

三年前,ChatGPT的发布引爆了全球生成式人工智能浪潮。如今,能说会道的机器正以翻译、写作、编程、绘画与音视频创作等各种形式融入人类社会的日常。此时,我们需要对目前以大语言模型为代表的机器语言智能予以冷静反思。

### 反思机器语言智能

2019年,英格兰银行发行的50英镑纸币上,印有计算机科学之父艾伦·图灵的一句话:“毫无见鸡影,滴酒醉流觞(This is only a fore-taste of what is to come, and only the shadow of what is going to be.)”。这句话是图灵1949年6月10日接受英国《泰晤士报》采访时,对曼彻斯特大学刚建成人类首台可存储程序计算机“机械大脑”(mechanical mind)的评论——它仅用几周就“解决”了人类持续计算了300年的梅森素数难题(判断 $2^n-1$ 这种形式的数是否为素数)。图灵预言:“依我看,机器可以进入人类智能的任何领域,并达到跟人类平起平坐的水平。”一年后,图灵在《心智》(Mind)杂志发表《计算机器与智能》,回应公众热议的问题——“机器会思考吗?”图灵的答案是用“模仿游戏”(imitation game)来证实或证伪:如果机器能与人笔谈而不被发现,就说明机器会思考。他甚至给出了时间表:“50年后,计算机器的信息存储量将达到大约 $10^{10}$ 比特,在提问5分钟后,一般提问者能区分人和机器的正确率不超过70%。”

现在回看,作为预言家的图灵犯了两个“错误”。一是他低估了计算机硬件的进步速度。2000年,计算机的内存已经可以超过 $10^7$ 比特。二是他高估了计算机软件的进步速度。2000年参加罗布纳奖(Loebner Prize)“模仿游戏”的程序,仍无法真正骗过评委。

不过,图灵的“错误”持续时间不算太长。2022年11月30日,ChatGPT横空出世,别说5分钟,即便与人聊上50分钟,ChatGPT也丝毫不露怯。图灵的机器智能预言在“迟到”20年后终于成真!机器真的“会思考”了?!人类还能稳坐智能之巅吗?一种更深的不安随之蔓延——人类是否会被AI主导,甚至完全替代?

有这类想法的不乏其人,甚至包括2024年诺贝尔物理学奖得主、被誉为“深度学习教父”的杰弗里·辛顿。作为这一轮AI浪潮中的旗手,辛顿却多次警示AI会对人类社会构成威胁,称AI不是基于统计的高级计算工具,而是可以理解、决策,甚至产生意识的全新智能形态。至于人类,则“只是智能演化的一个过渡阶段”。

本文无意作宏大叙事,但希望在ChatGPT问世三周年之际,对目前以大语言模型为代表的机器语言智能,做一点冷静的反思:一方面结合我们对AI基本工作原理的认识,探究AI的能与不能;另一方面则尝试跳出图灵的“模仿游戏”框架,探究智能的标准是否应该提高到能在无路可走时开辟新径、在无路可循时创建新规?

### AI的能与不能:成也“模仿”,败也“模仿”

今天的AI为什么看上去这么能干?一言以蔽之:强大的模仿能力!图灵提出的“模仿游戏”是为计算机设计的“考试”方式,但实际上,“模仿游戏”同时也是计算机的“学习”方式。今天的计算机模仿的对象是互联网上的海量数据(即人类智能的产出物)。具体可分两个层次看:一是“模仿”什么——把每个语言符号都放到语流中正确的位置上,二是怎么“模仿”——通过反复地预测被隐去的语言符号,不断降低预测错误率。

以“阿Q是县长派来的”这个句子为例,把其中的“派”隐去,变成“阿Q是县长( )来的”,然后让计算机去猜这个括号里应该填什么词。

当前,文化“走出去”背景下的中国文学外译方兴未艾,但多年来,面对主动外译的困境,学界认识不一,建立基本的价值共识势在必行。对中国文学外译的评价应以马克思主义哲学的实践论、历史观、价值论为指导,以人的社会历史实践来把握中国文学外译的基本性质,以确立评价的认识基础、逻辑起点。

**把握中国文学外译的主体能动性** 人的实践活动具有主体能动性。翻译活动及其产品是翻译主体精神客体化的过程与结果,凝聚着翻译主体的意志与价值观念。中国文学外译作品并非自在性存在,而是自为性存在、价值性事实。因此,对于中国文学外译批评的定位与评价,应认识到其主体依据,即认识翻译活动承载的意志性、能动性。文化“走出去”下的中国文学外译并非纯粹的市场行为,而是以国家作为价值主体的国家翻译行为,肩负国家形象建构与中国文化传播的使命任务,其主体能动性表征为向世界展示中国优秀精神文明成果,促进文学与文化平等交流,同时在对话中巩固中华民族文化主体性,提升国家文化软实力。因此,就中国文学“走出去”的基本性质而言,中国文学外译批评应把握主动外译的国家主体性、文化主体性。

**把握中国文学外译的社会历史性** 人的实践活动具有社会历史性。主体与客体构成的价值关系是在人具体的社会历史实践中生成与实现的。作为人的社会性与历史性活动,翻译发生于特定的时空语境并受制于特

### 深度

# 机器语言智能:“模仿游戏”只是序曲,远非终章

北京大学中文系教授 詹卫东

如果它填“弯、胖、细……”就是错的,填“派、请、叫……”就是对的。这个猜词填空的游戏就是在模仿人类使用语言的规律。通过对海量语料的反复模仿,大语言模型可以得出句中词语间准确的语义关系,学到这句话底层的语义模式;句子开头出现的人名是第一个动词的被动接受方,句中第二个指人名词是第一个动词的主动发出方,两个指人名词不是同一个人,即“县长派阿Q来”“阿Q不是县长”“不是县长来是阿Q来”等。跟区分哪些词能填入句中括号位置、哪些不能一样,模型也能把同样包含“阿Q是县长”这个片段的句子分为两类,一类句子中“阿Q≠县长”,另一类句子中“阿Q=县长”(后者如“虽然阿Q是县长但还是不管用”)。

机器观察了“阿Q、是、县长、派、来、的”这些基本语言单位几乎所有可能的位置,掌握了词语之间的意义关系模式,并将这些信息以相互关联的高维向量形式记录在模型参数中,从而能生成和理解自然的句子。但是,问题来了!如果一种语言现象从未在训练时出现过,模型会如何应对?它还能正确理解吗?答案是大概率不能。请看两个小测试。一个是对话合理性测试。下面两人的一问一答是不是意思合理的自然对话?甲:博士研究生今年找教职工作困难吗?乙:连博士研究生今年找教职工作都很困难。另一个是空间方位词异形同义测试。下面两句话中的同一位置有两个不同的方位词:前、后。这两句话描述的空间场景是否相同?第一句话:至今菲律宾的土著居民在见面时,握过手后还要转身向前走几步,意思是向对方表明背后没有藏刀,是真诚地迎接对方。第二句话:至今菲律宾的土著居民在见面时,握过手后还要转身向前走几步,意思是向对方表明背后没有藏刀,是真诚地迎接对方。

上述例子的共同特征是:人类能轻易理解作答,而大模型在训练时可能没有见过这类“罕见”文本,无从模仿,就容易答错。有海量数据可以仿效,大模型就“能说会道”;无先例可循,大模型就“不懂装懂、似是而非”。

### 智能的内插与外推:超越模仿,迈向创新

在当下AI火热的社会氛围中,我们很容易混淆两个概念:“能干”与“智能”。大语言模型“遍访人类名师”学到的表象式能干,可能在很大程度上掩盖了其内核的并不那么智能。与其说一个AI很能干就是智能,倒不如说,只有当一个AI不能干,不知道该怎么干时,它所表现出来的解决问题的语言能力,才是真正的智能。毋庸讳言,大语言模型通过模仿人类语言,确实能表现出某种创新能力,这像是统计学中的“内插”(interpolation)。如果将人类历史积累的既有知识比作一个巨大的几何“凸包”,大语言模型便是一位在这个闭环内登峰造极的舞者。它极其擅长在旧有的知识节点间搭建新的连线,比如将“莎士比亚的十四行诗”与“量子力学的波粒二象性”嫁接,生成一首惊艳的量子力学十四行诗。但这本质上只是旧元素的新排列,并未跨出人类经验的边界。

人类的创新智慧则是统计学中的“外推”(extrapolation)。无论是哥白尼针对地心说的破局,还是爱因斯坦折叠时空的曲率,都是不同历史时期人类智能跳出“凸包”,向着未知虚空勇敢的纵身一跃。这不是对历史数据的模仿,而是创造全新的范式。

试想将同一道题摆在两个人面前:对张三而言,这可能是道旧题或者一些不同旧题的新组合,他只需“不假思索地”从记忆库中“调用工具”就能完成任务;但对李四而言,这是一道从未见过的陌生难题,他必须苦思冥想才能找出解法,这才是“玩转智能”。目前的AI更像张三,是在已知数据的海洋里做完美的“内插

式”拟合;而李四的模式是在没有数据可用的荒滩上做艰难的“外推式”探索。这种探索“未知”的能力才真正触及智能的核心。

目前机器模仿的对象,实际上是人类智能的产出物(即语言数据),而非智能的策源地(即人脑本身)。机器能极致地复现人类已有的知识版图,却很可能还无法复刻产生知识的思维机制。模型把人类说过的话“学了个遍”,但并不等于它拥有了产生这些话的那种心智。AI能在现有地图上找到最优路径,但它还难以开疆拓土,画出新大陆的版图。

围棋世界冠军柯洁的两段视频让我印象深刻。一段视频中,他把围棋AI叫作“弟中之弟”,说AI“并不聪明”;另一段视频中,他又表示“就算AI让三子,自己也很难赢AI”。从“AI不过如此”到“根本赢不了AI”,世界冠军对待AI的这种“意难平”的矛盾心态非常有代表性。本文探讨AI的能干与不那么智能,目的

也是提倡建立人与AI更和谐的共处关系:AI很能干,所以人类应该与AI共舞,熟练地驾驭这位“超级人类模仿者”,让它成为人脑记忆与效率的放大器;AI还不够智能,所以人类应更积极地探索智能到底是什么,应借助AI之力,在探寻智能原理的路上勇于创新。

《计算机器与智能》的结尾写道:“初见前路近可至,细思百事竟徒忙(We can only see a short distance ahead, but we can see plenty there that needs to be done.)”。(引自尼克著《人工智能简史》)AI模仿智能虽然看上去“近在眼前”,但人类智能的门槛一直在“水涨船高”:当一项本来被认为是智能的能力被机器掌握后,人类一般就会把它从“智能”清单中剔除出去,把这项能力“贬低”为“只是计算而已”。毕竟,谁会觉得计算机有智能呢?今天的AI从模仿起步,未来终将实现创新。到那个时候,人类对自身智能原理的认识也必然会达到一个新的高度。



艾伦·图灵

### 视域

康德伦理学倡导一种有德、有爱、有乐的道德品格,也开辟了人们通向幸福的可能道路。

## 充满温情的康德伦理学

河南师范大学政治与公共管理学院副教授 惠永照

康德曾经认为,对知识的纯粹探究本身就具有价值,并因此轻视无知的大众,是“卢梭纠正了我,这种盲目的优越感消失了。如果这样的考虑在建立人性的权利时没有赋予所有其他人以价值,那么我会发现自己远不如一个普通劳动者”。这一著名的自白反映了康德思想的一个重要倾向,即知识要为人服务,一种知识无论多么高远玄妙,最终都以某种方式表现为对人的接纳、尊重和关怀。正是在这一思想背景下,面对众多对康德伦理学的诟病,我们才能发现,在康德伦理学“僵硬”“冰冷”的外表下隐藏着一颗充满温情的内心。

### 为人的尊严奠基

康德自白中“建立人性的权利”的工作正是通过理性的实践兴趣完成的,因而康德的伦理学不单纯是提供规范的伦理学,而且是人的权利和尊严奠基的伦理学。也就是说,康德伦理学要为一切道德上的“应该”确立一个坚实的基础,使道德免于被个人私利、专断意志和偏见左右。如此,道德所具备的普遍性才是真正的普遍性。普遍性和强制性是一般道德的要求,并非康德伦理学所独有。康德的独特性在于指出,道德所要求的普遍性不可能通过人们所追求的某种经验目的,或者人们所具有的某些天然的品质或情感而得到,因为这些东西都是个人偶然具有的,并不对所有人都有效。以往的道德原则,不论是建立在自然情感上还是建立在道德情感上,不论是建立在个人的完善上还是建立在个人的幸福上,都从一些特殊的经验内容出发,并试图将只具有特殊性的内容普遍化,将道德的强制变成一种纯粹外在强制,即他律。他律的道德不是真正普遍性的道德,它所具备的普遍性只是一种伪装起来的特殊性。而真正的普遍性只能通过自我强制而具备,这种自我强制在康德这里就表达为一条绝对命令。

虽然自黑格尔以来,对康德的道德命令

是形式主义的批评不绝如缕,但对命令并不只是从形式出发来规定道德,它同样包含了道德的实质内容。这就意味着,一条纯粹形式化的命令与将人性本身当作目的的具有实质内容的命令,以及意志自律的道德理念是一致的。三者一致性的基础便是人格。理性不以外在的某个客体为目的,但这并不意味着理性没有目的,理性以其自身,因而以每一个具有人格的理性存在者作为目的。只有作为自在目的的人格,才是道德上的普遍立法者,是只服从自我立法的自律者和有尊严者。

绝对命令通过对每个人私利私好的限制,将对人的尊重提升到经验条件之上。一个人被尊重并不是因为其社会地位、权力、财富等经验的东西,也不是因为其碰巧具有的可爱气质或同情心,而是基于其人格。所以绝对命令是要求接纳、关怀和尊重每一个人,但不是将每一个人从其偶然性方面,即出身、地位、财富等方面,而是从其作为一个人格方面来接纳、关怀和尊重。从这个角度看,康德伦理学就不与其他伦理学对立,而是为一切伦理学理论奠基。因为它回答了一个所有伦理学理论共同建基于其上的问题:为什么要关怀人?美德伦理学对美德和幸福的追求、功利主义对“最大多数人的最大幸福”的追求,都是对人的关怀提供了最终的根据,在这一根据的基础上才能说,人是值得被关怀的。

### 对情感的接纳

康德伦理学最为人诟病的,除了形式主义,就是其理论中无处不透露着的“冷漠”。这种冷漠感主要源自康德伦理学对情感的拒绝。他限制情感在伦理学中的作用,拒绝自然情感作为道德的合法动机;并且去除了爱和同情等情感的情感性一面,而使之以“实践的爱”“实践的同情”的面目重新出现。其实,实践的爱,实践的同情更多只是一种助人的行动,而

不是一种情感。康德伦理学拒斥情感的印象很大程度上与对康德文本的片面阅读有关。很多人只阅读或者只重视那个在《道德形而上学奠基》《实践理性批判》中谈论道德性的康德,而没有阅读或忽视了在《道德形而上学》和《实用人类学》等偏后期的著作中谈人性的康德。人是有限的,这是人性的一种根本事实,所以在后期的一系列作品中,康德在人性名义下论述了爱和同情等多种情感的作用。

事实上,康德从未要求一个人应该是冷漠的、无情的,哪怕对“不动心”的倡导也只是表明,人不应陷入激情之中,而应让自己处于理性的控制中,达到内心的平静安宁。而且我们不能脱离伦理学的语境看康德所说的实践的爱和同情的同情。这些论述的主旨是,如果将爱、同情等自然情感纳入伦理学,那么采取一种什么样的形式才是合适的?追问这一问题初衷不是要在道德上拒绝情感,而恰恰是要在道德上接纳情感。因为自然意义上的爱和同情不能被命令,无法成为一种道德义务,所以如果要自然情感意义上的爱和同情纳入伦理学,放任它们处于自然的状态就是不够的。康德在伦理学语境中对情感的论述皆在表明自然情感在伦理学中的这种不足,而不是要拒斥自然情感。

康德看到了作为自然情感的爱和同情在伦理学中的价值。如果一个人想要做出助人的行动(即实践的爱),那么离开了对于相关情境的敏感性就是不可能的,但对于情境的敏感性只能通过自然意义上的爱和同情做出,因而康德虽然认为拥有自然意义上的爱和同情不是义务,但培养自然意义上的爱和同情则是一项间接义务。所以,通过将实践的爱和同情的同情确立为一项道德义务,同时通过将培养自然意义上的爱和同情确立为间接的义务,康德的伦理学就以一种不同于诸如情感主义伦理学的方式接纳了情感。

### 有德、有爱、有乐的品格

很多人觉得康德伦理学对人提出了过高的要求,要人做“道德圣贤”,然而事实上,正是因为人的有限性,康德明确提出,没有人能够真正达到纯粹道德,因而没有人能够成为道德上的圣人。康德伦理学并不以纯粹道德为目的,而仅仅以培育德性为道德的目标。康德把德性看作一种意志的道德力量,让人在道德动机和非道德动机对立的时候选择道德的动机。德性在与恶习的战斗中表现出来并通过与恶习的斗争而不断增长,最后成为一个人道德品格中稳定和持久的方面。所以康德意义上的有德之人并非像很多人认为的那样时时在与偏好斗争,时时会陷入善与恶的抉择与挣扎之中。恰恰相反,一个有德之人会拥有一种稳定的道德品格,会让主体出于一贯的准则而行动,并且获得切实的满足。这种道德上的满足是一种道德快乐,它不同于经验意义上的快乐,但二者具有相似性。

通过道德品格的塑造,康德式的有德之人并不是冷漠的人,而是具有道德力量且具有仁爱之心的人,他在道德行为中也能切实地感受到快乐。这样,康德式的有德之人就迥异于传统的刻板印象:有德性、有仁爱、有快乐,构成了康德式有德之人的核心方面。这样一个面对一个道德的情境时,不会因同情心被激发而一同痛苦,也不会对他人的痛苦无动于衷,而是尽可能地理解他人的困境并力所能及的方式帮助他人。而当自己的帮助解除了他人的困境,他会由衷地感到满足;而当他人无法提供帮助或自己的帮助最终无法使他人脱离困境,也不自责和懊恼。

综上,康德伦理学通过为人的权利和人性尊严奠基,在一种超越的维度上确立了对人关怀的终极根据;与此同时,基于人性的有限性,康德伦理学以一种新的形式接纳了爱和同情之类的情感,在一种经验的维度上确立了对人关怀的新形态;最终,康德伦理学倡导一种有德、有爱、有乐的道德品格,也开辟了人们通向幸福的可能道路。“道是无情却有情”,康德伦理学看似无情,实则有情,康德伦理学的情是深沉的、不外显的,但仍然是温暖人心的。[本文系国家社科基金项目“康德道德心理学研究”(22BZ0073)成果]

### 见地

## 中国文学外译批评的价值共识

广东外语外贸大学高级翻译学院博士生 邱 瑾

定的主客观条件。实践、价值的社会历史性意味着翻译批评要秉持唯物史观,才能把握翻译活动及其主客价值关系的历史面目,进而评价其是否符合社会历史的发展趋势。就中国文学外译的必要性而言,中国文学外译批评只有将文学外译置于当代国内外发展大势的时空语境中,才能把握中国文学外译的定位与价值。如果仅以西方市场需求与读者接受程度为标准,就极易否定文学外译的历史必然性。当前,全球化深入推进,中国日益走向世界舞台中央,中外交流越发密切,中国文学翻译须自觉肩负起变革不平等价值关系、争取价值认同的历史使命。对此,中国文学外译批评应当发挥其价值引领与导向的力量。

**把握中国文学外译的实践发展性** 人的实践活动具有现实性与发展性。翻译活动是翻译主体改造翻译客体,将主体的价值需求变成客观现实的创价活动。翻译主体价值需求的满足包括通过翻译活动实现自我确证,也包括通过满足、创造读者需要获得者确证。因此,中国文学外译批评应从人类创价

活动的现实性与发展性角度来理解和评价中国文学外译的策略方法与读者接受问题。一方面,中国文学外译批评须认识到文学外译对本国价值需求的实现,以及对变革、引导、创造西方读者对中国文学需求的现实意义。另一方面,中国文学外译批评要把握文学外译的阶段性特征与长远性目标。实践的发展性表明,对中国文学外译成效的评价须以历史、发展的眼光来衡量。中国文学外译批评应克服对西方读者价值需要的单向度迎合以及对读者接受的功利性期待,不以当前阶段的译介效果而否定文学外译的长远价值,以及将“连译带改”模式视为文学外译的终极答案。如今,文学外译对变革、引导、创造西方读者需求的价值正在逐渐显现,当前西方读者越来越关注中国文学的文学性与艺术性就是例证。

**把握中国文学外译的内在规定性** 人的实践活动具有合规理性。马克思指出,“动物只是按照它所属的那个物种的尺度和需要来进行塑造,而人则懂得按照任何物种的尺度来进行生产”。由此,翻译批评应自觉认识和

把握翻译客体、翻译活动的内在规定性。一方面,作为翻译客体的中国文学作品的文学性、异质性构成翻译批评的核心维度。解构作品的特性便消解了其存在与价值。另一方面,翻译的发生、存在、发展皆系于“异”。语言、文化、思想间的差异呼唤翻译出场。取消差异等同于消解翻译存在的必要性,以及翻译之于人类思想文化交流互鉴的价值。因此,翻译的“忠实”原则构成了翻译存在的实践品格、伦理规约。只有尊重文学作品的异质性,推动人类文化交往与文明互鉴,翻译才能彰显其存在价值,这便是文学翻译的内在规定性。因此,中国文学外译批评须引导文学外译尊重中国文学的主体性,遵循翻译活动的规律性。

**把握中国文学外译的价值合理性** 人的实践活动具有合目的性。马克思指出,人“能用内在固有的尺度来衡量对象”。由此,主体的价值需求构成翻译批评的核心维度。需求有合理与不合理之分,翻译批评需要评估翻译活动是否具有价值合理性,即合乎人类历史发展的价值追求与根本利益。中国文学外译是否自觉立足于推动多元文化尊重差异与平等对话、共同维护语言与文化多样性,这将确定其合目的性。因此,中国文学外译批评应引导文学外译促进文化价值和谐共存与共享。中国文学外译批评应警惕以西方为中心的价值取向,指引中国文学外译以平等对话的姿态让世界认识、理解和欣赏中国文学与文化的独特价值,推动多元文化形成价值共识。