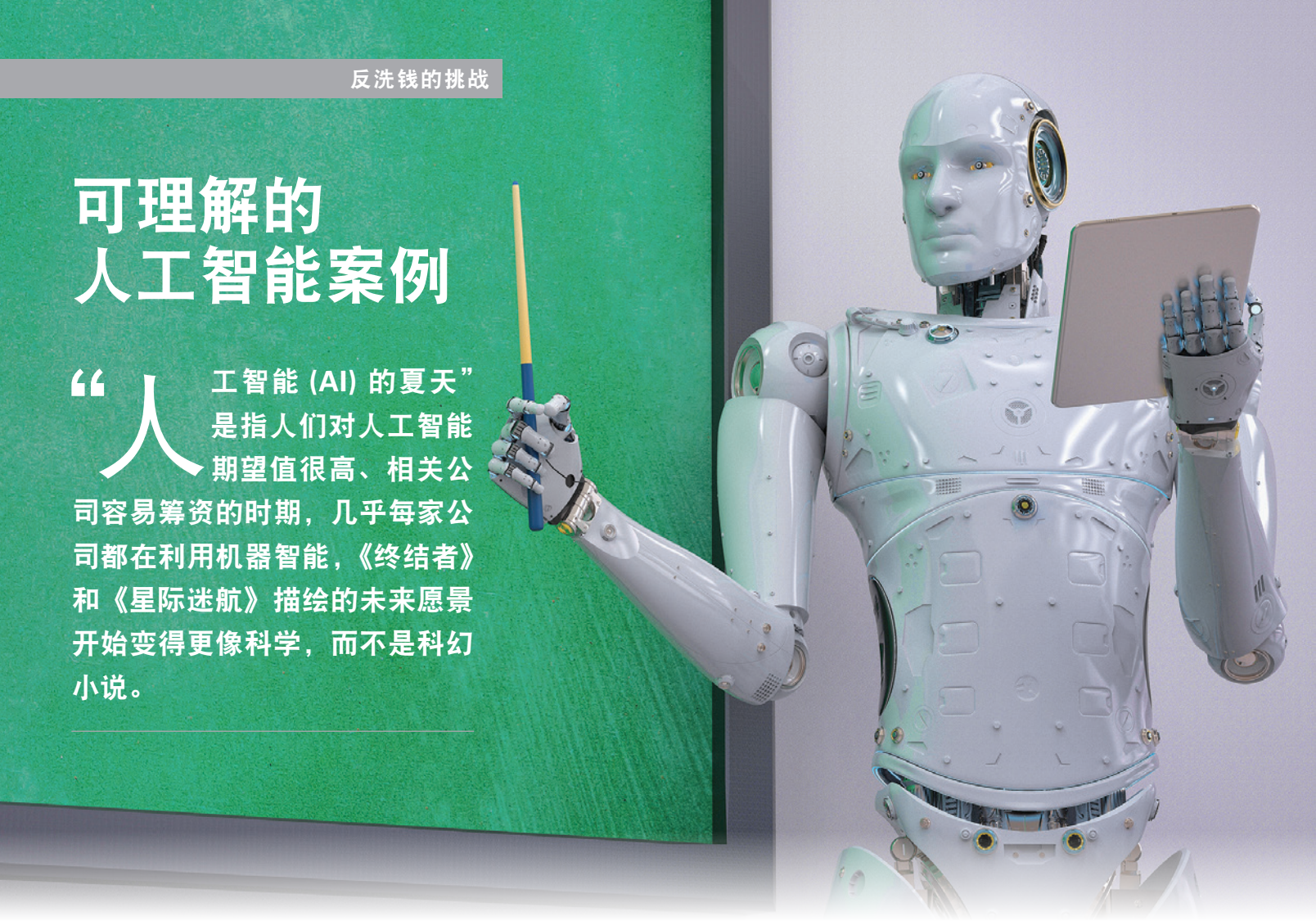
# 可理解的人工智能案例

“人工智能 (AI) 的夏天”是指人们对人工智能期望值很高、相关公司容易筹资的时期，几乎每家公司都在利用机器智能，《终结者》和《星际迷航》描绘的未来愿景开始变得更像科学，而不是科幻小说。



正如人们所感知到的，目前世界正在经历人工智能的夏天。几十年前马文·明斯基 (Marvin Minsky) 等先锋人物提出的美妙预言，随着科技的进步逐渐变为现实。当梦想的魅力再也无法弥补技术局限性之时，过去的“夏天”结束了。但这个“夏天”则完全不同。与前两次人工智能热潮不同的是，这次人工智能够完成具体的工作。人工智能可以驾驶汽车、创作歌曲、帮助整理全球信息。人工智能真实而强大，因此，人们有理由相信，期望不会在遭遇现实之后落空。

另一个更有趣的问题已经让人生出几分忧虑，但与人工智能的能力（或缺乏能力）无关。相反，它与人工智能的可理解性有关，即“黑箱”问题。如果反洗钱和防范金融犯罪技术要充分利用最新的人工智能技术，就必须理解和解决它的可理解性问题。

## 你永远无法破解魔术表演

机器学习是很多被称为“人工智能”技术的创新基础。

在机器学习出现之前，技术人员和领域专家直接为人工智能系统编程。要让人工智能系统执行任何任务，都必须先编制正式、明确的规则，而这些规则必须由人编写。虽然这些规则完全可以理解且有理有据，每项业务背后都有专家提供的专业知识支持，但不幸的是，这些规则建立的系统存在缺陷。

这些系统非常脆弱（对大多数任务而言，系统只能处理一小部分允许输入值，否则就会中断），并且应用范围极为有限（脆弱性缩小了应用范围）。简而言之，人们甚至专家都很难制定足够复杂的规则，来应对纷繁复杂、不可预测的输入类别。

这就是机器学习扭转局面的地方。使用机器学习的应用程序，其运行不是依靠明确清晰的输入输出链接库，而是构建于模型之上，而这些模型是所谓的训练过程的结果。训练过程将向系统提供大量输入 / 输出匹配，然后要求系统构造模型来描述它们之间的关系。在这个过程中，机器能够自行推断出变量之间的核心关系。

好消息是，这种方法允许自动化系统处理更具多样性的输入值。坏消息是，由于这种映射关系完全依靠人工智能生成，不具备人类推理所具备的直观性，而且机器学

习技术越先进（如深度学习），映射关系就越不直观。因此，除最资深的领域专家以外，最复杂的人工智能应用程序背后的逻辑对所有人而言都是完全不透明的。

这显然是一个严重的问题，尤其是在合规方面。毕竟，对于无法理解的信息，监管机构要如何监管呢？

## 与历史背道而驰

Matthew Van Buskirk 在《错位：为什么监管与创新相互矛盾》[1] 一书中写道“我们的监管体系是建立在模拟信号时代”，他继续写道：

> “许多现代监管体系的法律是 40 到 50 年前制定的，远在计算机成为基本商业资源之前。”正因如此，监管技术融合的法规植根于一个早已不复存在的世界：在那里，人类（且只有人类）负责填写表格和整理文件夹。”

当前监管政策所依据的假设，进一步加剧了“黑箱”问题带来的困扰。当前人工智能处理的许多基本分析任务，过去全部由人类负责，而当前的监管法规是在过去的时代制定的，因此，这些政策架构的预期更适合人类而非机器。

与当前人工智能革命相对立的另一个政策历史产物是：注重过程而非结果。[2] 直到最近，还很难根据结果来衡量特定政策的有效性。[3] 过去，得出这些结论的数据和计算能力根本不存在。[4] 由于这些限制，监管当局将合规设计为遵循已批准的流程，而非实现期望的结果，而最初开发那些理想流程时（读者可能已经猜到了）并没有考虑人工智能。[5]

监管传统对人工智能融合带来的挑战还有更深层的内涵，但核心问题很简单：目前，人工智能的内部运作机制必须是可理解的，才能实现合规性。

## 缩小差异，跨越障碍

尽管监管机构正积极努力理解和适应新的技术体系，但业界不能等到合规监管要求转变才解决黑箱问题。行业需要尽力将人工智能的作用发挥出来，让人工智能变得可理解。

虽然实际上有很多策略可以实现可理解的人工智能，但对这些备选方案进行有意义的探索，将涉及太多的树（个例），而缺少足够的森林（集合）。除了提供列表外，还有一种特别有趣的方法：学生 / 教师方法。

在《透明度还不够》一书中，计算数学家 Gurjeet Singh 探讨了如何使用学生 / 教师方法来满足合规性：

> “这种方法通过复杂的人工智能应用程序来解码数据集。它生成了关键洞察和关系；它创建了从输入到输出的基本映射。一旦这个阶段完成，就会部署另一个更简单的应用程序。但是，这个应用程序不是从数据中学习，而是从另一个模型中学习，生成另一个模型创建的输入 / 输出关系的简化版本。换句话说，更简单的模型提取了另一个模型发现的‘规则’，称为‘规则提取’，该模型更容易理解、更合理，可实际应用于生产技术中。”[6]

过度的（或巨大的）复杂性是黑箱问题的核心，学生 / 教师方法能够通过降低生产模型的复杂性而从根本上解决问题。这种方法通过使用两个独立的应用程序，改进高级人工智能的问题解决能力，同时没有不透明性，这是一举两得的解决方案，值得有关人员的关注。

## 理解行业所面临的风险

作为一种公众现象，人工智能曾经历过类似的热潮——但也经历过明显的低谷。然而，这次人们明显感到人工智能的时代真正来临。从数十亿美元的商业决策到杂货店采购，这项技术对社会产生了深远的影响。尽管如此，这项技术也必须满足合规标准。

在反洗钱和防范金融犯罪领域尤其如此。在这些受到严格监管的领域，学生 / 教师方法等方法将成为决定技术是否可部署，甚至是决定行业飞速发展还是止步不前的关键。 A

*Alex Detmering，Basis Technology 市场战略主管，美国马萨诸塞州剑桥市，adetmering@basistech.com*

[1] Matthew Van Buskirk，“The Mismatch：Why Regulation & Innovation Clash”（《错位：为什么监管与创新相互矛盾》），第 36 页。
[2] 同上。
[3] 同上。
[4] 同上。
[5] 同上。
[6] Gurjeet Singh，“Transparency Isn't Enough”（《透明度还不够》），49-50 页。