• 专题：重构机器智能图景 •

编者按：

在人工智能技术迅猛发展的当下，尤其是大语言模型如 GPT-4 的广泛应用引发了关于“机器是否能思维”“大模型是否具有因果推理或者反事实推理能力”等诸多哲学与技术层面的争论。本专题辑录三篇文章，试图从不同角度为这些问题提供更深层的思想资源与反思框架。“对大语言模型之经验论前提的反思”借助三木清对经验论的批评，深刻揭示当前大语言模型构建逻辑中的还原主义偏向。文章指出，大模型所仰赖的“语元 – 词嵌入”机制不过是对休谟式经验论的技术延续，而缺乏对构想力、文化结构和反事实认知能力的真正模拟，从而导致“机器幻觉”问题的哲学根源。“大语言模型与因果之梯”以珀尔的因果三阶梯理论为背景，讨论了结构因果模型与大语言模型之间的张力与互补性。文章认为，大语言模型虽未明确嵌入因果模型，却在表现上“越级”模拟了因果与反事实任务，从而挑战了珀尔的原初批评。然而，这种挑战并不意味着结构因果模型的过时，相反，它提供了拓展模型因果解释力的新机遇。“机器智能何以可能——图灵的智能概念研究”回溯了图灵关于机器智能的真正设想，指出主流对图灵测试的理解存在严重误读。文章强调图灵并未将测试视为智能的定义，而是作为探索机器是否有可能展现出类人智能的一种实验方式，其智能概念本质上是一种响应依赖的情感性概念。这三篇文章所展开的讨论，既涉及人工智能的哲学基础，也映射出当前智能技术发展中的深层张力。在图灵设想与因果建模之间、在经验论路径与构想力理论之间、在统计关联与因果推断之间，我们重新被迫思考一个老问题的现代版本：机器智能何以可能？

（专题策划：吴小安）

对大语言模型之经验论前提的反思
——从三木清哲学的视角看

Reflections on the Empiricist Assumption of Large Language Models:
A Study from the Perspective of Miki Kiyoshi’s Theory of Imagination

徐英瑾 /XU Yingjin

（复旦大学哲学学院，上海，200433）
(School of Philosophy, Fudan University, Shanghai, 200433)

摘　要：在日本哲学家三木清看来，经验论所提出的“经验”概念过“薄”，无法承载时间的延展与

基金项目：国家社会科学基金一般项目“对于通用人工智能与特定文化风土之间关系的哲学研究”（项目编号：22BZX031）。
收稿日期：2025 年 1 月 13 日
作者简介：徐英瑾（1978–）男，上海人，复旦大学哲学学院教授，研究方向为人工智能哲学与英美分析哲学研究。Email: yjxu@fudan.edu.cn

对外部对象的指涉，同时，也无法说明个体与环境之间的互动关系。三木的上述批评只要稍作改动，就可以被移用到对于方兴未艾的“大语言模型”上去。与经验论一样，大语言模型的构建者通过对于文本的“语元化”处理而人为构造出了休谟式“印象”的数码对标物——语元——并仿照休谟主义者通过累积“印象”而构造“观念”的思路，通过统计学手段构造出了语元的基本语义学参数：词向量嵌入。然而，由于这种还原论思维浓重的构造思路几乎绕开了心智运作的各种中层与高层架构，它也就很难复刻人类心智在输入数据不足时体现出的创造力与反事实条件下的推理力。这也便是所谓“机器幻觉”问题产生的哲学根源。

关键词：构想力　经验论　大语言模型　语元　词向量嵌入

Abstract: From the Japanese philosopher Miki Kiyoshi’s point of view, the empiricist notion of “experience” is too thin to accommodate temporal extension, reference to external objects, as well as the interplay between experience-owing agents and their environments. Once properly modified, the foregoing Miki’s criticism can be easily applied to the still-fashionable approach of Large Language Models (LLM) in AI. Paralleled with the empiricist philosophy, the LLM-builders construct the digital counterpart of the Humean notion of “impressions”, namely, “tokens” by treating the text as tokens. Moreover, similar to the Humean route-map of reconstructing “ideas” from the accumulations of “impressions”, LLM-builders also intend to reconstruct the semantic features of tokens via a proper statistical treatment of them, namely, a treatment routinely under the label of “word embedding”. Nonetheless, since such reductionism-oriented route-map has deliberately bypassed nearly all middle/high-level architecture required for a full-fledged notion of “cognition”, such rout-map could hardly account for why human’s cognitive machine can deliver creative decisions and be competent in counter-factual reasoning even when the size of the training data is much smaller than that is required by an empiricist theory. And the LLM-builders’ incompetence of accounting for all of this in turn philosophically explains the origin of the so-called “machine hallucination”.

Key Words: Imagination; Empiricism; Large language models; Tokens; Word-Embedding

中图分类号：B085; N031　DOI: 10.15994/j.1000–0763.2025.08.001　CSTR: 32281.14.jdn.2025.08.001

导　论

众所周知，经验论（empiricism）与唯理论（rationalism）的斗争，乃是近代哲学的一大主题。根据唯理论，人类知识的主要来源乃是理性推理，而根据经验论，人类知识的主要来源乃是经验归纳。1956年人工智能（以下简称AI）学科的确立通过美国达特茅斯会议而得到学界普遍承认后，通过“基于规则的AI”这一技术进路，唯理论曾长期成为AI界的主导哲学范式。而在上世纪80年代“基于统计的AI”这一进路经由“神经元网络技术”而开始被学界重视之后，经验论的思想预设则在AI学界开始收获更多的拥趸，与之相较，唯理派的思想预设则开始式微。而到了本世纪，特别是在神经元网络技术升级为深度学习技术后，这一“经（验）升理（性）降”的趋势变得更加明显，以至于有人认为作为深度学习技术最新发展成果的“大语言模型”（Large Language Model, LLM）技术就是“经验论复兴的标记”。[1]但需要注意的是，在人工智能哲学研究的领域内，对大语言模型技术的经验论前提的哲学反思似乎都未参考哲学史中既有的经验论批判资源，甚至还在基本概念层次上存在着一些严重误解。譬如，罗伯特·朗（Robert Long）最近就在英语世界重要哲学杂志《哲学探究》（*Philosophical Studies*）上发表论文，为AI研究中的经验论复兴趋势鼓与呼。[2]然而，朗在具体立论中不以认识论脉络中的“唯理论”为敌，而是以心理学脉络中的“先天论”（nativism）为敌，也几乎不提德国古典哲学、现象学、实用主义等资源对经验论的批评，显示出了一种经验论者的理论傲慢。

为了收敛本文论述的焦点，本文对于LLM所蕴含的经验论预设的批判，将主要依赖日本哲学家三木清（1897–1945，名字的罗马音写法为Miki Kiyoshi）在《构想力的逻辑》（構想力の論理）一书[3]中的相关资源。三木清熟悉欧陆哲学传统与美国实用主义，也熟悉英国经验论传统。故此，带上三木哲学的眼镜去看待经验论，能够使得我们以较小的哲学史阅读成本看到问题的肯綮。换言之，本文对于三木哲学的解读，其首要目的并非普及三木哲学，而是借由其哲学观点所提供的“思想X光”去透射经验论的理论之躯，并由此延伸到对于LLM技术的哲学基础的批判上去。

一、三木哲学经纬中的英国经验论之弊

三木清讨论经验问题的大背景，是他对于“构想力”问题的讨论。这一讨论与他对于康德哲学的解读有关。我们知道，作为近代经验论的一个重要的批判者，康德试图综合经验主义所强调的直观与唯理派所强调的概念思维，由此将其综合入一个新的哲学体系。但考虑到直观与思维之间的异质性，这就需要某种中介来促进二者之间的彼此配合，以使得完整的知识得以形成。这一中介就是康德所说的“Einbildungskraft”以及作为其运作结果的“图型”（Schema）。（[4],p.272）“Einbildungskraft”这词一般被翻译为“想象力”，而三木则将“Einbildungskraft”翻译为“构想力”（構想力），因为这既能凸显德文原文所带有的“图像构成”的意蕴，又能通过汉字“构”的“构筑”之义暗示三木赋予“Einbildungskraft”一词的客观知识向度。三木写道：

> 在构想力中，知性成分是与感性成分结合在一起的。……构想力总是包含知性要素和感性要素，是两者的内在的统一。在构想力自身之中，包含着内在性且生成性的知性的要素，在这一点上，构想力是同感情有别的。因此，构想力的哲学，既非单纯的理性主义，也非单纯的非理性主义。构想力的逻辑，与其说是感情的逻辑，毋宁说是形象的逻辑……。（[3]，p.35）

与康德原始版本的“想象力”概念不同，这段引文很清楚地体现了三木版本的“构想力”所试图综合的直观不仅含有知识论的面相，而且还带有情感的面相。此外，三木版“构想力”所试图综合的知识因素，也不再仅仅是康德式的静态知性范畴，而且还包括形象的动态发展要素。因此，套用分析哲学的术语来说，三木的“构想力”概念要比康德的“想象力”概念来得“厚”（thick）得多。为了彰显“构想力”概念的这一“厚”的意义，三木还特别提到了“构想力”对西方文化中两大要素的综合意义：其一乃是“逻格斯”（希腊文“λόγος”，日语假名写法“ロゴス”，意思是“理论思维力”，亦与语言有关）；其二则是“帕索斯”（希腊文“πάθος”，日语假名写法“パトス”，意思是“情绪感染力”）。在此基础上，三木还在《构想力的逻辑》中分别提到了构想力的三种构像方式：神话（对应人类学哲学）、技术（对应科学技术哲学）与社会建制（对应政治哲学）。

而与三木对“构想力”的“厚解读”相对应，他对于“经验”概念的解读也是偏“厚”的。他写道：

> 语词“经验”有着双重的意义。首先，它指涉某种客观物。当我们说到经验的时候，我们意指的乃是我们的确遇到过的事情，某种在客观上被给定的事情。倘若我们被允许说错觉或幻觉也是经验，那么，“某事被经验到了”这一事实就不再是一个客观事实了——而即使被经验到的乃是错觉或幻觉，情况也是如此。不过，另一方面，我们对“经验”的讨论也总是与一个认识主体相关联。经验总是某个经验拥有者的经验，而假若经验的主体不存在了，经验也就变得不可设想了。因此，经验便是某种既主观又客观之物，或某种既客观又主观之物。（[3]，p.123）

与之相对应，则是三木所概括的英国经验论视野中的“经验”的“薄”概念：“关于经验的正统观点——也就是英国经验论的观点——也将经验视为某种心理学事物，并因此是属于意识的。”（[3]，p.123）现在问题来了：凭什

么三木就说他自己对于“经验”的厚定义，要比经验论对“经验”的薄定义更有道理呢？我在三木本人的叙述中找到了四点根据。

第一，基于日常语言用法的考量：只有对“经验”的厚解读（也就是允许其既包含其客观维度，又包含其主观维度的解读）才能符合我们对“经验”一词的日常用法，因为该用法不允许我们在客观对象缺席的意义上去使用“人生经验”之类的表达式（参考前文他对“经验”的解读）。

第二，基于认识论的考量：假若我们遵循英国经验论的思路，将客观性剥离于“经验”，那么，如下的认识论困惑就会无法解决：基于主观经验的认识主体是如何获得客观知识的呢？由此，根据三木的描述，英国经验论者就陷入了三种尴尬境地：（甲）或者像贝克莱那样，先是执着于经验的主观性，后来又放弃该立场导向客观唯心论；（乙）或者像洛克那样，做一个不彻底的经验论者，即不自洽地去承认心灵之外的外部事物的存在；（丙）或者像休谟那样倒向对于外部世界的怀疑论。（[3]，p.124）既然这三种立场都不能让人满意，那么，经验论者对“经验”的定义就需要被修正。

第三，基于经验的时间整合性的考量：就我们一般人的现象学体验而言，几乎没人能感受到刚才逝去的旋律中的一个个音符的，毋宁说，我们只能听到一整段已被高度整合的旋律。但按照英国经验论的薄“经验”概念，经验本身也必须被原子化以便卸载任何一种客观性。不过，这种极端的观点就使得如下两个问题变得难以被解答了：“共相何以存在？”；“单个的感觉与观念如何在一个特定的关系中被互相捆绑？”（[3]，p.131）与之相较，三木自己的厚“经验”观则完全可以通过援引与之贯通的“想象力”概念来为上述难题提供思路。换言之，既然“回忆”与“幻想”是“通过被视为想象力的运作方式而存在于经验之中的”，而且，既然“对于未来的想象力投射乃是通过对于过往的想象力重建才得以可能的，且反之亦然”，（[4]，p.130）那么，想象力自身所自带的时间延展力就会使得经验的客观性延展有了基本的保证。

第四，基于进化论的考量：根据进化论，生物体的任何内部表征能力的存在，在相当程度上都是为了“适应”特定的外部环境因素，因此，一种脱离了环境因素考量的“经验”观（如英国经验论的经验观）就是达尔文主义所无法消化之物。而三木则在同样深受进化论震动的杜威的影响下，接受了这种经由外部环境要素反思“经验”的新视角。不过，与生物学意义上的环境适应说不同，三木将“经验”视为一种“在我们自己与环境之间达成的主动的妥协”，并认为“经验”作为“事件”是自带“历史性”的。（[3]，p.125）这就意味着三木版的“经验”还带有鲜明的“人文历史面相”。

三木的经验观以及其对英国经验论的批评，对当代LLM技术又有何借鉴意义呢？

二、三木对休谟的批评对大语言模型的辐射意义

三木对于休谟经验论的下述批判，只要稍作改动，就能被沿用到对当代AI主流技术的批判上去。三木写道：

> 正如我们在休谟哲学里所看到的那样，经验论赋予了“习惯”以很大的理论权重。更重要的是，经验论的习惯理论乃是建立在一种对于“观念之联想”的机械主义解释的基础之上的。习惯被视为观念与行为之间的一种联系，而这种联系本身则通过重复得到构造与强化。然而，被联系到的词项、知觉与行为，其就自身而言却是不变的，而且，经验论关于“观念的联想”的教条是在精神事实中将“要素”视为不变的东西的，而这些要素本身只能通过机械的方式而得到逐步的构建。（[3]，p.129）

三木笔下的休谟对于“习惯”之形成过程的描述，貌似仅仅是在心理学范围内进行的，但也完全可以被挪用到对于一个人工神经元网络（Artificial Neural Network, ANN）的运作的解释上去（顺便说一句，ANN技术是今天的LLM技术的技术前提之一）。我们知道，任

何一个ANN架构都由通过数学建模产生的大量“人工神经元”构成，它们被排列为一层输入层、若干层中间层与一层输出层，而这些层次彼此之间又存在着大量的潜在联系通道，而每个通道又都有待通过“训练”而被赋予一个合适的权重值。至于对于一个ANN的“训练”，在本质上就是调整这些人工神经元之间的联系权重的过程，以使得整个网络能够以用户所期望的方式将特定类型的输入映射为特定类型的输出。这里需要注意的是，当ANN执行诸如图像识别之类的任务的时候，“输入”就往往是特定图像的数码化表征，而“输出”就是特定图像的语义标签的数码化表征。因此，假若我们将“输入”视为休谟所说的“印象”的数码化，而将“输出”视为休谟所说的“观念”的数码化，那么，ANN中的“输入–输出”联系，就是休谟笔下的“观念–印象”联系。至于这两类联系的建立方式，大体而言，也都是三木所说的“机械式”的方式，只是其技术展开方式会有很多种具体的可能性。就ANN而言，假若其构建过程走的是“监督式学习”（supervised learning）的进路，那么，人类工程师就会主动干预系统自动生成的大量可能的“输入–输出”映射方式，并仅仅从中遴选出那些符合人类用户期待的映射方式。假若相关的构建方式走的是“非监督式学习”（unsupervised learning）的进路，那么，系统就会在相对缺乏人类干预的情况下，从环境中自主提取有关特定输入与特定输出之间的映射关系，并以此作为自身训练的依据。需要注意的是，诸如ChatGPT这样的LLM模型所经常采用的“预训练”（pre-training）技术，在本质上就是一种非监督性学习。而休谟所理解的习惯形成过程，也更多的是一种非监督学习的产物。由此，我们也就可以对休谟所说的“习惯”进行一种基于ANN技术的新定义：所谓“习惯”，便是一个ANN系统在完成特定的学习过程中，将一类特定输入映射为特定输出的倾向或禀赋（disposition）。

但需要注意的是，在三木看来，以上所展现的，恰恰就是以休谟为代表的英国经验论者所误解的习惯形成方式。用他自己的话来说：

> 经验论忽略了一个关键要点，即知觉与行为恰恰会随着习惯的形成而发生改变。而且，我们在当下与自身亲密接触的环境中感知事物的方式，是与我们初次与该事物接触时对它们的知觉方式大相径庭的。对于一个行为展现来说，在练习开始时掌握事物所需要的知觉方式，乃是迥异于在练习结束时的知觉方式的。（[3]，p.129）

换言之，三木对休谟的批评所聚焦的乃是其一个重要理论前提，即印象是可以在被个体化的前提下而被设定为恒常不变的对象，因此，印象的加强，在本质上就是同一种质的印象在量上的累积的结果。而三木之所以说休谟式的习惯形成说是“机械式”的，也恰恰是因为，典型的机械操作方式往往得预设被操作对象的原子化与恒定性。然而，依据三木本人的经验观，真实的知觉过程，恰恰会随着任务进程的改变而产生质的差异，而不仅仅是量的差异。因此，休谟的习惯形成说是不能被运用于人类认识世界的真实过程的。

有三类实证心理学的证据能进一步证明三木的观点要比休谟的观点更符合实情。第一，根据韦伯–费希纳法则（Weber-Fechner law），感觉量与物理量的对数值成正比，也就是说，感觉量的增加幅度会大大落后于物理量的增加幅度——物理量成几何级数增长，而心理量成算术级数增长。这自然就会导致一种结果，即我们非常难用机械的方式去调整感觉量的大小，因为作为“自变量”的物理刺激的变化幅度实在太剧烈了，而很难被用以精确调整作为“应变量”的感觉量的微小波动。这也就是说，我们实际上很难在技术上保证休谟的习惯学说之前提的有效性，即特定的印象原子的“感觉量”可以像一块泥砖的“含泥量”那样通过外围的物理手段而被精确地制造出来。第二，根据特沃斯基（Amos Tversky）与卡内曼（Daniel Kahneman）发现的“锚定效应”，[5] 先被心理主体所获取的证据在锁定相关决策方向时所做出的贡献，要明显高于后被主体所获得的证据所做出的贡献。这也就是说，与休谟主义者所设想的不同，知觉类证据在量上的积累，并不

会线性地导致相关观念联系的强化。第三，根据罗克（Irvin Rock）提出的“智能知觉”理论，[6]低阶层的感知觉活动会在相当程度上受到高层次的认知活动的牵引（比如，一个中国人对汉字文本的视觉扫视方式，就会迥异于一个不懂中文的外国人的扫视方式）。因此，低层次的印象输入对于知识输出所做出的贡献，也就很容易被高层次的理论输入所做出的贡献给“对冲”掉，而休谟的“经验论加还原论”的复合立场却使得他很难认真对待高层次的认知活动的独立价值。

而对于休谟主义不自觉的继承者的当代LLM研发者来说，上述这三点批评中貌似最容易被打发掉的是第一点。其道理是，LLM研发者根本不需要预设我们需要通过物理方式制造感觉印象，毋宁说，我们处理的基本材料乃是作为语言的最基本技术单位的“语元”（token）。由于语元本身非常容易被数码标签所锁定，因此，我们也就不必担心“如何保持语元自身的同一”这一问题。

但这一处理方案所付出的代价是什么呢？站在三木哲学的立场上看，这首先当然是指：LLM由此绕过了如何让机器具有与人类类似的感觉印象的问题，而仅仅成为某种语言处理机制，并因此在根底上无法具有三木所看重的作为“构像之力”的“构想力”。其次，LMM技术语境中的“语元”的实例，往往不是一个单词，而是一个单词的一部分（一个单词约等于0.75–0.9个单词）。因此，“语元”之被人为构造出来，纯粹就是为了方便切割文本以便进行统计学处理。但三木的“经验”观既然是由“构想力”为基底的，而“构想力”又自带在时间的“过去”与“未来”两个维度中所拓展的能力，三木版的“经验”观就不可能容忍LLM技术进路中的“语元”观所自带的原子主义倾向。或挪用三木自己本用来批评经验论的话来说，“经验论之被构造出来，其方式往往是有赖于‘被给与者’这个概念。但经验就其本质意义而言，乃是实验性质的，是一种对‘被给与者’进行改变的努力”。（[3]，p.128）而这句话只要被稍微改变措辞，就能变成这样一个更适合AI时代的版本：LLM之被构造出来，是高度依赖于对于“语元”的不变性的假设的。但语用经验就其本质而言，乃是实验性质的，是一种对“语元”的不变性假设进行挑战的努力。

既然在“语元”这一LMM技术的基本出发点上，三木的观点与作为经验论继承者的LLM进路有如此大的分歧，那么在语义学层面上，二者的分歧也只能变得更大。对于LLM进路来说，一种基于语元不变性假设的语义学，也同时是统计学性质的，这种统计学的思维方式的具体技术体现方式，便是所谓的“词向量嵌入”（word-embedding）。

那么，什么叫“词向量嵌入”呢？假设你是一个排版工人，需要以最快的速度将成千上万个放在超级“语元”备用板上的“语元活字”置于特定的排字板上，以便印刷出明日的社论。那么，你该怎样设计这一语元备用板以提高工作效率呢？很显然，如下措施是很容易被想到的：（甲）将常用的语元放到最容易被获取的位置；（丙）将彼此有亲缘关系的语元就近放置。而这里的“亲缘关系”，则完全可以用“语义距离”这一量化标准来加以标注。由此一来，任何一个语元就得到了一个在巨型数字矩阵中的“经纬度”，根据此坐标，你就能知道它与谁是远亲，与谁是近邻。而此类坐标信息，就是该语元的“词向量嵌入”的产物。

但从三木哲学的角度看，这一做法的问题就是：既然“词向量嵌入”数据的来源本身是某种对所有输入语料都“一视同仁”的统计学处理，这些语料先后进入系统的时间以及自身的权威性，就会被高度边缘化了。由此，前文提到的特沃斯基与卡内曼发现的“锚定效应”，以及罗克发明的“智能知觉”理论，都无法在词向量嵌入的模型中得到应有的尊重。由此付出的代价则是，这样的自然语言处理模型，是无法刻画人类在基于小数据背景下所做出的相对剧烈的语义学知识修正的。与之相较，对于社会权威的罗克式认知，能够帮助认知主体主动地挑战历史所构成的习惯，由此大大提升知识修正的效率。很显然，这样的“挑战习惯的机制”，是作为休谟主义的继承者的LLM进路

所无法包容的。

而也恰恰是因为LMM式的语言处理技术无法容纳对于休谟式习惯的挑战，它也就无法实现三木清赋予构想力的两大面相：帕索斯与逻各斯。从表面上看，帕索斯所代表的诗性思维与逻各斯所代表的科学思维是彼此对峙的，但二者却在某个问题上达成了隐秘的共识：二者都不能是统计学意义上的平均化处理的产物。诗性思维需要别出心裁，需要巧思妙想，而这就意味着对统计学平均的背叛；而科学思维需要对“反事实条件下的法则”的虚拟式思考，而此类思考也未必会得到来自现实世界的既有数据的支持。特别需要注意的是，法则支持下的科学表述往往还具有“边角分明”的特征，而统计学视角下的LLM输出则往往缺乏是非分明的“界限感”，这两者之间的自然落差，便是被学界诟病多日的LMM的“机器幻觉”的重要来源之一——在这里我们不妨参考一下苹果公司的扎哈德（Iman Mirzadeh）团队对LLM的数学–逻辑推理能力的诟病。[7]需要指出的是，虽然由华裔计算机科学家魏杰森（Jason Wei）所开发的思维链（chain-of-thought）技术通过“分解复杂推理”这一措施，的确有限提高了LLM的科学推理能力，但就连魏杰森本人也在相关论文的末尾处承认，该技术只有在科学推理所涉及的那个领域所具有的训练语料足够大、且相关的LLM模型规模也足够大的前提下，才能有效地遏制“机器幻觉”的产生。[8]因此，即使是在思维链技术的加持下，LLM也无法复刻出柏拉图在《美诺篇》中所写到的那个小奴隶在小数据环境下具备的推理能力：他在毫无教育背景的情况下，仅仅通过苏格拉底区区几分钟的点拨，就能从事初步的几何学研究。

需要补充指出的是，在上文中被反复提及的“小数据环境”，虽然未在三木的文本中被明确提到，但也透过其本人的意义理论得到了烘托。与LLM基于统计学的语义学相对应，三木本人对意义的定义如下：“行动的形式，乃是通过对整体性环境的回应而产生的整体性行为的形式。如此看来，这样的形式也就表达出了意义。意义乃是整体与部分之间的内在性的功能性关系。”（[3]，p.128）

三木的这种意义观显然是基于小数据的，因为他所说的“部分”与“整体”之间的关系，并不是LLM背景中的“被聚焦的当下语料”与“曾被训练的整体语料”之间的关系，而是一个有血有肉的行动者与其直接面对的微观环境之间的关系。由于语词的意义与一套弓箭之类的器具一样，都是为了行动者为了适应特定的微观环境而被发明出来的，因此，行动者当然有权为了方便（convenience）而去修订他的语言工具，由此改变语义。而当相关的行动者具有强大的社会权力的时候，为了满足他自己的政治目的，他甚至可以强迫别的社会共同体去迅速改变既有的语义（如中国古代复杂的避讳文化所展现的）。所以，当LLM的构建者在面对语义变迁时不得不去追索相关的统计学数据的时候，三木式的哲学思考者所追索的，则是那些具体的语言使用者的意志力与构想力之所指。

面对三木哲学发起的这种挑战，经验论与LLM进路的拥护者或许会说：三木的哲学所最后导出的结果，乃是“通用人工智能理想的不可能实现性”，而并不仅仅是针对LLM进路。不过，三木哲学的追随者完全可以既否认LLM进路的“技术”成色，但同时不去否认通用人工智能理想的可行性。

三、大语言模型的可疑的“技术”成色

三木认为，“‘技术’这词就其广义而言，指涉用以获取一个确定目标的任何举措、手段（或手段的组合方式）与体系”。（[3]，p.94）虽然三木的这种宽泛的技术定义允许他像古希腊人那样将任何一种“制作”（poiesis）都视为一种技术工作，但他依然将技术与其前身——神话——相区分。大致而言，以集体表象为基础的神话并不要求神话的信仰者清楚地意识到他为何选择这一神话体系（毋宁说，对于神话的盲信态度往往会压抑这种自觉意识），而技术的使用者却肯定要清楚地知道他为何选择这

一特定的技术手段。因此，在技术语境中，对于此类“为何”的辩护（如，为何要造一条长城，而不是挖一条宽大的战壕，等等）永远是可被期待之事。

倘若三木能活着看到今日的LLM技术的话，那么他很可能会认为这种技术路径在本质上并不是一种真正的技术，而是一种“类技术存在”，因为与标准的技术产品不同，LLM的设计具有“模糊对于使得技术产品得以运作的科学规律的聚焦”与“模糊对于技术的服务对象的聚焦”这两个特征。

先来看第一个特征。LLM进路的维护者可能会反驳说，对于“规模律”（scaling law）的追求，恰恰说明LLM的研究并非真地悬置了使得LLM得以工作的科学规律的探究。但笔者并不认为这一反驳是有效的。“规模律”所试图刻画的，乃是与LLM的运作有关的四个参数之间的关系：（甲）模型自身的大小（类比于炼钢炉的大小）；（乙）训练的数据集的大小（类比于铁矿石的量）；（丙）训练代价（类比于炼钢成本）；（丁）训练的后呈报出的错误率（类比于炼钢后的废品率）。

很显然，上述前三个参数是LLM的研发者所直接获取的信息，而最后一个参数则可以在用户端被显现出来。因此，所谓“规模律”所试图聚焦的问题，便是如何调整（甲）–（丙）这三大参数以便使得（丁）这个参数能尽量符合人类的期待（即降低错误率）。一般而言，业界都同意这样的一个观点，即前三个参数的规模越大，最后一个参数的表现也会更好，或简言之，“规模越大，表现越好”（这也便是“规模律”这一说法的由来）。

然而，在（甲）–（丙）这三大参数中，哪些参数对提高LLM的表现更有帮助呢？如何对它们各自所做出的贡献进行更精确的数学刻画呢？一旦触及到这一层面，“规模律”自身的空洞性就立即浮现了出来。大致而言，OpenAI公司的卡普兰（Jared Kaplan）团队认为，模型参数的规模（即“炼钢炉”的大小）为达到一定“输出合格率”所做出的正面贡献，乃是训练数据量(即“铁矿石”的数量)为达到同一目标而做出的贡献的约3倍，[9] 而根据DeepMind公司的霍夫曼（Jordan Hoffmann）团队的研究成果，以上二者为达到类似输出质量而做出的贡献是大约彼此相当的。[10] 这显然是两种差异很大的估计，分别会导致各自相信这两种不同估计结果的LLM研发者再分别去采纳彼此非常不同的投资策略。而且，从科学哲学的角度看，任何一种科学法则的表达都不能允许在量化表达方面如此马虎（比如，没有一种成熟的科学规律能允许包含一个未得到精密测算的常数）。从这个角度看，“规模律”要么就是某种真正的律则在被发现之前所出现的征兆，要么就连某种真正的律则的征兆都算不上，而只能被视为对于既有的机器学习数据的某种“事后诸葛亮式”的函数拟合。

再来看前述第二个特征。根据卡耐基梅隆大学的迪亚茨（Fernando Diaz）与谷歌公司的马代奥（Michael Madaio）的研究，[11] 目前主流的规模律探索模式缺乏正常的社会科学研究所应该具备的一项基本条件，即在统计学意义上对于不同人群间差异性的照顾（比如，在统计选民的政治倾向的时候，必须注意性别、种族、教育背景、年龄、母语等因素，而不能在一个固定人群中采集数据）。毋宁说，主流的LLM研究者往往从特定网络平台所搜集的大数据作为评价数据集，却罔顾如下事实：特定平台自身的特性就可能使得相关用户具有了与之对应的特异性，并因此失去了对于全体人口的统计学代表性。从这个角度看，既然LLM的研发者对自己产品的输出的评价群体都没有搞清楚（即将特定平台上的评价群体误认为是具有统计学代表性的），这就使得LLM的研发失去了一般意义上的技术研发所应有的目的论特征：对于技术服务对象的清晰意识。

不过，上面的讨论并不意味着从三木哲学的视角看，任何一种通用人工智能技术都是不可能出现的。三木从来没有做出过类似塞尔那样的理论承诺，即：即使某种AI装置对于语言的处理会让人类用户彻底满意，它也不具有真正的心灵。[12] 毋宁说，三木式的哲学家只愿意做出这样的承诺：第一，我们目前尚且没有

看到塞尔所描述的这样的机器问世，而既有的LLM显然还没达到塞尔的“中文屋”思想实验所给出的相关概念标准；第二，假若这样的机器会在未来问世，其问世也必须以厘清这两个要点为前提：（甲）它是为谁服务的（尽管这个“谁”未必要具体落实到特定个人，但至少也要与对特定人群的画像相联系）；（乙）它的技术构造是以关于“心灵如何运作”这一点的怎样的理论构想为内核的。第三，无论此类关于心智的理论构成是怎样的，它肯定不能是一种经验论式的心灵理论，而必须容纳一种对于“经验”与“想象力”概念的“厚”理解方式。

总　　结

三木哲学乃是日本哲学家在上世纪三十年代完成的对于德国古典哲学、德国现象学、美国实用主义等思想资源的一种日本式综合。三木哲学所包含的这些资源虽然彼此有异，但都在反对古典经验论的薄“经验”概念上达成了一致。而三木在综合这些要素的基础上所做出的独特性贡献，乃在于将在康德哲学中相对边缘的“想象力”概念升级为作为他本人思想体系枢纽的“构想力”概念，并在此基础上开发出了一种新的技术哲学。非常令人惊讶的是，尽管任何一个初读三木哲学的读者，或许都会由于他所援引的思想资源的强大而至少被他的一部分论点所说服，当代的主流AI研究却恰恰就是在绕开这些20世纪的哲学洞见的前提下进行的。而与之相较，被20世纪哲学主流所抛弃的休谟式的原子式经验观，却成为了LLM研究中的“语元化”操作的哲学前提，由此戏剧性地造成了主流AI研究的哲学前提与三木哲学之间约一个世纪半的时间差（根据休谟与三木各自卒年之间的差值计算）。

因此，与很多人的估计相反，LLM技术或许并不代表AI技术的最先进的部分。从哲学角度看，这只是穿上了数码化外衣的休谟的幽灵在游荡罢了，这样的幽灵其实早就在哲学殿堂中被康德、黑格尔、三木清、海德格尔、塞拉斯等大哲学家赶下了宝座，而仅仅凭借某些幸运的机缘而在AI的庇护下找到了新的关注度。但随着LLM自身的运作所需要消耗的巨量资源（特别是较易获取的来自互联网的数据资源）自身的耗尽，这个休谟的幽灵或许迟早会开始其新的流浪，而AI的殿堂也迟早会意识到他们的佛龛一直供错了菩萨。

［参 考 文 献］

[1]Xin, E. 'The RenAIssance: Why AI Marks a Resurgence of Empiricism'[EB/OL]. World Economic Forum, https://www.weforum.org/stories/2023/11/why-ai-marks-a-resurgence-of-empiricism/. 2023–11–12.

[2]Long, R. 'Nativism and Empiricism in Artificial Intelligence'[J]. *Philosophical Studies*, 2004, 181(4): 763–788.

[3]Miki, K. *The Logic of Imagination*[M]. Bloomsbury: Bloomsbury Academy, 2024.

[4]Kant, I. *Critique of Pure Reason*[M]. Cambridge: Cambridge University Press, 1998.

[5]Tversky, A., Kahneman, D. 'Judgment under Uncertainty: Heuristics and Biases'[J]. *Science*. 1974, 185(4157): 1124–1131.

[6]Rock, I. *The Logic of Perception*[M]. Cambridge (MA): MIT Press, 1983.

[7]Mirzadeh, I. 'GSM-Symbolic: Understanding the Limitations of Mathematical Reasoning in Large Language Models'[J]. ArXiv Preprint ArXiv: 2410. 05229, 2024,

[8]Wei, J. 'Chain-of-Thought Prompting Elicits Reasoning in Large Language Models'[J]. *Advances in Neural Information Processing Systems*, 2022, 35: 24824–24837.

[9]Kaplan, J. et al. 'Scaling Laws for Neural Language Models'[J]. ArXiv Preprint ArXiv: 2001.08361, 2020.

[10]Hoffmann, J. 'Training Compute-Optimal Large Language Models'[J]. ArXiv Preprint ArXiv: 2203.15556, 2022.

[11]Diaz, F., Madaio, M. 'Scaling Laws Do Not Scale'[J]. ArXiv Preprint ArXiv: 2307.03201v2, 2024.

[12]Searle, J. 'Minds, Brains and Programs'[J]. *Behavioral and Brain Sciences*, 1980, 3(3): 417–457.

［责任编辑　王巍　谭笑］