哪朝的剑，哪朝的官？

新背景下的科学规范和诚信

崔博涵

（南京大学 计算机学院，江苏 南京）

摘要：在后人工智能时代、学术成果贬值和科学研究高度职业化的新背景下，学术诚信问题面临全新的挑战。本文基于社会建构论、教育信号理论和博弈论等理论基础，结合教育和研究的双重视角，探讨了当前学术诚信所遭遇的困境。首先，分析了人工智能生成内容的泛滥对学术诚信审查机制的影响，指出了对人工智能生成内容进行检测的技术是无效的。其次，探讨了在学术成果贬值的背景下，学术诚信审查可能导致的资源浪费和不公平现象。最后，从科学研究职业化的视角出发，重新审视了“灌水”现象的社会功能和存在合理性。本文认为，在推进当前的学术诚信审查前，需要准备与之适应的新的社会环境和技术背景，才能有效发挥其作用，避免资源浪费，并促进科学研究的可持续发展。

关键词：科学诚信；社会建构；学术诚信审查

引言

科学研究的学术诚信问是一个长期存在且非常受人关注的问题。自现代科学的诞生之初，剽窃、抄袭就是非常严重的指控。学术诚信对于维护现有的基于引用的人类知识体系的架构是至关重要的。在高等教育中，学术诚信问题则是科学研究学术诚信的一个预演和前提。教育中对学术诚信的强调使得正在受教育的、未来的研究者们养成对学术诚信的自我要求和自觉遵守，从而在进入学术界后成为诚信体系的维护者。同时另一方面，学术诚信还可以维护教育制度对于不同学术能力者的分层和筛选。

本文将在新背景下，针对一些社会现象对科学规范和诚信结合教育和研究两个视角的讨论。

1 学术诚信问题的背景

1.1 科学诚信的“旧定义”

简单来说，科学诚信的核心内容就是科学研究者应当发表正确的、原创的研究结果。发表的内容有相似、相同之处就会有剽窃的嫌疑。学术诚信的最大问题就是其模糊性：学术的诚信和不诚信并没有的清晰、准确的界定。

1.2 “新背景”

在本文中，“新背景”着重强调讨论在后人工智能时代，后疫情时代全球经济低迷，学术界扩张导致的科学成果贬值以及科学研究高度职业化甚至产业化的今天，科学诚信的现状与问题。

2 研究视角和理论基础

下面介绍本文在对诚信制度的讨论和构想中遵循的理论基础。

2.1 社会建构论

学术诚信似乎是一个，独立存在的、理想的、道德约束层面的的观念。但现实中，其实是一个变化的、且不得不受到客观世界和社会现实制约的概念。由于学术诚信的模糊性，由人定义和判定的的学术诚信必然会存在灰色地带和监管偏差，因此不可避免的受到社会的直接影响。这与社会建构论的思想是一致的。

2.2 教育的信号理论

由于本文也涉及了对教育的探讨，我们此处使用教育的信号理论来作为对教育问题的建模。这种建模存在过度简化等问题，但在科学研究职业化的背景下却能比较精确的反映和解释社会现实。

2.3 博弈理论

目前，学术诚信尚未上升到法律层面，因而有其非强制性。同时，虽然科学研究者往往具有较高的社会地位，可观的收入水平和远高于平均的受教育水平，但是这并不代表其道德水平高于其他群体。因此我们无法假设或要求每一个研究者个体成为诚信体系自觉的遵守者和维护者。“玩家”与“玩家”，“玩家”与“裁判”之间，可以认为是部分信息博弈的关系。这是本文重要的假设。

3 人工智能时代下的科学诚信

随着生成式人工智能模型使用的泛滥，多家机构都给出了相关的应对措施来防止人工智能生成内容造成低质量信息的污染。比如国外著名的技术问答论坛 StackOverflow 直接禁止了任何人工智能生成内容的存在[1]。而计算机领域的学术会议使用的规则相对保守，一般是禁止使用大语言模型直接生成文章的正文，比如国际会议 CVPR[2]与 ICML[3]。但是这种规定实则存在严重的问题。

首先，因为技术的迅速演进和多样性，这种规则是模糊的、缺乏实际意义的。比如“生成正文”：使用人工智能搜索辞典检索词汇短语，其他语种的研究人员通过翻译工具写作论文，或者使用一些自动语句补全工具进行写作。这些都是生成正文的行为，是否应该被禁止？ 又比如“‘大’语言模型”：人工智能专家 Yann Lecun 就提出质疑（或者说讽刺），如果使用的是“中小”模型，那么应当如何判定？这些极为模糊的表示实际上导致判断最终还是完全由人的主观判断来决定。

其次，人工智能生成内容实际上是无法识别的。如果无法识别，任何关于生成内容的规则都可能是“无效”的。

这里的“是否有效”是一个有趣的问题。如果我们进行人工智能审查的目的是筛选人类的知识，提高知识的整体质量，那么实际上是有一定效果的。因为判定算法可以筛选一部分“明显是人工智能生成的”有明显的不正当企图的内容，从而减少人类审稿的工作压力；但是，如果我们的目的仅仅是防止出现人工智能生成的内容本身的话，这种审查将会适得其反。

为什么会适得其反？目前主流的的预训练大语言模型(GPT, generative pre-trained transformer)均采用了经过清理的互联网数据。而生成的内容也会以各种形式流入互联网。比如发布在预印本网站上但最终被退稿的文章或者是根本未被发现的内容。在机器学习领域有一种模型为对抗生成模型，即对于某种生成任务，可以构造一个对手模型，对生成模型生成的质量进行评估，并反馈给生成模型，生成模型和对

手模型互相迭代，最终就会收敛成为一个可以产生高质量输出的模型。而我们目前为了一定程度上发现经过人类调整的生成内容，不得不采用人工智能模型来对文本的隐藏特征进行分析。这样，审查机制就成了一种对手模型，审查过程实际上相当于在对生成手段进行训练，反而促进了更逼真的生成内容的出现，从而使得人工智能生成内容和人类所写的内容越来越接近，难以区分。如果审查的结果是的使得生成内容与人类所写的内容越来越接近，生成的手段越来越高超，那么是否意味着：审查是完全没有意义的？

当然这取决于：这些制度希望“防止出现人工智能生成的内容本身”的目的是什么？

其中一个比较形而上的原因可能是：隔离机器对人类知识探索的影响。也就是说杜绝科学研究本身使用的工具或者手段对于科学研究的影响，避免科学研究受到他们的影响而形成某种偏见。相比于电脑，互联网等技术，这种担忧为什么在人工智能技术上得到了空前的重视？最主要的区别是以前的技术几乎无法生成任何人类思维之外的内容，基本上是对人类思维的一种翻译，而人工智能模型是生成式的，其生成内容可能会对研究的结论产生引导。但这个理由也有逻辑问题。在其语境下，既然存在隔离的必要，那么科学研究是一种独立存在的、抽象的独立于工具和手段的人类活动。这就需要使用可知论的假设，同时必须要能够证明人类思考对于真理获取的独特性，即为何人类思维具有在高于机器等的“思维行为”的层面至上判断正误的能力。而这是很难给出解释的。

同时，这种隔离是不可能的，因为从社会建构论的角度上来看，任何这些技术都会从各种角度影响人类使用者的思维方式，即使不直接生成内容，也会对科学研究产生间接的影响。因此“隔离”这个原因的基础是不牢靠的。

另一个比较现实的原因可能是：避免生成内容的知识产权归属争议。也就是说，生成的内容的知识产权不属于或者部分不属于对其进行提示的人。这种考虑其实来源一个假设，就是将知识产权与一种科学研究过程中的“无差别的人类劳动”划上了约等号，而知识产权制度存在的目的就是保护思维劳动的劳动者从其劳动成果中获得回报的权利。生成模型的使用者（提示者）某种程度上使用了一种“轻松的”、“取巧的”劳动方式，因而不应该具有完整的从中获利（如发表学术论文并从中获得间接收益）的权利。但这也是难以服人的。首先，使用者对人工智能模型的设计者支付了使用费用。其次，使用生成模型的使用可以认为是提高劳动生产率的方式，因而是没有理由拒绝的，而是应该寻找相适应的生产关系。最重要的，这个假设的“约等号”并不稳固，因为在科学领域，“付出的劳动”往往并不正比于研究结论的价值，不能将劳动的多少作为衡量。如果科研人员直接在睡梦中凭空想出了一个新公式，是否应该将知识产权归结为“上帝”？

从上文的层层深入的质疑可以看出，相比于纠结于“是否是人工智能模型生成”这个标准，更好的选择是坚守内容的质量的这个底线。也就是只判断这个结果是否值得发表，而不去判断这个结果的来源是否为人工智能模型。同时要坚守“人类审查”的底线，也就是不可以完全由人工智能进行稿件审查。（这主要是因为人工智能模型的正确性和可靠性还没有得到证明，可能会导致无法发现错误的内容）这一原则也

得到了实践[4]，利用审稿人的“裁判”和“玩家”的双重身份，用重复博弈的方法来限制审稿人使用 AI 审稿的行为。

如果人工智能模型的出现使得值得发表的论文数量增多，也不应该感到担忧，而是应该接受这种进步。至于论文研究内容的价值，经过审稿初步判断后，更多是经由时间来证明的，论文数量增多并不会影响优质科研成果的沉淀。因此相比于畏惧量的增加，更重要的是提高学术文献的搜索和检索效率，以便于在更多的论文的背景下保持科学研究结果有效的复用性。

而对于教育领域，这个问题反而非常的简单，因为这完全取决于高等教育的目的。如果高等教育的目的是为了使人获得知识，也就是“教育人”，那么没有必要去解决这种问题，因为是否去选择真正获得知识完全是受教育者的自由选择。但如果高等教育的目的是为了筛选人，也就是类似于上文信号理论的模型，那么公平就是最重要的，由于上文所说的检测“是否为生成”的困难性，就必须放弃课程论文，书面作业等无法充分控制来源的考核形式。

4 学术成果贬值视角下的科学诚信审查

不再考虑人工智能，下面我们探讨一个更加现实、更加消极的问题，即：我们是否应该打击学术不端？表面上，这个问题的答案是显然的。但我们再次回到学术诚信审查的根本目的，一方面是研究结果的正确性和可复用性。但在如今的环境下，比如说生物、医药、化学等领域，一种非常普遍的现象是研究人员对非顶级期刊上的他人的研究结果完全不相信（尽管他们的部分成果也会发表在这些期刊上），甚至说将其当作完全不存在，也就是某种意义上说，它们的正确与否并不重要。那么，我们对这些占学术成果的绝大多数的正确性审查是没有意义的。如果结果不被信任和直接复用，那么就不必提供一种额外的检查。

审查的目的的另一方面是公平，我们对科学研究成果的归属权的问题之所以非常敏感，是因为我们希望维护公平正义，使得研究者可以从他们的成果中正当的、恰当的获得回报，而不能侵吞和抢占他人的回报。这种回报当然可以是直接的经济回报，当然也可以是间接的，如评选职称、获得职位，满足毕业条件或者获得学术声誉和社会声誉等等。所以公平，也就是“一致的获利，一致的惩罚”。

我们先讨论“一致的获利”。我们可以发现，这个目的建立在一个重要的前提上，就是“回报”。但是在如今的社会环境中，这个回报却是在不断降低。由于论文数量的不断增加，学历的社会认可度不断降低，上文的所述一些回报都不再称之为回报。首先科学研究可以商业化的部分很少，也就很难谈得上经济回报。对于获得职位，由于论文的整体质量降低，公司不再倾向于以论文为标准招收学生，而是选择以实习经历等更加具体的方式了解学生，因此论文对于想要求职的学生价值不断降低。对于评选职称，在学术界越来越显示出“帽子”、职称的评选与学术能力和学术成果量的关系越来越低，而是与被评选者的社会关系，师徒门派，利益交换等强相关。也就是论文对于科研人员的价值也在降低。对于申请教育机会，由于越来越多论文量，教授在筛选学生时往往更希望了解学生的工作能力而不强调本身发表的论文数。对于声誉，尤其是在我国，用于利益交换的声誉往往是和“帽子”本身相关的，一个高引用学者依然可能默默

无闻。因而我们可以看出，“论文”的价值在不断下降，也就是说，其逐渐在变成一种没有流通性的、空中楼阁的“虚拟货币”。这种货币因为丧失了流通性也就是交换利益和回报的能力，那么就是没有“物质的”价值的。那么，如果我们投入越来越大量的资源人力去检查是否有人偷窃了这种越来越没有流通价值的物品，是否是一种资源的浪费呢？就像如果我们对一间办公室的所有人进行隔离审讯，目的只是为了找出谁从一个没来的人工位的纸抽上抽走了几张卫生纸，一定会被认为是小题大做。

我们继续讨论“一致的惩罚”。不同人从学术诚信审查中获得的惩罚是否是相当的？很遗憾，也不是。目前这种审查其实是一种“只许州官放火”的状态，原因是：处于优势地位的人的利益交换与学术成果的关联性比较低，而地位较低的人的利益交换与学术成果的关联性非常高。举一个例子来说，最常被发现的一个学术造假就是将他人的实验数据进行“二次创作”或者直接伪造实验的图像和结果。这种行为被发现之后往往会造成论文被撤稿。那么对于一位院士来说，这种行为并不会对其造成实质性的影响。近年来有相当数量的中国科学院或科学院院士被发现学术造假，但是一般的处理方法仅仅是发表一个“图片误用”的声明，并无视社会舆论即可，从未有任何一位院士因为学术不端而被吊销头衔或者获得任何附加的惩罚。同时，在双轨制的背景下，院士头衔的利益交换能力并不会因为其学术能力被发现有水分而下降，因为头衔的流通价值实际上并不是作为交换物本身，而是作为一种其他资源交换的担保，和学术能力是没有什么关系的。而对于一位正在非升即走考察期的助理教授，退稿的学术污点会导致其转正的机会严重降低，还可能会由于学校希望保全名声而失去教职。对于一位研究生而言，一篇论文的撤稿则可能会导致延毕，甚至是无法获得学位，对其个人生涯造成极其严重的影响。相同的量的惩罚内容对于一个院士（“学阀”），一个教授，一个学生造成的影响是完全不同的，因此就造成了“只许州官放火”的现象。更深远的讲，由于科学界的“金字塔结构”，这种惩罚效果的不平衡就会导致“劣币驱逐良币”，从而使得地位较高者的资源越发集中，造假的成本越发低廉，是一种学术资源的兼并。因此“惩罚”需要的不是平等，而是“公平”，也就是当事人需要受到与之相适应的惩罚。更不用说一些“程序无问题”的学术不端，比如抢夺学生的一作，或者审稿人拒稿后洗稿自投。

另外，从研究者的视角来看，每一个学术成果是有时间和精力成本的，如果在某种社会背景下，一个普通的研究者诚实的完成一个缺乏回报的科学研究造成的损失已经超过学术造假受到惩罚的损失，那么我们的学术诚信审查是否还能认为一种公平正义？

综上所述，我们遗憾的发现，我们的学术诚信审查正在越来越成为一种资源的浪费。

5 学术职业化背景下的“灌水”

近年来有很多关于学术界的批评聚焦于学术研究的灌水现象，即“正确的”，“诚实的”废话。从传统的科学精神来看，这些研究当然也是不合格的，因为其研究的原初目的往往不是探究未知扩展未知领域，而是为了研究本身而研究。

从成果本身来看，这类研究本身确实是“几乎没有价值的”。首先他们往往缺乏方法论上的创新，而

是凭空制造了一些“不同”，本身无法扩展人类知识的边界（相比于常见的气泡比喻，人类知识可能更近似于一块海绵，同样是获得“新的事实”，对于科学体系真正有价值的研究相当于海绵的外表面的扩展，而灌水相当于是在填补内部的空泡，而这种空泡的多数部分其实是没有必要填补的，有无限的排列组合也是无法填补完的）。另外由于缺少计划和这些故意的“无中生有”的差异性，大量的这种低价值的研究并不能堆砌成一个有价值的成果，是无法累加的。（与之相对的是诸如人类基因组计划，每一次测序当然都没有方法论的创新，可以说是“没什么价值的”，但由于集中的统筹规划，一致的方法和结果管理，这些“低价值研究”最后会形成一个高价值的数据库，并用于产生更加高阶的结论或者支持进一步的研究，但这种松散的“灌水”缺乏这种一致性和集中性，也就没有办法积累价值）。

但是，我们是否应该打击这种行为？下面我们将提供一个新的视角。

在当今的社会，有大量的学术研究已经高度职业化了。也就是研究者其实是类似于一种雇员，而科研机构则是以类似于公司的形式运转，也就是相比于研究和探索，学术成果更多是被“生产”出来。这些科研成果作为一种产品，用以交换更多的经费投入，申请更多的基金以维持科研机构的自身运转和扩张。类似的，某种“学术资本”也就在剥削一线科研人员的“剩余价值”。在这个视角下，这种“灌水”就不再是值得批评的，毕竟不会有任何人批评工厂生产产品是“灌水”。

那么我们能否退回 18、19 世纪及以前的形式，直接取消几乎所有科学研究的获利形式，让科学研究回归兴趣驱动呢？是不能的。

原因之一是“开倒车”绝对不是简单的回到过去。经过上百年的发展，学术界早已成为了一个与经济文化教育紧密关联的巨大的领域，其中的从业者体量极大。以我国为例，根据中国科技部 2022 年的统计数据，我国的科研人员全时当量为 635.4 万人年[5]，可以说是“百万漕工衣食所系”，而这其中的绝大多数从事的是这类的研究，如果取消这一部分，将会造成严重的社会问题。

更重要的原因是，科学界是一种对工业界的制衡，对于人类社会的进步发展具有重要意义。这些研究者虽然进行的是“灌水”研究，但实际上是具有一定的解决现实的问题的能力和相应的方法论的，是实打实的人才。而科学界与教育界相互融合，最大程度的吸纳了“人才”这种最重要之一的生产要素。科学界实际上是以人才生产要素作为筹码，掌握了一定的与工业界的谈判能力，从而支撑起了科学事业的发展。也就是说科学界通过大体量的“低价值”甚至是“无价值”的研究吸纳人才，并用之与工业界交换，从而换取来自工业界的资本，而这部分资本就可以一部分被分流到高价值研究中，从而促进人类科学技术的发展。通俗地说就是，大量的低价值研究“养活”了高价值研究。可以设想，如果科学界的制衡不存在，人才就失去了选择的空间，都必须在工业界的模式中生存。那么高风险、高成本的探索性科学研究是不符合市场规律的，也就无法获得资本，人类的科学研究就会陷入严重的停滞。对科学人才的压榨也会增加。同时，相比于遵循市场规律的工业界，科学界更容易接受政府的控制，也就可以作为政府平衡和调控工业界的一种中间手段。

从我们的新视角可以看出，“灌水”并不是一种简单的缺乏“科学精神”的行为，而是与复杂的社会因素相交织。相比于对“灌水”进行批评和打击，一个更有效和更现实的做法是做好这类学术成果和探索性的学术成果的隔离，避免因其巨大体量对探索性研究构建的知识体系造成的冲击和混淆。同时这种隔离一定程度会影响科学界的上述的谈判力，因此更多是在“台面之下”。因而从事探索性的研究的研究者自身也应当有一定的驱动力，去建立更有效的检索机制，保护自身构建的知识体系的可持续性。

6 结论

本文对三种新背景下的科学诚信进行了评估和分析。在后人工智能时代的背景下，分析了人工智能生成内容检查的“不可为”。在学术成果贬值的背景下，我们得出了科学诚信审查并没有起到改善科学界的环境，而是起到了破坏的作用。在科学研究职业化的背景下，我们提供了一个新视角，即在现有社会环境中“灌水”是一种社会化的，不可轻易铲除的集体行为。

本文并不是鼓励学术不诚信和“灌水”的行为，也不是认为这些行为是正当的，更不是宣称要放弃对于科学诚信的检查制度，而是说明了在现有社会环境不变的前提下，加大科学诚信审查和打击“灌水”的投入只能是一种资源浪费或者装点门面。

科学诚信审查不仅要有可行的技术前提，更重要的是相适应的社会土壤。因此，应该有与其相适应的制度改变与分配方式改变走在前面。

参考文献：

[1] Stackoverflow. Policy: Generative AI (e.g., ChatGPT) is banned [EB/OL]. [2025-02-04]. https://meta.stackoverflow.com/questions/421831/policy-generative-ai-e-g-chatgpt-is-banned.

[2] Clarification on Large Language Model Policy [EB/OL]. [2025-02-04]. https://icml.cc/Conferences/2023/llm-policy.

[3]2024 Author Guidelines [EB/OL]. [2025-02-04]. https://cvpr.thecvf.com/Conferences/2024/AuthorGuidelines.

[4] CVPR 2025 Changes [EB/OL]. [2025-02-04]. https://cvpr.thecvf.com/Conferences/2025/CVPRChanges.

[5] 中国科学技术部. 中国科技人才发展报告. 2022[M]. 北京: 科学技术文献出版社, 2023.

[6] Spence, Michael. Job Market Signaling[J]. The Quarterly Journal of Economics, 1973, 87(3): 355－374.