

哲学家会被取代吗?

——从胡塞尔现象学角度看ChatGPT等大型语言模型的发展

Will Philosophers Be Replaced: A Discussion of the Development of ChatGPT and Other Large Language Models from the Perspective of Husserlian Phenomenology

吴程程 / WU Chengcheng

(山东大学人文社科青岛研究院, 山东青岛, 266237)

(Qingdao Research Institute of Humanities and Social Sciences, Shandong University, Qingdao, Shandong, 266237)

摘要: 以ChatGPT为代表大语言模型最近获得了重大进展, 它由于在生成文本方面的出色表现而获得了人们的关注, 也引起了哲学研究者们的担忧。ChatGPT的基本架构决定了它的工作原理是以统计为基础的, 这使它遭到了很多批评, 其中主要的批评是: 只随机处理数据而没有涉及世界模型、缺乏创造性以及大数据主义的路线错误。但从胡塞尔现象学的角度来看, ChatGPT表现出的缺陷另有原因。ChatGPT面临的挑战是如何建立一个可生成、能流动的世界模型和时间结构, 它所运用的大数据主义路线, 从胡塞尔的视域理论角度来说也可以获得支持。

关键词: ChatGPT 世界模型 生活世界 视域 大数据路线

Abstract: Large language model, represented by ChatGPT, has recently gained significant momentum, not only gaining attention for its outstanding performance in generating text, but also raising concerns among philosophical researchers. The basic architecture of ChatGPT dictates that it works on a statistical basis, which has led to a number of criticisms, the main of which are that it only randomly processes data without involving a world model, that it lacks creativity, and that it has a wrong way of Big Dataism. However, from a Husserlian phenomenological point of view, ChatGPT exhibits flaws for other reasons. The challenges for ChatGPT are the need to solve the problem of building a generative and fluid world model and building the temporal structure. Moreover, way of Big Dataism it employs can also be supported from a Husserlian perspective of the theory of horizon.

Key Words: ChatGPT; World model; Lifeworld; Horizon; Big dataism

中图分类号: TP18; O225 DOI: 10.15994/j.1000-0763.2025.03.007 CSTR: 32281.14.jdn.2025.03.007

大型语言模型 (large language model, LLM) 作为人工智能的成果之一, 在多年发展的基础上获得了质的飞跃, 由微软公司的OpenAI开发的ChatGPT引起了人们的高度关注。“它们在数据的质量和规模上都得到了大幅度的提高, 并且可以根据人类的反馈校准这些模型。它是新一代的搜索引擎”。^[1]它在多个领域都显示出了强大的功能, 特别是在问答

的质量上, 相比于其他的聊天机器人, 它似乎显示出了一种新的智能。最近围绕ChatGPT的一个实验引起了笔者的注意: 埃里克·施维茨格贝尔 (Eric Schwitzgebel) 等人以丹尼尔·丹尼特 (Daniel Dennett) 的著作作为数据样本, 对ChatGPT3进行了微调, 训练ChatGPT3模拟丹尼特的风格来回答哲学问题, 并且将丹尼特本人的回答混在机器给出的回答中, 测试人

收稿日期: 2023年8月4日; 返修日期: 2024年12月19日

作者简介: 吴程程 (1989-) 女, 山东曲阜人, 山东大学人文社科青岛研究院助理研究员, 研究方向为现象学认识论。Email: wuchengcheng1989@outlook.com

们是否能够区分真正哲学家的回答与机器人的回答,来检测ChatGPT是否具备回答、甚至理解哲学问题的可能。^[2]实验结果证明,在机器人的回答和真正哲学家的回答之间,并不存在可靠的区分。也就是说,机器人的回答(至少在这十个问题上)可以高度模拟,或者无限接近真正哲学家的回答,即使是研究丹尼特的专家,也并不能完全正确地区分两者。这个实验确实在某种程度上证实了笔者的担忧:如果让ChatGPT获得现有哲学研究的数据,它会不会在哲学研究上超越大多数的研究者?这个实验显然在小范围上证实了ChatGPT在简短回答哲学问题方面是非常出色的。

当然简短的回答哲学问题与真正的哲学研究还相差甚远,但是这个实验确实提醒哲学研究者们要认真对待ChatGPT这一大语言模型。它在生成文本方面表现出色,但其缺陷也很明显,因而并没有取得大多数哲学研究者的认可,^[3]这与人工智能研究者的乐观态度形成了鲜明对比。但是这些流行的关于ChatGPT的批判是否真的切中要害,以至于可以让哲学研究者们高枕无忧?本文试图在澄清ChatGPT基本架构的基础上,简要讨论当前流行的反对ChatGPT的意见;并且从胡塞尔的现象学观点出发,指出这些反驳意见的不可靠性,从建设性的角度指明ChatGPT真正需要突破的方向;最后文章将表明,能够超越ChatGPT生成的普通文本而创造出真正具有创造性的研究才是哲学研究者们要面临的真正挑战。

一、ChatGPT的基本架构

1. 大型语言模型在人工智能发展中的位置

ChatGPT基本上可以视为深度学习技术的成果,在深入解释这一技术之前,回溯一下人工智能发展史,我们可以对大型语言模型所在的位置有更宏观的了解。

人工智能产生的标志性事件是1956年的达特茅斯会议,随后人工智能经历了早期的无限期望期(50-60s)——受挫低潮期(70-80s)——回归期(90s)——快速发展

期(至今)这几个阶段。从人工智能发展伊始,它就分为了两个主要的学派,即联结主义(Connectionism)与符号主义(Symbolism)。联结主义可以简单总结为试图通过数学模型来模拟人脑的神经网络。神经网络是通过相互连接的节点来发挥作用,这些节点分为输入节点、输出节点以及中间的隐藏节点。这一学派的开端可以追溯到沃伦·麦卡洛克(Warren McCulloch)对大脑神经活动的建模研究,^[4]同时1957年弗兰克·罗森布拉特(Frank Rosenblatt)则发明的叫做“感知机”(Perceptron)的神经网络模型,这些都是早期联结主义的重要成就。但“感知机”的研究受到符号主义的领军人物马文·明斯基(Marvin Lee Minsky)的猛烈批评,因此联结主义在七十年代转入低潮期。^[5]

作为人工智能研究的另一大学派符号主义,其指导思想的前身则可追溯至19世纪逻辑的发展,特别是数理逻辑的发展,因此符号主义又称逻辑主义。这一学派的主要路径是依据逻辑原理,通过符号的运算来表达思维过程。艾伦·纽厄尔(Allen Newell)、赫伯特·西蒙(Herbert Simon)在1976年提出的著名的物理符号系统,认为“任何显示出智能的系统(人类或机器)必须通过操作由符号组成的数据结构来运行”,^[6]但是八十年代中期沿此路线开发的专家系统上遭遇了很大的失败,人工智能的开发进入一个寒冬。人工智能的发展遇到瓶颈,研究者们开始反思人工智能的路线,这些反思促使了联结主义的复兴。“专家系统的脆弱性导致了一种新的、更科学的方法,结合了概率而不是布尔逻辑,基于机器学习而不是手工编码,重视实验结果而不是哲学主张”。^[6]联结主义者们的几项技术突破——重新发明的反向传播算法(back-propagation)、贝叶斯网络(Bayesian network)以及卷积神经网络(Convolutional Neural Network)等——叠加在一起,使得以神经网络为架构的机器学习取得了长足的进步,这些都为深度学习技术的产生奠定了基础,在两千年之后,深度学习真正的流行起来。

在人工智能发展的起起落落中，对自然语言的处理（Natural Language Processing, NLP）一直都在人工智能发展中占有重要地位。著名的图灵机假设就是通过语言来检测智能，这也被视为是自然语言处理的开端，机器翻译、语音识别、信息检索、问答等等，都是其中要处理的任务。与人工智能发展的两大流派一样，对自然语言的处理方法也可分为两种：一种是依据规则（如句法、语法规则等），一种是依据统计。自然语言处理经历了从规则方法（50-60s）向统计方法（70-80s）的转变，这一转变同上文提到的专家系统的失败紧密相关。但随着联结主义模型的重新兴起，人们也将这些模型应用于自然语言的处理上，罗素指出，“事实上，我们在语言中命名的许多概念，经过仔细检查后，都未能获得早期人工智能研究人员希望以公里形式描述逻辑定义的充分条件。联结主义模型可能以一种更流畅和不精确的方式形成内部概念，更适配真实世界的混乱。它们还具备从样本中学习的能力，它们可以将它们的预测输出值与问题的真实值进行比较，并修改参数以减少差异，使它们在未来的样本中更有可能表现良好。”^[6]这一看法也得到了验证，将深度学习应用于自然语言处理中，一种功能强大的大型语言模型 ChatGPT 诞生了。

2. ChatGPT 的架构

通过 OpenAI 的创始人之一安德烈·卡普西（Andrej Karpathy）对 ChatGPT 的介绍，我们大致可以了解这一语言模型的基本架构。^[7] ChatGPT 的训练简单来说包括了四个阶段，第一步是预训练（pretraining），第二步是监督改进（supervised finetuning），第三步是奖励模型（reward modeling），第四步是强化学习（reinforcement learning），后面两个步骤是联系在一起的。预训练阶段集中了大量的数据和绝大部分的计算，后面三个阶段都是在第一个阶段的基础上做出调整，这些调整虽然涉及的数据量并不是很大，但是却发挥着非常重要的作用。

首先是预训练阶段，简单来说就是收集大量的数据，形成 GPT 的神经网络的训练集。其

中在处理这些数据之前，要对数据进行标记（tokenization）。通俗的理解，就是对每个单词都贴上标签，每个标签都是可以随时调取的。这些数据被放入一个叫 Transformer 的神经网络中，然后获得更多的训练。在这个阶段，它认识了文本的各个部分，比如说单词，标点符号的位置等等。然后它能够形成一个看起来没有什么错误的句子。第二个阶段就是监督改进阶段，就是在训练中给模型设定一个目标函数值，对它进行特定的训练，比如你用莎士比亚的作品来训练它，它就能模仿莎士比亚的风格输出文本。这一阶段的特征是根据少量、但是高质量的数据来建立一个模型。第三和第四个阶段就是在第二个阶段建立的模型基础上，对它进行奖励建模，在这里奖励建模和强化学习是联系在一起的。奖励模型的基本原理是，通过不断的反馈调整，它能够选择一些成功率更高行为，来完成相关的任务。在奖励模型中，他们引入了一对概念：状态（state）与行为（action）。通过规定状态，模型就知道采取何种行动可以完成任务。在完成了奖励模型后，最后再对它们的完成度进行打分，进而不断优化这个模型，最终形成一个最好的模型。

二、ChatGPT 面临的批评

如果我们用简化的语言来概括训练 ChatGPT 的这个过程，那么其原理就是通过收集海量的文本数据，然后给每个数据一个标记，把它放到一个神经网络里，它能够根据这些数据发现规律，知道如何回答问题，然后通过奖励它，让它知道哪些回答是更高质量的回答，最后它在问答中就能显得比较聪明。如果对比人类掌握语言的过程，那么很难说它“理解”语言，因为 ChatGPT 的整个架构中，都没有任何一处涉及了语义、或者说语言的意义部分，它的全部可以说都是数学公式。因此在了解了 ChatGPT 的基本架构后，我们对 ChatGPT 面临的批评会有更深刻的理解。基本上，以 ChatGPT 为代表的大型语言模型所面临的批评主要有两类：第一类是从根本上否定 LLM 处理

路线的反对意见;第二类是接受LLM的路线,但是认为ChatGPT在很多方面需要改进。

1. “随机的鹦鹉”

艾米丽·班德(Emily Bender)将以ChatGPT为代表的大型语言模型称为“随机的鹦鹉”,^[8]这个模型并没有任何迹象显示了它理解了语言的“意义”(meaning),并且这些“语言模型生成的文本并没有基于任何交流意图、世界模型或者任何人类心灵状态的模型。”^[8]查默斯将其批评总结为“只是在模仿文本,而不是思考世界”。^[9]这一批评基本上代表了主流批评的中心思想,即使ChatGPT能够回答“你如何思考世界”这一问题,它也不是在真正的思考世界。班德认为这一大型语言模型具有非常多的潜在危险性,除了需要耗费大量的资金,它还会在加强偏见、改变社会观点等方面推波助澜。根据它的工作方式,这种语言模型即使规模非常大,也不能保证产生多样性,它可能只会加强已经偏狭的观念。^[8]考虑到ChatGPT也属于NLP中的统计派,统计派的特征就是通过计算来合成文本,那么从规则派的角度来说,它对语义的处理——后面关联着意义——肯定有着巨大的缺陷,因此“随机的鹦鹉”这一标签指出了ChatGPT在处理语言方面最不能被人们接受的一个方面。

如果沿着班德的批判进行回溯,我们还可以看到对这些大型语言模型基本路线的更彻底的否定。休伯特·德雷福斯(Hubert Dreyfus)在其早期著名的《计算机不能做什么》一书中就批评过计算机对语言的处理方式,他认为“让机器给出正确答案的困难不在于排除词汇的歧义,也不在于缺乏异常有力的逻辑工具上。主要的问题是还没有人构造出适用于理解故事的、某种知识体的元素。”^[10]著名语言学家乔姆斯基也对ChatGPT持激烈的否定态度,他指出人脑对语言运用是一种创造性运用,而不是海量数据的集合。国内学者徐英瑾教授对此也提出过类似的批评,即以ChatGPT为代表性成果的深度学习技术,在路线上(大数据主义路线)就有天然的缺陷,^[11]如海量数据的获得并不是人类的常态,这些海量数据的收集同时

也会带来非常多的伦理问题等等。^[12]以上这些批评从根源上来说,都是对ChatGPT在路线问题上的根本否定,这种以统计为特征、利用大数据来处理语言的方式很难得到批评者们的认同,因此在他们看来,ChatGPT可以说并没真正理解过任何事情,更遑论思考世界。

2. “只想模仿,不想成功”

另外一个值得重视的批评是由ChatGPT的开发者提出的。卡普西曾提到的它有一种不想成功的怪癖,它要做到什么,必须是在接收到明确要求时才会去行动,否则它就只想模仿。这一表现从技术角度来说是由于ChatGPT的工作原理是为单词设置很多的标记(tokens),当它被要求作答时,它只关心自己的回答中的下一个标记是什么,而并不关心回答的质量,如果数据库中有对同一个问题的两个答案,它会随机模仿其中一个回答,并不会主动选取质量更高的回答,除非要求它给出一个更好的回答,它才会重新检查自己的标记,给出另外的答案。这一工作原理也让它表现出了一些让人难以接受的缺陷,比如非常低级的错误答案。正如斯蒂芬·沃尔夫拉姆(Stephen Wolfram)提到的,它在一些没有“标准答案”的问题上可以给出“像人一样的”回答,但却在某些具有明确答案的问题上经常失灵,但是这一问题从技术角度来说是可解决的。^[13]但是它的这一特征也产生了一个巨大的问题:缺乏创造性。如果从哲学的角度来理解这一表现,也可以解释为ChatGPT没有表现出主动动机,它只会被动接受指令,并且无法从这种被动中产生出主动动机,这一点可能也是大型语言模型乃至整个人工智能的研究要获得突破的地方。

当然,ChatGPT在伦理、法律等方面也面临很多需要解决的问题,但文章接下来将指出,ChatGPT之所以表现出这些缺陷,并不是上述批评者们给出的那些理由,如果从胡塞尔现象学角度来看的话,这些缺陷另有原因。

三、胡塞尔现象学的启示

人工智能大约只有在产生之初,同当时的

分析哲学家们有过建设性的对话之外，在其后的发展中，哲学基本上贡献了最尖锐的批评，德雷福斯就是其中的代表。也正由他起，现象学与人工智能问题就深刻地联系在了一起，并且随着人工智能的发展，不断地变换、产生新的主题。但本文试图站在对话的立场，从胡塞尔现象学理论的角度出发，认真思考 ChatGPT 的缺陷，这些反思能够让我们更谨慎地给出意见，为 ChatGPT 的未来发展提供一些潜在思考路径与指南。

1. ChatGPT 摆脱“随机的鹦鹉”这一标签所面临的真正挑战

面对大型语言模型只处理文本的看法，查默斯提出了不同的意见。他指出在训练 ChatGPT 的环节中，很可能就涉及到世界模型，至少是一种有限的世界模型。^[9]世界模型简单来说，就是整体的世界图景、生活背景以及各种常识等等组合成关于完整的世界背景。在人工智能发展史上，通过建构和完善世界模型来推进人工智能的想法并不是新近的事，在专家系统如火如荼的八十年代，“包括麻省理工学院的尤金·查尔尼克（Eugene Charniak）及耶鲁大学的罗杰·尚克（Roger Schank）在内的几位研究人员一致认为，强大的语言理解需要关于世界的一般知识以及使用这些知识的一般方法”，^[6]在那时科学家们已经认识到了世界模型对于发展真正智能的重要意义。最新基于世界模型的应用已经取得了一些突破，在一些游戏中，人工智能已经可以在有限的计算资源下生成关于相关环境的动态模型，从而生成合理的对战策略。^[15]正如查默斯宣称的，建构一种具有深度的、稳健的世界模型在原则上是可能的，并且已经在实践中。^[9]因此断言 ChatGPT 只处理文本，不理解世界只具有有限程度的可信性，至少从技术角度来说，构建一个有限的世界模型已经成为可能。

但就此认定世界模型的构建是一片坦途也为时过早。在这里，我们如果用胡塞尔的生活世界理论来理解世界模型，那么我们会看到大型语言模型面临的挑战，还不仅仅是能不能建构一个完整的世界模型，而是这个模型是否具

有生成性、边界是否具有开放性等。作为最具启发性的理论之一，胡塞尔的生活世界理论从它最初的解释科学危机的作用中延展开来，在理解人以及理解人类生活的各个方面都发挥了更大的作用。在胡塞尔对生活世界的揭示中，其中有两个特点值得注意，这些特点可以启发世界模型的建构。

首先是生活世界的预先给予性。“生活世界没有‘被给予’意识，而是‘预先-被给予’意识。对胡塞尔来说，一个被知觉的对象‘被给予’意识——它在一个关于意识活动的意向性关系的重点显现出来。然而，生活世界是被预先-给予的——在你认识它之前，它就已经先于我们的意向活动在那里存在了。它作为事物在其中给予的世界始终并已经在那里为你而存在了。”^[15]在这一点上，以 ChatGPT 为代表的人工智能显然遇到一个难题：ChatGPT 的世界模型只能是对人类的生活世界的模仿或者复制，因为它没有自己可以预先给予的世界模型（它所在的人工智能领域的各种成就并不能成为它的“世界模型”或“生活世界”），它只有再次根据对人类生活世界的理解来建构起的世界模型。在这一点上，ChatGPT 的世界模型就永远落后“生活世界”一步。

其次是胡塞尔的生活世界理论还揭示了非常重要的一点，那就是生活世界里的具体内容可以不断地发生变化，但是作为“整体”的生活世界依然保持着它自己。“更确切的说，是一作为习惯的获得物从很早以来我们就具有的，并且作为我们可以随意再次实现的如此这般内容的有效性，在我们内心中所包含的那种有效性的形式而具有存在意义的。当然，所有这些都经历了多种多样的变化，而‘这个’世界作为统一存在着的、只不过在内容方面进行了修改的世界，仍然保持着自身。”^[16]生活世界的这一特点一方面表现为边界的开放性，虽然围绕每个主体的整体生活世界重要背景已经固定，但是许多新事物都可以在产生之后沉入生活世界中，成为我们整体的背景，正如 ChatGPT 的出现，在我们研究它的同时，它也会慢慢融入生活世界，成为我们整体生活背

景的一部分,但是作为“整体”的生活世界,还是保持着它自己。对比来看,这一点对于ChatGPT来说也是重大挑战,它们的世界模型一旦形成就已固定,新的背景和内容只能不断的添加进去,因此它的世界模型不具备生成性。因此,如果它要模仿或复制一个“整体”世界,那么它的整体内容就无法发生新的变化,如果它要拆分部分内容构建有限的“世界模型”,那么它就很难形成一个“整体”的生活世界的模型,这应该是ChatGPT在建构世界模型时面临的两难选择。

2. 被动发生理论对ChatGPT主动动机产生的启示

在前面我们提到ChatGPT的一个缺陷,就是它只想模仿已有的数据集,只有被要求时,它才会选择可以完成目标的路径,它对自己的缺陷一无所知,因此它们总是需要被告知,才会寻求最佳解决方案,因此ChatGPT确实缺少一种主动动机。在某种程度上,人工智能的发展对于主动建构与被动接受之间的动机转换兴趣甚少,因为总体来说,科学家们在努力使得人工智能的成果显得很聪明的方法,基本上是用人类自己的智慧,去要求人工智能完成某些工作,人工智能产品本身却不能从海量的接受性材料中主动生成智能活动,被冠以主动学习(active learning)名称的技术也并非真正的主动性,它们依然需要人类专家的协助。当然“主动”完成一个智能活动一方面可以从意向性角度进行思考(在这一方面人工智能和哲学产生了最紧密的联结),但另外一方面,胡塞尔的发生现象学中对一个意识活动的主动性所包含的复杂结构的分析,特别是对“被动发生”的解释,恰好可以为ChatGPT摆脱完全被动的境况提供启发。但胡塞尔的发生现象学是一个庞大的理论结构,其中包括经验的发生结构、自我的发生结构以及内时间意识的发生结构等等,在这里我们只能概括式的提及。

胡塞尔指出,一个主动的意识行为产生之前,它已经包含被预先给予的对象,这些对象本身就包含着一种“诱导”,即从意识中“凸显”出来,随时在同一性的构造中形成一个对

象,在这里这些“材料”为什么不仅仅是单纯的材料,而是一种具有“诱导”的材料呢?胡塞尔指出,这些材料在给予之初就已经展现出最低层次的统一性,因为最初的材料是在一个感觉场域中给予的,比如一块红色的颜色,它是作为某物的颜色被给予的,而不是一个单纯的颜色被给予的,但是这种综合最初的发源地是时间意识,“时间意识是建构一般统一体的策源地”。^[17]因此“尽管这种把握具有素朴性,但它绝不是一种简单的素材;它表明自身具有多种多样的结构,而在这些结构中,它本身是作为内在时间统一性而建构起来的。”^[17]正是由于时间结构,这些素朴的材料能够形成一种将过去“仍然保持在手”、对当下的把握以及对未来的前瞻结合在一起的统一性,这一时间意识将这些被动的材料变成了一种主动把握中的被动性。

如果对比ChatGPT的运行模式,这种对过去材料的“仍然保持在手”是ChatGPT最擅长的,因为只要对相应的材料进行了标记,它能够随时调取这些材料。但相比于胡塞尔揭示的主动把握中的被动性的“保持在手”,它的“保持在手”只是完全被动给予的材料,所有这种复杂结构的综合,都被标记所取代,但是这种标记自身不会对模型本身发出刺激,因为ChatGPT并没形成主动性的时间结构。它拥有完美的记忆力,但它缺少的是主动调取这些记忆的动机,因此它只能表现出纯粹的接受性。故而从现象学的角度来看,ChatGPT表现出的绝对被动性,一个可能的原因就是缺乏真正的流动的时间结构。如果计算机科学家们能够为大型语言模型搭建一个真正流动的时间结构,那么ChatGPT可能会表现出非常大的创造性。

3. 视域理论下的大数据路线的合理性

以ChatGPT为代表的大语言模型所面对的最激烈的批评,就是对其基本路线的批评,这种以海量文本为基础、以统计为特征的大数据路线,被认为是一种低效的浪费、同人类运用语言方式相去甚远的一种方法。但如果我们结合胡塞尔的视域理论,这一路线却显示出很大

程度的合理性。

胡塞尔在论证谓词判断在前谓词经验中的起源时，提出了两种不同的视域，其中一种是同生活世界相关的视域，亦即作为所有经验与判断基础的世界信念，在任何目光转向它之前，已经预先被给予我们了。“周围环境作为预先被给定性领域，作为一种被动的预先被给定性领域而在此同在，它是指这样一个领域，即没有在任何添加、没有借助任何把握的眼光、没有唤起任何兴趣的情况下，它就一直在此了。”^[17]另一个视域则是判断活动的前谓词领域，这一领域是指在在一个谓词判断发生之前，与此判断相关的共同给予也预先被给予了我们。“任何在本来意义上总是有所经验的经验，即当某物自身被观看（Gesicht）时的经验，都不言而喻、必然地具有正是对此物的某种知识和共识（Mitwissen），也就是对于此物的这样一种在经验尚未观看到它时即为它所固有的特性和某种知识和共识。”^[17]

虽然人脑工作的方式看起来是支持小数据主义的，但胡塞尔的任何一个视域理论都表明：心智在其活动以前，已经存储了大量数据，它是一个隐性的大数据库。前谓词领域的视域概念从深度层面指出，一个最简单的谓词判断包含了海量的数据，而生活领域的视域概念则从广度方面说明了更复杂的思维活动（情绪，道德判断等等）也包含了无数潜在信息。看似“小”的数据其实并不“小”，它们只是作为预先被给予、潜在的数据存在于我们的知识网络中。因此ChatGPT的大数据主义路线也显示出了其合理性。

总之，从胡塞尔现象学角度来看，我们能够对ChatGPT表现出的缺陷提供一种建设性的意见。ChatGPT表现出对文本的随机处理、常识的低级错误的缺陷，并不是因为它真的不涉及世界模型，而是需要解决如何建立一个可生成、能流动的世界模型；它没能表现出创造性的原因，可能是因为缺乏一个形成主动性中的被动发生的时间结构；如果细致分析围绕我们意识活动的视域中的材料，那它也包含一个海量的庞大数据结构，因此对其基本路径的全盘

否定也是有待商榷的。

结 语

以上论证表明，即使大型语言模型的发展面临着多重挑战，这并不意味着它们的路线错误，胡塞尔的视域理论为大数据路线提供了某些支持证据。但是由于ChatGPT在世界模型以及主动动机这两个方面的缺陷，距离它真正代替哲学研究者的工作还有很长的距离。最后应当指出的是，当ChatGPT以一种完全不同于人类思考语言的策略，实现了人类在语言方面某些相同的功能时，人们确实给出了许多相互矛盾的反应：其中一个就是当我们发现它在生成文本方面表现出色时，我们做出了对其做出了相应的限制，比如说利用ChatGPT的生成论文的行为将被判为抄袭（许多大学和期刊杂志都最先给出了反应），因为确实很有可能人类在文本上无法识别出机器的回答，埃里克·施维茨格贝尔的实验就说明这一点，即使在哲学领域也不例外；但另一方面，我们又认为ChatGPT的模式无法真正完成创造性的内容，它并没有我们想象中的那么智能。这一矛盾反应至少说明，ChatGPT可以完成一些普通的文字工作，但它缺乏创造性，那么它的这一特征可以产生一个附加功能：为人类提供一种从平庸的文字工作中解放出来的可能性，一种从无价值的文字劳动中解放出来的自由。也正是在这意义上，它对哲学工作者们提出了真正的挑战：督促研究者们生产真正具有创造性的研究。如果我们认为ChatGPT在很长时间内不可能真正从事哲学研究，那么我们的表达至少应该超越ChatGPT目前在哲学上能表达出的东西。

[参 考 文 献]

- [1] Van Dis, E. A. M., Bollen, J., Zuidema, W., et al. 'ChatGPT: Five Priorities for Research'[J], *Nature*, 2023, 614(7947): 224-226.
- [2] Schwitzgebel, E., Schwitzgebel, D., Strasser, A. 'Creating a Large Language Model of a Philosopher'[J]. *Mind & Language*, 2024, 39(2): 237-259.
- [3] Bouzid, A. '24 Philosophy Professors React to ChatGPT's

- Arrival'[J]. *Social Epistemology Review and Reply*, 2023 12(3): 50-65.
- [4] McCulloch, W. S., Pitts, W. 'A Logical Calculus of the Ideas Immanent in Nervous Activity'[J]. *Bulletin of Mathematical Biophysics*, 1943, 5: 115-133.
- [5] Minsky, M., Seymour, A. P. *Perceptrons: An Introduction to Computational Geometry*[M]. Cambridge: The MIT Press, 2017.
- [6] 斯图尔特·罗素、彼得·诺维格. 人工智能: 现代方法[M]. 张博雅、陈坤、田超等译, 北京: 人民邮电出版社, 2022, 18-22.
- [7] Karpathy, A. 'State of GPT'[EB/OL]. <https://karpathy.ai/stateofgpt.pdf>. 2012-07-18.
- [8] Bender, E. 'On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?'[EB/OL]. <https://doi.org/10.1145/3442188.3445922>. 2023-07-18.
- [9] Chalmers, D. J. 'Could a Large Language Model Be Conscious?'[OL]. arXiv:2303.07103. <https://doi.org/10.48550/arXiv.2303.07103>. 2023-07-18.
- [10] 休伯特·德雷福斯. 计算机不能做什么[M]. 宁春岩译, 北京: 生活·读书·新知三联书店, 2002, 15.
- [11] 徐英瑾. 人工智能十五讲[M]. 北京: 北京大学出版社, 2021, 139-156.
- [12] 徐英瑾. ChatGPT时代我们的“脸面”何在——多维视野下的颜面现象学[J]. 探索与争鸣, 2023, (2): 55-64; 178; 2.
- [13] Wolfram, S. 'What is ChatGPT Doing ... and Why Does It Work?'[OL]. <https://writings.stephenwolfram.com/2023/02/what-is-chatgpt-doing-and-why-does-it-work/>. 2023-07-18.
- [14] Matsuo, Y., LeCun, Y., Sahani, M., et al. 'Deep Learning, Reinforcement Learning, and World Models'[J]. *Neural Networks*, 2022, 152: 267-275.
- [15] 肖恩·加拉格尔. 现象学导论[M]. 张浩军译, 北京: 中国人民大学出版社, 2021, 137-138.
- [16] 胡塞尔. 欧洲科学的危机与超越论的现象学[M]. 王炳文译, 北京: 商务印书馆, 2017, 133-134.
- [17] 胡塞尔. 经验与判断[M]. 邓晓芒、张廷国译, 北京: 生活·读书·新知三联书店, 1999, 47-127.

[责任编辑 王巍 谭笑]

