

# 数字孪生算法黑箱的生成机制与治理创新

刁生富<sup>1</sup>, 李思琦<sup>2</sup>

(1.佛山科学技术学院 法学与知识产权学院, 广东 佛山 528000; 2.佛山科学技术学院 马克思主义学院, 广东 佛山 528000)

**摘要:**把“物理实体、连接、数据、虚拟模型、服务”作为数字孪生模型的五个维度,与“输入、处理、输出”算法运作的三层过程结构相结合,提出一个数字孪生算法运作的“五维三构”模型。这一模型有助于理清数字孪生算法黑箱的生成机制并探寻其有效治理的路径。数字孪生机器学习算法的复杂性、数字孪生数据交互融合的动态性、数字孪生技术内嵌价值非中立性等因素,是造成其算法黑箱的内外部原因。数字孪生算法黑箱本质上是由于技术进步所带来的不确定性风险,人们对数字孪生算法决策的担忧也正是源于这种不确定风险的存在。因此,采取以算法解释权的运用应对数字孪生机器学习算法的复杂性、以交流与协作的方式应对数字孪生数据交互融合的动态性、以伦理与法律的规制应对数字孪生技术内嵌价值非中立性的治理路径,以有效减少数字孪生算法黑箱带来的不确定性风险,增强人们对数字孪生算法决策的可信任度。

**关键词:**数字孪生;算法黑箱;“五维三构”模型

中图分类号:TP391

文献标识码:A

文章编号:1008-7699(2022)06-0025-09

作为新一代信息技术的代表性产物,数字孪生(Digital Twin)集成与融合了互联网、大数据、人工智能、物联网、可视化技术等多种新兴技术,愈来愈成为当下及未来科学决策和工程实践的重要依据,并深刻地影响着人们的生产和生活。在依托数字孪生形成决策的过程中,算法运算是核心和关键之所在。然而,算法所具有的“不透明性”所带来的“算法黑箱”问题,使得人们对数字孪生算法形成的决策产生隐忧。本文提出了一个理解数字孪生算法黑箱的“五维三构”模型,并分析了数字孪生算法黑箱的生成机制与相应的治理路径,以期帮助人们增强对数字孪生算法决策的可信任度。

## 一、数字孪生算法黑箱的生成机制

数字孪生是指基于现实世界中的物理实体,在数字化空间中构建其完整映射状态下的全生命周期的虚拟模型,通过集成多学科、多物理性、多尺度的仿真过程,有效实现物理实体与虚拟模型之间的交互反馈与虚实融合,从而达到以虚控实,优化现实物理世界的目的。<sup>[1]</sup>数字孪生具有精准映射、交互融合、动态仿真等特征,使得其算法黑箱的研究更具独特性。

### (一)算法与算法黑箱的界定

算法最开始被界定为“算术运算法则”,<sup>[2]</sup>随后被理解为是包括输入、输出和隐层的一套程序指令或步骤。<sup>[3]</sup>一直以来,人们将算法的概念界定为进行某项工作或解决某种问题时的步骤和方法。<sup>[4]</sup>而随着新一代信息技术的发展,算法概念的内涵和外延有了更深层的发展,算法越来越成为一种决策参考的依据与技术力量的化身。

“黑箱”概念源于1956年英国学者艾什比在《控制论导论》一书中阐述的“黑箱也称闭盒、暗盒或黑

收稿日期:2022-06-08

基金项目:广东省哲学社会科学规划项目(GD22CZX04);广东省哲学社会科学规划项目(GD21CZX03)

作者简介:刁生富(1964—),男,河南南阳人,佛山科学技术学院法学与知识产权学院教授,哲学博士;李思琦(1997—),女,河北唐山人,佛山科学技术学院马克思主义学院硕士研究生。

匣、暗匣,它是指这样一个系统,我们只能看到它的输入值和输出值,而不知道它的内部结构是什么,只有从观察输入变化引起的输出反应,来认识其功能和特性。”<sup>[5]</sup>学者们将其中难以窥探到的隐层称为“黑匣子”。<sup>[6]</sup>自此,“算法黑箱”被界定为算法设计者运用不透明的程序将输入转换为输出的过程,人们只能通过输入和输出来进行理解,而不知道其内部工作原理。<sup>[7-8]</sup>

学术界对算法黑箱的研究大多聚焦在新闻传播算法、人工智能算法等领域,而对于当下及未来的研究热点“数字孪生”较少有专门的“算法黑箱”研究。数字孪生作为一种集成式的技术,呈现出与上述领域不同的算法黑箱新特点。对于这些特点的具体分析不仅有助于更好地认识算法黑箱,而且能够推进数字孪生技术的进一步落地应用。

(二)数字孪生算法黑箱的生成机制

按照著名学者陶飞教授等人在《数字孪生五维模型及十大领域应用》一文中提出的数字孪生五维结构:物理层—连接层—数据层—模型层—服务层,<sup>[9]</sup>有:

$$M_{DT} = (PE, CN, DD, VE, SS) \tag{1}$$

再结合数字孪生算法(A<sub>DT</sub>)的数据处理过程:输入(Ia)—处理(Pa)—输出(Oa),即:

$$A_{DT} = (Ia, Pa, Oa) \tag{2}$$

综合(1)和(2),从系统架构的角度,提出数字孪生与算法之间运作过程的“五维三构”模型,如图1所示。

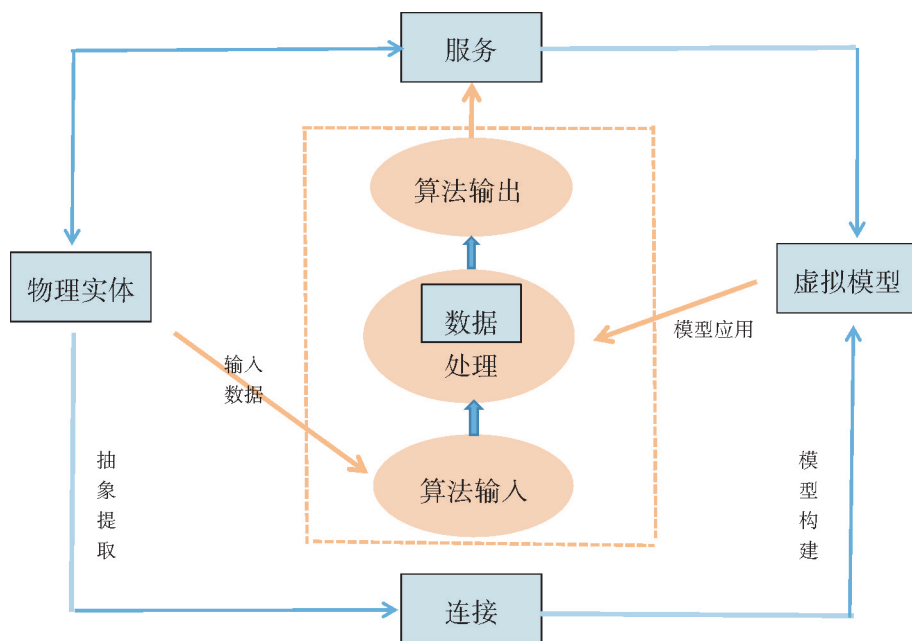


图1 数字孪生算法运作过程——五维三构模型

注:就数字孪生算法黑箱而言,涉及孪生模型构建和孪生模型应用过程中的算法黑箱问题。

从图1可以看出,物理实体、连接、数据、虚拟模型、服务构成了一个闭环的信息物理系统。而在这个闭环的信息物理系统之内,输入、数据(处理)与输出的算法运作过程始终相伴存在。“五维三构”模型从数字孪生模型五个维度和算法运作过程三个层次结构,清晰地表达了数字孪生的算法运作过程,为进一步探究“算法黑箱”问题提供了模型化的表达。

首先,物理实体维度可作为数字孪生算法输入的初始起点环节。虚拟孪生模型构建所需要的各种数据信息和特征集合皆是对物理实体对象的“精准映射”,数字孪生算法初始输入的各种参数信息正是经过对物理实体对象进行特征选择和特征提取之后的结果。这部分初始数据后续会作为基础数据与物理对

象生成的实时数据一同为数字孪生算法处理过程提供数据支撑。同时,数字孪生中的连接维度能将物理实体对象的各种信息和实时动态数据传输给虚拟孪生模型以供建模。因此,来源于现实物理实体的数据信息准确传输或精准映射到计算机数据平台的过程,既是数字孪生算法输入的过程,同时也是相应孪生模型构建的过程。

其次,虚拟模型维度与数据维度共同作为数字孪生算法处理的环节。二者为数字孪生算法处理过程提供了可视化的虚拟载体和必要的的数据支撑。孪生数据是数字孪生运行的核心驱动,从算法运作的初始输入环节——物理实体数据,到算法处理过程中的虚拟模型数据,以及虚实融合、动态交互过程中生成的融合数据,正是这些数据的支撑,保障算法运作最终能够到达算法输出环节,从而生成服务系统数据。数字孪生依托物理实体的各种几何、物理、行为以及规则等数据信息,孪生模型才得以构建。模型构建完成之后,最为关键的便是虚拟模型数据及动态融合数据的算法处理、自动化决策生成的过程,这个过程对于人类目前来说仍处于“未知地带”,而这种“未知地带”被人们形象地称为“算法黑箱”。

最后,服务维度可作为数字孪生算法输出的最后终点环节。数字孪生本质上是一种集成式的技术应用,其最终目的是为了满足不同用户的应用需求,提高事物的运行效率。数字孪生的真正功能在于,在数字化空间中完成对虚拟孪生模型的构建,从而搭建起虚拟孪生世界同现实物理世界之间的桥梁,并在物理实体对象全生命周期内进行数据反馈与虚实融合,经过大量的数据累积与迭代优化,数字孪生能够基于模型优化物理对象,提高物理世界资源配置效率,赋能基于模型的正向研发和协同创新。就数字孪生的具体应用而言,依据不同用户的选择和应用需求,数字孪生的输出服务层可提供仿真模型、智能运算、运行监控、故障诊断等方面的服务供给;还能将用户所需的各类数据、模型、算法、仿真、结果等进行服务化封装,并以应用软件或移动端 App 的形式提供给用户,实现对服务的便捷与按需使用。<sup>[10]</sup>因此,数字孪生模型通过算法运作生成算法决策的过程是为算法输出环节,最终形成服务并应用到现实物理实体中,即为算法运作的终点环节。

实质上,数字孪生的算法输入、处理、输出等环节的界限并非如此分明,它们是相互渗透存在的,且依靠连接维度实现实时动态连接与信息传递,依靠数据与模型双驱动实现数字孪生的虚拟模型建构与智能算法运行。但总体而言,数字孪生的五维模型与算法的三个层次结构紧密相关。其中的数据不仅是孪生模型的核心驱动,更是算法运作过程中的重要数据支撑,而算法数据处理过程又伴随着不为人知的“算法黑箱”。本文接下来将重点深入探讨数字孪生算法黑箱形成的原因与治理路径。

## 二、数字孪生算法黑箱的成因分析

数字孪生算法黑箱的生成,既有源于数字孪生算法本身的机器学习技术的复杂性和数字孪生虚实融合的动态变化等内部原因,又有诸如数字孪生内嵌了人类非向善的价值观念这一外部因素。

### (一)数字孪生机器学习算法的复杂性

在数字孪生算法黑箱的成因分析中,数字孪生本身所运用的机器学习算法的复杂性是造成黑箱生成尤为重要的原因。数字孪生模型构建的初衷是将物理世界中存在的各种数据信息同步于虚拟孪生世界,同时在数字孪生模型中进行各种参数变动与未来预测,从而以较低的成本、更快捷的速度作出符合当下情境的最优决策。其中最重要的一环就是数字孪生虚拟模型的构建与应用,与之始终相伴存在的即为机器学习算法。

机器学习通常被描述为“计算机模拟并实现人类的学习的行为过程,通过向系统提供训练数据或是学习数据,能够自动确定系统的参数。”<sup>[11]</sup>换言之,机器学习通过训练集,不断识别特征与建模,最后形成有效的模型,以达到能够像人类一样作出决策的目的。但正是由于机器学习能够根据训练数据或者学习数据自动计算确定系统运行所需要的参数,在提高决策效率与决策准确率的同时,也不可避免地带来了算法黑箱的问题。

在实现数字孪生的过程中,无论是虚拟模型的构建环节,还是虚拟模型的应用环节,都离不开机器学习技术的应用。机器学习按照训练方法,可分为监督学习、无监督学习、强化学习三种类别,与之相对的数字孪生的算法黑箱的“黑箱程度”也在不断加深。

对于监督式机器学习而言,其对数字孪生虚拟模型的构建发挥着重要作用。数字孪生模型不同于之前的实体模型或数字化建模,它有着更高且更真实的“仿真性”,不仅需要物理实体对象的几何数据、物理数据进行模型构建,更需要对其行为数据、规则数据进行虚拟映射。监督式机器学习是指,人们事先已经明确了自己想要的目标结果,在系统对训练集数据形成自己的判断方法后,就能够对新的测试集进行解答,得到人们想要的结果。<sup>[12]</sup>尤其是针对数字孪生模型构建时的大规模数据采集的过程,借用监督式机器学习算法,能够高效率地将数据类型进行分类,准确地将这些数据信息归为几何类、物理类、行为类、规则类。在这个过程中,算法的输入端和输出端在一定程度上都是能够被确定的,或者说处于人们可掌控的范围之内,此时的算法黑箱可以理解为是从输入值映射到期望输出值的函数处理过程。

但是,对于无监督式机器学习而言,无论是输入数据,亦或是输出数据,乃至中间的机器学习计算过程,人们都无法形成明确的认识。数字孪生在模型建构与决策应用环节,采用的无监督式机器学习算法所造成的算法黑箱的影响尤为严重。由于数字孪生所面对的数据集规模相当之大,若采用监督式机器学习算法,所需要负担的成本相应也会很高。无监督学习的任务是从给定的数据集中挖掘出潜在的结构,相较于监督式机器学习,无监督式机器学习是一个没有标签的数据集,是基于数据之间的相似性进行聚类分析学习。<sup>[13]</sup>这在降低数字孪生模型初始数据的分类成本的同时,不可避免地在系统的输入端与输出端以及中间的学习计算过程形成了一个全流程的闭环,这时的算法黑箱表现为人们无法窥探到其中的任何一个环节,黑箱程度明显加深。

## (二)数字孪生数据交互融合的动态性

当我们从人类认知角度对数字孪生进行界定时,会发现它是一种全新的认识客观世界与人类自身的方法,而这种方法能够得以实现的一大原因是数据的表征与计算。数字孪生虚拟模型的构建实现了数据的表征化,这一过程将现实物理世界中繁杂的特征与信息转换成计算机可以读取和处理的数据,并同步于孪生世界之中(即图1算法输入及孪生模型构建过程);数字孪生虚拟模型的应用实现了数据的计算化,而算法在其中起到了关键性作用(即图1孪生模型运作及算法输出过程)。尤其是学习型算法,它能够在无需人为干预的自然状态下,利用数据进行自动学习,从而具有调整系统运行参数与形成决策规则以及进行未来预测的能力,最终生成各种服务以优化物理世界的运行。

当我们对数字孪生算法黑箱的成因进行深度分析时,就会发现,数字孪生数据交互融合的动态性不仅是数字孪生技术的显著特征,更是造成其算法黑箱形成的另一重要原因。就强化学习而言,数字孪生算法黑箱的黑箱程度已经完全脱离了人类能够掌控的范围。数字孪生在虚拟孪生模型构建完成后,就需要和现实物理世界进行实时的动态交互与虚实融合,在这个过程中,数字孪生系统内会生成大量以供决策制定的经验参考。强化学习是一种常用于“时变系统控制规则构建”的机器学习方法,这种学习方法一般是将机器置于一个能够不断进行反复试验的环境当中,通过对经验的自动学习与重复试错,最后形成一套属于自己的决策方法,同时还能够从经验中学习最新的知识以对未来情况作出预测判断。<sup>[14]</sup>强化学习重点关注的是机器与环境之间的交互与反馈,系统会根据强化学习的效果给予正奖励或负惩罚。

数字孪生虚拟模型在进行应用之时,即通过虚拟孪生数据与现实物理数据进行的实时交互以作出决策的过程中,强化学习正是通过学习这些从连续实时监控中收集到的几何、物理、行为与规则数据,并已从环境和系统中得到的正奖励作为生成动态决策的依据。数字孪生数据交互融合的动态性使得算法运行的难度较之前更为复杂化,数字孪生算法在应用过程中不仅要考虑系统之内的虚拟孪生数据的状态与特征,更要结合系统之外的客观物理世界中发生的实时动态变化的数据信息,如此才能作出符合当下情境、适合数字孪生主体的准确决策。但也正是由于这种数据交互融合的动态性,使得数字孪生算法黑箱



的黑箱程度更为加深。

例如,当数字孪生应用于人类自身构建数字孪生人之时,算法运行的过程贯穿于生命体“全生命”周期之内。现实世界中的生命体一定会随着环境的变化而发生变动,随之而来的就是与生命体相关的数据的变动与更新。这种变动会随时传送到已经构建的数字孪生虚拟人中,当算法想要作出符合数字孪生主体的决策之时,就不得不考虑二者之间的融合与交互。正是这种交互融合,大大加深了算法决策的复杂性,从而进一步加深了算法黑箱的黑箱程度。

### (三)数字孪生技术内嵌价值非中立性

如果将数字孪生机器学习算法的复杂性以及数据交互融合的动态性所造成的数字孪生的算法黑箱问题归结为是数字孪生技术及算法本身所造成的,那么数字孪生技术内嵌的价值非中立性则可归结为是外部人为因素加深了数字孪生算法黑箱的黑箱程度。由于人之决策不仅容易受到“判断误差”的影响,更容易受到诸如认知程度、情绪波动、环境变化等“判断噪声”的影响,带有极大的主观性;而算法决策通过算法程序的运行,具有高度的专业性与客观程序性以及极强的工具理性等特征。正是凭借着算法具备的决策高效率与高度精准化的特征,现如今越来越多的决策是依靠算法形成而非人为进行,人之决策逐渐让位于算法决策。<sup>[15]</sup>当下的数字孪生技术更是将这种算法决策推向了更高的维度,通过构建现实物理世界中的虚拟孪生体,依托二者之间进行的实时的数据交互与虚实融合,人们期待数字孪生算法能够作出符合现实物理世界中主客体需要的决策。但我们往往会发现,在实现这种“准确且无偏”决策的过程中,工具化和技术化的算法决策往往由于内嵌了人类的价值观念而与人们最初的期待背道而驰。

数字孪生作为一种集成式的技术应用,本应该遵循着“技术中立”的原则,却因为种种原因而无法实现其自身的“中立属性”。“技术中立”这一概念主要包括功能中立、责任中立和价值中立三个方面。<sup>[16]</sup>我们围绕数字孪生算法决策,针对其中的“价值中立”,进一步探究数字孪生算法黑箱形成的原因。

数字孪生在对现实物理客体进行模型构建之时,首先需要的就是大规模原始数据的输入,在对数据进行处理的过程中,特征选择与特征提取是必不可少的环节。而人们对这些规模庞大、纷繁复杂的数据进行处理,势必要按照“人为的”价值观念和要求进行筛选与判定。由此,数据在初始输入环节就已经内嵌和负载了人为的价值选择与价值判断。从数字孪生技术的算法设计者角度而言,若出于商业利益考量,他们在很大程度上会出于保护商业秘密而在特征提取之际选择不公开这部分数据信息。由此一来,这在一定程度会由于人为因素而加深数字孪生算法黑箱的黑箱程度。

数字孪生在对虚拟模型进行算法应用的过程中,虽然从表面看来,人之决策确实让位于算法决策,但从更深层次来看,这个过程中恰恰内嵌了人类主体的价值选择。就数字孪生机器学习方式而言,无论是监督学习,还是无监督学习,乃至强化学习,在其背后都隐藏着要将模型训练成为符合人类意志要求、作出满足人类需求决策的期许。而当这些人为赋予的价值观念一旦失去约束与掌控,其带来的不仅是加剧数字孪生算法黑箱的黑箱程度,更会是不可想象的风险挑战。

## 三、数字孪生算法黑箱的治理路径

在以往人工智能的算法黑箱治理分析中,人们通常向往采取“打开黑箱”的方法来解决这一问题,但这一方法存在着很多难以克服和解决的实操性难题。因此,对于数字孪生算法黑箱问题来说,本文关注的重点不是要去如何“打开黑箱”,而是将视线投放于人们依据数字孪生进行决策时如何减少对不确定性风险的担忧以及如何增强对决策的信任度上。

### (一)以算法解释权的运用应对数字孪生机器学习算法的复杂性

在上述讨论数字孪生算法黑箱的生成原因时,首先归因于数字孪生机器学习算法的复杂性,这是由于数字孪生应用算法技术本身造成的技术黑箱。我们分析了对于监督式机器学习而言,无论是数据的输入还是输出,都尚且在人们可掌控的范围之内,此时的数字孪生算法黑箱被界定为“从输入值映射到期望

输出值的函数处理过程”,其黑箱带来的“不确定性风险”尚不足以构成人们的担忧与恐惧;而到了无监督式机器学习,由于算法输入的是“不带标签”的规模数据,且人们不明确算法的数据输出,更不知晓算法运作过程的具体机制,此时的数字孪生算法黑箱被界定为“从输入端到输出端的全流程闭环的未知地带”,其黑箱带来的“不确定性风险”足以让人们们对数字孪生算法决策产生隐忧与畏惧。<sup>[17]</sup>

数字孪生算法能够从数据本身出发,构建虚拟孪生模型,通过深度学习与迭代优化,自动生成高级的认知结果,从而作出能够影响现实物理世界的决策。面对这种算法技术本身所造成的“不确定性风险”,就目前人类的认知水平与应对能力而言,不得不承认,人们无法打破黑箱去窥探其内里,但我们不能任由算法肆意运行。数字孪生算法作出的决策与人们的生产生活息息相关,甚至在某种程度上将起到决定性作用,人们自然会想要了解其是“如何作出决策的?”以及将“如何影响我们的生产生活?”这时,算法解释权的运用在一定程度上将有助于消解人们对数字孪生算法的疑团,缓解数字孪生因机器学习算法的复杂性而带来的黑箱问题。

数字孪生技术下的算法解释权指的是,当算法依靠虚拟孪生数据作出的自动决策应用到客观物理实体时,与该客观物理实体相关的任一人类主体都有权要求算法设计及实施者对这一过程作出“适当的解释”。我们在前面已经论述过,人们目前其实无法真正破解算法黑箱难题,即使人们采用将算法决策过程中的所有数据源及代码都公开的方法,受影响的相关方由于个人认知能力局限,也不能完全理解算法的运算与决策过程。事实上,只有从“适当的解释”这一思路出发,才能够提高人们对数字孪生算法决策的认识与理解程度。算法解释权所遵循的“适当的解释”是以一种能够被受影响的相关方理解的方式,<sup>[18]</sup>解释数字孪生算法决策到底是如何形成,以及对人们的生产生活又会造成哪些影响。其本质意义上是在追求一种理性的结果,是在与自身利益进行权衡之后,对算法决策的理解与接受,这在很大意义上能够减少人们对数字孪生算法所作出的自动决策带来的不确定性风险的担忧。从某种程度而言,算法解释权的运用就已经缓解了因机器学习的复杂性而造成的数字孪生算法黑箱问题。

具体而言,数字孪生算法解释权“适当的解释”的解释范围应当以虚拟孪生体与客观物理实体之间的虚实融合为重点而展开,人们期待知晓数字孪生算法从数据输入到算法运作及算法输出过程环节的简单操作,期待了解数字孪生是如何通过虚拟孪生模型的构建来影响人们的生产生活的。因此,这种“适当的解释”应该涵盖对整个算法过程的解释,以及对算法具体决策的解释。<sup>[19]64-65</sup>当受影响的相关方对于算法所作出的决策难以理解或者与自己预期不符,甚至出现严重损害自身利益的情况时,就需要数字孪生技术的算法设计与实施方能够针对种种疑惑提供相应的解释,这一方面有助于消解公众对于算法作出的适当决策的质疑而不至于否定其正确性;另一方面也有利于对算法决策确实出现偏颇与失误的情况下能够采取及时的救济。就数字孪生算法解释权“适当的解释”的展开形式而言,可以采用可视化、可交互的举例解释。<sup>[19]65</sup>尤其是对于涉及个人切身利益时,如当数字孪生虚拟孪生人所作出的决策应用到客观生命体自身时,人们往往会更加谨慎也更倾向于得到更清晰准确的解释。此时就可以通过举例解释的方式,采用诸如图形文字、动画演示等可视化的形式,具体生动地将虚拟孪生体与客观生命体进行虚实融合的交互过程展现给受影响的相关方。这种具体的举例解释其实是帮助人们能够以一种更加理性的方式对数字孪生算法作出的决策进行利益权衡,有助于增加人们对算法决策的可信任度,缓解人们对数字孪生算法黑箱的担忧。

## (二)以交流与协作的方式应对数字孪生数据交互融合的动态性

无论是现实中的客观物理实体,还是系统内的数字孪生虚体,乃至其中算法的设计与实施过程,都处于一个不断变化与更新的状态之中。物理实体和虚拟孪生体之间的虚实交互与融合是数字孪生技术的一大显著特征,正是这种动态性变化加重了算法运算的复杂程度,进而加剧了算法黑箱的黑箱程度。那么如何才能更好地缓解这种因虚实融合的动态性变化而造成的算法黑箱呢?加强物理实体与数字孪生体之间的交流与协作不失为有效且可行的方法之一。

数字孪生作为一项技术手段,其想要达到的效果是将现实物理客体的特征与信息统统转换成计算机可读取的数据。当数字孪生技术在未来实现全面落地应用之时,人们对现实物理客体的研究通常不需要直接面向对象本身,而是在数字孪生虚拟模型中进行数据分析与推演运算,依靠数字孪生算法得出准确且可靠的结论。而如何保证该研究结论的可信性,则需要以现实世界中物理实体的变动情况为衡量标准,即当依据数字孪生算法作出的决策应用到现实物理实体,能够促使其向好的方向发展时,便说明该算法决策是值得信任的。这种信任度的达成始终是围绕着现实物理实体与虚拟孪生体之间的交流与协作展开的。这可以从两个方向维度来具体理解:一是从现实物理实体到虚拟孪生体数据输入的过程(即图1所示的算法输入过程);二是从虚拟孪生模型的算法决策输出到应用于现实物理实体的过程(即图1所示的算法输出过程)。

首先,从物理实体到虚拟孪生体这一方向维度来看,在现实物理世界中的数据和信息传递到虚拟孪生体的过程当中,对于关涉人类隐私与切身利益的数据特征提取,就需要现实客体与虚拟孪生体之间的交流与协作。通过充分的交流与沟通,找到一种能够在维护人们隐私权的同时尽可能地提取到能够依据其进行算法决策的数据特征。与此同时,也需要提高人们的算法素养,此时的算法素养尤其是指人们在多大限度内能够允许涉及个人切身利益的算法数据的搜集,这个过程需要进行不断的交流沟通与利益博弈。

另一方面,从虚拟孪生模型的算法应用这一方向维度来看,首先要明确数字孪生技术所作出的算法决策的最终目的是为现实世界中的主客体服务,是为满足人类的发展和社会的进步。因此,数字孪生算法在形成决策过程中,一定要同现实世界中的实体进行充分的交流与数据反馈。因为现实世界总是处于动态变化过程之中,如果算法决策忽视了这一点,那么其作出的算法决策就会因为失去时效性而不能满足人们的需求。与此同时,在数字孪生算法决策形成之后,其能否满足现实客体需要,能否促进人们生产生活的优化提升,是对算法决策制定好坏进行评判的标准。当应用效果显著时,就形成了正向的积极反馈,将这种正向反馈继续应用到日后的算法决策当中,那么数字孪生这种虚实融合的动态性特征将会提升人们对算法的信任程度,从而减轻因其造成的算法复杂性而加深的算法黑箱的影响程度。

### (三)以伦理与法律的规制应对数字孪生技术内嵌价值非中立性

数字孪生技术本身内嵌了人类与社会所期许的价值观念,而一旦这些人为赋予的价值观念出于商业利益的考量或是被个人私欲所利用,就会加剧算法黑箱程度,从而导致数字孪生算法作出的决策损害公众利益,此时就需要人们采取一些行动来减少此类行为的发生。

数字孪生技术赋能的实现离不开伦理底线的构建。对数字孪生算法技术及算法黑箱的探究,恰恰就是在厘清技术发展的旨趣,构筑技术与善的内在一致性。<sup>[20]</sup>从宏观角度来看,数字孪生归根结底是一项能够帮助人们更便捷认识世界、更快捷作出决策的技术手段,这种技术手段的应用不能突破社会伦理的底线。而在数字孪生技术的应用实践过程中,往往会出现出于商业利益考量而植入“非向善”的价值观念的情况。这种情况的出现“并非简单的是技术不够成熟的表现,究其本质而言,是由技术工具论的伦理维度缺席所致”。<sup>[21]</sup>而对数字孪生算法技术及黑箱问题的伦理反思和治理,也在推动着数字孪生技术及服务的不断优化与迭代升级。

因此,针对数字孪生技术内嵌的“非向善”的价值观念,要发挥机器算法在效率上的长处和人脑算法在结合价值计算后的审时度势优势。值得注意的是,这里的“人脑算法在结合价值计算”,指的是符合人类合理利益需求与社会发展趋势的“底线价值”。这种技术价值寻求的是机器算法在事实上的求真与人脑算法在道德上的求善之间的最佳协调。<sup>[22]</sup>

对于数字孪生技术而言,在可以预见的未来,当与人们息息相关的决策都依托数字孪生虚拟模型作出的时候,由于算法黑箱的客观存在,导致人们对数字孪生算法决策已经产生了很大程度上的担忧,我们更不能任由人类因个体私欲而在数字孪生技术中内嵌自己非向善的价值观念。对于数字孪生算法技术



的设计与实施者而言尤为重要,因为他们不仅承担着要帮助人们作出高效而准确决策的责任,更承担着不加刷算法黑箱程度的责任。因此,对于数字孪生算法技术的从业人员以及相关利益方而言,守住道德伦理底线是数字孪生技术在未来落地应用的题中应有之义。

此外,还可以采取“算法标准化”等专项法律法规的形式,对数字孪生算法的日常运作进行合理监督与检查。我们可以类比上市公司的审计制度与药品食品的生产流程,无论是上市公司,还是药品食品生产过程,其背后都隐藏着“黑箱”的含义。但通过第三方公允专业的审计以及药品食品监察机构对标准流程的监督的方法,在一定程度上是可以消解人们的质疑,取得公众信任的。因此,对于在全流程中始终伴随着人类价值观念的数字孪生技术而言,也可以采取这种标准化的手段,对算法运作从初始环节到形成决策的全流程建立算法标准,以及采用第三方监督的形式。<sup>[23]</sup>这在一定程度上可以减少人们非向善的价值观念的嵌入,缓解因人为因素而加剧的算法黑箱的黑箱程度。

### 参考文献:

- [1]GLAESSGEN E H, STARGEL D S. The digital twin paradigm for future NASA and U.S. air force vehicles[C]// Proceedings of the 53<sup>rd</sup> structures dynamics and materials conference, Special session on the digital twin. Reston, Va., USA: AIAA, 2012: 7.
- [2]张淑玲. 破解黑箱: 智媒时代的算法权力规制与透明实现机制[J]. 中国出版, 2018(7): 49.
- [3]卢西亚诺·弗洛里迪. 第四次革命: 人工智能如何重塑人类现实[M]. 王文革, 译. 杭州: 浙江人民出版社, 2016: 107.
- [4]仇筠茜, 陈昌凤. 基于人工智能与算法新闻透明度的“黑箱”打开方式选择[J]. 郑州大学学报(哲学社会科学版), 2018(5): 84.
- [5]W. R. 艾什比. 控制论导论[M]. 张理京, 译. 北京: 科学出版社, 1965: 53.
- [6]陶迎春. 技术中的知识问题——技术黑箱[J]. 科协论坛, 2008(7·下半月刊): 54.
- [7]BURRELL J. How the machine “thinks”: Understanding opacity in machine learning algorithms[J]. Big data & society, 2016(1): 2.
- [8]VON ESCHENBACH W J. Transparency and the black box problem: Why we do not trust AI[J]. Philosophy & technology, 2021(4): 1607.
- [9]陶飞, 刘蔚然, 张萌, 等. 数字孪生五维模型及十大领域应用[J]. 计算机集成制造系统, 2019(1): 4-6.
- [10]陶飞, 张贺, 戚庆林, 等. 数字孪生十问: 分析与思考[J]. 计算机集成制造系统, 2020(1): 7.
- [11]BOTTOU L. From machine learning to machine reasoning: An essay[J]. Machine learning, 2014(2): 133.
- [12]PRAVEENA M, JAIGANESH V. A literature review on supervised machine learning algorithms and boosting process[J]. International journal of computer applications, 2017(8): 32.
- [13]USAMA M, QADIR J, RAZA A, et al. Unsupervised machine learning for networking: Techniques, applications and research challenges[J]. IEEE Access, 2019, 7: 65580.
- [14]LI Y, DU X Y, XU Y H. Research and analysis of machine learning algorithm in artificial intelligence[J]. Artificial intelligence advances, 2020(2): 89.
- [15]汪德飞. 算法伦理争论的六重维度及其走向[J]. 科学技术哲学研究, 2021(4): 59.
- [16]丁晓东. 论算法的法律规制[J]. 中国社会科学, 2020(12): 141.
- [17]谭九生, 范晓韵. 算法“黑箱”的成因、风险及其治理[J]. 湖南科技大学学报(社会科学版), 2020(6): 93.
- [18]姜野, 李拥军. 破解算法黑箱: 算法解释权的功能证成与适用路径——以社会信用体系建设为场景[J]. 福建师范大学学报(哲学社会科学版), 2019(4): 91-92.
- [19]许可, 朱悦. 算法解释权: 科技与法律的双重视角[J]. 苏州大学学报(哲学社会科学版), 2020(2): 64-65.
- [20]闰宏秀. 数据赋能的伦理基质[J]. 社会科学, 2022(1): 137.
- [21]FLORIDI L. Infraethics-on the conditions of possibility of morality[J]. Philosophy & technology, 2017(4): 392.
- [22]肖峰. 人工智能与认识论新问题[J]. 西北师大学报(社会科学版), 2020(5): 40.
- [23]袁康. 可信算法的法律规制[J]. 东方法学, 2021(3): 16.



## Generation Mechanism and Governance Innovation of Digital Twin Algorithm Black Box

DIAO Shengfu<sup>1</sup>, LI Siqu<sup>2</sup>

(1. School of Law and Intellectual Property, Foshan University, Foshan, Guangdong 528000, China;

2. School of Marxism, Foshan University, Foshan, Guangdong 528000, China)

**Abstract:** A five-dimension three-layer model for digital twin algorithm operation is proposed by combining five dimensions of “physical entity, connection, data, virtual model and service” and three-layer architecture of “algorithm input, processing and output”, which is helpful to clarify the generation mechanism of digital twin algorithm black box and explore its effective governance path. The complexity of the digital twin machine learning algorithm, the dynamics of the digital twin data interaction and fusion, and the non-neutrality of the embedded value of the digital twin technology are the internal and external reasons for the black box of the digital twin algorithm. Therefore, in order to effectively reduce the uncertainty of digital twin algorithm black box and enhance the reliability of digital twin algorithm decision, we can adopt algorithm interpretation right to deal with the complexity of the digital twin machine learning algorithm, adopt the communication and collaboration approach to deal with the dynamics of digital twin data interaction and fusion, and adopt the ethical and legal regulation to deal with non-neutrality of the embedded value of the digital twin technology.

**Key words:** digital twin; algorithm blackbox; a five-dimension three-layer model

(责任编辑:江 雯)

(上接第 24 页)

## TripleLogic of PostAcademic Science and Its Integration

XUE Guibo

(College of Marxism, Nanjing Forestry University, Nanjing 210037, China)

**Abstract:** In the postacademic science era, science has assumed an important social function as a key resource for market interests and major political decisions and diversification of scientific research types tends to be more prominent. Academic logic, business logic and policy logic are deeply intertwined and embedded in the whole production and practical application of scientific knowledge, and become the internal logical dimension and powerful driving force of scientific progress in the postacademic science era. In order to promote the development of science and better serve the society, it is necessary to coordinate and integrate the “triple logic” of post-academic science. In particular, with the rapid development of emerging technologies, scientific progress is deeply related to social factors to an unprecedented extent, and various scientific risk issues and scientific ethics issues are becoming increasingly complex. Attaching importance to the coordination and integration of the “triple logic” of post academic science and improving the system and mechanism of scientific research and innovation will help optimize the governance of science and technology ethics, which will provide support for further promoting science and technology and realizing high level self-reliance in science and technology.

**Key words:** post-academic science; academic logic; business logic; policy logic

(责任编辑:江 雯)