

# 人机智能的分界

刘伟

北京邮电大学人工智能学院, 北京 100876

**摘要** 迄今人工智能在智能水平和能力范围上与人类相比仍存在极大的差距,这使得它的发展受到了阻碍。究其原因,人类智能和机器智能之间存在无法跨越的界限。依赖于符号指向对象的机器只能进行封闭环境下的形式化计算,却无法像人那样实现开放环境下的意向性算计,在这种背景下,人机优势互补显得尤为重要。而当前人机的关系主要是硬性的功能分配,未来数字世界中人机关系更可能是柔性的能力分工。探讨了人工智能的瓶颈、智能的第一原理等问题,思考了其对人机问题的影响,指出人机混合智能才是人工智能未来的发展方向的观点。

**关键词** 人工智能;功能分配;算计;人机混合智能

1997年深蓝的胜利,重燃人们对人工智能的兴趣。2006年,杰弗里·辛顿(Geoffrey Hinton)提出深度置信网络,使深层神经网络的训练成为可能,也使深度学习迎来了春天。2011年,国际商业机器公司(International Business Machines Corporation, IBM)的Watson参加“危险边缘”问答节目,并打败了2位人类冠军,轰动一时。2012年,辛顿的学生Alex Krizhevsky使用AlexNet以大幅度的优势取得ImageNet图像分类比赛的冠军,深度神经网络开始大放异彩。同年,运用深度学习技术的谷歌大脑(Google Brain)通过观看数千段视频后,自发找出了视频中的猫。2016年,Google DeepMind的AlphaGo(阿尔法狗)战胜了世界顶级围棋高手李世

石,推动人工智能再一次发展。此后,“AlphaZero”“MuZero”“AlphaFold”等一系列算法陆续出现,引发人工智能将如何改变人类社会生活形态的话题。目前正处于人工智能发展的第3次高潮期。当前国内对于人机混合智能的研究与应用如火如荼,但对于人机智能分界的研究却寥寥无几。

智能不是人脑(或类脑)的产物,也不是人自身的产物,而是人、物、环境系统相互作用的产物,正如马克思所言:“人的本质不是单个人所固有的抽象物,在其现实性上,它是一切社会关系的总和。”比如狼孩尽管具有人脑的所有结构和组成成分,但没有与人类社会环境系统的交流或交互,也不可能有人类的智能和智慧。事实上,真实的智能同样蕴含

收稿日期:2023-02-06;修回日期:2023-03-08

作者简介:刘伟,特聘研究员,研究方向为人机混合智能、认知工程、人机环境系统工程、未来态势感知模式与行为分析/预测技术等,电子邮箱:twlw@163.com

引用格式:刘伟. 人机智能的分界[J]. 科技导报, 2023, 41(13): 118-128; doi: 10.3981/j.issn.1000-7857.2023.13.012

着人、物、环境3种成分,随着科技的快速发展,其中的“物”逐渐被人造物——“机”所取代,简称为人机环境系统(为简化,后面均以“人机环境系统”一词代替“人物环境系统”一词)。平心而论,人工智能要超越人类智能,在现有数学体系和软硬件设计模式基础之上,基本上不大可能,但在人机一体化或人机环境系统中却有着可能性<sup>[1]</sup>。人工智能是逻辑的,智能则不一定是逻辑的。智能是一个非常辽阔的空间,它可以随时打开异质的集合,把客观的逻辑与主观的超逻辑结合起来。

研究复杂问题是困难的,但把它分解成人机环境系统问题就相对简单一些,至少可以从人、机、环境角度去思考理解;研究智能这个复杂问题也是困难的,但同样可把它分解成人机环境系统问题研究分析处理,“人”所要解决的是“做正确的事(杂)”,“机”所要解决的是“正确地做事(复)”,“环境”所要解决的是“提供做事平台(复杂)”。郭雷指出:“复杂性和智能化是系统科学发展的两个永恒的主题。复杂性主要体现的是系统科学认识世界的一面,而智能化主要体现的是系统科学改造世界的一面。”

2020年和2021年注定是人类难忘的年份,这2年除席卷全球的新冠疫情外,还出现一个奇特的现

象,即人件、软件、硬件、环件等智能化条件均属世界第1的美国新冠感染者世界排名第1,死亡人数也是世界排名第1。不难看出,对于人+机+环境系统而言,美国不但不是第1,而且是规模性失调。所以,中国和美国角力的焦点不仅仅是人、机、环境每一项或所有单项人工智能的领先优势,更重要的是人、机、环境系统混合智能的整合。下面针对人机智能分界问题展开分析和探讨。

## 1 人工智能的瓶颈:总想用逻辑的手段解决非逻辑问题

现在经常有人问这样的问题:未来数字世界中,人与智能机器是何种分工模式?人与机器的边界将如何划分?

实际上,当前人机的关系主要是功能分配,人把握主要方向,机处理精细过程;而未来的人机关系可能是某种能力的分工,机也可以把握某些不关键的方向,人也可以处理某些缜密的过程。人机的边界在于should——“应”和change——“变”,即如何实现适时的“弥”(散)与“聚”(焦)、“跨”(域)与“协”(同)、“反”(思)与“创”(造)(图1)。

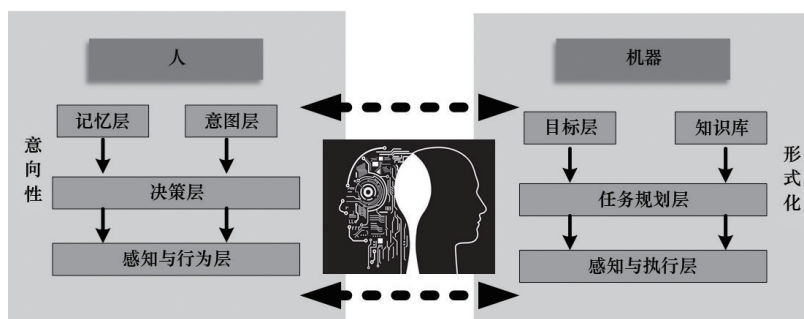


图1 人机融合智能结构

人类学习的秘密在于数据信息知识的弥散与聚焦(弥聚),人类使用数据信息知识的秘密在于跨越与协同(跨协),人类智能的核心在于反思与创造(反创)。人由内外2种态势感知系统(situation awareness, SA)耦合而成,共振时最强,抵消时最弱,另外还有一个非智能(即智慧)影响决策系统:想不想、愿不愿、敢不敢、能不能……这些因素虽在

智能领域之外,但对智能的影响很大。外在的SA是联结客观环境的眼耳鼻舌身等客观事实通道,内在的SA是联结主观想象环境的知情意等主观价值通道,阿尔法狗试图完成主观价值的客观事实化,可惜只完成了封闭环境下的形式化计算,没有完成开放环境下的意向性算计,原因在于传统映射思想是确定性的同质对应,远没有不确定性异质散射、

漫射、影射的跨域变尺度的对应机制出现。

智能领域的真正瓶颈和难点之一是人机环境系统失调问题,具体体现在跨域协同中的“跨”与“协”如何有效实现的问题,这不但关系到解决各种辅助决策系统中“有态无势”(甚至是“无态无势”)的不足,而且还涉及到许多辅助决策体系“低效失能”的溯源。也许需要尝试把认知域、物理域、信息域构成的基础理论域与陆海空电网构成的技术域有机地结合起来,为实现跨域协同中的真实“跨”与有效“协”打下基础<sup>[2]</sup>。人工智能中的强化学习无法实现人类强化学习后的意图隐藏(比如小孩被强制学习或受到惩罚后,表面上顺从但实际上是隐匿想玩的意图);另外,那些因为做了一项任务而得到奖励的人,可能没有那些因为做同样的任务而没有得到奖励的人愉快,这是因为他们把他们的参与仅仅归因于奖励而不是情感与体验。机器深度学习容易实现局部优化却很难实现全局优化和泛化等等。

电脑先驱阿兰·凯伊(Alan Kay)指出:“预测未来的最好办法就是创造未来。”判断力和洞察力,是广域生存最核心的竞争优势。判断力和洞察力,常基于“直觉”。正是这样的直觉,使“企业家”完全不同于“管理者”,使“军事家”完全不同于“指挥员”,使“优秀者”不同于“普通者”。

## 2 智能的第一原理

### 2.1 计算与算计

休谟认为:“一切科学都与人性有关,对人性的研究应是一切科学的基础。”任何科学都或多或少与人性有些关系,无论学科看似与人性相隔多远,它们最终都会以某种途径再次回归到人性中。科学尚且如此,包含科学的复杂也不例外,其中真实的智能有着双重含义:一个是事实形式上的含义,即通常说的理性行动和决策的逻辑,在资源稀缺的情况下,如何理性选择,使效用最大化;另一个是价值实质性含义,既不以理性的决策为前提,也不以稀缺条件为前提,仅指人类如何从其社会和自然环境中谋划,这个过程并不一定与效用最大化相关,更大程度上属于感性范畴。理性的力量之所以有

限,是因为真实世界中,人的行为不仅受理性的影响,也有“非理性”的一面<sup>[3]</sup>。人工智能“合乎伦理的设计”很可能是黄粱一梦,原因很简单,伦理对人而言还是一个很难遵守的复杂体系。简单的伦理规则往往是最难以实现的,比如应该帮助处在困难中的人,这就是一条很难(遵守者极容易上当受骗)操作的伦理准则。对于AI这个工具而言,合乎伦理设计应该科幻成分多于科学成分、想象成分多于真实成分。

当前的人工智能及未来的智能科学研究具有2个致命的缺点:(1)把数学等同于逻辑;(2)把符号与对象的指涉混淆。所以,人机混合深度态势感知的难点和瓶颈在于:(1)(符号)表征的非符号性(可变性);(2)(逻辑)推理的非逻辑性(非真实性);(3)(客观)决策的非客观性(主观性)。

智能是一个复杂的系统,既包括计算也包括算计,一般而言,人工(机器)智能擅长客观事实(真理性)计算,人类智能优于主观价值(道理性)算计。当计算大于算计时,可以侧重人工智能;当算计大于计算时,应该偏向人类智能;当计算等于算计时,最好使用人机智能(图2)。费曼认为:“物理学家们只是力图解释那些不依赖于偶然的事件,但在现实世界中,我们试图去理解的事情大多取决于偶然。”但是人、机两者智能的核心都在于:变,因时而变、因境而变、因法而变、因势而变……

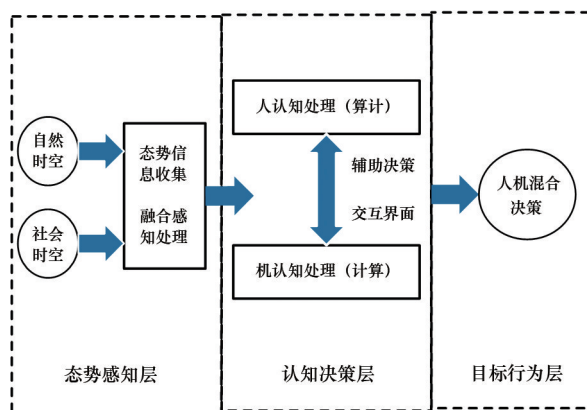


图2 计算—算计模型

如何实现人的算计(经验)与机的计算(模型)混合后的“计算计”(计算+算计)系统? 太极八卦图就是一个典型的“计算计”系统,有算有计,有性

有量,有显有隐,计算交融,情理相依(图3)。其中的“与或非”逻辑既有人经验的、也有物(机)数据的,即人价值性的“与或非”+机事实性的“与或非”,人机混合智能及深度态势感知的任务之一就是要打破“与或非”的狭隘,比如大与、小与,大或、小或,大非、小非……大是(being)、大应(should)、小是、小应。人的经验性概率与机器的事实性概率不同,它是一种价值性概率,可以穿透非家族相似性的壁垒,用其他领域的成败得失结果影响当前领域的态势感知,如同情、共感、同理心、信任等。



图3 人机环境系统智能“计算”机理示意

人类智能的核心是意向指向的对象,机器智能的核心是符号指向的对象,人机智能的核心是意向指向对象与符号指向对象的结合问题<sup>[4]</sup>。它们都是对存在的关涉,存在分为事实性的存在、价值性的存在、责任性的存在。

一般而言,数学解决的是等价与相容(包涵)问题,然而这个世界的等价与相容非常复杂,客观事实上的等价与主观价值上的等价常常不是一回事,客观事实上的相容与主观价值上的相容往往也不是一回事,于是世界应该是由事实与价值共同组成的,也即除了数学部分之外,还由非数之学部分构成,科学技术是建立在数学逻辑(公理逻辑)与实验验证基础上的相对理性部分,人文艺术、哲学宗教则是基于非数之学逻辑与想象揣测之上的相对感性部分,二者的结合使人类在自然界中得以不息地存在着。

某种意义上,数学就是解决哲学上“being”(是、存在)的学问(如 $1/2, 2/4, 4/8$ ……等价、包涵问题),但其远远没有,也不可能解决“should”(应、义)的问题。例如,当自然哲学家企图在变动不居的自然中寻求永恒不变的本原时,巴门尼德却发现,没有哪种自然事物是永恒不变的,真正不变的只能是“存在”。在一个判断中(“S是P”),主词与宾词都是变动不居的,不变的惟有这个“是”(being)。换言之,一切事物都“是”、都“存在”,不过其

中的事物总有一天将“不是”“不存在”,然而“是”或“存在”却不会因为事物的生灭变化而发生变化,它是永恒不变的,这个“是”或“存在”就是使事物“是”或“存在”的根据,因而与探寻时间上在先的本原的宇宙论不同,巴门尼德所追问的主要是逻辑上在先的存在,它虽然还不是但却相当于我们所说的“本质”。这个“是”的一部分也许就是数学。

人机环境之间的关系有有向闭环、无向开环、有向开环、无向闭环,自主系统大多是一种有向闭环行为。人机环境系统混合的“计算”系统也许就是解决休谟之问的一个秘密通道,即通过人的算计结合机器的计算实现了从“事实”向“价值”的“质的飞跃”。

有人认为:“全场景智慧是一个技术的大混合。”实际上,这是指工程应用的一个方面,如果深究起来,还是一个科学技术、人文艺术、哲学思想、伦理道德、习俗信仰等方面的人物环境系统大混合,如同这次抗疫<sup>[5]</sup>。较好的人机交互关系如同阴阳图一样,你中有我,我中有你,相互依存,相互平衡。简单地说,目前人机关系就是两条鱼,头尾相连,黑白相间。

每个事物、每个人、每个字、每个字母……都可以看成一个事实+价值+责任的弥聚子,心理性反馈与生理性反馈、物理性反馈不同。感觉的逻辑与知觉的逻辑不同,易位思考,对知而言,概念就是图型,对感而言,概念就是符号。从智能领域上看,没有所谓的元,只有变化的元,元可以是一个很大的事物,比如太阳系、银河系都可以看成一个元单位,称为智能弥聚子。

科学家常常只是力图解释那些不依赖于偶然的事件,但在现实世界中,人机环境系统工程试图去理解的事情大多取决于一些偶然因素,如同人类的命运。维特根斯坦(Ludwig Wittgenstein)就此曾评论:“在整个现代世界观的根基之下存在一种幻觉,即:所谓的自然法则就是对自然现象的解释。”基切尔也一直试图复活用原因解释单个事件的观点,可是,无穷多的事物都可能影响一个事件,究竟哪个才应该被视作它的原因呢?更进一步讲,科学永远都不可能解释任何道德原则。在“是”与“应”

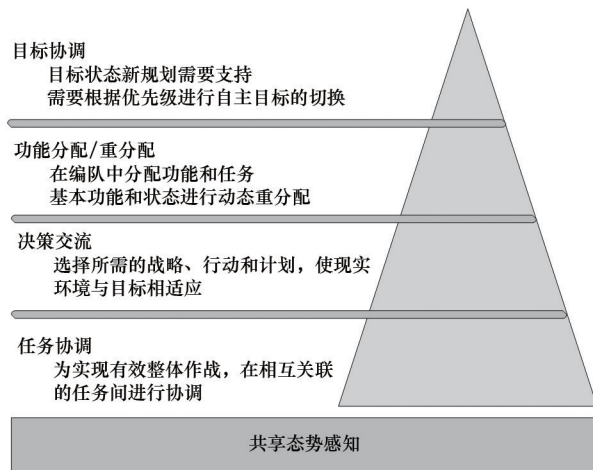
的问题之间似乎存在一道不可逾越的鸿沟<sup>[6]</sup>。或许我们能够解释为什么人们认为有些事情应该做,或者说解释为什么人类进化到认定某些事情应该做,而其他事情却不能做,但是对于我们而言,超越这些基于生物学的道德法则依然是一个开放的问题。牛津大学教授彭罗斯也认为:“在宇宙中根本听不到同一个节奏的‘滴答滴答’声响。一些你认为将在未来发生的事情也许早在我的过去就已经发生了。2位观察者眼中的2个无关事件的发生顺序并不是固定不变的;也就是说,亚当可能会说事件P发生在事件Q之前,而夏娃也许会反驳说事件P发生在事件Q之后。在这种情形下,我们熟悉的那种清晰明朗的先后关系——过去引发现在,而现在又引发未来——彻底瓦解了。没错,事实上所谓的因果关系(causality)在此也彻底瓦解了。”也许有一种东西,并且只有这种东西恒久不变,它先于这个世界而存在,而且也将存在于这个世界自身的组织结构之中:它就是——“变”。

某种意义上讲,智能是文化的产物,人类的每个概念和知识都是动态的,而且只有在实践的活动中才可能产生多个与其他概念和知识的关联虫洞,进而实现其“活”的状态及“生”的趋势<sup>[7]</sup>。同时,这些概念和知识又会保持一定的稳定性和继承性,以便在不断演化中保持类基因的不变性。时间和空间是一切作为知识概念的可能条件,同时也是许多原理的限制:即它们不能与存在的自然本身完全一致。可能性的关键在于前提和条件,一般情况下,人们常常关注可能性,而忽略关注其约束和范围。我们把自己局限在那些只与范畴相关的原理之上,很多与范畴无关的原理得不到注意和关涉。实际上,人机环境系统中的中态、势、感、知都有弹性,而关于心灵的纯粹物理概念的一个问题是,它似乎没有给自由意志留多少空间:如果心灵完全由物理法则支配,那么它的自由意志就像一块“决定”落向地心的石头一样。所有的智能都与人机环境系统有关,人工智能的优点在于缝合,缺点在于割裂,不考虑人、环境的单纯的人工智能软件、硬件就是刻舟求剑、盲人摸象、曹冲称象……简单地说,就是自动化。

人的学习是由初期的灌输及更重要的后期环境触发的交互学习构成,机器缺乏后期的能力。人的学习是事实与价值的混合性学习,而且是权重调整性动态学习。人的记忆也有自适应性,随人机环境系统而变化,不时会找到以前没注意到的特征。通过学习,人可以把态转为势,把感化成知,机器好像也可以,只不过大多是脱离环境变化的“死”势“僵”知。“聪明反被聪明误”有时是因为人的因素,有时是因为环境变化的因素。我们生活在一个复杂系统(complex system)中,在这种系统中有许多互相作用的变主体(agent)和变客体。人机混合中有多个环节,有些适合人做,有些适合机做,有些适合人机共做,有些适合等待任务发生波动后再做,如何确定这些分工及匹配很重要,如何在态势中感知?或在一串感知中生成态势?从时间维度上如何态、势、感、知?从空间维度上如何态、势、感、知?从价值维度上如何态、势、感、知?这些方面都很重要。

那么,如何实现有向的人机混合与深度的态势感知呢?一是“泛事实”的有向性。如国际象棋、围棋中的规则规定、统计概率、约束条件等用到的量的有向性,人类学习、机器学习中用到的运算法则、理性推导的有向性等,这些都是有向性的例子。尽管这里的问题很不相同,但是它们都只有正、负2个方向,而且之间的夹角并不大,因此称为“泛事实”的有向性。这种在数学与物理中广泛使用的有向性便于计算。二是“泛价值”的有向性,亦即我们在主观意向性分析、判断中常用到的,但不便测量的有向性。这里的向量有无穷多个方向,而且2个方向不同的向量相加通常得到1个方向不同的向量。因此,称为“泛价值”的有向量。这种“泛向”的有向数学模型,对于我们来说方向太多,不便应用(图4)。

然而,“泛价值”有向量的可加性与“泛物”有向性的二值性,启示我们研究一种既有二值有向性,又有可加性的认知量。一维空间的有向距离,二维空间的有向面积,三维空间乃至一般的 $N$ 维空间的有向体积等都是这种几何量的例子。一般地,把带有方向的度量称为有向量。态势感知中,态一般是“泛事实”的有向性,势是“泛价值”的有向性,感



一般是“泛事实”的有向性，知是“泛价值”的有向性。人机关系有点像量子纠缠，常常不是“有或无”的问题，而是“有与无”的问题。有无相生，“有”的可以计算，“无”的可以算计，“有与无”的可以“计算”，所以未来的军事人机混合指控系统中，一定要有人类参谋和机器参谋，一个负责“有”的计算，一个处理“无”的算计，形成指控“计算”系统。既能从直观上把握事物，还能从间接中理解规律。

西方发展起来的科学侧重于对真理的探求，常常被分为2大类：理论的科学和实践的科学。前者的目的是探索知识及真理，后者则寻求通过人的行动控制对象。这两者具体表现在这样一个对真理的证明体系的探求上：形式意义上的真理（工具论——逻辑），实证意义上的真理（物理——经验世界），批判意义上的真理（后物理学——形而上学）。俞吾金认为：“迄今为止的西方形而上学发展史是由以下3次翻转构成的：首先是以笛卡尔、康德、黑格尔为代表的‘主体性形而上学’对柏拉图主义的‘在场形而上学’的翻转；其次是在主体性形而上学的内部，以叔本华、尼采为代表的‘意志形而上学’对以笛卡尔、康德、黑格尔为代表的‘理性形而上学’的翻转；再次是后期海德格尔的‘世界之四重整体（天地神人）的形而上学’对其前期的‘此在形而上学’的翻转。”通过这3次翻转，可以引申出这样的结论：智能是一种人机环境系统交互，不但涉及理性及逻辑的研究，还包括感性和非逻辑的浸入，

当前的人工智能仅仅是统计概率性混合了人类认知机理的自动化体系，还远远没有进入真正智能领域的探索。若要达到真正的智能研究，必须超越现有的人工智能框架，把西方的“真”理同东方的“道”理结合起来，形成事实与价值、人智与机智、叙述与证明、计算与算计混合的“计算”系统。

自此，真正的智能将不仅能在叙述的框架中讲道理，而且还应能在证明的体系中讲真理；不仅能在对世界的感性体验中言说散文性的诗性智慧以满足情感的需要，而且能在对世界的理智把握中表达逻辑性的分析智慧以满足科学精神的要求，那时，智能才能真正克服危机——人性的危机。

当前制约机器人科技发展的瓶颈是人工智能，人工智能研究的难点是对认知的解释与建构，而认知研究的关键问题则是自主和情感等意识现象的破解。生命认知中没有任何问题比弄清楚意识的本质更具挑战性，或者说更引人入胜。这个领域是科学、哲学、人文艺术、神学等领域的交集。意识的变化莫测与主观随意等特点有时严重偏离了追问人工智能：科学技术的逻辑实证与感觉经验验证判断，既然与科学技术体系相距较远，自然就不会得到相应的认同与支持，这一点顺理成章，进一步说理应如此！然而，最近科技界一系列的前沿研究正悄悄地改变着这个局面：研究飘忽不定的意识固然不符合科技的尺度，那么在“意识”前面加上“情境”（或“场景”“上下文”“态势”）呢？人在大时空环境下的意识是不确定的，但“格物致知”一下，在小尺度时空情境下的意识应该有迹可循。自古以来，人们就知道“天时地利人和”的小尺度时空情境对态势感知及意识的影响，只是直至1988年，才出现了明确用现代的科学手段实现情境（或情景）意识的研究，即米卡·安德斯雷（Mica Endsley）提出的态势感知概念框架。但这只是个定性分析概念模型，其机理分析与定量计算还远远没有完善。

在真实的人-机-环境系统交互领域中，人的情景意识、机器的物理情景意识、环境的地理情景意识等往往同构于同一时空中（人的5种感知也应是并行的），人注意的切换使之对于人而言发生着不同的主题与背景感受/体验。在人的行为环境与

机的物理环境、地理环境相互作用过程中,人的情景意识被视为一个开放的系统,是一个整体,其行为特征并非由人的元素单独决定,而是取决于人-机-环境系统整体的内在特征,人的情景意识及其行为只不过是这个整体过程中的一部分。另外,人机环境中许多个闭环系统常常是并行或嵌套的,并且在特定情境下这些闭环系统的不同反馈环节信息又往往交叉混合在一起,起着引起兴奋或抑制的作用,不但有类似宗教情感类的柔性反馈,称为软调节反馈,人常常会延迟控制不同情感的释放;也存在着类似法律强制类的刚性反馈,称为硬调节反馈,常规意义上的自动控制反馈大多属于这类反馈。如何快速化繁为简、化虚为实是衡量一个人机系统稳定性、有效性、可靠性大小的主要标志,是用数学方法的快速搜索比对还是运筹学的优化修剪计算,这是一个值得人工智能领域深究的问题。

人-机-环境交互系统往往由有意志、有目的和有学习能力的人的活动构成,涉及变量众多,关系复杂,贯穿着人的主观因素和自觉目的,所以其中的主客体界限常常是模糊的,具有个别性、人为性、异质性、不确定性、价值与事实的统一性、主客相关性等特点,其中充满了复杂的随机因素的作用,不具备重复性。另外,人-机-环境交互系统有关机(装备)、环境(自然)研究活动中的主客体界限分明,具有较强的实证性、自在性、同质性、确定性、价值中立性、客观性等特点。在西方国家,无论是在古代、中世纪还是在现代,哲学宗教早已不单纯是意识形态,而是逐渐成为各个阶级中的强大政治力量,其影响不断渗透到社会生活的各个领域,甚至哲学、政治、法律等上层建筑都处于宗教控制之下。总之,以上诸多主客观元素的影响导致了人-机-环境交互系统的异常复杂和不确定。所以,对人-机-环境交互系统的研究不应仅仅包含科学的范式,如实验、理论、模拟、大数据,还应涉及人文艺术的多种方法,如直观、揣测、思辨、风格、图像、情境等,在许多情况下还应与哲学宗教的多种进路相关联,如现象、具身、分析、理解与信仰等。

在充满变数的人-机-环境交互系统中,存在的逻辑不是主客观的必然性和确定性,而是与各种

可能性保持互动的同步性,是一种得“意”忘“形”的见招拆招和随机应变能力。这种思维和能力可能更适合人类的各种复杂艺术过程,也恰恰是人工智能所欠缺的地方。

人机智能是人-机-环境系统相互作用而产生的新型智能系统。其与人的智慧、人工智能的差异具体表现在3个方面:首先,在混合智能输入端,把设备传感器客观采集的数据与人主观感知到的信息结合起来,形成一种新的输入方式;其次,在智能的数据/信息中间处理过程,机器数据计算与人的信息认知相混合,构建起一种独特的理解途径;最后,在智能输出端,它将机器运算结果与人的价值决策相匹配,形成概率化与规则化有机协调的优化判断。人机混合智能是一种广义上的“群体”智能形式,这里的人不仅包括个人,还包括众人,机不但包括机器装备,还涉及机制。此外,还关联自然/社会环境、真实/虚拟环境、网络/电磁环境等<sup>[8]</sup>。

## 2.2 有关人机几个问题的思考

1) 人-机-环境交互系统是不是要先考虑任务目标,任务的模型该考虑哪些关键要素?

从多维度到变维度,从多尺度到变尺度,从多关系到变关系,从多推理到变推理,从多决策到变决策,从多边界条件到变边界条件。计算—算计相互作用的整合法则(线性与非线性的整合)。神经中的序可以装任何东西,并可进行泛化成新的序。任务需求是智能的目的,一切行为都是任务和目标驱动的。任务的模型最基础的是5W2H(who、where、when、what、why、how、how much),并结合各服务领域的关键要素展开,进行事实性与价值性混合观察、判断、分析、执行。

2) 人机混合是不是要对人、机建模?若是,人和机的模型,要考虑哪些关键因素?

人和机的混合基于场景和任务(事件),要考虑输入、处理、输出、反馈、系统及其影响因素等,具体如下:(1) 客观数据与主观信息、知识的弹性输入——灵活的表征;(2) 公理与非公理推理的有机混合——有效的处理;(3) 责任性判断与无风险性决策的无缝衔接——虚实互补的输出;(4) 人类反思与机器反馈之间的相互协同调整;(5) 深度态势感

知与其逆向资源管理过程的双向平衡;(6) 人机之间的透明信任机制生成;(7) 机器学习范围/内容与人类学习的不同。

3) 衡量人机混合(人机高效协作)的关键指标有哪些?

粗略地说,可分别从人、机和任务3个方面研讨:人机环境系统高效协同的关键指标在于三者运行绩效中的反应时、准确率,具体体现在计划协同,动作协同,特别是跨组织实现步调上的协同,还有资源、成本的协同等方面。比如人的主动、辩证、平衡能力,机的精确、逻辑、快速功能,任务的弹性、变化、整体要求。如何有机地把人、机、任务的这些特点融入到系统协同的反应时、准确率2大指标之中,也是一个关键问题。

4) 从认知工程的智能系统框架,以及中西方的基础理论来看,哪些是未来认知功能具备可工程化的能力框架?哪些是尚不具备工程化的认知功能?

以上问题简单地回答就是:计算部分与算计部分之分。未来认知功能具备可工程化的能力框架在于软硬件计算功能在快速、精确、大存储量等方面的进一步提升,尚不具备工程化的认知功能在于反映规划、组织、协同算计谋划能力的知几趣时变通得到明显改善。智,常常在可判定性领域里存在;能,往往存在于可计算性领域。认知工程的瓶颈和矛盾在于:总想用逻辑手段解决非逻辑问题,例如试图用形式化手段解决意向性的问题。不同的人机任务上下文中的上下程度弹性也是不同的。计算是算计的产物,计算常是算计的简化版,不能体现出算计中主动、辩证、矛盾的价值。计算可以处理关键场景的特征函数,但较难解决基本场景的对应规则,更难对付任意场景的统计概率,可惜这些还仅仅是场景,尚远未涉及情境和意识……计算常常是针对状态参数和属性的(客观数据和事实),算计则是一种趋势和关系之间的谋划(根据主观价值的出谋划策),所以态势感知中,态与感侧重计算推理,势和知偏向算计谋划。“计算”最大的特点就是异、易的事实价值并行不悖。人类的符号、联结、行为、机制主义是多层次、多角度,甚至是

变层次、变角度的,相比之下,机器的符号、联结、行为、机制主义是单层次、单角度,而且是固层次、固角度的。人类思维的本质是随机应变的程序,也是可实时创造的程序,能够解释符号主义、联结主义、行为主义、机制主义之间的联系并能够打通这些联系,实现综合处理。达文波特认为:人类的某种智能行为一旦被拆解成明确的步骤、规则和算法,它就不再专属于人类了。这就涉及一个基本问题,即科学发现如何成为一个可以被研究的问题<sup>[9]</sup>。

### 3 人机混合智能是人工智能未来的发展方向

人机混合智能有2大难点:理解与反思。人是弱态强势,机是强态弱势,人是弱感强知,机是强感弱知。人机之间目前还未达到相声界“一逗一捧”的程度,因为还没有单向理解机制出现,能够具有幽默性的机器出现依旧遥遥无期。乒乓球比赛中运动员的算到做到、心理不影响技术(想赢不怕输)、如何调度自己的心理(气力)生出最佳状态、关键时刻心理的坚强、信念的坚定等,这都是机器难以产生出来的生命特征物。此外,人机之间配合必须有组合预期策略,尤其是合适的第二、第三预期策略。自信心是匹配训练出来的,人机之间信任链的产生过程常常如下:陌生—不信任—弱信任—较信任—信任—较强信任—强信任,没有信任就不会产生期望,没有期望就会人机失调,而单纯的一次期望匹配很难达成混合,所以第二、第三预期的符合程度很可能是人机混合一致性的关键问题。人机信任链产生的前提是人要自信(这种自信心也是匹配训练出来的),其次才能产生他信和信他机制,他信和信他就涉及到多阶预期问题。若 being 是语法, should 就是语义,二者中和相加就是语用,人机混合是语法与语义、离散与连续、明晰与粗略、自组织与他组织、自学习与他学习、自适应与他适应、自主化与智能化相结合的无身认知+具身认知共同体、算+法混合体、形式系统+非形式系统的化合物。反应时与准确率是人机混合智能好坏的重要指标。人机混合就是机机混合,器机理+脑机制;

人机混合也是人人混合,人情意+人理智。

人工智能相对是硬智,人的智能相对是软智,人机智能的混合则是软硬智。通用的、强的、超级的智能都是软硬智,所以人机混合智能是未来,但是混合机制还远未清晰,更令人恍惚的是一不留神,不但人进化了不少,机又变化得太快。个体与群体行为的异质性,不仅体现在经济学、心理学领域,而且还是智能领域最为重要的问题之一。现在主流的智能科学在犯一个以前经济学犯过的错误,即把人看成是理性人,殊不知,人是活的人,智是活的智,人有欲望、动机、信念、情感、意识,而数学性的人工智能目前对此还无能为力。如何混合这些元素,使之从冰冻的、生硬的状态转化为温暖的、柔性情形,应该是衡量智能是否智能的主要标准和尺度,同时这也是目前人工智能很难跳出人工的瓶颈和痛点——“只有钢筋没有混凝土”。经济学融入心理学后即可使理性经济人变为感性经济人,而当前的智能科学仅仅融入心理学是不够的,还需要渗入社会学、哲学、人文学、艺术学等,方能做到通情达理,进而实现由当前理性智能人的状态演进成自然智能人的形势。智能中的意向性是由事实和价值共同产生出来的,内隐时为意识,外显时为关系。从这个意义上说,数学的形式化也许会有损于智能,维特根斯坦认为:形式是结构的可能性。对象是稳定的、持续存在的东西;而配置则是变动的、非持久的东西。维特根斯坦还认为:我们不能从当前的事情推导出将来的事情。迷信恰恰是相信因果关系。也就是说,基本的事态或事实之间不存在因果关系。只有不具有任何结构的东西才可以永远稳定不灭、持续存在;而任何有结构的东西都必然是不稳定的,可以毁灭的。因为当组成它们的那些成分不再依原有的方式组合在一起时,它们也就不复存在了。事实上,每个传统的选择(匹配)背后都隐藏着2个假设:程序不变性和描述不变性。这两者也是造成期望效用描述不够深刻的原因之一。程序不变性表明对前景和行为的偏好并不依赖于推导出这些偏好的方式(如偏好反转),而描述不变性规定对被选事物的偏好并不依赖于对这些被选事物的描述。

人机混合智能难题,即机器的自主程度越高,人类对态势的感知程度越低,人机之间接管任务顺畅的难度也越大,称为“生理负荷下降、心理认知负荷增加”现象。对于如何破解这一难题,有经验的人常常抓关键任务中的薄弱环节,在危险情境中提高警觉性和注意力,以防意外,随时准备接管机器自动化操作,也可以此训练新手,进而形成真实敏锐地把握事故的苗头的能力,恰当地把握处理时机,准确地随机应变,并在实践中不断磨砺训练增强。即便如此,如何在非典型、非意外情境中解决人机交互难题仍需要进一步探讨<sup>[10]</sup>。

计算与算计,“合久必分,分久必合”。算计需要的是发散思维,计算需要的是缜密思维,这是2种很不一样的思维方式,这2种方式同时发生在某个复杂过程中是小概率事件,由此带来的直接后果就是,复杂领域的突破也只能是小概率事件。对待场景中的变化,机器智能可以处理重复性相同的“变”,人类智能能够理解杂乱相似性(甚至不相似)的“变”,更重要的是还能够适时地进行“化”,其中“随动”效应是人类“计算计”的一个突出特点,另外,人类“计算计”还有一个更厉害的武器——“主动”。

有人说:“自动化的最大悖论在于,使人类免于劳动的愿望总是给人类带来新的任务。”解决三体以上的科学问题是非常困难的,概念就是一个超三体的问题:变尺度、变时空、变表征、变推理、变反馈、变规则、变概率、变决策、变态势、变感知、变关系……犹如速度与加速度之间的关系映射一般,反映着智能的边界。有效概念的认知是怎样产生的,OODA还是OAOODDDAA?亦或是OA?这是一个值得思考的问题。多,意味着差异的存在;变,意味着非存在的有;复杂,意味着反直观特性;自组织/自相似/自适应/自学习/自演进/自评估意味着系统的智能……人机环境网络中重要/不重要节点的隐匿与恢复是造成全局态势有无的关键,好的语言学家与好的数学家相似:少计算多算计,知道怎么做时计算,不知道怎么做时算计,算计是从战略到策略的多逻辑组合,人机混合“计算计”机制犹如树藤相绕的多螺旋结构,始于技术,成于管理。如果说计算是科学的,算计是艺术的,那么“计算计”就是

科学与艺术的。

价值不同于事实之处在于,可以站在时间的另一端看待发生的各种条件维度及其变化。仅仅是机器智能永远无法理解现实,因为它们只操纵不包含语义的语法符号。系统论的核心词是突显(整体大于部分),偏向价值性 should 关系;控制论的核心词是反馈(结果影响原因),侧重事实性 being 作用。耗散结构论的核心词是开放性自组织(从非平衡到平衡),强调从 being 到 should 过程。控制论中的反馈是极简单的结果影响(下一个)原因的问题,距离人类的反思——这种复杂的“因果”(超时空情境)问题很遥远。计算是关于人机环境体系功能(功能+能力)价值性结构谋划,而不是单事实逻辑连续的计算,计算—计算正是关于正在结构中事实—价值—责任—情感多逻辑组合连续处理过程,人机混合智能难题的实质也就是计算—计算的平衡。

人机融合智能可以理解为人类和机器系统之间的协同工作,以达到更高效、更准确的智能处理。在这个过程中,清晰和模糊是2个重要的概念。首先,清晰指的是事物具有明确的定义和边界,能够被准确地描述和处理。例如,数字、逻辑等概念可以被机器系统精确地处理和计算,而人类也可以准确地理解和运用这些概念。其次,模糊指的是事物具有不确定性和模糊性,难以被精确地定义和处理。例如,人类的语言表达和情感体验等就具有一定的模糊性,而这些方面的处理对于机器系统来说则相对困难。因此,人机融合智能是清晰和模糊的综合过程,需要人类和机器系统共同合作来处理各种问题。在处理清晰问题时,机器系统可以提供高效、准确的计算和处理能力;在处理模糊问题时,人类则可以提供更好的语言理解、情感体验等方面的能力。两者相互补充,形成一种协同工作的模式,从而实现更高效、更智能的处理。若不考虑人机各自智能程度的计算,人机融合智能系统中人、机协同的智能程度,其大小与2个智能体的智能程度成正比,与两者一致性的大小成反比。即

$$AI_{\text{人机协同}} = e \times AI_{\text{人}} \times AI_{\text{机}} / K \quad (1)$$

式中, $e$ 为比例系数,其结果取决于交互环境的复杂性; $K$ 为 $AI_{\text{人}} \times AI_{\text{机}}$ 的同向程度,即一致性程度(人、机智能具有共识性的大小)。

人机混合智能是人工智能发展的必经之路,其中既需要新的理论方法,也需要对人、机、环境之间的关系进行新的探索。人工智能的热度不断升高,越来越多的产品走进人们的生活。但是,强人工智能依然没有实现,如何将人的计算智能迁移到机器中去,这是一个必然要解决的问题。我们已经从认知角度构建认知模型或者从意识的角度构建计算—计算模型,这都是对人的认知思维的尝试性理解和模拟,期望实现人的计算能力。计算—计算模型的研究不仅需要考察机器技术的飞速发展,还要考虑交互主体,即人的思维和认知方式,让机器与人各司其职,互相混合促进,这才是人机混合智能的前景和趋势。

#### 参考文献(References)

- [1] Lee D. Birth of intelligence: From RNA to artificial intelligence[M]. Oxford: Oxford University Press, 2020.
- [2] Segal L X. The dream of reality: Heinz von Foerster's constructivism[M]. New York: Springer Science & Business Media, 2001.
- [3] Manuel B, Lenore B. A theoretical computer science perspective on consciousness[J]. Journal of Artificial Intelligence and Consciousness, 2021, 8(1):1-42.
- [4] Samuel B. Wintermute abstraction, imagery, and control in cognitive architecture[D]. Michigan: University of Michigan, 2010.
- [5] Pat Langley, John E Laird, Seth Rogers. Cognitive architectures: Research issues and challenges[J]. Cognitive Systems Research, 2009, 10(2): 141-160.
- [6] Wlodzislaw D, Richard J Oentaryo, Michael P. Cognitive architectures: Where do we go from here[C]//Proceedings of the 2008 Conference on Artificial General Intelligence. Netherlands, 2008: 122-136.
- [7] Baars B J. In the theater of consciousness: A rigorous scientific theory of consciousness[J]. Journal of Consciousness Studies, 1997, 4(4): 292-309.
- [8] 刘伟. 追问人工智能:从剑桥到北京[M]. 北京: 科学出版社, 2019: 10-17.
- [9] 刘伟. 人机融合:超越人工智能[M]. 北京: 清华大学出版社, 2021: 29-37.
- [10] Cummings M L, Bruni S, Mercier S, et al. Automation architecture for single operator, multiple UAV command and control[J]. International C2 Journal, 2008(2): 1-29.

## Demarcation of human-machine intelligence

LIU Wei

School of Artificial Intelligence, Beijing University of Posts and Telecommunications, Beijing 100876, China

**Abstract** The victory of Deep Blue against Kasparov in 1997 and the victory of Alpha Dog over Lee Sedol in 2016 set off the trend of artificial intelligence, however, artificial intelligence still has so far a huge gap behind humans in terms of intelligence level and range of capabilities, which hinders its development. The reason is that there is an unbridgeable boundary between human intelligence and machine intelligence. In this context, machines that rely on symbols pointing to objects can only perform formal calculations in a closed environment but cannot realize intentional calculations in an open environment like humans. Therefore, the complementary advantages of man and machine are particularly important. The current relationship between man and machine is mainly a hard function allocation, whereas the relationship between man and machine in the future digital world would more likely be a flexible division of capabilities. This paper discusses in detail the bottlenecks of artificial intelligence, the first principle of intelligence, and other issues, and gives more in-depth thinking about human-machine problem amid these problems. It is proposed that human-machine hybrid intelligence is the future development direction of artificial intelligence.

**Keywords** artificial intelligence; function distribution; suanji; human-machine hybrid intelligence ●



(责任编辑 陈广仁)