯-大卫德维兹著：《数据、谎言与真相》，陈琇玲译，我国台湾地区商周出版社 2017 年版，第 147-153 页。

[9] See Margaret A. Boden, *Artificial Intelligence: A Very Short Introduction*, Oxford University Press, 2018, p. 3.

[10] 参见陈景辉：《人工智能的法律挑战：应该从哪里开始》，《比较法研究》2018 年第 5 期，第 143 页。

[11] 参见[英]维克托·迈尔-舍恩伯格、肯尼思·库克耶著：《大数据时代：生活、工作与思维的大变革》，盛杨燕、周涛译，浙江人民出版社 2013 年版，第 195-197、202-203、210-214 页。

[12] 参见赵汀阳：《人工智能“革命”的“近忧”与“远虑”》，《哲学动态》2018 年第 4 期，第 7-10 页。

非常强烈的“人与人工智能”相互比较的色彩,不安地审视着人与人工智能之间的区别和差距还有多少,怀疑地反问一个会自我学习、自我反思、自我决策的工具是不是只是人类的工具,^[13]如果不只是工具,人与人工智能之间还可以存在哪些可能的相互关系,并视这类问题为 AI 科技对人类社会提出的最大挑战。

(三)错在哪里?

但上述理解和提问方式就回答新科技的社会挑战来说是错误的。主要的原因在于,它建立在对技术的预测而非技术的现实的基础之上。换言之,它构筑的是一个前瞻式的远景:倘若人工智能以现在的态势顺利地、高速地发展下去,它有可能达到或突破某个智能状态,譬如说它最终获得了“自我设定目标的能力”,然后社会的生态与人类的处境就全部改变了。达到这种发展水平的人工智能,学界称之为“强人工智能”。人们认为“强人工智能”可能到来,并对到来之后的场景忧心忡忡,这种担忧是合理的。基于这种担忧尝试重新审视、规划和调整 AI 科技的发展可能是有裨益的,在该前瞻远景的刺激之下对人性、理性、尊严等概念重新进行分析和论证也是有价值的。但是,一来,人工智能能够顺利、持续、高速发展并最终取得临界突破,这是一个科学技术层面的事实性判断,前述担忧建立在这一事实性判断的基础之上,这类考量的质量便也高度依赖这一事实性判断的可靠程度;^[14]二来,就这一前瞻远景带来的认知冲击来说,该冲击的性质并不属于 AI 科技带来的社会挑战,通过重新反思上述基本概念也无法算作是在把握和回应它的社会挑战。

就科技层面的事实来说,目前的人工智能无法理解因果关系,^[15]没有直觉,没有感性能力,没有自我意识,无法提出假设,无法自主选择,离推理、想象等人类具有的高级智能能力还相距甚远。即便是从远景视角来看,就考虑 AI 是否可能经由发展具备自我意识来说,仅仅要达到人脑具备的 1000 亿个神经元数量,按照目前的发展速度,也要至少三十年左右的时间才能让机器的神经网络达到人脑的连接数。但“神经元与连接数只是一个必要条件,即使达到或超过人脑的数量也未必产生自我意识”。^[16]根据瑞典乌普萨拉大学应用数学系桑普特教授的研究,目前人工智能的发展程度,就其智慧能力来说,最接近的生物是肚子里的细菌。根据现有证据,AI 能模拟的智能层级和细菌差不多。^[17]

这里需要说明的是,人工智能现实的智慧等级比较初步,不代表其构不成重大社会挑战,不代表人类社会可以对与细菌处于同一个智力水准的人工智能科技等闲视之。新冠病毒就不是什么智慧病毒,但它造成的全球危机是险峻又直接的,AI 科技的出现与应用也是同样的道理:它不需要多么智慧、多么高级自主,就可以也已经深刻地塑造和改变了

[13] 参见林彤:《全球人工智慧法规与治理的发展趋势》,我国台湾地区《台湾经济研究月刊》2020 年第 2 期,第 39 页。

[14] 很多哲学分析都犯了这样的错误,最近的一个错误例证是纽约大学的大卫·查尔默斯教授。See David Chalmers, *Reality + : Virtual Worlds and the Problems of Philosophy*, W. W. Norton & Company, 2022, pp. 287 - 293.

[15] See Brian Bergstein, What AI Still Can't Do, MIT Technology Review, February 19, 2020, <https://www.technology-review.com/2020/02/19/868178/what-ai-still-cant-do/>,最近访问时间[2020-12-18]。

[16] 参见[美]王维嘉著:《暗知识:机器认知如何颠覆商业和社会》,中信出版集团 2019 年版,第 294 页。

[17] 参见[英]大卫·桑普特著:《演算法的一百道阴影》,赖盈满译,我国台湾地区猫头鹰出版社 2020 年版,第 251 - 260 页。

整个世界。重新回到人工智能的技术现实上来,尊重科技发展的技术事实,在技术事实的现实与前瞻视角之中,人类还有非常长的一段时间是唯一拥有人类智能的物种,至少就未来五十年来说,强人工智能的出现还不具有现实的可能性。也正因如此,以“人与人工智能的关系”用作分析 AI 科技之社会挑战的视角也就被抽掉了坚实的基础,主要剩下概念思辨的意义。如果做概念分析,欠缺现实基础并不必然构成分析的缺陷,因为概念分析可以脱离现实的束缚,以预设的条件或某个理想情境作为分析的前提;但在实践分析中,对社会挑战的立论与说明高度依赖现实性描述的准确性与可靠程度,虚假的事实性前提会过宽地且不当地拓展理论认知的概念空间,进而损伤相关论证的质量。

与三、五十年间也无法出现的强人工智能所可能造成的挑战相比,现有的人工智能已经实实在在走进了我们的生活,对个体和共同生活制造着切实的影响。我们的生活正在被人工智能科技塑造着,也由此带来了重要性程度上绝不应忽视的道德困境,构成了下文将具体展开证明的大挑战,急迫地要求我们在认知上做出实践性的分析与回应。之所以这些道德困境构成的挑战可以被归诸现有的人工智能技术,是基于该技术内置的社会属性。下文先完成对其社会属性的说明,再在此基础上阐述该技术的社会挑战是什么。

三 个人数据与 AI 科技的社会属性

回到现有“弱人工智能”应用的技术特征,作为一种信息处理、应用的机制,弱人工智能的社会属性主要体现在两方面。一是,这种信息处理的机制受制于并反映现实社会的道德境况,AI 科技可以说是所在社会的道德镜像;二是,这种信息处理的机制广泛应用于社会各个领域,被用作实现各类目的的手段,AI 科技可以说是所在社会的工具媒介。

(一) AI 科技的道德镜像属性

主张 AI 科技是所在社会的道德镜像,反映了所在社会的道德境况,该主张立基于以下几个技术现实。

第一, AI 科技作为一种信息处理机制,它的运作建立在收集来的由个人信息汇聚而成的大数据之上,而大数据本身要么不是道德中立的,很多数据是对信息主体价值偏好的收集;要么受到了现有社会分配结构的渲染。

譬如,当美国的警察局将人工智能技术部署于犯罪预测与刑事执法,人工智能的有效运算便需要采集该地区过去的犯罪信息、种族信息、学历信息、性别信息等等进行数据化,并在分析相关数据的基础上得出预测性的或决策性的结论。学界已有大量的研究表明,美国的刑事实践呈现着严重的种族不公正的事实结构,^[18]这个不公正的执法结构的量化表达便体现为各种数据。换言之,不公正的社会结构会生产不公正的数据,对历史执法信息的收集所得到的数据本身是有失公正的,带着种族歧视的偏见,呈现着某一种族、某一

[18] See Will Douglas Heaven, Predictive Policing Algorithms Are Racist. They Need to be Dismantled, MIT Technology Review, July 17, 2020, <https://www.technologyreview.com/2020/07/17/1005396/predictive-police-algorithms-racist-dismantled-machine-learning-bias-criminal-justice/>,最近访问时间[2020-12-18]。

社区区域被过度刑事执法的事实。这些事实作为人工智能分析的数据会引导人工智能进一步得出带有偏见的预测结论。进一步得出带有偏见结论的人工智能既是对现有不公正的执法结构的反映,又会对现有的结构性不公产生进一步的叠加影响。

以美国佛罗里达州布劳沃德县的刑事执法实践为例。自2008年开始,布劳沃德县采用Northpointe公司开发的算法程式COMPAS对犯罪嫌疑人的人身危险性与再犯可能性进行测评打分,以帮助确定哪些人需要入监收押,哪些人可以保释。根据朱莉娅·安格温(Julia Angwin)与杰夫·拉森(Jeff Larson)等学者的研究,^[19]Northpointe测评打分的结果很遗憾地呈现了以下的结构:在所有的犯罪嫌疑人当中,被算法标记为高危险性但其实并没有再犯罪的非洲裔美国人高达44.9%,与此相对,标记为高危险性但不再犯的白人只有23.5%;标记为低危险性却再次犯罪的白人有47.7%,与此相对,标记为低危险性却再犯罪的非洲裔美国人只有28%。这样的算法测评结果显示,算法更偏好白人,更倾向于做出不利于非裔的结论。^[20]可以看到,这样的算法结果是对布劳沃德县以往刑事执法中存在的种族不平等问题的呈现,这样的算法应用也进一步维系和加深着既有的执法不公,并没有像我们直觉想象的那样,以为应用算法会提升刑事执法的公正水平。

第二,AI科技作为一种信息处理机制,其处理数据分析的算法本身也具有非中立性。现阶段的人工智能既不能自主地理解因果关系,又不能自主地选择或者预先设定算法所依赖的核心概念与变量。事实上,是程序开发者或数据科学家在对相关的概念进行选择和界定,这样的初步设定使得算法无法摆脱程序开发者或拥有者的主观偏好,^[21]隐默地内嵌着社会价值观点的渲染。在这个议题上,关键的问题不是人们常说的程序开发者的控制或人工智能算法黑箱的问题。^[22]就程序开发者与数据科学家的控制来说,即使再公正再高超再努力的科学家也无法完全预测或控制算法,这不是说算法很智能,很不好控制,而是说大数据运算本身带有随机性与一定程度的不可理解性,再加上科学家无法摆脱自身的主观偏好,无法根除算法初步设置中的非中立状态;就算法黑箱的问题来说,人脑的运算也是黑箱的状态,这一状态本身不是要求矫正人脑运算的充分理由,算法的完全透明化与要求人脑运算的透明化一样,几乎是不可能的。尽管这些并不意味着科学家不需要努力提升和规训算法,^[23]也不意味着算法不需要满足合理的公开、透明等要求。

对此,一个常见的主张是说,要对算法进行技术上的修正和完善,以及,要加强技术外的规范与约束。本文想指出,技术内外的应对和完善是必要的,但这样的应对策略存在一个误区,即以为人工智能不公的问题主要是个技术性、科学性的问题,也主要应当通过技

[19] See Julia Angwin, Jeff Larson, Surya Mattu & Lauren Kirchner, Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And it's Biased Against Blacks, ProPublica (23 May 2016), <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>, 最近访问时间[2020-12-18]。

[20] See Julia Dressel and Hany Farid, The Accuracy, Fairness, and Limits of Predicting Recidivism, 4 *Science Advances* 1, 1-5 (2018).

[21] See Ninareh Mehrabi, Fred Morstatter, Nripsuta Saxena, Kristina Lerman & Aram Galstyan, A Survey on Bias and Fairness in Machine Learning, 54 *ACM Computing Surveys* 1, 7-8 (2021).

[22] See Michael Kearns and Aaron Roth, *The Ethical Algorithm: The Science of Socially Aware Algorithm Design*, Oxford University Press, 2019, pp. 7-8, 10-11.

[23] 参见郑玉双:《计算正义:算法与法律之关系的法理建构》,《政治与法律》2021年第11期,第99页。

术性、科学性的方法来解决。这种应对策略之所以错误,是因为它忽视了人工智能所应用的目标社会其自身的公正状况对新信息科技应用的决定性影响,忽视了人工智能不公的社会属性。用朱迪亚·帕尔的话来说:“因果问题无法只依靠数据而得到解答,……解释数据意味着要对真实的世界是如何运作的进行把握。”^[24]可以这样说,任凭人工智能技术如何高速发展,都无法否认人工智能技术应用的非中立属性。倘若现有的实践中,修正得再完善的算法也会带有特定主体的事实想象与道德信念,就像数据自身带有的价值立场使得算法处理的数据自身存在好坏之分一样,算法偏见、算法不公的问题就是人工智能应用中的一个技术事实,呈现和承受着现实世界的道德境况与影响。

(二) AI 科技的社会工具属性

AI 科技的社会属性还体现在它的工具属性之上。AI 科技是社会主体实现各自目的的手段,是不同主体之间交往、博弈的工具,这一判断作为事实判断同样建立在社会实践的可验证性之上。

就人工智能被工具化并带来不确定的社会影响进行列举:人工智能这件事本身既可以被诸如美国政府拿来用作博弈、限制高新科技企业的事由,又可以被用作重返或推动计划经济的手段,还可以被用来实现监控、维稳等社会、政治目标;^[25]同样,人工智能既可以被苹果公司用来实现残障人士的出行自由,^[26]又能够被美团用来挤压市井摊贩的经营空间,^[27]既可以被部署来抓捕罪犯,^[28]又可以用于犯罪,^[29]还有人试图以之提升学生上课的效率与质量,^[30]抑或用于识别同性恋或进行心理治疗,^[31]等等。总之,人工智能可以被用作各种想象得到和想象不到的目的,以媒介的方式影响和塑造着人际交往与共同体的道德境况。需要补充交待的是,尽管人工智能作为工具媒介发挥着功用,但与以往不同的一个情形是:以往倘若产生损害,工具本身是绝缘于责任的,责任回溯至工具的使用者。但在人工智能应用的情境之中,因为多种原因,作为工具的人工智能本身是否依旧绝缘于责任变得模糊起来,同时还使得工具使用者的责任也模糊起来。这主要是因为,人工智能技术应用具有新颖性与特殊性,既有的法律治理体系有一个利益竞争、规范调整和实践试

[24] Judea Pearl and Dana Mackenzie, *The Book of Why: The New Science of Cause and Effect*, Penguin Books, 2018, pp. 350 - 351.

[25] See Edward Snowden, *Permanent Record*, Metropolitan Books, 2019, pp. 174 - 180.

[26] Devin Coldewey, iPhones Can Now Tell Blind Users Where and How Far Away People Are, TechCrunch (31 October 2020), <https://techcrunch.com/2020/10/30/iphones-can-now-tell-blind-users-where-and-how-far-away-people-are/>,最近访问时间[2020-12-18]。

[27] 参见常盛:《“社区团购”争议背后,是对互联网巨头科技创新的更多期待》,人民日报评论微信公众号,2020年12月11日。

[28] See Ying-Lung Lin, Tenge-Yang Chen & Liang-Chin Yu, Using Machine Learning to Assist Crime Prevention, in *Presented at the 2017 6th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI)*, 2017, pp. 1029 - 1030.

[29] See Nekfaria Kaloudi & Jingyue Li, The AI-based Cyber Threat Landscape, 53 *ACM Computing Surveys* 1, 1 - 34 (2020).

[30] See S. Dent, A Paris School is Using AI to Monitor Distracted Students, 26 May 2017, <https://www.engadget.com/2017-05-26-nestor-ai-paris-inattentive-students.html>,最近访问时间[2021-01-18]。

[31] See Paul Lewis, “I Was Shocked It Was So Easy”: Meet the Professor Who Says Facial Recognition Can Tell If You’re Gay, 07 July 2018, <https://www.theguardian.com/technology/2018/jul/07/artificial-intelligence-can-tell-your-sexuality-politics-surveillance-paul-lewis>,最近访问时间[2021-01-18]。

错的不确定期,在此期间,法律责任与人工智能的创造者、支配者以及使用者之间的联结通常被悬置了起来。

回到人工智能的社会属性。由上可见,如果人工智能技术的应用本身就会承载和反映现有社会的道德境况,也会影响和塑造个体与共同生活的道德境况,那么人工智能就会带来可以归诸于它的挑战。此外,正如上文已经阐明的,人工智能带来的这些挑战不适合采取“人与人工智能的关系”的分析视角去审视和回答人与人工智能之间现有的以及可能的相互关系,而是应该采取“智能系统冲击下的人际关系维度”的分析视角去审视和回答人工智能如何改变了我们的共同生活,制造了怎样的实践困境,我们又应当如何努力做出恰当的应对等等,以此确保人工智能是在促就更加公正的社会和更加自主、美好的个人生活,而不是相反。

四 双重不对称性难题

当我们的视角终于回到“人工智能冲击之下的人际关系”,我们便能直面以下问题:人工智能技术给社会带来了怎样的变化,在其中谁得到了裨益,谁受到了损失,这样的得失之间怎样的道德困境最为紧要?本文力图主张,人工智能科技冲击形成的最重要的道德困境是制造了双重不对称性难题,在事实层面改写了权力正当性的结构,改变了和持续改变着政治社群实质上的共同生活样态。

(一) 多重现实描述的困境

基于回溯视角的特性,这里的重点在于分析 AI 科技的挑战而非贡献。常见的主流认知是肯定 AI 科技的重要性以及它的积极意义,^[32]与惯常的类似理解不同,近年来有重要的研究指出,AI 科技的重要性还在于它现实上以及可能带来的社会风险与危害。^[33]人们逐渐意识到,人工智能科技带来的不一定是高效率,不一定是专业性,也不一定是中立性和公正。以美国斯坦福医疗中心最近的一次公共事故为例。2020 年年底,美国的疫情居高不下,美国 FDA 经过安全专家委员会的表决,批准了辉瑞公司与德国 BioNTech 公司共同研发的新冠病毒疫苗的紧急使用授权,斯坦福医疗中心首批分到了 5000 支疫苗。如何分配这 5000 支疫苗呢?斯坦福医疗中心决定借助算法来实现这一稀缺医疗资源的分配,以期最高效也最公正地实现医疗利益的最大化。但算法给出的分配方案非常令人震惊:5000 支疫苗中只分配了 7 支给一线医护人员。这迅速引起了强烈的抗议,压力之下,院方出面道歉并承诺更改分配方案。^[34]类似这样的例证每天都在很多地方出现,使得人

[32] See Iain M. Cockburn, Rebecca Henderson and Scott Stern, The Impact of Artificial Intelligence on Innovation, in Ajay Agrawal, Joshua Gans & Avi Goldfarb eds., *The Economics of Artificial Intelligence*, The University of Chicago Press, 2019, pp. 115 - 146.

[33] See Meredith Broussard, *Artificial Unintelligence*, The MIT Press, 2018; Safiya Umoja Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism*, New York University Press, 2018.

[34] See Caroline Chen, Only Seven of Stanford's First 5,000 Vaccines Were Designated for Medical Residents, ProPublica (18 December 2020), <https://www.propublica.org/article/only-seven-of-stanford-s-first-5-000-vaccines-were-designated-for-medical-residents>,最近访问时间[2021-01-19]。

们越发以疑惑和审慎的眼光看待 AI 科技。

AI 科技不只是令人疑惑,它也令人担忧。这种担忧主要不是担心人工智能是不是人,有没有可能超越或统治我们,而是指向人工智能科技应用造成的社会困境与隐患。想要清楚又全面地呈现这些困境与隐患是困难的,因为它涉及到很多层次与面向,彼此之间似乎也没有统一性可言。

以 AI 科技中发展迅猛的人脸识别技术应用为例。^[35] 第一,人脸识别的技术发展非常迅速,其应用门槛也很低,任何能够安置普通摄像头的介质都可以用作采集人脸信息的传感器,理论上以非常低的成本就可以实现对不特定主体人脸生物信息的采集,这样的科技不仅很难封锁,而且很容易被滥用;第二,正是因为只需要具备可以成像的传感器再加上面部识别的软件和算法就可以远距离采集和应用人脸信息,这类技术应用具备了非接触的特质,从而使得对人脸生物信息的采集可以静默无声,在技术上不再需要被采集主体的授权或同意,排除了对象的积极参与,如果不明确告知,被采集人甚至不知道脸纹被扫描了的事实;第三,人脸信息的获取在人工智能时代结合大数据分析,可以做到对脸纹主体的深度扫描与刻画,提供脸纹就意味着提供了脸纹主体的身份信息、家族信息、性别信息、种族信息、性取向信息、资产信息、医疗信息、心理偏好信息、人格特质信息等等,人脸识别技术的应用可以让被采集的对象成为透明人,隐私虽然在规范的维度还存在,但在技术上可以被消解。

上述三点对应的每一个问题都很严重,都不太容易被 AI 科技应用带来的好处所遮盖。那么,如何能够更好地归纳和理解这些困境与隐患,以便准确把握挑战的核心呢?如上文所说,人工智能的问题不可能仅仅通过科学的方式独自加以解决,^[36] 因为人工智能被嵌置在社会生活之中,受到社会公正结构与道德境况的影响,因此,重要的便是审视人工智能嵌入社会关系网络之后怎样改变了主体间的交往状态。对此,本文主张,不对称性 (asymmetry) 是一个适当的概念工具。

(二) 不对称性作为分析概念

还是以人脸识别技术应用为例,回到上述列举的问题,我们可以看到:第一,技术的开发者与应用者之间存在着不对称性。尽管技术的普及使得技术运用的门槛降低,但技术的使用者如政府部门与技术的开发者如科技公司之间的交往结构是不对称的。这意味着围绕技术的开发与应用,政府部门与科技公司之间存在着或竞争或合作的关系,但就技术本身来说,是科技公司占据着不对称性的优势,^[37] 与之相对,政府的不对称优势在于它的政策与执法功能;第二,技术的应用者与用户之间也存在着不对称性。无论人脸识别技术的应用者是科技公司、政府部门还是其他社会主体,技术应用的属性使得采集者与被采

[35] 参见杨建:《人脸识别技术应用的法律规制》,《中国社会科学报》2020年3月25日第6版。

[36] See Annette Zimmermann, Elena Di Rosa, Hohan Kim, Technology Can't Fix Algorithmic Injustice, Boston Review (09 January 2020), <http://bostonreview.net/science-nature-politics/annette-zimmermann-elena-di-rosa-hochan-kim-technology-cant-fix-algorithmic>, 最近访问时间[2021-01-19]。

[37] 有关技术公司的竞争力与宰制力,参见[美]露西·格林著:《硅谷帝国:商业巨头如何掌控经济与社会》,李瑞芳译,中信出版集团2019年版。

集者之间的权能结构是不对称的,如上所述,被采集者甚至不知道自己的生物信息被扫描了的事实。此外,就应用者与用户之间的权益结构关系来说,也是高度不对称的。一次给出脸纹,在技术上便等于终身给出了个体生物信息。并且,现实生活中同意面部识别大多只是为了便捷的目的,也许在日常能够直接感知的范围内确实提升了便捷度,但商业机构与政府部门借此可以获得的裨益却是高度不对称的。个体实际交出去的并脱离自己掌控的综合信息不仅对应着不成比例的商业收益,而且可能意味着自由度与自主性的减损。由此可见,AI 科技引致的诸多困境似乎都可以经由“不对称性”概念得到妥当的说明。

在进一步阐述 AI 科技带来的不对称困境之前,有必要停下来做几点概念与事理上的澄清。第一,不对称性指向的是这样一种状态:本来被理解为性质上相同或对等,或者构成一对恰当范畴的两个对象,事实上处在性质上不同或不对等或者并不当然恰当匹配的状态。譬如说,责难与赞赏通常被理解为一组对称概念,但有学者指出,事实上它们是不对称的,责难在功能上要远比赞善来得丰富与深刻。^[38] 同样,有不少伦理学学者表明,善与恶、好与坏、快乐与痛苦在道德上其实是不对称的,对它们进行量化衡量和比较在根本上是行不通的。^[39] 不仅道德领域是这样,而且政治与法律领域也是这样,例如,布鲁姆教授指出,同样是歧视,但在美国歧视黑人与歧视白人是**不对称的,它们是性质不同的两个事件。^[40]

第二,不对称性本身并不必然是不好的、道德上错的,换言之,不对称性并不必然是道德性的,具有道德性的不对称性也不必然就构成道德困境。一般而言,理科生比文科生更会维修电器,更擅长动手,有更强的生存能力,他们之间也可以说是不对称的,但这种不对称并不是不道德的。同理,科学家比普通人更会编程,在 AI 科技的开发与研究上具有很强的不对称优势,这本身是不需要也不应当被矫正的。

第三,现代社会的共同生活中既存在着相互性关系,又存在着互惠性关系,但相互性不等于互惠性。^[41] 在科技公司与用户所缔结的契约关系之中,他们之间存在着道德性的相互关系,但权责、利益的分配却可能是高度不对称的,用户以极大的付出获得的只是极少的裨益,在政府与公民的相互性关系之中也可能是这样。在这一维度,人工智能既有可能是朝强势一方增进不对称性,也有可能是朝弱势一方增进不对称性,还有可能是调和了原有的不对称性,当然也可能是多种程度多种方向的增进与调和的交织。AI 科技带来的困境与机遇并存,但现有人类社会的不公正结构往往使得 AI 科技制造困境的现实性要整体上盖过它带来积极裨益的可能性。

[38] See Syeve Guglielmo, Bertram F. Malle, Asymmetric Morality: Blame is More Differentiated and More Extreme Than Praise, 14 *PLoS ONE* 1, 1–20 (2019).

[39] See Knut Erik Tranøy, Asymmetries in Ethics, 10 *Inquiry: An Interdisciplinary Journal of Philosophy* 351, 351–372 (1967); Jamie Mayerfeld, The Moral Asymmetry of Happiness and Suffering, XXXIV *The Southern Journal of Philosophy* 317, 317–338 (1996).

[40] See Lawrence Blum, Moral Asymmetry: A Problem for the Protected Categories Approach, 16 *Lewis & Clark Law Review* 647, 647–655 (2012).

[41] See Robert Gibbs, Asymmetry and Mutuality, 23 *Philosophy & Social Criticism* 51, 51–63 (1997).

AI 科技引致的不对称困境究竟应当怎么把握? 倘若科技公司的技术优势构成的不对称性并不必然属于需要矫正的道德问题,何以 AI 科技依旧带来了社会大挑战? 下文依次从一阶行动维度与二阶结构维度予以阐述。

(三) 一阶与二阶不对称性困境

从一阶行动的维度来说,相较于普通用户,各大科技公司与政府部门整体上是占据不对称优势的强势主体。无论是科技公司还是政府部门,都在事实上使得原先的不对称优势演变成了需要矫正的道德困境,或者具备了触发这些道德困境的高度可能性。

第一,对于科技公司来说,首先,认知与科技优势并不必然是负道德性的,但倘若科技公司将这些不对称的技术优势用在与普通用户形成的各类相互性关系之中,就可能借助不对称的优势谋取超额利润,或者利用技术上的能力突破用户在规范上的隐私权、知情权、决定权等基本权利,^[42]若如此,技术上的不对称性便不当地介入科技公司与用户之间的关系,演变成需要纠正的道德困境。其次,高科技公司主导的 AI 科技迅猛发展,使得很多商业公司的运作越来越人工智能化,不仅商业行为而且整个社会的消费行为与生活方式也越来越人工智能化,^[43]在某种宏观视野下,可以说,是高科技公司在决定、塑造和培养着我们的交往方式与生活选项,^[44]强化着我们对新信息科技的依赖。基于 AI 科技的特性,这种塑造是大规模的、精准的和隐默的,往往体现的是高科技公司的商业逻辑与利益意志,整个框架是基于便于它们行动和利益最大化的目标设置的。

第二,对于政府部门来说,整体上 AI 科技的迅猛发展不成比例地提升了政府的认知能力、管控能力,使得权力的触角无远弗届,这一发展本身便构成并加重了公共生活中不对称的道德困境。

首先,当下以及未来公共服务的提供和保障都需要公民向政府这一最大的公共服务提供者持续地提供自身的信息数据,并保持这些信息数据的更新。即使在公民不感知的情况下这一切也在有序地推进,这让政府具备了在技术上实现对公民进行全方位了解和监控的可能,^[45]并借此获得更为强盛的政治权力。政府很难抵制这种让权力“全知全能”化的诱惑,美国政府开发和部署“棱镜计划”以及“上游收集计划”就是一个例证。

其次,借助信息新科技,政府已经可以做到强力贯彻执行法律与政策,往往也有动力使用“AI 科技 + 刑法”的组合去治理社会。不像从前,如今一点小错都有可能受到严惩,本来的民事案件、行政案件都可能刑事化,现代社会整体治理结构的泛刑罚化是人工智能时代高效、便捷生活的另一个面向。^[46]但法律不应当也不可以以强制的方式贯彻到底,

[42] See Martin Moore, Tech Giants and Civic Power, Working Paper of King's College London (April 2016), <https://www.kcl.ac.uk/policy-institute/assets/cmcp/tech-giants-and-civic-power.pdf>, 最近访问时间[2021-01-20]。

[43] See John Horgan, Big Tech, Out-of-Control Capitalism and the End of Civilization, Scientific American (07 October 2020), <https://www.scientificamerican.com/article/big-tech-out-of-control-capitalism-and-the-end-of-civilization/>, 最近访问时间[2021-01-20]。

[44] See Brad Smith & Carol Ann Browne, *Tools and Weapons: The Promise and the Peril of the Digital Age*, Penguin Press, 2019, pp. 128-173.

[45] See Paul Bernal, Data Gathering, Surveillance and Human Rights: Recasting the Debate, 1 *Journal of Cyber Policy* 243, 243-264 (2016).

[46] 参见杨建:《刑罚规制个体自决事务的限度》,《学海》2018年第6期,第210-216页。

法律之治的一个居于核心地位的典型特征就是它的“自我应用性”:绝大多数法律不需要触发保障机制就被绝大多数民众自觉遵守了。^[47] 这是因为,人们是具备实践理性能力因而具有尊严的生命,他们对自己的生活能够进行自我控制、自我反思,能够理解法律规范并据此决定自己的行动。借助 AI 科技将法律落实到位,让违法违规无所遁形,则会与法律之治的这一典型特征相冲突。这也是为什么即使技术上可行,现实中高度可能,大多数国家也要尽力对借助 AI 科技武装刑事执法以治理社会的做法表达谦抑的重要原因。

再次,人工智能提高政府办事效率这件事并不当然构成优势,仍有可能加重了共同体的道德困境。一个重要的原因在于,早在人工智能技术成熟之前,现代社会就已经有了许多方法可以让政府的行动变得快捷又高效,但法律创设的核心指向之一就是刻意降低政府的效率,让政府的行动慢下来。为政府实现职责目标设立很多“规矩”与“限制”,这是现代社会组织、运作的基本法则。只是,顶着进步的光环,政府已经在很大程度上打破了这一基本法则,使得法治的面向越来越复杂、陌生。

第三,对于普通民众来说,他们在行动的维度上也陷入了多重的道德困境。首先,普通民众内部是分化的,不对称的,不仅与科技公司、政府部门无法并称“我们”,而且在其内部,简单的、同构的“我们”也不存在。他们中的大部分人逐步顺应了新的、被给定的生活框架,还有一部分人则不太跟得上这个新世界。这部分落伍者的处境就比较依赖所在社会原有社会结构与道德境况的良善与稳健程度,依赖于所在社会的纠偏与矫正能力。在普通民众内部,AI 科技应用生产着不对称的收益与损害分布,但利益受损的群体往往也是消失在社会公众视野中的或话语能力比较低的那群人。在现代社会,不被知晓与关注,通常就进入不了共同体的议事日程,不公的状况就容易被忽视、被延续、被恶化。

其次,AI 科技带来了普通民众之间信息茧房、另类事实、群体抱团、封闭自足、对抗敌视等问题,这些问题集中在普通民众交往所形成的公共生活之中,将更为强烈地凸显出社会宽容或社会自由的问题。密尔在《论自由》中开创性地提出了“社会自由”的概念,阐述了公共舆论对个性可能的不当压制。这过去了不过一百六十多年,AI 科技的普及与应用已经让密尔最大的担忧变成了社会现实:社会宽容变得越来越困难,个体的一切言行均有迹可查,在政治正确和道德正确的框架之下,观念表达的豁免性已全面让位于实质审查,并对应着工作以及社会认可等方面权益的不当剥夺,“社会性死亡”已经以其不确定性、不可预知性、难以补救性等特征参与构筑了当今社会的公共生活。的确,现有的 AI 科技也可以在这个泛政治正确的时代被部署、应用于维护和促进社会宽容,譬如,美国佛罗里达州圣彼德斯堡的非营利性机构波因特研究所(Poynter Institute)就开发出了 PolitiFact 网站,应用 AI 科技进行事实查证,以尽力化解政治人物的虚假言论对普通民众产生的不当影响,破解群众斗群众的社会困境。^[48] 尽管如此,但普通民众所处的整体上对其不利的

[47] See Jeremy Waldron, *Dignity, Rank, & Rights*, Oxford University Press, 2012, p. 52.

[48] 该网站的运作及其功能介绍, see <https://www.politifact.com>, 最近访问时间[2022-03-10]。

一个结构性社会现实是：暂且不谈不断强化的认知偏好的束缚，他们既缺乏自行鉴别与查证的能力，又缺乏追踪与更新的时间和精力。

此外，本文还要简要提及普通民众面临的另一层无力单方面化解的技术困境：元数据牢笼。很多人存在一个直觉式的误解，以为人工智能时代重要的是确保社会交往的内容不被未经允许地观察、记录、分析，如我们的购物车里买了哪些东西，在短信和电话里说了什么，某天在哪里见了谁等等。这些确实都在发生，也并非不重要，但同样重要的是我们的元数据(metadata)被收集、整合、分析与部署的事实。与有实质内容的数据相比，元数据是数据背后的数据，是数据制造的数据，是普通用户感知不到的有关自身的信息，它是自动制造的，脱离用户掌控的，只要我们使用智能设备就无时无刻不在生产的。它确实不是实质数据，但假如将实质数据比作肉身，用实质数据画像可以描绘出我们活生生的模样，元数据就好比是心智，用元数据可以拼凑和测算出我们的行为模式乃至内部世界，使得整个大数据分析既见树木又见森林。^[49] 这是无法不关注但并没有被广泛关注的道德困境。

二阶结构维度是一阶行动维度所处的结构与语境，是一阶维度何以呈现如此道德困境的重要原因。如果说，一阶行动维度是在新给定的生活框架之下，由新兴的科技权力与新科技武装过的政府权力对社会资源动态分配的话，那么，二阶结构维度则是对这些分配社会资源的权力本身的分配。就政府权力的分配而言，原本建立在各种不同形式的代议制基础之上的权力正当性分配模式并没有消失，权力的产生与延续还是需要得到普通民众的认可、同意与确认。只是，在 AI 科技带来的更为不对称的相互性关系之下，新信息科技的运用使得在政治权力的产生与延续上，普通民众越来越可以被规划和安排，^[50] 尽管越来越多的民众对此越来越不安，但在目前的境况之下也都无能为力。

就高科技公司的新兴科技权力来说，并不存在政府权力产生和更替所依赖的那种分配，但无论是科技公司的政策决定所能涉及的规模、产生的影响，还是违背这些政策会导致的不利后果，^[51] 都与法律的治理非常相似，甚至有过之而无不及。为了阻止澳大利亚政府通过一项待表决的法案，高科技巨头 Meta 公司以通知的方式出台了一项临时政策，直接暂停了澳大利亚境内所有民众和媒体在其平台发布“朋友圈”的权限，最终成功迫使澳大利亚政府搁置法案的进程，并与其重新展开谈判。^[52] 高科技公司凭借信息新科技掌握的这种新兴能力，就其实质而言，当然是一种公共权力。^[53] 但这种“权力”以及它发布的“命令”，形式上看仍然属于私主体在私人领地发布的自家“决定”，享有既有法律体系

[49] See Edward Snowden, *Permanent Record*, Metropolitan Books, 2019, pp. 178 – 180.

[50] See Vincent C. Müller, Ethics of Artificial Intelligence and Robotics, in Edward N. Zalta ed., *The Stanford Encyclopedia of Philosophy (Winter 2020 Edition)*, <https://plato.stanford.edu/archives/win2020/entries/ethics-ai/>, 最近访问时间[2021-01-19]。

[51] See Gabrielw Buchholtz, Artificial Intelligence and Legal Tech: Challenges to the Rule of Law, in Thomas Wischmeyer, Timo Rademacher (eds.), *Regulating Artificial Intelligence*, Springer, 2020, pp. 176 – 198.

[52] See Bobby Allyn, Facebook Blocks News In Australia Over Government Plan To Force Payment To Publishers, NPR (17 February 2021), <https://www.npr.org/2021/02/17/968723929/facebook-blocks-news-in-australia-over-government-plan-to-force-payment-to-publi>, 最近访问时间[2021-02-19]。

[53] See Ro Khanna, *Dignity in a Digital Age: Making Tech Work for All of Us*, Simon & Schuster, 2022.

安排之中私主体的积极行动权利与消极豁免权利。并且,它发布影响甚广的政策还不需
要像法律那样满足“必须事前拟定并公布”“必须清晰”“必须确定”“必须一致”等一系列
的法治要求,要求高科技公司像政府那样确立起“依法行政”的约束框架与规范原则似乎
更是无从谈起。

正是在这样的一阶行动维度与二阶结构维度之中,科技时代的不对称难题在两个维
度都蔓延开来。要让民众相信,尽管政府部门与高科技公司能够并且事实上在收集、积
累、分析他们的元数据与实质数据,但政府与高科技公司不会越过其选择与决定权,不会
侵犯他们的权益,只是,如何确保并让民众相信这一点?同样,如何让民众相信,AI 科技
的大规模部署与应用不会强化社会原有的不公,不会制造新的更稳固的阶层固化,不会进
一步扩大整个社会的贫富差距,相反,它是矫正并提升社会道德境况的希望?如何让民众
相信,社会的人工智能化不会减损民众对个人生活的支配,他们的人生仍然是自主的、自
由的、可错的,个体和社会生活的人工智能化会保障这一点,甚至可以让人们更加自主、更
加体面和有尊严地生活?如何让民众相信,AI 科技介入公共生活不会罔顾个体的意志与
责任而强行推行某种集体的利益与意志?

以上所述,即是 AI 科技社会应用带来双重不对称性困境之后出现的公共信任危机。
这种信任危机的解决,^[54]有赖于认真处理多个层次、多个面向的多个具体问题,但整个危
机的根源还是系于这个结构性的问题:进入人工智能时代之后,在重新塑造的相互性关系
之中,民众是否还是他自己、他的公共生活、他的国家的主人?换言之,社会权力与国家权
力是否还是来源于民众,服务于民众,受制于民众?AI 科技的介入之所以构成的是社会
大挑战就在于,权力分配或者权力正当性这个根源性问题在新科技不对称的影响之下陷
入了危机之中。AI 科技的研发与应用还将继续推进下去,权力正当性的危机没有一次
性的、统一的解决方案,更不会在短期内解决。在这个不对称的、动态的社会困境之中,
无论是科技权力还是国家权力,其不对称的性质,其强制的正当性程度,都取决于应对和处
理上述危机的状况。

五 代结语:法理作为前共识应对

个人信息保护为何如此特殊和重要不是没有原因的,它恰恰是 AI 科技冲击之下重塑
社会结构之后的结果。个人信息权益在人工智能时代的保护应该朝着化解而不是堆积双
重不对称性困境的方向进行。在这样的实践语境之中,个人信息权益保护的应有之义,是
重申个体在人工智能时代对于自我的伦理生活和公共的道德生活的自主性。共同生活中
的双重不对称性困境虽然是 AI 科技介入社会之后带来的,但它不只是技术问题,无法单
靠技术的继续进步来加以解决,AI 科技的应用受制于并呈现着目标社会的正义状态与道
德境况;在高科技公司与政府共同塑造的新的生活框架与分配结构之中,民众对这一框架

[54] See Woodrow Hartzog & Neil Richards, *Privacy's Constitutional Moment and the Limits of Data Protection*, 61 *Boston College Law Review* 1688, 1688 - 1761 (2020).

结构有着持续且渐进的信任危机,对此,高科技公司与政府担负着重新宣誓忠实、重新赢得信任的道德与政治义务;^[55]不论是科技能力的性质,还是公共责任的大小,抑或巨额利润的再分配或者数字税的设置等等,这些都是共同生活之中对高科技公司的重新定位需要回答的基本问题束;AI 科技这一社会大挑战对应着大讨论、大变革,在 AI 科技运作着的新的生活框架之中,没有任何人能够真的避免或摆脱不公正的共同体结构所带来的或内嵌的负面影响,共同体中的每一位成员都有重要的理由去关心和参与这一大变革。

如何应对上述 AI 科技时代共同体内部的实践大挑战呢?本文主张将“法理应对”作为一个概括的实践方案。法理应对也许是重要的,^[56]我们不难列举出它如何重要的理由,就像我们同样可以说“立法应对”“政策应对”“伦理应对”“哲学应对”很重要一样,^[57]但这会消解这一实践方案的严肃性与说服力。本文的主张与全文的分析紧扣在一起,直指 AI 科技构成的社会大挑战的性质:这是一个法理性质的挑战,因此需要法理性的回应方案。

具体一点说,它首先要与作为学术研究的法理区分开来,后者以追求真理为活动的本质,逻辑和真值是它的生命,主张作为学术研究的法理追求共识,就像主张这一活动应当在不同的竞争学说之间和稀泥一样,这类奇怪的主张是对法理专业本身的否定与消解。与之相对,这里所指的是作为公共实践的法理,它是实现共同体合理共识的优质方案。^[58]其次,之所以人工智能应用形成的是法理挑战,需要法理方案,有两大原因:一是要想妥当说明这一实践挑战的性质,就必须准确抓取这一社会实践之中的核心要素,它们至少是个体权利、强制力与合法性等,可以说法理内嵌于这一社会实践之中;二是这一实践挑战所形成的实践困境具有开放、不确定、无法彻底解决的特征。

也正因如此,作为概括的实践方案,公共实践的法理无法也不应当化身成为一种大型的决策咨询活动,致力于给出一次性的对策建议,无论这些对策建议是立法的、政策的还是政治的。它是对话、竞争、协作等形式,呈现社会现实性与社会道德镜像的长期活动,是坚持重塑公共信任与公共正义的不懈努力,是守住个体自主性与美好生活的持续试错。它追求结果意义上的成功,但结果的成功与否并不界定它的性质,也无法决定它的生命。AI 科技时代的法律创制与实践不应当被视为单纯的立法、行政与司法问题,更不是单纯的利益权衡与争夺,而是应当被视为大挑战之大讨论的一部分,是政治共同体真诚实践试错的一部分,需要持续地在公共法理之中接受检验,进行纠错与革新。

[本文为作者主持的 2018 年度教育部青年基金项目“美好生活的法理意涵与法理论述体系建构”(18YJC820081)的研究成果。]

[55] See Woodrow Hartzog & Neil Richards, *Privacy's Constitutional Moment and the Limits of Data Protection*, 61 *Boston College Law Review* 1688, 1688 - 1761 (2020).

[56] 参见陆幸福:《人工智能时代的主体性之忧:法理学如何回应》,《比较法研究》2022 年第 1 期,第 33 页。

[57] See Mark Coeckelbergh, *The Political Philosophy of AI*, Polity Press, 2022.

[58] See Nigel E. Simmonds, *Central Issues in Jurisprudence*, Sweet & Maxwell, 2018, p. 6.

[**Abstract**] China provides a rather broad policy space for the application of AI technologies, so as to ensure their rapid development. After the accumulation of mega-practices, China has started to actively use laws to ensure its leadership in the AI industry. This paper is devoted to a general moral analysis of the social structure under the influence of this widespread use of AI technology. In terms of analysis, there are two types: conceptual and practical. Many scholars in China confuse the two types and draw a number of conclusions that either exaggerate or miss the focus. Conceptual analysis has cognitive priority, and it is also the main and primary work of jurisprudence. However, conceptual analysis cannot replace practical analysis, nor can it be used directly as an answer to practical questions; these two types of analysis are distinct and relatively independent of each other. Mixing them or restoring practical analysis to conceptual one will undermine the quality of the argument and lead to the most common and fatal fallacy of practical analysis - false factual premises. This paper is a practical analysis of the practical challenges posed by AI technologies. As such, it answers the questions of how the structure of human interactions and social relations has changed under the impact of AI technologies, and what issues in such changes are most important and need to be redressed most urgently. It argues that the “comparative perspective of human and AI” commonly adopted in existing studies is wrong because it is based on false factual premises, which expands the conceptual space of practice analysis too broadly for the practice argument to be really sound and real. This paper turns to “interpersonal relationships under the impact of AI technologies” as the analytical perspective, and uses “asymmetry” as the analytical concept, claiming that the application of AI technology is subject to and reflects the overall morality and fairness of the society in which it is used. Just as the big data used by AI technologies is itself value-biased and can be evaluated as good or bad, algorithmic bias and algorithmic injustice are a technical fact in AI applications. The most jurisprudential social challenge posed by AI today is that it alters the actual state of power domination in democratic relations, which in turn makes the exercise of coercive power less justifiable. In the face of a growing crisis of public trust, both tech giants and governments need to renew their oath of fidelity and win back public trust. This paper concludes by distinguishing between the two kinds of jurisprudence, elaborating the latter as a generalized solution to the grand challenge of the AI era.

(责任编辑:田 夫)