借助人工智能重新思考发现性研究：超学科性的前景与还原论的危险

Iulia Georgescu[a] 和 Venkatesh Narayanamurti[b]

理论上，跨学科研究旨在解决现实世界问题，不受学科边界和方法论的限制。但在实践中，这并非易事，正如那句名言所说：“世界有问题，大学有院系。”作为一种强大的整合工具，人工智能或许能帮助我们摆脱传统的还原论方法，实现真正的超学科性。本文将探讨这一切如何实现，以及为何在当前的激励体系下，这一切并未发生。

arXiv:2611.29729v1 [physics.soc-ph] 22 Nov 2026

人工智能技术正在改变我们开展科学研究的方式，但原因并非仅仅是人们最初可能想到的那些。人工智能本质上是一种整合工具，能够帮助我们摆脱传统的发现性研究还原论方法，实现真正的跨学科性；而这正是应对一些复杂社会问题的关键。尽管人们很容易这样认为，实际情况却恰恰相反。现在正是重新思考支撑发现性研究的激励体系的关键时刻，以确保人工智能促进而非损害科学和对理解的追求。

问题还是学科?

1952年，哲学家卡尔·波普尔指出：“学科的划分，一部分出于历史原因和行政便利（例如教学和职位任命的组织安排），另一部分则是因为我们为解决问题而构建的理论往往会发展成统一的体系。”[1]

事实上，历史原因与行政便利性息息相关。直到19世纪，科学都没有明确的划分。在自然科学领域，物理学与化学之间几乎没有区别，基础科学与应用科学之间也没有区别，贡献者之间亦无区分，业余人士与职业科学家之间同样如此[2]。随着知识语料库

[a] 物理学会。以个人身份撰稿。

[b]技术与公共政策、工程与应用科学及物理学名誉教授，任职于哈佛大学。

不断扩大，将其划分为易于管理的部分变得势在必行，这也促成了专业化。例如，18世纪中叶时，每年发表的科学论文总数还不到或约为100篇（人们可以轻松读完）；到了1800年，每年发表的论文约有900篇（已经多到难以跟上）。

其他因素也促成了从自然哲学向现代科学的转变。科学职业化将业余科学家推到了边缘。大学重新调整课程体系和讲席，以反映这些变化，由此在十九世纪上半叶改革后的德国大学中形成了新的行政单位 [3]。这一模式迅速传播到欧洲及其他地区。与此同时，面向不同领域和群体的专业学术期刊也逐渐围绕这些领域形成。

纯科学／基础科学与应用科学的分野随着工业革命而出现。1883年，美国物理学家亨利·奥古斯塔斯·罗兰发表了一场题为《为纯科学辩护》的演讲，区分了物理学科学与其应用，例如“电报、电灯之类的便利设施” [4]。由于范内瓦·布什在1945年的报告《科学：无尽的前沿》，这一二分法在二战后被奉为定论 [5] ，尽管布什曾提醒说：“应用研究与纯研究之间的区别并非截然分明。”

科学学科的涌现和基础／应用分类，都没有解决知识量不断增长的问题。恰恰相反，专业科学知识的数量仍在持续增长。“过去二十年来，化学取得了惊人的快速进展，以至于即使是最勤奋的科学研究者，也难以跟上当下的知识发展，”化学家尤斯图斯·冯·李比希在 1889 [2] 中指出。同样的话在 20 世纪 60 年代的研究者听来想必也深有共鸣，直到今天仍然如此。事实上，1850 年发表的文章超过 12,000 篇，到 1900 年增至 47,000 多篇。到 1950 年（当年发表的文章超过 200,000 篇），人们已经无法跟上一门学科的出版物，而许多子学科也已涌现。2000 年发表的文章超过 150 万篇，2025 年则达到 670 万篇（数据来自 Dimensions，不包括预印本和会议论文集）。

划分方式有很多，相应的分类体系也多种多样。例如，澳大利亚和新西兰标准研究分类 ANZSRC（Dimensions 使用的分类体系）列出了约 2,000 个子领域。以物理学为例，其子领域包括核物理学或粒子物理学等宽泛领域，而这些领域还可以进一步细分为研究领

域。专业化程度如此之深，以至于有些粒子物理学家对其他粒子物理学家所研究的内容几乎不熟悉。这些学科边界是人为划定的，并且在交界处变得模糊。事实上，即便对于物理学这样一门历史悠久的学科，也很难划定清晰的界限，并绘制出准确的领域图谱 [6]。

那么，除了传统和行政便利（包括信息管理）之外，我们真的还需要学科吗？波普尔曾指出：“但所有这些分类和区分都相对无关紧要且流于表面。我们不是某一学科领域的研究者，而是问题的研究者。问题可能会跨越任何学科领域或学科的边界。”[1]换句话说，这就是超学科性的理念：研究超越学科边界和方法论，并以解决现实世界问题的需求为驱动 [7]。波普尔的观点虽然颇具吸引力，却与科研生态系统的现实形成了鲜明矛盾，正如这句话所概括的那样：“世界面临问题；大学设有院系。”[8]

在当今这个复杂的生态系统中，激励机制强化了学科壁垒，并加深了基础研究与应用研究之间的二分法，我们能否成为问题的研究者？在尝试回答这个问题之前，我们将探讨一个秉持波普尔理想、在科学与技术领域留下非凡遗产的组织实例。

贝尔实验室模式

过去100年来，AT&T贝尔实验室为科学与技术作出了巨大贡献：从塑造当今世界的发明（晶体管、激光器、光伏电池和Unix操作系统），到开创射电天文学、信息论等新领域。关于这一切如何成为可能，已有大量论述 [9–11], ，但有几点尤为突出。贝尔实验室不拘泥于学科之间的界限，也不拘泥于所谓“基础”研究与“应用”研究之间的界限。团队成员往往并没有相关研究领域的特定资历，而方法和背景上的多样性恰恰促成了突破。例如，研发可靠的固态放大器（晶体管）这一挑战，既需要理论物理学的新见解，也需要材料科学与工程学的进步，以实现对电子的控制和信号放大[12]。

尽管实验室设有部门（出于行政管理的便利），但有意采用了不致形成孤岛的组织方式。在物理空间上，默里山园区的建筑设计体现了这一点：长长的走廊和开放式办公室让人们难免相遇；办公室与实验室交替分布，使理论学家、实验学家和技术人员得以共处一室。在心理层面，这则通过平等精英制得以实现：无论背景和学历资历如何，每个人的想法和贡献都同样受到尊重。

在贝尔实验室，新入职者会得到刻意保持含糊的指引，并受到鼓励去追随自己的兴趣。约翰·霍普菲尔德便是如此。他后来在诺贝尔奖演讲中回忆道，自己在贝尔实验室的研究生涯始于寻找一个好问题，这一寻找将他带入固体物理学领域，继而又走向生物学。“我当时在寻找一个重大问题，它的解决与理解将产生远超其名义学科边界的意义，并重塑它们所源自的领域。” [13]贝尔实验室的研究人员能够享有长期思考的自由，这意味着他们常常思考“更长远的问题”（或“下一个相邻可能” [12]）；这些问题本质上超越了现有的学科边界。因此，超学科性深深融入了他们的工作方式，也融入了他们所寻求解决的问题之中。

一个机遇

如今，要重建贝尔实验室或许颇具挑战 [10], ，但也许还有另一种方式可以打破学科边界，成为问题的探索者。这就是运用霍普菲尔德的研究助力开发的技术：人工智能。

回到学科最初出现的原因，如今似乎终于有办法跟上不断增长的科学文献数量。大语言模型搜索知识语料库并检索信息的能力，表明我们可以借助它们跨领域获取前所未有的广度（适用条款与条件，见下文）。从哲学角度看，科学仪器是认知增强工具，能够扩展人类有限的感官。望远镜和显微镜让我们得以观察遥远之物和微小之物，而其他仪器则让我们得以窥见不可见之物，例如探测辐射或磁场。与计算机

模拟一样，人工智能技术可用于外推、转换和增强，但它们还提供了一种新的认知增强方式：整合并综合大量迥异类型的信息。

作为一种整合工具，人工智能似乎为实现超学科性提供了一条途径。此外，人工智能也存在于不同的交界处。

学科之间

人工智能作为一个研究领域，始终处于不同学科的交汇处。想想1956年达特茅斯研讨会的参与者吧，这次研讨会被视为人工智能的创立时刻。他们当时还不是计算机科学家（计算机科学刚刚兴起），而是数学家、电气工程师、物理学家、心理学家和经济学家。此后，该领域的发展仍不断源自这些学科及其他领域。因此，即使学科边界将人工智能划归为计算机科学的一个分支学科，这种划分在一定程度上也是任意的；学科之间模糊的边界上，存在许多具有影响力的研究成果。

Hopfield于1982年发表的文章《具有涌现集体计算能力的神经网络与物理系统》[14]，是人工智能发展中的一项里程碑式成果，并使他获得了2024年诺贝尔物理学奖。这就是一个例子。该文章被归类为生物物理学，这或许与其关键词不太相符：联想记忆/并行处理/分类/内容寻址存储器/软失效设备。Hopfield隶属于加州理工学院化学与生物学部及贝尔实验室。该文章引用的文献中，8篇来自神经科学，3篇属于多学科研究，3篇属于生物数学，2篇属于心理学，2篇属于生物科学，2篇属于物理科学，1篇属于计算理论。这是一篇真正的跨学科论文（并且在 [6]的分析中也被如此认定），它彰显了各种分类和区分的表面性。

介于基础研究与应用研究之间

就分类而言，人们很可能会将人工智能技术归入工程研究。然而，这些方法的发展得益于更多基础性研究，并推动科学与技术进一步演进。前者的一个例子是

扩散模型，它能够实现图像生成。这些模型的灵感源自非平衡统计物理学。至于后者，事实证明，神经网络理论对弦理论等物理学纯理论研究领域也很有用。

一种工具，也是一个研究对象

如今，凝聚态物理和材料物理可能是物理学中最大的分支学科，但在20世纪50年代以前，它还被视为边缘领域，甚至令人不屑（沃尔夫冈·泡利称之为“肮脏的物理学”）。它之所以成功，是因为为开发新的主动工具和器件提供了一个舞台（想想新材料的各种应用），而这又使人们得以观测新的现象（例如高温超导、分数量子霍尔效应以及其他涌现集体现象）。同样，人工智能是一种能带来新能力的工具（想想 AlphaFold），但与此同时，它也是一个值得研究的迷人复杂系统。工具同时也是研究对象的情况相对少见；若要概括而言，这类情况对科学和技术进步都大有裨益。

综上所述，人们很容易得出这样的结论：人工智能技术提供了一个实现真正跨学科研究的机会，让研究以问题为中心，以整合的方式开展。但与任何技术一样，重要的是要深入了解并仔细审视其中的细节。

风险

在科学研究中使用人工智能工具会带来多种风险。一方面，存在不准确性、偏见、剽窃和蓄意虚假信息等问题。另一方面，人们担忧研究主题范围缩窄、科学单一文化的涌现，以及“解释广度幻觉” [15], 、意外发现的丧失、深刻理解的缺失和创造力的流失。本文不讨论这些问题（参见 [16]综述），而是补充另一种风险：加剧还原论。

我们已经看到还原论的萌芽，而不是在培育一种超学科方法。科学人工智能（意指“发现性研究中的人工智能工具，以及用于研究人工智能的科学方法”）已经开始被拆分为“化学人工智能”“数学人工智能”等。

这表明人们采取的是学科导向的方法，而非问题导向的方法（用于药物和材料发现的人工智能是例外）。此外，自动化端到端科学是一种极端的还原论，因为它将研究拆分为独立模块，并将这些模块的输入和输出连接成线性图。这为什么是个问题?

还原论是一种强大的方法。将复杂系统分解成较小的组成部分，再用模型和理论加以描述，许多科学进步正是由此取得的。但还原论虽已证明行之有效，却并非万能。它只能单向发挥作用。还原论并不意味着建构主义，因此，仅凭较小的部分，未必就能构建出复杂系统；复杂系统的行为可能与各个组成部分之和所预期的不同（这就是所谓的涌现）。

这一简单却深刻的观察，出自贝尔实验室研究员、诺贝尔奖得主、物理学家 Phil W. Anderson 于1972年发表的论述：“能够将一切还原为简单的基本定律，并不意味着能够从这些定律出发重构宇宙” [17]。几十年过去，这一点已广为人知，却也常常被遗忘。例如，学科之间以及基础研究与应用研究之间的人为界限，便是还原论思维的产物。

在发现性研究中，还原论与建构主义缺一不可：既要用它们界定问题，也要用它们解决问题。探索答案与问题的过程，便展示了两者如何协同发挥作用。 [12]提出的技术科学方法认为：“科学借助技术发现新事实，而技术则借助科学发明新的形式，以实现人类期望的功能。”从这一视角来看，技术科学知识由相互关联的问答对构成；这些问答对彼此结合并不断演进，形成新的问答对，从而拓展已知领域。

“答案发现始于还原论，将问题分解为构成功能，并分析这些构成功能是否能够得到满足或解决，最终以整合告终，将这些构成功能整合为综合解决方案。”相比之下，“问题发现始于合成，取用已有的答案或解决方案并将其整合，以发现它们可能回答的新问题或可能解决的问题。随后，一旦初步发现了一个问题，就必须通过还原论分析来验证答案／解决方案是否确实回答了该问题或解决了该问题。”这种不对称使问题发现更加困难。“因为问题发现

着眼于知识网络的上层，因此无法像答案发现那样，依靠学科内部系统性的还原论分析来深入搜寻答案。问题的搜寻通常发生在学科之外，因此更加依赖意外发现和思想的偶然碰撞。” [12]

由于答案发现可以通过系统化的还原论分析方法来解决，因此在现有激励结构中，它比问题发现更受青睐。整合方法成本高、风险大。此外，答案比问题更容易验证，而评估问题是否优质则更加困难（优秀的问题最初往往看起来荒诞不经）。偏向效用而非理解的倾向（例如将科学问题用作评估人工智能模型能力的基准）使这一问题更加严重，也再次使寻找答案而非提出新问题更受青睐。这就是在研究中使用人工智能工具可能强化还原论的原因。

前进之路

显然，在复杂激励因素交织驱动下，人工智能技术将继续迅速发展，并显著改变发现性研究的开展方式。我们必须认真考虑在系统层面和个人层面需要作出哪些调整，才能确保这种变化有益于科学界和社会。我们认为，如果使用得当，人工智能工具可以助力实现超学科方法。然而，仅凭人工智能技术本身还不足以做到这一点。

系统

在系统层面，我们需要在效用与理解之间取得平衡。为此，有必要明确：科学旨在理解新事物，而工程学旨在构建新事物。发现与发明是科学与技术协同作用所产生的离散且极具价值的成果。然而，科学和工程学中的研究都是持续、动态且复杂的过程，其内涵不只是各个部分的简单总和。而且，这一过程与最终成果同样重要。狭隘地针对成果进行优化，不太可能获得理想的结果。

近期的例子说明了这一点。人工智能工具已被非常成功地用于预测具有各种实用特性的新材料的结构。然而，这一成功也带来了其他挑战：如何真正制造出这些材料，以及如何测试数百万种可能的化合物？随着工作转向利用人工智能解决新材料的合成问题，人们可以想象，未来在如何表征和理解这些材料方面还会出现更多挑战。这种人工智能方法与晶体管的开发形成了鲜明对比；在晶体管以及随后许多其他材料的开发过程中，具有物理学、化学和工程学背景的实验学家、理论学家和技术人员团队紧密合作，在各个方面不断取得进展。人工智能取得巨大成功的另一个领域是数学，但近期的进展引发了人们的担忧：大量依赖人工智能辅助证明的证明过程，可能掩盖证明发现过程所揭示的洞见与理解。此外，尽管数学研究的诸多目标在历史上一直呈正相关，但过度针对其中一两个目标进行优化，可能会使它们彼此分化 [18]。

这些例子表明，过程的重要性不亚于，有时甚至超过最终结果。往往仅仅得到最终答案并不足够。无论是在实验研究还是纯理论研究中，这一过程可能并非一帆风顺，却是获得扎实理解所必不可少的。这些例子也为将研究拆分成模块的还原论方法敲响了警钟。风险在于，优化各个部分并不能保证整体得到改善（这一教训在将人工智能应用于各种领域时日益显而易见，不仅限于数学）。

值得记住的是：“科学中的创造力并非源自一种静态、重复且界限分明的过程，正如流水线制造中所见的那样。相反，创造性成果是在一个动态、不断变化且定义模糊的过程中涌现出来的”[19]。或者，正如 Julio Ottino 和 Brian Uzzi 最近所说：“科学工作流程的结构可以被视作一只时钟。但作为一个发现系统，科学却是一团云。优化每个组成部分，并不能保证整体的完整性。事实上，优化可能会破坏整体。” [20]

个人

在个人层面，人工智能能够为人类研究人员提供前所未有的广度。然而，它无法直接提供深度，因为深度很可能只能

通过实践才能获得。深度源于主动知识，而主动知识是在不断提问和大量辛勤工作中形成的（获取隐性知识是构建深度的重要组成部分）。单有广度或深度都无法实现超学科性。两者缺一不可，“因为深度是广度产生最大影响的必要条件，所以知识广度似乎存在一个能够最大限度激发创造力的‘最佳点’” [12]。

我们可以把广度看作一个概念网络。在这一空间中运作，意味着创造新的连接、重新组合想法。深度则是结晶知识，而这一维度上的操作包括创建和使用表征与模式，重新配置想法，使其适应新的组合。不过，广度和深度都并非完全静态：概念网络不断增加连接，深度则会调整并完善表征（用物理隐喻来说，这些表征是不同的晶体结构）。我们很容易认为，广度更为通用，而深度则更贴合个体研究者的思维方式（这与从事同一课题的科学家可能有截然不同的思维方式这一事实相符）。

Hopfield 的研究是一个很好的例子，展示了如何平衡深度与广度，从而达到科学创造力的最佳点。但他的故事也表明，另一个重要要素同样不可或缺：探寻好问题。我们将这种寻求问题的心态称为做一名知情的逆向思考者。这意味着通过换个角度看问题、质疑既有定论、思考已知领域之外的新方向，来拓展理解的边界。知情的逆向思考者依靠广度和深度，窥见“下一个邻近可能性” [12]。

我们看到，人工智能工具可以帮助拓展广度（近期的例子表明，如何借此应对不同领域的问题 [21]）。或许如今比以往任何时候都更需要专业深度以及批判性的质疑心态，而这两者都离不开持之以恒的努力。人工智能技术或许也能以某些方式助力培养专业深度和知情的逆向思考心态。然而，到目前为止发现的风险，加上我们指出的还原论加剧问题，似乎都指向相反的方向。更明确的是，除非科学界批判性地评估人工智能在发现性研究中的作用，并探讨现有激励体系需要如何调整，否则既无法充分利用人工智能带来的机遇，也无法应对其所带来的风险。

[1] K. R. Popper, 《哲学问题的本质及其科学根源》, 《英国科学哲学杂志》**3**, 124（1952年）。[2] A. G. Gross、J. E. Harmon 和 M. S. Reidy, 《科学传播：从*17*世纪至今的科学论文》（牛津大学出版社，2002年）。[3] R. Stichweh, 《科学学科社会学：现代科学学科结构的起源与稳定性》, 《科学与语境》**5**, 3（1992年）。[4] H. A. Rowland, 《为纯科学请愿》, 《科学》**2**, 242（1883年）。[5] V. Bush, 《科学：无尽的前沿》, 技术报告（美国科学研究与发展办公室，1945年）。[6] R. Sinatra、P. Deville、M. Szell、等, 《一个世纪的物理学》, 《自然·物理学》**11**, 791（2015年）。[7] G. H. Hadorn 等, 《超学科性作为一种研究形式的涌现》, 载于《超学科研究手册》（施普林格，2008年）。[8] 凯洛格州立大学与赠地大学未来委员会, 《回归我们的根源：学生体验》, 技术报告（凯洛格州立大学与赠地大学未来委员会, 1997年）；另见 Gary Brewer, “跨学科性的挑战”, 《政策科学》, 第32卷, 第327–337页（1999年）。[9] V. Narayanamurti 和 T. Odumosu, 《发明与发现的循环：重新思考无尽的前沿》（哈佛大学出版社，2016年）。[10] I. Georgescu, 《重拾贝尔实验室的黄金时代》, 《自然综述：物理学》**4**, 76（2022年）。[11] V. Narayanamurti 和 J. Y. Tsao, 《技术科学知识如何进步：受贝尔实验室启发的架构》, Research Policy **53**, 104983（2024年）。[12] V. Narayanamurti 和 J. Y. Tsao, 《技术科学革命的起源：重新思考研究的本质与培育》（哈佛大学出版社，2021年）。[13] J. J. Hopfield, 《诺贝尔演讲：物理学是一种视角》, 《现代物理学评论》**97**, 030501（2025年）。[14] J. J. Hopfield, 《具有涌现集体计算能力的神经网络与物理系统》, 《美国国家科学院院刊》**79**, 2554（1982年）。[15] L. Messeri 和 M. J. Crockett, 《人工智能与科学研究中的理解幻觉》, 《自然》**627**, 49（2024年）。

[16] S. Bianchini等，《科学中的人工智能：何时以及何地会产生影响》，Research Policy **55**, 105478（2026年）。[17] P. W. Anderson，《越多越不同》，Science**177**，393（1972年）。[18] T. Tao，《人工智能时代的数学》（2026年），arXiv:2608.16753。[19] D. K. Simonton，《科学创造力》（剑桥大学出版社，2004年）。[20] J. M. Ottino 和 B. Uzzi，《没有进步的演进》，Science **392**，791（2026年）。[21] C. Zhao，《22个科学领域中的问题，其解决方案一直近在眼前。人工智能找到了它们》，Science 10.1126/science.zwu3e54（2026年）。