

人工智能与国际关系学:重新找回“人”

唐士其 庞 珣

内容提要 人工智能发展对国际关系研究形成冲击的一个重要面向是迫使学科对“非人化”和“拟人化”双重思维惯习进行反思和反观。这一技术革命提供了一个契机来重新找回“人”在学科中的核心地位。“仁”与“智”构成了人性的两个关键方面,由人的自由意志及其带来的复杂性、不确定性和可能性构成。人工智能模仿人类的智能活动和仁爱表达,但却无法进行仁与智的内在体验,是“类人而非人”的存在。这一点直击国际关系学以理性假定简化研究对象进行的理论构建和学科关切时面临的学术局限和困境。在人工智能时代,国际关系学需重新审视“目中无人”的传统,转向以“人”为本的研究路径,关注个体的福祉、情感和道德,超越传统的理性假设,探索技术与人文的平衡,确保人工智能成为捍卫人性而非削弱人性的技术力量。

关键词 非传统安全 人工智能 国际关系学 人 仁 智

* 唐士其:北京大学国际关系学院教授。(邮编:100871);庞珣:北京大学国际关系学院教授。(邮编:100871)

** 感谢《国际政治研究》匿名评审专家的意见和建议,文责自负。

国际关系学的传统研究对象是国家及其在各领域中的互动关系。国家是高度复杂的组织,其互动也是战略互动而非简单的“刺激—反应”。因此,国际关系学本质上是关于复杂性和不确定性的学问。然而,国际关系学中长期占据主导的却是追求高度简化和不变铁律的学术旨趣和研究路径。对此的反思和批判也催生了不同学派和视角,但它们却都在不同程度上承袭了主导路径的简化策略,即采取物质理性(material rationality)假定来实现政治行为体的“非人化”,剔除人性的复杂性和不确定性。在这种策略下,无论是个体、群体或组织都可以同一化为“质点人”^①,研究目的是发现和揭示所有“质点人”在国际关系中都必然遵循的客观规律和呈现的稳定模式。

值得注意的是,这种“非人化”策略在国际关系学中一直和“拟人化”倾向并存融合。“拟人化”是以谈论“人”的方式和态度来言说和理解国际关系行为体,如国家等政治实体的生存、成长、欲望、偏好等。尽管“拟人化”看似与“非人化”之间相互矛盾,但两者又构成了国际关系学既有分析框架的同一硬币的两面。“拟人化”所“拟”之“人”并非有血有肉、有社会上下文的人,而是“质点人”,这不仅合理化了“非人化”策略,而且强化了这样一种认知,即剥离了人性的理性行为体不但是“人”,而且是理想类型的“人”。同时,也正是“非人化”策略让以“拟人化”方式谈论政治行为体显得极为合理和自然,因其让人性中的重要面向成为国际关系分析中的冗余。

这种矛盾性“和谐”如今正面临着一种始料未及的力量所带来的巨大冲击。近年来,人工智能迅速发展并被日益广泛地部署在各个社会领域,国际关系学也不例外。由此,国际关系学也面临着新的挑战 and 机遇:一方面,国际关系学者将人工智能作为研究对象,延续“非人化”和“拟人化”的双重策略;另一方面,人工智能也作为追求知识的工具和辅助,在方法论和技术层面进入学科。然而,这一新兴技术在本体论层面上对国际关系学的重要性,学者们迄今为止还鲜有触及,缺乏以人工智能来反观“人之为人”的学术自觉,反思这种以

^① 唐士其:《理性主义的政治学:流变、困境与超越》,北京大学出版社2022年版,第86页。

“拟人又非人”的方式来观察和理解研究对象的传统。

人工智能作为一种颠覆性技术,其迅猛发展带来普遍兴奋的同时,也引发了具有远见的担忧,尤其是关于它对人类可能造成的“存在性”风险。^①“存在性风险”中的存在,不仅指人的生物性存在,而更重要的意涵是关于人性的特征和本质。与其他科技不同,人工智能旨在对人类智能进行效仿和“制造”,其发展离不开对人进行深入思考和不懈理解。这个过程同时也是对人的反观和思索,延续、改变和挑战了人类长期以来对自身的思索和困惑、骄傲和沮丧、误解和见地,拷问智能为何、心灵为何、意识为何。近年来,人工智能在逻辑推理、文字创作、知识运用等各个方面以惊人的速度超越人类的平均水平,甚至在一些领域胜过人类中的佼佼者,使“人何以为人?”的问题变得史无前例地紧迫和尖锐。如果高级智能不再独属人类,那它就无法成为人的根本性特征。如此,在人工智能时代,又该如何定义和实践“人”的存在?

这一根本性问题在国际关系学中表现为“理性人”传统陷入极为尴尬的境地。“理性人”原本是一种虚拟的、用以激励现实决策者向其逼近的理想类型,但人工智能这种建立在最优化逻辑基础之上的复杂体系产生了惊人的理性智能,让理想中的“理性人”从遥不可及的虚拟世界突然降临现世。人们在惊叹之余,也立刻发现理想变为现实后的大量问题及其可能产生的风险。当前,控制人工智能风险的基础路径恰恰是用具有种种人性“缺陷”的人类来对这一最优化系统进行反馈强化训练^②,让“理性人”在言行上变得更具人性。

可见,人工智能对“现实之人”和“理想之人”产生双重冲击,这给国际关系

① See Toby Ord, *The Precipice: Existential Risk and the Future of Humanity*, New York: Hachette Books, 2020; Stuart Russell, *Human Compatible: Artificial Intelligence and the Problem of Control*, New York: Viking, 2019.

② Iason Gabriel, “Artificial Intelligence, Values, and Alignment,” *Minds and Machines* 30, No.4, 2020, pp. 411-437; Paul Christiano, et al., “Deep Reinforcement Learning from Human Preferences,” *31st Conference on Neural Information Processing Systems (NIPS 2017)*, Long Beach, CA, USA. Long Ouyang, et al., “Training Language Models to Follow Instructions with Human Feedback,” *arXiv preprint* (2022), arXiv:2203.02155.

学者以契机,来直面一个从前羞于启齿的问题:这个学科是否不明智地遗忘了“人”?对这个问题进行思考,就是在人工智能技术迅猛发展的背景下,去重新定位国际关系学的关注与关切。学科的核心问题从未如此醒目地呈现出来:如何理解和处理国际关系,从而让新技术成为捍卫人性的工具?围绕这一问题展开的反思和探索,是为国际关系学寻求一种融合技术与人文的视角,将“人”置于学科的中心和重心。这不仅是学科对人工智能发展的回应,更是对人工智能时代的国际关系学发挥价值和做出贡献的探寻。

一、“类人而非人”的人工智能

人工智能掀起了一场空前的技术革命,在其技术层面表现出强大的计算和分析能力,更是引发了深刻的哲学、伦理和社会思考。一直以来被认为只有人类才具有能力处理的复杂任务,正在一个又一个被人工智能以高速、高效和高度的准确性加以完成,从语言理解、图像识别、创意生成到学术研究,人类智能的高超之处和独特之处正在迅速消失。人工智能在越来越多的领域显示出远超人类的智能,如在极短时间内分析海量数据,发现人类无法察觉的隐藏模式,在高维空间中进行抽象推理,琴棋书画和诗词歌赋也同样才情出众。

人工智能的智能不仅媲美人类的个体智能,在社会智能方面也突飞猛进,具有社会性交互能力,承担起一些原本独属于人类的社会角色。人工智能高度“人性化”,对人类在社会交往中的语言、语气和情感进行模仿,与人类的互动越来越自然,建立起原本只在人际中才能产生的社会纽带,如在复杂任务上人对机器的信任和依赖,人类在获取知识中将其置于权威地位,以及对其倾吐心声建立起情感依恋。

人工智能对人类的模仿和超越及其与人类的互动,引发了对“人类中心

论”崩塌的担忧。^① 人类作为一个整体,在历史上第一次面对以下问题:人工智能在越来越多的领域平替甚至超越人类,是否会让人类陷入“无用”境地? 当社会意义和社会价值不再独属人类时,人的自我意义和自我价值又该如何持续? “人之为人”何去何从的问题从来没有像今天这样具有整体性和紧迫性,迫使我们重新审视“人”的内涵。

(一) 关于人性:人何以为“人”?

中国的传统思想体系和古希腊罗马的思想传统都对人性进行过深刻讨论。东西哲思的一个共识是,“仁”与“智”构成人性的两个关键方面。中国的儒家思想是以“仁学”为核心的思想体系,强调“仁者爱人”。其中,“仁”是对人给予的一种普遍性的同情和关怀。如何能够以适当的方式为“仁”? 致力于回答这一问题的儒家思想,提出一系列的方法和途径,比如,义、礼、智都是达致“仁”的途径。其中,“义”和“智”与人的判断有关。同时,“义”是为人处世的根本原则。孔子说:“富与贵,是人之所欲也,不以其道得之,不处也;贫与贱,是人之所恶也,不以其道得之,不去也。君子去仁,恶乎成名? 君子无终食之间违仁,造次必于是,颠沛必于是。”^②孟子说:“非其道,则一箪食不可受于人;如其道,则舜受尧之天下不以为泰。”^③但“义”也是在特定环境下对行动原则和行为方式的选择,所以说“义者宜也”^④,需审时度势、慎思明断。“言必信,行必果,硁硁然小人哉!”^⑤,“言不必信,行不必果,唯义所在。”^⑥可见,“义”是一种特殊的智,“仁者爱人”在实践“仁”时,须以“智”与“义”进行判断,正所谓“智者利

① Joanna J. Bryson, “Patience Is Not a Virtue: The Design of Intelligent Systems and the Future of Moral Theories,” *Ethics and Information Technology*, Vol.20, No.1, 2018, pp. 15-26; Luciano Floridi, *The Fourth Revolution: How the Infosphere Is Reshaping Human Reality*, Oxford: Oxford University Press, 2016.

② 《论语·里仁》。

③ 《孟子·滕文公下》。

④ 《礼记·中庸》第二十章。

⑤ 《论语·子路》。

⑥ 《孟子·离娄章句下》。

仁”。尤其值得注意的是,“智”的判断并不存在一般性的标准,而是要在特定时间、特定场合、针对特定事件进行恰当判断。“恰当”与否不由一套确定标准来加以评判,而是“知进退存亡而不失其正者其唯圣人乎?”理想中的恰当是“夫大人者,与天地合其德,与日月合其明,与四时合其序,与鬼神合其吉凶。先天而天弗违,后天而奉天时。天且弗违,而况於人乎?况於鬼神乎?”^①

相对于儒家思想而言,古希腊思想并不把“仁”放在如此显要的位置,但在“与人为善”这一人性基本特征上,双方有着强烈的共鸣。柏拉图强调正义之人不应伤害他人,而亚里士多德在《伦理学》中用大量篇幅来讨论友谊,认为友谊的基本特征是相互为善。诚然,古希腊人把理性视为人的基本特征,但现代人对理性的理解具有本质不同,视理性为智慧的直觉(nous)和逻辑的思考(logos)而非利益算计。在古希腊人看来,理性更多是人看穿事物本质形式(form)的能力。例如,亚里士多德认为人的智慧有理论智慧(sophia)和实践智慧(phronesis)两个方面,前者使人看透事物的本质,后者帮助人们在实践中做出正确判断。可见,这两种智慧其实就是中国古人所言的“智”与“义”。尤其值得一提的是,与中国古代思想相应和,在亚里士多德看来,所谓的美德(virtue)也是人在特定时间、特定场所、针对特定事件进行恰当判断的能力,而“恰当”也没有确定标准,是一种“可能”。

与古希腊人不同,卢梭并不把理性视为人的根本标志。在他看来,理性与语言一样,都是人在应对环境变化的过程中发展而来的。人的基本特征在于三个方面,包括自由意志、可完善性和同情自爱。这三个方面在道德意义上相互关联。正因为具有同情和自爱的本性,人才希望对别人和对自己“更”好,也才能够对别人和自己“更”好。这表明,人可以改善自己的思想与行为,而不是随自然欲望而为,这正是人的自由的体现,也是卢梭所言自由与霍布斯所言自由的根本区别。

中西思想的交汇和呼应,强调了人之为人并不在于某些显现于所有人身

^① 《周易·乾卦·文言》。

上的普遍性因素,而在于人具有某种潜在能力,可称之为“可能性”或“可普遍化的可能性”(universalizable possibility)。什么是“可普遍化的可能性”?颜回对孔子的评价是“仰之弥高,钻之弥坚,瞻之在前,忽焉在后”^①,子贡的评价是“夫子之不可及也,犹天之不可阶而升也”,又说“夫子之墙数仞,不得其门而入”。^②司马迁说:“高山仰止,景行行止,虽不能至,心向往之。”^③可见,“可能性”并非每个人都能做到的“可能”,而一种每个人都能理解和向往的“可能”,是一种草上之风的“可能”。可能性的实质在于,人可以与人为善,也可以判断如何与人为善。这并非所有人在所有时候都能够做到,也并不按照某种外在的标准加以实践。每一个具体的实践,都具有无限丰富和完善的可能。因此,人性具有特殊性、可变性与可能性的特点。

(二) 人工智能的“类人”与“非人”

何处是人工智能和人之间的界限,是当前技术哲学和技术风险学关注的重大问题。这一问题不仅涉及技术的可能性,更是要求思考人类自我认知的深层边界和存在意义。关于人工智能的公开讨论显露出一种认识倾向,即对人机边界的模糊甚至消融的预期,相信随着技术的指数级进步,人工智能不仅能够模拟人类智能,还可能发展出真正的意识和自我认知。^④这种倾向的代表性观点是“奇点理论”,声称“人性”不一定要建立在生物性基础上,而是可以在复杂的信息处理中实现意识涌现。因此,人工智能完全有可能形成一种新的存在形式,并与人类融合。^⑤一些哲学家从意识哲学角度认同了这种可能性,

① 《论语·子罕》。

② 《论语·子张》。

③ 《史记·孔子世家》。

④ Giulio Tononi and Christof Koch, “Consciousness: Here, There but Not Everywhere,” *Philosophical Transactions of the Royal Society B: Biological Sciences*, Vol.370, No.1668, 2015, 20140167, <https://royalsocietypublishing.org/doi/epdf/10.1098/rstb.2014.0167>, 2025-06-10; Susan Schneider, *Artificial You: AI and the Future of Your Mind*, Princeton: Princeton University Press, 2019.

⑤ Ray Kurzweil, *The Singularity Is Near: When Humans Transcend Biology*, New York: Viking, 2005; Ray Kurzweil, *How to Create a Mind: The Secret of Human Thought Revealed*, New York: Viking, 2012.

认为如果意识是信息处理的结果,那它就可以与生物体相分离。^① 对这些激进观点的批评认为,人工智能无论如何复杂和强大,都无法理解或拥有意识。例如,曾任加州大学伯克利分校教授的约翰·塞尔(John Searle)提出“中文房间”(Chinese Room)思想实验,指出人工智能的行为只是对输入数据的机械反应,缺乏真正的理解和意义。^② 牛津大学教授罗杰·彭罗斯(Roger Penrose)从量子力学的角度推导出人类意识与物理世界深层结构具有密切相关,而人工智能无法通过算法模拟建立起这样的关系。^③ 需注意的是,人工智能是否能产生意识并不是问题的全部,重要的是它已经开始改变和改造人类的意识。人性具有不确定性和可能性,人工智能已经开始操控这种不确定性和可能性,假以时日便可深刻改变人的存在和重新定义“人”的概念。例如,加州大学圣克鲁斯分校教授、后人类主义者唐娜·哈拉维(Donna Haraway)提出,随着人工智能的发展,生物技术、脑机接口及人工智能的广泛部署,在生理和社会层面上,技术日益与人交融,技术增强下的人类可能变成一种“赛博格”式的存在,在方式、意义、身份、伦理等各方面发生改变。^④

可见,人工智能和人之间的界限,既是技术问题,也是哲学和伦理问题。既有讨论基本上围绕着意识展开,而较少从人性角度来进行讨论。以人性之核心的“智”与“仁”作为参照,能够帮助捋清人工智能和人之间的相似与区别:对人进行模仿的人工智能可以称得上是“类人”,但其“非人”本质却难以掩藏。

1. 人工智能的“智”。如前所述,人性之“智”是在特定时间、场合和事件中进行恰当判断的能力,这种判断没有固定的标准,要依赖于个体的实践智慧和

① David J. Chalmers, *The Conscious Mind: In Search of a Fundamental Theory*, Oxford: Oxford University Press, 1996; David J. Chalmers, “Could a Large Language Model Be Conscious?” *arXiv Preprint*, 2023, arXiv:2303.07103.

② John R. Searle, “Minds, Brains, and Programs,” *Behavioral and Brain Sciences*, Vol.3, No.3, 1980, pp. 417-424.

③ Roger Penrose, *The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics*, Oxford: Oxford University Press, 1989.

④ Donna J. Haraway, *Simians, Cyborgs, and Women: The Reinvention of Nature*, London: Routledge, 1991.

理论智慧,区别于遵循特定程序和标准的最优化能力。从某些方面来看,人工智能的智能与人性之“智”具有关联性和相似性,在能够剥离社会上下文和忽略道德审视的一些判断和决策场景中,两者之间的区别尤其模糊。在纯技术领域的决策,“恰当”可能等于最优化,但这些场景并不是人性之“智”所关切之“智”。人工智能的智能模仿也涉及理论智慧的应用,通过大量的数据训练来学习复杂的模式和规律,根据特定任务和环境来做出类似于人类的“定制化”判断。人工智能模型还能够对数据进行抽象和泛化,在一定程度上展现出理论智慧所要求的对事物本质进行理解和洞察。

相较于人工智能对人性之“智”中的理论智慧的模仿,它对实践智慧的模仿还显得十分笨拙。人工智能基于数据和算法进行判断和决策,缺乏人类的道德和伦理考量。例如,自动驾驶汽车在遇到道德困境时,无法像人类一样自行做出符合伦理、超越伦理或者违背一般伦理的决策。它的道德选择来自人类积累的已有信息和强化反馈学习,并没有内在的道德感。人工智能在表达方面已经可以惟妙惟肖地模仿人类的情感和直觉,但并没有发自“内心”的情感、道德和直觉。与之不同,人类在做出判断时,情感和直觉因素会从潜意识层面涌出,并非由意识进行控制。^①最优化(即理性)不是人类决策的唯一机制,甚至可能不是最重要的机制。即使在重大决策中,决策者虽然会进行反复斟酌和理性讨论,尽量搜集信息来计算预期成本收益。但最终决策是把不确定性的选项变为再也无法更改的事实,此刻决策者内心的渴望和呼唤往往以“非理性”方式来给予人告别可能性的勇气。换言之,即使对决策者本人而言,决策很可能都是一个难以完全用理性加以解释的“黑箱”。

柏拉图把人的灵魂分为三个部分,包括理性、激情和欲望;亚里士多德也

^① Jonathan Haidt, “The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment,” *Psychological Review*, Vol.108, No.4, 2001, pp. 814-834x; Joseph LeDoux, “Emotion Circuits in the Brain,” *Annual Review of Neuroscience*, Vol.23, No.1, 2000, pp. 155-184; Jonathan Haidt, *The Righteous Mind: Why Good People Are Divided by Politics and Religion*, New York: Pantheon Books, 2012; Joseph LeDoux, *The Emotional Brain: The Mysterious Underpinnings of Emotional Life*, New York: Simon & Schuster, 1996.

认为植物灵魂、动物灵魂和理性灵魂共同构成了人的灵魂；而弗洛伊德则认为自我、本我和超我都是“我”。这些都表明理性决策不过是人的智慧的浅层部分和局部构成。在东方,《中庸》有言:“君子之道,本诸身,征诸庶民,考诸三王而不缪,建诸天地而不悖,质诸鬼神而无疑,百世以俟圣人而不惑。质诸鬼神而无疑,知天也;百世以俟圣人而不惑,知人也。”^①这里的决策标准不是基于客观信息分析来对客观概率的估计和对预期结果进行最优化。相反,决策可以也应当被合理化为一种伦理或者美学的创造,与物质利益无关,甚至可以背道而驰,倡导通过这样的决策实现人的不断自我超越,最终得到天地神人的认可。

2. 人工智能的“仁”。“仁”是一种普遍性的同情与关怀,是人性中重要的道德维度,强调人与人之间的情感联系和道德责任,表现为对他人痛苦和福祉发自内心的关注和关怀。在人工智能的发展中,尤其在关于人工智能风险管理问题上,人们付出了大量而持续努力,让人工智能的行为和判断表现出这种“仁”,来对齐人类价值观和道德观。人工智能正在被设计为践行“仁”的应用系统,部署在心理健康等领域,为孤独或抑郁的人提供情感支持。

如果说人工智能系统可以具有一定程度和不同类型的“智”,但至少到目前为止,基于算法和数据的人工智能无法产生出人性之“仁”。“仁”是一种道德实践,需要人类在具体情境中做出道德判断和选择。人工智能的“关怀”缺乏道德判断,无法通过“询问良心”来做出道德选择,对其选择产生的后果也不会感受到道德上的自我谴责和自我反省。这种局限性使得人工智能无法实践“仁”的道德维度。

建立在人类情感和道德基础上的“仁”,和人的“肉身”无法分割。真实的爱与恨、荣誉与羞耻,必然是人的内在体验,而不一定要对外进行表达。情感 and 道德体验总是和人的生理反应同步,即弗洛伊德所言的“本我”(id)。当且仅当它们是理性的策略表达和沟通技巧时,才会与人的生物基础相分离。人工

^① 《中庸》第二十九章。

智能的“关怀”仅是对“仁”在行为层面的模仿,它没有内在的情感和道德体验,其语言和行为完全依赖于预设的程序和数据的输入。例如,一个人工智能医疗诊断系统可能会根据数据推荐最佳治疗方案,也可能用语言表达出“医者仁心”,但它无法对患者的痛苦或焦虑感同身受,缺乏人类情感中的自发性和深度。由于“仁”强调人与人之间的互动和情感联系,人类对“仁”的实践是通过互动和共情来建立情感纽带。但人工智能的“关怀”则缺乏这种互动性。一个聊天机器人可以模拟共情,也可以让人类对其产生出感情和关爱,但这种情感纽带是单向而非互动的。

3. 人工智能无法模仿的人性本质。人工智能在“智”和“仁”两方面体现出了其类人而非人的特性,根源在于它可以无限接近地模仿人类“智”和“仁”的表象,但缺乏内在的自由意志及意志的可完善性,而后者才是人性的核心。人类具有通过自我反思和自我改进来提升思想和行为的能力,这种自由意志不仅体现在个体对目标和价值观的选择能力,还在面对复杂情境时能够进行道德判断和自我修正。人工智能的学习,包括自我强化学习,都建立在人类的设计和数据的输入基础上,尤其在目标和价值观方面,无法由内而外地产生出来和自主决定。换言之,人工智能的学习是从数据中进行模式提取而不是基于自主意识的自我完善。人工智能的“可完善性”是技术层面的进步,而非哲学意义上的自我超越。这将“人”和人工智能彻底区分开来。

自由意志的独特之处不在于对规律的发现和运用,而在于对规律的突破和对可能性的探求。这决定了人性的特殊性与可变性。尽管不乏有人担忧在未来发展中人工智能有可能涌现出意识,但这种“意识”显著不同于自由意志。人工智能的可变性基于算法和数据,可以对外界做出多样化的反应,如大语言模型的温度参数上调能够使其生成的文本具有更大的随机性和“创造性”。这些都是受操控的自由度,从本质上与自由意志背道而驰。人性的特殊性和可变性意味着,每个人在不同的情境下产生的不同反应和不同行为来自人的自由意志。孟子讲到四端中的恻隐之心,说:“所以谓人皆有不忍人之心者,今人

乍见孺子将入于井,皆有怵惕惻隐之心,非所以内交于孺子之父母也,非所以要誉于乡党朋友也,非恶其声而然也。”^①此处的“怵惕惻隐”,人们在决策和行动中都有着切身体验,不是来自理性的计算,而是一种不受人意识控制的生理反应以及同步产生的行为冲动。这种特殊性和灵活性源于人类复杂的情感、认知和道德体系,这正为人工智能所缺乏。

因此,从智与仁两个方面来看,人工智能都是类人而非人。人性中的“智”超越了观察、理解、学习和行动的层次,而是要求达到自主地取舍进退。人类所仰佩的“智者”,绝不是事事精打细算,也不是能够最优化效率和完美执行计划,而是如《周易》所言的“知至至之,知终终之”,“知进退存亡而不失其正”。^②在“仁”的方面,人工智能的模仿更是流于表面,无法拥有“仁”的体验,缺乏内在情感、道德和自由意志。人工智能即使能在智力方面媲美甚至超过人类,也无法产生“人工人性”,不可能复制或替代人性的复杂和深度。

二、“目中无人”的国际关系学遇见“似人非人”的人工智能

有关人工智能与人之间本质差异的既有讨论,集中于管控技术风险,为人类在人工智能时代锚定和保留具有尊严和意义的位置。这也是作为人类文明传承和发展的每个学科,在人工智能冲击下不得不思考的重大问题。当然,这样的思考首先应该以对学科自身反思为前提。在人工智能时代之前,很少有人意识到,国际关系学实际上是一门研究“人工智能体”的学问——在国际关系研究中,国家、国家组织、超国家组织甚至作为决策者的个人都被剥离了“人性”,从而被处理为与人工智能在本质更为接近的“智能体”。在理论的简化下,这些研究对象甚至是比当今的人工智能更为单薄和简单的“智能体”——它们的“智”高度扁平化和单一化,而它们的“仁”则因被视为软肋或欺诈而不

① 《孟子·告子上》。

② 《周易·上经·乾卦·文言》。

值一提。

(一) 国际关系行为体的“非人”假定

“非人”假定是国际关系学主流理论的核心方法论工具,将国家和其他行为体视为单一、理性和非人格化的决策主体。这一假定是现代国际关系学之父汉斯·摩根索(Hans Morgenthau)坚定明确的理论主张。他的思想始于对“人性”的悲观认识,追求权力源于人生而有之的欲望。由人构成的国家具有与人性欲望同样的权力渴望,但在摩根索奠定的理论传统中,国家的“类人化”只保留了权力欲而摒弃了人性的其他面相,更重要的是,权力欲在其中也由欲望转化为理性,即由权力所定义的利益,其最大化即为国家理性。^① 这里,国家作为单一行为体的“拟人化”和“理性化”是同一个过程,共同导致将国家行为体的理性动机锚定在对权力的追求上,将复杂的国际关系简化为权力逻辑,人性问题则从此与国际关系不再相干。

肯尼思·沃尔兹(Kenneth Waltz)提出的国际关系结构现实主义理论进一步将“非人化”的理性假定推向极端。在他这里,经典现实主义以人性为起点的思考毫无必要,国家的理性来自国际体系的基本结构特征,而非国家的内部构成,因此对“人”的讨论是冗余的。^② 国际体系的无政府状态是国际关系行为体在利益和行动上的主宰,而体系中国家之间的实力分布决定了国家间的互动关系,即如何达到、维持和破坏势力均衡,进而决定了体系在战争与和平意义上的稳定性。这种外部结构决定的行为体理性,使得行为体本身的意义被最小化为对结构性逻辑的表达和呈现,其国内的政治、文化、历史等因素对国家的对外行为毫无影响,在此基础上建立的理论具有极致的“非人化”和“理性化”特色。

① [美]汉斯·摩根索:《国家间政治:权力斗争与和平》,[美]肯尼思·汤普森、戴维·克林顿修订,徐昕、郝望、李保平译,北京大学出版社2006年版。

② [美]肯尼思·N.华尔兹:《人、国家与战争》,倪世雄、林至敏、王建伟译,上海译文出版社1991年版;[美]肯尼思·华尔兹:《国际政治理论》,信强译,上海人民出版社2008年版。

国际关系新自由主义学派和建构主义学派的出现和发展,伴随着对现实主义尤其是结构现实主义的尖锐批评,在一定程度上找回了国家行为体的能动性、差异性和可能性。然而,它们的批判无论显得多么离经叛道,但却几乎全盘接受了现实主义对国际关系行为体的“似人非人”处理方式。例如,亚历山大·温特(Alexander Wendt)强调观念和认同的建构,国家行为体的身份和利益并非固定不变,而是在社会互动和文化实践中不断重塑,但建构的主体及“主体间性”中的“主体”仍是被剥离了人性的抽象存在。这些理论在打开国家“黑箱”之后引入了其他次国家和超国家行为体,但无论它们是国际组织、跨国企业还是政治家或选民,仍然都是抽象化的理性行为体。^①

进一步看,即使是那些对理性假设进行批判的学者,也依然把理性假定下的决策视同决策的理想基准,探讨的重点在于揭示现实对理想的偏离,而这种偏离就是对理性的违反。塞缪尔·亨廷顿(Samuel Huntington)的“文明冲突论”强调了文化在国际关系中的重要性,将文化的差异性作为冲突的来源和形式,但他的分析仍然停留在宏观层面。^② 罗伯特·杰维斯(Robert Jervis)在其著作《国际政治中的知觉与错误知觉》中指出,国家行为体的决策并非完全理性,而是受到领导人个性、社会情绪和认知偏差的影响。这些因素用来解释外交决策中的错误和无效为何产生。^③ 近年来发生在国际关系领域中的“行为转向”,呼唤对认知情感等因素进行高度关注,但目的也最终落脚在解释这些因素为何让当今国际关系乱象丛生,让决策漏洞百出。^④ 换言之,一个普遍的观念在学科中依然根深蒂固:对理性假定的偏离在现实中普遍发生,这种偏离就是偏差和错误,人性中非理性部分对国际关系形成干扰。

① [美]亚历山大·温特:《国际政治的社会理论》,上海人民出版社 2014 年版。

② [美]塞缪尔·亨廷顿:《文明的冲突与世界秩序的重建》,北京:新华出版社 2010 年版。

③ [美]罗伯特·杰维斯:《国际政治中的知觉与错误知觉》,上海人民出版社 2015 年版。

④ Emilie M. Hafner-Burton, et al., “The Behavioral Revolution and International Relations,” *International Organization*, Vol.71, Special Issue 1, 2017, pp. S1-S31.

(二) “非人”假定的理论和政策的后果

“非人化”理性假定产生了具有高度逻辑性和普适性的国际关系理论,也促进了形式逻辑在学科中的应用发展,例如,通过数学模型和博弈论分析国家行为。然而,这一路径忽略或极大简化了决策的复杂性和不确定性。例如,博弈论模型虽然为理解国际冲突提供了重要工具,但其忽略动态进程和高度简化还原,提供的往往不过是“后知之明”,而缺乏基于可能性、灵活性和复杂性的前瞻性智慧。

在政策实践中,“非人”假定对现实的过度简化造成更为严重的问题。专注于“利益最大化”的单一理性,忽视历史记忆和文化认同,让“利益”变为既空洞、又极易作为一个口号遭到滥用的概念。在单一理性下,国际冲突显得不可避免,而国际合作极为艰难,因为高度同质化的利益想象,人为增加了国际关系中的“稀缺性”。理性行为准则的最优化原则,让国际关系中的妥协更为困难,且妥协被错误地等同于在权力胁迫下的退让。剥离了人性的“智”,必然导向“斗智斗勇”的“智”,而摒弃了人性的“仁”,则让妥协精神和礼让行为落为“软弱可欺”的笑柄。这种对“人”的彻底忽视让国际关系学者常常不自知地陷入“国家理由”(raison d'état),为国家反道德的行为提供正当性和合理性,使国家利益成为“集体的自私”。^①

“非人”假定导致国际关系学科对人性关怀的长期缺失,忽视了“人”作为国际关系核心主体地位,成为一个“目中无人”的学科。在这样的学术传统下,国际关系研究在面对国家利益、国家安全、全球安全和全球发展这些宏观而重大的议题时,自然地忽略对个体、群体及人性的关切,在工具和目的、手段和价值问题上陷入混乱。比如,“非人”假定让国际关系学者对国家利益缺乏深入

^① 列奥·施特劳斯在《马基雅维利思想》一书中说:“即使我们被迫承认马基雅维利本质上是一位爱国者或者科学家,我们也不能在压力之下忽视他是一名邪恶教师。马基雅维利理解的爱国主义是一种集体的自私。因为对自己国家的热爱而对正当与错误的区别无动于衷,要比因为一味计较自己的安逸与荣耀而对这种区别无动于衷更令人反感。但也正因为如此,它更具诱惑而且更危险。”参见 Leo Strauss, *Thoughts on Machiavelli*, Chicago: University of Chicago Press, 1958, p. 11.

思考,而想当然地认为它是一种外生给定的原动力,国际关系学的任务只是去探讨如何对其最大化,从而过度强调权力争夺而忽视国家行为与个体命运之间的相互作用。正如玛莎·芬尼莫尔(Martha Finnemore)在其著作《国际社会中的国家利益》中指出的,国家利益并非一成不变,而是受到来自外部的国际规范和来自内部的社会和个人价值的塑造。^① 这些规范和价值的核心正是“人”的需求和期望。如果缺乏对个体命运的关切,国际关系学就难以深刻地理解国家行为的道德和社会后果。

在诸如战争、冲突或经济制裁等国际博弈中,个体的生命、尊严和福祉往往成为牺牲品。国际关系学者倾向于津津乐道博弈本身的输赢,对宏观博弈下个人被迫承担的代价视而不见。这种“目中无人”的研究取向,让国际关系学在应对全球性问题、国际危机和攸关人类未来的重大风险时显得冷漠和无力。正如牛津大学教授阿奇·布朗(Archie Brown)所言,国际关系研究如果仅仅关注国家利益而忽视个体福祉,便无法为全球治理提供真正有效的解决方案。^②

尽管全球治理在近些年来成为国际关系研究的新议题,但“非人”假定仍然主导着研究议程,并对分析角度有着根深蒂固的重要影响。国际关系学对全球治理的理论贡献集中于“理性制度设计”,关注国家间如何达成合作以应对诸如气候变化等全球性问题。^③ 其中,国家间的权力争夺和利益分配仍然主导研究兴趣和占据学术重心,倾向于将全球治理视为国家间博弈的新领域。然而,正如阿马蒂亚·森(Amartya Sen)在其著作《以自由看待发展》中指出,发展的核心是扩大个体的自由和能力,而不仅仅是经济增长或制度完善^④,国

① Martha Finnemore, *National Interests in International Society*, Ithaca: Cornell University Press, 1996.

② Archie Brown, *The Myth of the Strong Leader: Political Leadership in the Modern Age*, New York: Basic Books, 2014.

③ Barbara Koremenos, et al., “The Rational Design of International Institutions,” *International Organization*, Vol.55, No.4, 2001, pp. 761-799.

④ Amartya Sen, *Development as Freedom*, Oxford: Oxford University Press, 1999.

际关系学的“目中无人”,在面对当今充满复杂性、可能性和可变性的国家关系和全球治理时,这个学科的适应性和贡献力正在显著降低。

(三) 人工智能与国际关系研究

在人们纷纷讨论人工智能的兴起对学科(包括国际关系学)进行替代、赋能、转型的可能性时,关注的重点通常聚焦在人工智能如何扩展研究对象和升级研究手段。但在人工智能对人性的镜鉴下,人们看到这场技术革命如何直击学科的终极关怀,提出如何进行相应调整和重塑的问题。对这个问题的回答将影响到国际关系学在人工智能时代的价值和意义,也决定了人们该如何采取行动来实现这些价值和意义。

自动化分析和建模让人工智能与“非人”假定在方法论上天然契合。以数据为驱动的人工智能分析方法,很容易顺着学科的惯性而强化“非人”视角:算法可以用来识别国家行为模式和预测冲突风险,对国际关系进行模式化整理和理性化解读提供了新工具。国际关系学者早就开始应用算法分析大规模的国际事件数据,预测冲突和合作的趋势。^① 人工智能迅速兴起和高速迭代,可以极大助力通过模式的识别和提取来简化复杂现象和趋势,鼓励国际关系学者把预测的雄心扩展到更广阔的领域和更丰富的议题。如果这种趋势成为主流,国际关系学的视野将进一步远离“人”的核心地位,加深以“非人”假定为基础的学术惯性和路径依赖。

然而,这并不是人工智能对国际关系学唯一可能的影响。当强大而“非人”的人工智能系统广泛进入到人们生活的方方面面时,人的价值问题被高度凸显出来,每个学科都需要反思以“能力”来定义“价值”的传统认知。人的价值遭遇技术的冲击和质疑时,也可能正是深刻理解和热情拥抱人之价值的契机,同时也可能将注意力进一步引向在哲学意义上、而非仅止于社会意义上的

^① 庞珣:《定量预测的风险来源与处理方法:以“高烈度政治动荡”预测研究项目的再分析为例》,《国际政治科学》2017年第3期。

人之价值。比如,哲学思考逐渐进入人工智能对齐领域,提供价值规范和理论框架。而触发和驱动这些思考和介入的,正是人工智能发展对作为“人性之人”的个体和群体可能造成存在性风险。

人工智能作为复杂系统,很可能帮助重拾国际关系学中的“人性”维度。在学科发展的漫长历史,科学化和技术化将政治哲学的思辨传统逼进边缘的角落。但在人工智能时代,学科亟需关于人与政治的哲学沉思,要求“以人为本”来调整学科,在人工智能席卷人类知识体系的浪潮之中,让学科因坚守社会科学的价值而屹立不倒。因此,人工智能的发展很有可能是一次独特的契机,促使学科探索技术与人文之间的平衡,以技术维护和促进复杂性和多样性下国际关系中人的发展可能,为在国际关系宏观结构和宏观大势中安放“人性”探索出路。

三、以“人”为本:国际关系学的新路径

在过去几十年中,随着国际关系实践的变化,国际关系学的观察和思考逐渐从宏观层面下移到微观层面,尤其是追踪了具有特质的个体在互动中通过制度来形成加总的利益和力量。越来越多的学者加入挑战传统整体论假设的队伍中,打开国家和其他组织的“黑箱”,将分析单元下沉到其组成部分。然而,这些努力却又被诟病为“还原主义”,因其在对微观层面加以关注时,仍然将微观单元视为组织的机械性“部件”。换言之,这些路径即使在讨论诸如劳工等“人”时,仍然将他们“非人化”,视为不同利益的载体,因其承载的利益进行理论演绎,得出其规定性行为。例如,个人被处理为不同资源禀赋的拥有者,根据其拥有何种资源禀赋而决定支持或反对特定的政策,其逻辑正是最大化自己的经济利益。^①

^① Jeffrey A. Frieden and David A. Lake, *International Political Economy*, 4th edition, Boston: Bedford/St. Martin's, 1999.

这些根深蒂固的“非人化”研究路径表明,国际关系学必须彻底审视国家中心主义的理论取向,放松在此基础上已经被固化的“非人”假定。这要求国际关系学者以高度自觉和艰巨努力,将在人性意义上的“人”重新找回,转向以“人”为本的研究路径,探索人的“智”和“仁”在历史、现实和未来如何塑造国际关系和全球政治。

(一) 以“人”为本的国际关系研究的主要特征

以“人”为本的路径要将注意力从宏大视野转向微观分析,尤其关注个体与国际关系现象之间的相互作用和彼此关联。在国家是理性单一行为体的传统假定下,决策者个体的信仰信念、实现其自身可能性的情感冲动、以及在道德取向下的妥协和内心对“恰当”的判断,这些丰富多彩的因素都被排除在了决策模型之外。这些因素即使在理性假定的掩盖下也仍然在国际关系学中偶尔可见。如格雷厄姆·艾利森(Graham T. Allison)在其经典著作《决策的本质》中,探讨了“古巴导弹危机”中美国决策者的认知偏差和组织惯性。^① 这样的讨论虽然以理性为参照,但也揭示出在危机决策中极为关键的“人性”要素。揭秘档案显示,“古巴导弹危机”之所以能够避免一场全球的灾难,并不是由于后来一度为人津津乐道的“边缘政策”理性讹诈的成功,也不仅是白宫和克里姆林宫里领导人的深思熟虑,而是得益于在海上和空中行动的众多个人在千钧一发之际遵从内心“智”和“仁”的判断。^② 然而,在国际关系学的主流叙事中,这个“个人”被忽略了,留下了高度简化的版本作为两极格局下势力均衡维持稳态的力证。^③

① Graham T. Allison, *Essence of Decision: Explaining the Cuban Missile Crisis*, Boston: Little, Brown, 1971; Graham T. Allison and Philip Zelikow, *Essence of Decision: Explaining the Cuban Missile Crisis*, 2nd edition, New York: Longman, 1999.

② Michael Dobbs, *One Minute to Midnight: Kennedy, Khrushchev, and Castro on the Brink of Nuclear War*, New York: Knopf, 2008; Sheldon M. Stern, *The Cuban Missile Crisis in American Memory: Myths versus Reality*, Stanford: Stanford University Press, 2012.

③ Kenneth N. Waltz, “The Origins of War in Neorealist Theory,” *Journal of Interdisciplinary History*, Vol.18, No.4, 1988, pp. 615-628.

以“人”为本的路径还要强调国际关系中的不确定性。但这种不确定性不仅来源于国际体系的无政府状态,更重要的是源自人类行为的复杂性和不可预测性。尽管国际关系学中不乏对不确定性进行考量,但不确定性是来自外部的,是决策的框架性条件,而不是人的可能性和可变性。基于冷战实践的威慑理论,关注竞争性的大国如何通过对对方意图的判断和误解来调整自己的行为。这种行为的变动来自客观现实的改变以及根据现实而调整的判断,其中的不确定性来自信息的不完美和不完全,尤其是关于对方意志和意愿的信息。如何传递信息成为理论中的关键。然而,决策者对现实的理解及其变化,并不一定完全来自关于现实的外部信息,内在感受到的不确定性及其引发的情绪和道德紧张状态,都可以形成决策者内在的不确定状态。在重大决策中“聆听内心的声音”是常见的说法和做法,而这个“内心的声音”之所以需要努力“聆听”,正是因为决策者具有内在的复杂性和不确定性,这甚至让来自外部的信息成为杂音和干扰。换言之,如果考虑到人从本质上和内心中的不确定性,对于国家关系和全球政治的分析将会揭示出与传统认知非常不同的逻辑。

除了关注深刻而内在的不确定性,未来导向也应是以“人”为本路径的重要特征。传统国际关系理论基本是“后知之明”,对当下判断时局限于权力博弈和功利主义的短期利益。而以“人”为本的路径要求对未来加以更具整体性和长远性的关切,如对全球性和长期性风险加以关注,考量那些关系到人类如何在未来有尊严地生存和有自由地发展的根本性挑战。人与环境以及人与人之间的复杂互动,应该在国际关系研究中占据更重要的地位。国际关系学追求的智慧应超越当下国家间的利益纷争和权力争夺,为人的未来共同承担跨国界、跨世代的全球责任。换言之,一方面,“以人为本”的研究路径要求从更长远的时间尺度和更广阔的空间维度来观察和处理全球问题;另一方面,这一路径要求将最终的关切落在个体的人。

国际关系学对解决全球性问题的切入点和思考方向应更明确地采取“人”的视角。无论是战争、贫困还是环境保护,全球性问题从根本上看是关于人的

命运和福祉。例如,人们在环境问题上高喊“保护地球”,而地球作为一颗孤独的星球并不需要人来保护,“保护地球”其实是保护地球上适合人生存的环境。全球减贫问题也需从教育、健康和社会保障等层面上改善个人的生活条件,而不是过度依赖宏观经济增长指标以及其他众多模糊人类个体的指标。全球问题的解决路径需要从改善个体之间的关系、增强人与环境的互动出发,推动人的可持续性发展。全球问题的解决离不开“人”的层面,只有将个体置于国际关系研究的中心,才能应对这些全球性挑战。

(二) 重新找回“人”:人工智能时代国际关系研究的核心问题

“以人为本”的国际关系研究,不仅是人工智能的发展带给学科自我反思和自我调适的契机,更是学科回应人工智能时代如何有所作为的重要命题。首先,人工智能正在模糊技术工具与政治目的之间的界限。技术是为人类服务的工具,而不应是政治决策和政治行动的目的。人工智能正在或即将被广泛应用于决策支持、军事战略规划、外交谈判和全球治理等关键领域。如果这些技术的应用不以人性之“智”和“仁”加以规范和约束,不以维护人之尊严和人之为人的价值为目的,而是将技术理解为实现抽象的宏大目的或利益关键,用作机械性地最优化效率和最大化决策控制。那么,世界将可能陷入空前的技术统治逻辑,落入以人的利益和自由为代价的技术军备竞赛。因此,以“人”为本的国际关系研究必须对技术的目的进行终极思考,关注国家之间的政治互动和全球治理的制度安排如何避免人工智能成为权力和控制的工具,避免“非人”行为体的“非人”代理人对人性施以强大技术下的否定和扭曲。

第二,人工智能引发的人类存在性风险应是国际关系学关注的另一核心问题。与传统技术不同,人工智能的自主性和不可预测性对人类构成文明级风险。这种风险不可以被限制在某个局部,而是从本质上具有全球性和人类整体性。例如,高度自动化的金融交易系统可能因算法失控引发全球经济危机,军事自动武器的广泛部署则可能导致战争的不可控升级。人工智能在这

些关键领域逐步取代人类进行决策,而“智”与“仁”逐渐退出决策环节,那么,全球政治将进入前所未有的机械性状态——人工智能决策可能会降低日常决策中犯技术型错误的频率,但几乎所有重要决策都是道德性的决策,人类不可担负在重大道德和人性问题上决策的失误。除关键性领域外,人工智能也广泛进入人类生活,接管阅读、消费、社会关系等日常决策,这种自动化将会在不自觉中让人失去或忘记自己的“智”与“仁”。人性是在日常而连续不断的选择中得以磨砺和绽放,当人工智能的引入让人不再主动选择而是消极接受时,人性将遭到湮没,由人组成的社会和践行的政治将是一场场僵死的戏剧。因此,国际关系学在思考如何应对人工智能带来的全球性风险时,不能停留在由其引发的新地缘竞争层面,而是要探索如何建立以人为本的人工智能风险全球治理体系。

第三,人性的不确定性、可能性和可塑性,意味着人所处的外部信息生态系统对人性发展的高度重要性,而人工智能正在制造一个鱼龙混杂的全球信息生态系统。在这个新的信息生态系统中,国际关系和全球政治如何发挥对人性的保护和引导,如何避免权力利用这个系统中的虚假信息来操控和利用人性中的脆弱性,也是“以人为本”的国际关系学应特别关注的问题。

人工智能正在重塑人类的社会体验,影响个体内部的主体感与相互间的情感互动。在政治传播和社会治理中,算法推荐和人工智能生成内容正逐渐成为主流。个性化新闻推送系统可以精准匹配用户兴趣,但同时也可能形成“信息茧房”,削弱公共舆论的多元性。在国际政治领域,外交谈判、政策传播甚至选举竞选活动等也越来越依赖人工智能技术,这可能导致政治沟通的“去人化”。当人工智能撰写的政治演讲、外交辞令和新闻报道成为常态,政治领导人与公众之间的情感联系被技术屏障所阻断,这会削弱公民对政治的真实体验。更极端的情况是,人工智能生成的深度伪造(deepfake),将进一步削弱国际关系和全球治理中原本就相当脆弱的信任机制。如何避免技术的使用削弱人类的主体感,如何确保政治沟通的真实性与透明度,是国际关系和全球治

理必须面对的重大问题。

第四,在人工智能时代,国际关系学的使命是探索如何超越已有规则和创造更美好未来的可能性。这要求关注如何通过构建更加综合和开放的知识体系,释放国际关系学在知识和实践上的可能性,实现对现有国际规则和框架的超越,在新技术革命机遇和挑战面前,维护、促进和创造全球合作与和平。随着人工智能等新兴技术的快速发展,国际关系的传统规则和权力结构正在受到挑战,全球治理的复杂性和不确定性进一步加剧。威慑、制衡和对冲等建立在“非人”假定基础上的战略和策略,试图控制风险的同时势必会制造出更多的严重风险。国际关系和全球政治在漫长的老路上行进到今天,历经了数不清的恐怖和灾难。在人工智能这样史无前例的新兴科技面前,如果再不释放国际关系和全球政治中的新的可能性,这条老路将很可能导致毁灭性的灾难。超越以利益分割为前提、以最优化为准则的框架,从人性出发而非机械的“理性”出发,关注个体性、特殊性和复杂性,追求人的价值和自由,促进群体的公平、正义和仁爱,而非追求普遍性的规律。唯有如此,才能实现以人性为纽带的全球共同体,在人工智能新兴技术的冲击下,保护和发扬来自复杂性、可能性和特殊性的人性光辉,用“仁”与“智”超越国际关系中的“力”与“诈”。

总之,在人工智能时代,国际关系和全球政治的核心任务是重新找回“人”的主体性。唯有如此,人工智能才能真正成为促进全球福祉的力量,而不是全球政治变迁中的主导者。在这一过程中,国际社会的合作与共识尤为重要,各国需要超越国家利益的短期考量,共同制定符合人类共同价值观的技术治理框架。探索什么是通往这些远景的路径,是“以人为本”的国际关系学所应肩负的任务。

结 语

人工智能作为人类能力和认知的增强者已经在多个领域得到展现。随着

人工智能成为人类能力的“外脑”，人类及作为人类的国际关系学者是否会过度依赖技术而逐渐丧失独立思考的能力？当人工智能的分析和建议成为政策决策的主要依据时，是否意味着人类决策者只是执行人工智能计算结果的行动“代理人”？是否意味着人类国际关系学者不再能够有自己的见地？除了增强人类能力，人工智能也正在塑造人类的行为、认知和互动方式。这种塑造是否意味着人类个体自主性的削弱？如果人类的思想、决策和行动模式越来越多地受到人工智能的影响，那么自由意志是否仍然存在？此外，人工智能还在逐步替代人类的一些传统角色，包括在政治和安全领域中的重要角色，如军事指挥官、外交决策者、全球治理技术官僚等。如果人工智能在这些领域的能力持续提升，我们是否能够防止它成为人类实际的统治者和治理者？

面对这些紧迫而深刻的问题，人们需要意识到，人工智能时代的国际政治是全球政治，更是“人”的政治。人工智能不仅提升了人类的能力，也正在塑造人类的思维方式，替代人类的部分生物功能和社会功能，甚至可能全面进入到未来的政治系统和治理体系。面对人工智能的快速发展和迅速扩散，人类正处在一个关键的历史节点，直面未来的全球秩序将何去何从的问题。未来秩序将由人类主导？人工智能主导？还是两者的某种混合？这不是关于人类与人工智能之间的权力争夺，而是关于“人之为人”如何经受住新的冲击。人工智能可以是人类的增强者、塑造者、替代者和主宰者，面对种种可能性，人的不确定性和可能性面临着空前的挑战和机遇。挑战和机遇具有全球性、整体性和未来性，但归根结底又是个体性的。这意味着人工智能时代的国际关系学，其核心是关于人的问题。