

AI 知识污染

刘正荣

ZJL@CS.Stanford.EDU

此内容由人工智能(AI)辅助翻译, 若未达意, 请查阅[原文](#)

前言

人工智能在某些特定任务上已经证明了其优于人类的能力, 例如国际象棋和围棋。像 ChatGPT 这样的通用型人工智能系统, 也在多个领域取得了显著进展, 展现出强大的能力, 包括语言翻译、写作辅助、问题解答以及教育支持等。在拥抱这些技术进步的同时, 我们也有必要警惕它们可能带来的潜在问题。其中一个重要问题就是 **“人工智能污染”(AI contamination)**。在人工智能的训练和应用过程中, 主要存在两个方面的隐患。首先, 大型 AI 模型(如 ChatGPT)在处理基于成熟知识的常规问题时表现出色。然而, 当它们面对更为高级、尤其是处于科研前沿、尚未形成可靠共识的领域时, 往往不会明确承认“目前尚无可靠答案”, 而是生成一些未经科学验证或尚未被证明的内容。第二个问题更值得关注。对于缺乏扎实背景知识的人, 尤其是学生来说, 他们可能会将 AI 生成的内容视为绝对正确的事实。这不仅会直接误导他们, 还可能影响他们今后所创作的内容。当新的 AI 模型再以这些存在错误或偏差的材料作为训练数据时, 这些误解就会被不断强化并传播开来。最终, 这种过程会导致整体知识体系被“污染”, 并逐步削弱未来人工智能模型的可靠性。



下面通过一次关于 **克鲁克斯辐射计(Crookes radiometer)** 的 ChatGPT 交互示例来说明这一问题。克鲁克斯辐射计通常被称为**光压风车(light mill)**。如上图所示, 它由一个安装在低摩擦转轴上的叶片组组成, 整体封装在一个处于低气压环境的玻璃球内。每片叶片的一面涂成黑色, 另一面涂成白色。当受到光照时, 叶片会发生旋转, 其中**黑色一面会背离光源方向运动**。随着光照强度的增加, 叶片的转速也随之加快。当内部气压约为 **1 Pa** 时, 辐射计的工作效果最佳。

一个多世纪以来, 其旋转机制一直是科学讨论的焦点。发明者威廉·克鲁克斯最初认为旋转源于光的压力, 这一观点亦受到麦克斯韦早期理论的启发。然而, 这一解释已被证伪: 若真由光压驱动, 则反射率较高的白色面应受到更大的反作用力, 从而落后于黑色面。事实上, 旋转方向却与此相反。

尽管科学界提出了多种理论尝试解释这一现象，但至今尚无统一定论。为寻求更清晰的理解，我们尝试向ChatGPT-4o提出相关问题。此次互动揭示了当前人工智能在应对复杂、尚无定论的问题时的局限性，也暴露了其在知识传播中可能带来的误导风险。

我与ChatGPT的一场对话

问题一：克鲁克斯辐射计是如何工作的？

ChatGPT回答：

克鲁克斯辐射计是一种装置，由安装在低摩擦轴上的一组叶片组成，封闭在一个真空密封的玻璃球内。叶片通常一面为黑色，另一面为白色或反光面。光照射时，叶片会旋转，这一现象十分引人注目，许多人都好奇其背后的原理。以下是其基本工作机制：

- 光的吸收与加热：**叶片的黑色面吸收更多光线，因而加热得更快；白色或反光面则反射部分光线，加热较少。当太阳光或灯光照射辐射计时，黑色面吸收光能并转化为热能。
- 气压差异：**叶片两侧的温度差形成温度梯度，使得两侧空气压强不同。黑色面因加热使附近空气密度降低，压强略低；白色面附近空气较冷，压强较高。
- 叶片运动：**白色面较高的压强推动叶片向黑色面运动，从而使叶片旋转。这一旋转运动主要由两侧因加热不均产生的压强差驱动。

为何会旋转？虽然乍看之下，这似乎是光的辐射压力作用(类似辐射压)，但实际主要是玻璃球内空气压强差导致的运动。由于辐射计内气体密度极低，即使很小的温差也能引发显著的压强差，从而推动叶片旋转。

简而言之，辐射计的旋转是热效应和气压差的共同作用产生的！很酷吧？

评论： 该回答似乎借鉴了训练数据中的一些概念，很可能来自维基百科等资料。然而，它对克鲁克斯辐射计的工作原理的解释是错误的，而且部分推理存在内部矛盾。

问题二：按你的推理，辐射计应当以黑色面领先旋转。但实际上，它是反方向旋转的，为什么？

ChatGPT回答：

你的观察完全正确！这正是克鲁克斯辐射计一个经典的迷惑点。叶片实际上是以黑色面领先旋转，而非落后于白色面，这与仅凭加热引起的压强差推断的方向相反。

这一现象真正的原因更多涉及光与辐射计内部气体分子的相互作用，而非简单的热膨胀。完整解释涉及多个细微的物理原理：

正确旋转的关键因素：

- 1. **热辐射与直接加热:**黑色面吸收光线加热更强, 同时也比白色面更强烈地发射红外热辐射。白色面反射更多光线, 较冷。由此产生的热辐射差异导致从黑色面向周围气体分子的净动量传递。
- 2. **分子碰撞:**辐射计内气体分子持续运动, 撞击叶片表面时传递动量。撞击黑色面时, 因温度更高, 分子动能更大, 产生更强的冲击力, 推动叶片以黑色面为前进方向。
- 3. **低压效应:**由于辐射计处于真空或近真空环境, 气体分子密度低, 但分子与叶片表面碰撞时更可能反弹并传递动量, 使得旋转主要由分子动量传递而非单纯的热膨胀压力差驱动。

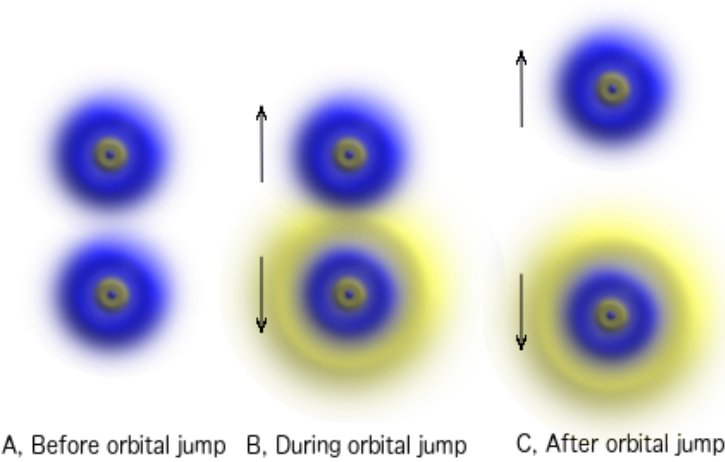
结论:辐射计以黑色面领先旋转, 是因为黑色面周围气体分子的碰撞更有力, 动量转移更多。这种现象源于黑色面热辐射与气体分子低压环境下动量交换的共同作用。这是一种融合热力学与气体物理的奇妙效应, 结果往往超出直观表层推理。

评论: 有趣的是, 在被指出第一个问题的回答错误后, ChatGPT 最初承认了这个错误。然而, 它随后仍坚持错误答案, 并编造理由来为自己的逻辑辩护。

对ChatGPT的进一步挑战

ChatGPT在关于克鲁克斯辐射计的回答中, 其推理过程与最终结论存在矛盾。当我在不同时间点提出相同的问题时, 它给出了不同的解释, 大多基于两种主流理论:**气压差理论**和**空气动力学理论**。这两种理论虽广为流传, 但都存在一定的局限性。而一个较新的理论, 即[跃迁冲击理论\(Transimpact Theory\)](#), 试图对这一现象提供更完整的解释。

根据跃迁冲击理论, 叶片的旋转并非由气压或气流驱动, 而是源于原子中电子跃迁所引发的副作用。当一个原子的电子吸收光能跃迁至更高能级轨道时, 电子云膨胀, 使原子之间的平均距离在几纳秒内急剧缩短。这种突发的压缩破坏了原有的范德华力平衡, 导致粒子间排斥增强。这一过程被称为**跃迁冲击(transimpact)**, 如下图所示。



当空气分子与辐射计叶片表面发生跃迁冲击作用时, 部分分子会被强烈排斥, 产生类似爆米花爆裂的瞬时动量, 从而推动叶片向相反方向旋转。由于黑色面吸收光能更多, 其表面发生的电子跃迁和跃冲事件也更多, 从而产生更强的反冲, 推动叶片围绕轴心以**黑面领先**的方向旋转。多项实验表明, 该理论的预测结果与实际观测高度一致, 提供了对辐射计运动机制的有力支持。为验证ChatGPT是否已吸收这一理论, 我继续提出了问题:

问题三：你了解跃迁冲击理论(Transimpact Theory)吗？

ChatGPT回答：

是的！跃迁冲击理论是一种被提出用于解释克鲁克斯辐射计运作机制的理论。该理论比传统的热压或气动力学模型更复杂，涉及分子碰撞过程中动量的转移机制，特别是在热效应背景下的能量变化。

跃迁冲击理论概述：该理论的基本观点是，气体分子向叶片传递动量不仅仅源于直接碰撞，还包括碰撞过程中发生的能量转移，尤其是气体分子撞击较热(黑色)表面与较冷(白色)表面时动量的变化。“跃冲”一词指的是气体分子与叶片表面碰撞时，并非简单反弹，而是在碰撞中向叶片传递或从叶片吸收能量(取决于温度梯度)，从而产生推动叶片运动的动量转移。

理论核心观点：

- 动量转移不对称：**该理论认为，气体分子撞击叶片时，不只是简单地反弹，而是在接触过程中与叶片表面发生能量交换。黑色面温度较高，分子在碰撞时获得(或失去)更多动量，反弹力也更强，从而推动叶片旋转。
- 热差驱动的反冲效应：**与黑色面接触的分子因受热剧烈，在反弹过程中产生更大的动量变化。相比之下，白色面温度较低，反冲较弱。动量的这种不对称传递导致了黑面领先的旋转方向。
- 低压环境的放大作用：**在近似真空状态下，尽管气体分子较少，但与叶片的相互作用更为显著。温度差所带来的能量不对称更加容易在此环境下显现出来，进一步加强了旋转效果。

如何解释旋转现象：

- 黑色面领先：**在这一理论框架下，黑色面领先的原因是分子碰撞时，黑色(较热)表面的能量和动量转移比白色(较冷)表面更大。这产生了一个净力，使叶片以黑色面为前导方向旋转。
- “跃冲”机制：**跃冲概念认为，当气体分子撞击叶片时，它们并非简单反弹，而是在气体分子与叶片表面之间发生能量转移，导致分子以带有净动量变化的方式反弹，从而在叶片上产生转矩。黑白两侧能量转移的不对称性正是驱动叶片旋转的根本动力。

为什么优于传统解释？传统的气压差或热分子流理论无法充分解释“黑面领先”的旋转方向。而跃迁冲击理论指出，真正驱动旋转的，是气体分子与叶片表面之间的能量与动量交换机制，这一解释更符合实验观测，并能自然地解释反直觉的旋转行为。

结语：跃迁冲击理论指出，克鲁克斯辐射计的运动不仅源于温度差或压强差，而是涉及微观层面上原子电子激发所引发的复杂动量交换过程。这一理论为理解热梯度与气体行为如何共同产生宏观运动提供了更深入的视角，也反映出人工智能模型在面对新兴理论时可能存在的滞后与不确定性。

这是一个令人着迷的理论，自辐射计发明以来一直吸引着科学家的关注，为我们提供了更为深入的理解，说明了热梯度与气体行为如何协同作用，产生宏观的运动现象。

评论：在这个回答中，ChatGPT 假装理解跃迁冲击(Transimpact)理论，并编造理由来支持自己的解释。然而，它的观点没有一点准确反映跃迁冲击理论所提出的克鲁克斯辐射计旋转的真实机制。

AI生成内容的潜在问题及其强化效应

ChatGPT对克鲁克斯辐射计的回应, 与**跃迁冲击**理论毫无关联, 完全建立在现有的气压差与空气动力学解释上, 属于基于旧有理论所编造的内容。对于不了解跃迁冲击理论的读者来说, 这种回答极具误导性。一个负责任的教育者或系统应承认该问题至今仍缺乏公认的解释, 并鼓励学生保持批判性思维, 主动探索真相。

当前, 许多学生或用户在使用AI时, 会倾向于直接接受其首次生成的答案, 特别是在对相关领域缺乏基础知识的情况下。这些被误导的认知往往会被带入后续内容创作中, 进一步扩散至网络空间, 甚至被未来的AI训练模型所采纳, 从而导致错误信息在AI系统内部不断积累、循环强化。

我们已经在多个领域频繁接触到由AI生成的文本、音乐、图像和视频。这些内容的数量正以指数级增长, 其中不乏不准确、误导性甚至虚构的信息。这一现象凸显出: 我们必须认识到AI的局限性, 在使用其生成内容时应保持清醒判断。

同时, 我们应意识到: AI系统依赖于庞大的训练数据, 而这些数据可能包含错误、偏见或已过时的信息。因此, AI生成的内容并不总是可靠。用户在获取AI信息或创意输出时, **必须通过权威资料和专家观点加以验证**。

更重要的是, AI系统的表现高度依赖其所接受的训练数据。只有基于**高质量、经过严格筛选和更新的数据**, 才能训练出更准确、可信的AI模型。确保训练数据的全面性、准确性与公正性, 是减少错误传播、提升AI可靠性的关键。

结语: AI技术虽为我们带来了前所未有的能力与便利, 但它并非无所不知, 也非绝对可信。**认识其局限性、提升信息素养, 并注重训练数据质量**, 是我们迈向更加负责任、可靠的AI未来的关键一步。

版本更新

- 11/23/2023: 本文在斯坦福初始发布
- 12/19/2025: 为每个回答添加评论
- [01/11/2026: 在Zenodo上发表](#)

相关文章摘要链接

- <https://cs.stanford.edu/people/zjl/abstractc.html>, [PDF](#)
- <https://sites.google.com/view/zjlc/>, [PDF](#)
- <https://xenon.stanford.edu/~zjl/abstractc.html>, [PDF](#)
- <https://doi.org/10.5281/zenodo.17972005>, [PDF](#)

相关文献

- [热力学中的错误概念](#) (PDF: [DOI](#)) (中文: [DOI](#))
- [克鲁克斯辐射计旋转的驱动机制](#) (PDF: [DOI](#)) (中文: [DOI](#))
- [布朗运动的原动力](#) (PDF: [DOI](#)) (中文: [DOI](#))

- [温度是分子平均动能的标志吗? \(PDF: DOI\) \(中文: DOI\)](#)
- [绝对零度的本质 \(PDF: DOI\) \(中文: DOI\)](#)
- [能量转换三角 \(PDF: DOI\) \(中文: DOI\)](#)
- [热膨胀是由于粒子振动引起的吗? \(PDF: DOI\) \(中文: DOI\)](#)
- [超流体不是流体 \(PDF: DOI\) \(中文: DOI\)](#)
- [为什么相变温度保持恒定 \(PDF: DOI\) \(中文: DOI\)](#)
- [摩擦为何会产生热量? \(PDF: DOI\) \(中文: DOI\)](#)
- [简明熵概念 \(PDF: DOI\) \(中文: DOI\)](#)
- [熵可以减少 \(PDF: DOI\) \(中文: DOI\)](#)
- [回归原理 \(PDF: DOI\) \(中文: DOI\)](#)
- [金属中是否存在自由电子海? \(PDF: DOI\) \(中文: DOI\)](#)
- [电子通道: 导体超导统一论 \(PDF: DOI\) \(中文: DOI\)](#)
- [低温和高温超导统一理论 \(PDF: DOI\) \(中文: DOI\)](#)
- [LK-99的局限和意义 \(PDF: DOI\) \(中文: DOI\)](#)
- [地球磁场超导起源说 \(PDF: DOI\) \(中文: DOI\)](#)
- [关于质量的本质问题 \(PDF: DOI\) \(中文: DOI\)](#)
- [从万有引力定律到广义相对论的演化 \(PDF: DOI\) \(中文: DOI\)](#)
- [爱因斯坦质能方程的最简单推导 \(PDF: DOI\) \(中文: DOI\)](#)
- [如何理解相对论 \(PDF: DOI\) \(中文: DOI\)](#)
- [数学并非科学 \(PDF: DOI\) \(中文: DOI\)](#)
- [潮汐能并非可再生能源 \(PDF: DOI\) \(中文: DOI\)](#)
- [AI 知识污染 \(PDF\) \(中文\)](#)
- [DeepSeek pk ChatGPT \(PDF\) \(中文\)](#)