人类水平的智力还是动物的能力？[[1]](" \l "_ftn1" \o ")

阿德南·达威奇

计算机科学系

加利福尼亚大学，洛杉矶darwiche@cs.ucla.edu

2017年7月8日

鹰和蛇的视觉系统比我们在实验室里制造的任何东西都要好，但是蛇和鹰却不能制造眼镜、望远镜或显微镜。

朱迪亚珍珠[2]

# 1简介

最近，神经网络在语音识别、视觉和自主导航等应用领域的成功，引起了人工智能（AI）社区成员和广大公众的极大兴奋。在相对较短的时间内，通过科学时钟，我们成功地实现了一些几十年来一直困扰我们的任务的自动化，并使用了人工智能研究中的一种更为经典的技术。

对这些成就的胜利使一些人把这些任务的自动化描述为达到了人类的智能水平。这一观点最初在学术界有所暗示，但现在已经得到了更广泛的发展，并导致了一些影响。例如，一种趋势正在出现，即机器学习研究正在被简化为神经网络研究，其新获得的标签是“深度学习”。这种看法也使一些人质疑继续投资于其他机器学习方法，甚至是人工智能的主流领域是否明智，如知识表示、符号推理和规划。公开场合对人工智能的一些报道，特别是一些看得见的人物的评论，导致了这种兴奋与对人工智能在未来可能给我们带来什么（即世界末日场景）的恐惧交织在一起

人工智能历史上的这一转变给更广泛的人工智能社区的研究人员造成了一个两难的局面。一方面，人们不得不对我们用神经网络所能完成的工作留下深刻印象并加以享受。另一方面，主流科学直觉阻碍了人们接受这样一种观点：一种不需要任何明确建模或复杂推理的方法，就足以再现人类的水平

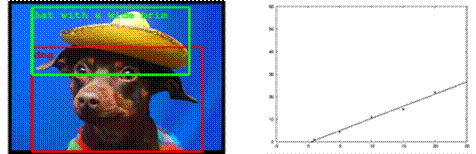


图1：左图：对象识别和定位（ImageNet）。右：将函数拟合到数据。

智慧。观察到最近的发展并没有最终形成一个明确特征和深刻的科学发现，例如通常要求对人工智能课程进行大规模更新的新思维理论，这进一步加剧了这种困境。来自人工智能和计算机科学之外的学者们经常感觉到这种困境，因为他们抱怨说，他们没有得到一个智力上令人满意的答案来回答“人工智能刚刚发生了什么？”？”

解决这一困境的答案在于仔细评估我们通过深入学习所取得的成就，并确定和欣赏这一研究领域最新发展的关键科学成果。不幸的是，这在很大程度上一直缺乏。我在这里的目的是引发这样一场讨论，受到对本文所表达想法的积极和好奇的反馈的鼓励。

# 2背景

为了为这个讨论打下基础，我将首先指出两种不同的方法来处理人工智能感兴趣的问题。我将称第一种方法基于模型，第二种方法基于函数。考虑图1中的对象识别和本地化任务。为了解决这个识别任务，基于模型的方法需要一个人来表示关于狗和帽子等的知识，并涉及到用这些知识进行推理。今天这种方法的主要工具是逻辑和概率（更一般的数学建模），它可以被认为是人工智能创始人最初设想的表示和推理方法。在某种程度上，这也是科学界知情人士通常期望的方法。另一方面，基于函数的方法将此描述为一个函数拟合问题，函数输入直接来自图像像素，其输出与我们正在寻找的高级识别相对应（函数必须具有能够有效评估的形式）。目前这种方法的主要工具是神经网络。许多大学生曾在物理或化学实验室中练习过这种方法，在那里他们将简单的函数与从各种实验中收集的数据相匹配；见图1。这里的主要区别在于，我们现在使用的是具有多个输入和输出的函数；这些函数的结构可以任意复杂；我们正在处理的问题往往与认知有关，而不是说，估计工程系统中体积和压力之间的关系

过去几年人工智能的主要观察是，基于功能的方法在某些人工智能任务上可以相当有效，比基于模型的方法更有效。这不仅让主流人工智能研究人员感到惊讶，他们主要实践基于模型的方法，而且也让机器学习研究人员感到惊讶，这些研究人员主要实践基于模型的方法，其中基于函数的方法只是其中之一。[5]这有许多含义，有些是积极的，有些是令人担忧的。

积极的一面是，越来越多的任务和应用程序现在触手可及，使用的工具可以非常熟悉的人只有广泛的工程背景，特别是谁习惯于估计功能和使用他们作出预测。然而，令人关注的是，目前一方面利用、享受和欢呼这一工具，另一方面考虑它之间的不平衡。这种思考不仅对充分发挥这一工具的潜力，而且对于科学地描述和理解其潜在的影响范围也很重要。缺乏这样的描述，或者至少是试图足够积极地寻求它，是当前对人工智能进步的误解的罪魁祸首，这可能会导致我们未来的发展。

3人工智能刚刚发生了什么？

我告诉我们，通过深入的研究，我可以更深入地了解他们在解决神经网络问题方面取得的进展。这些网络是参数化的函数，具有足够的表达能力，可以捕捉输入和输出之间的任何关系，并且具有一种可以有效评估的形式。这一点我们已经知道了几十年，并且在各种人工智能教科书中都有详细说明。是什么导致了现在的事态发展？

为了阐明这个问题，让我再陈述一下我们最近的发现。也就是说，通过简单地将函数拟合到数据中，就可以在合理的程度上捕捉和再现一些通常与认知有关的看似复杂的能力，也就是说，不需要对环境进行明确的建模或进行象征性的推理。虽然这是一个了不起的发现，但它强调的问题和门槛比它突出的技术。让我解释一下。每一种行为，无论是否聪明，都可以被一个函数捕捉，该函数将输入（环境感知）映射到输出（思想或行为）。不过，两个关键问题如下。这些函数是否足够简单，以允许一个紧凑的表示，允许像神经网络一样有效地将输入映射到输出（即，无需推理）？如果答案是肯定的，我们现在是否有能力从输入输出对（即数据）中估计这些函数？

人工智能最近发生的事情有三个直接影响到这些问题的发展。第一个发展是我们改进了将函数拟合到数据中的能力，这是由于（a）大量数据的可用性；（b）我们现在拥有的计算能力的增加；以及（c）用于拟合函数的日益复杂的统计技术。第二个进展是，我们已经确定了一类实际应用程序，它们对应的函数，我们现在知道，这些函数非常简单，能够有效地评估紧凑的表示（同样，无需推理），其估计值在当前收集数据的阈值范围内，计算速度和估计技术。这包括识别和定位某些图像类别中的对象以及与自然语言和语音相关的某些任务。第三个基本上没有被注意到的进展是，我们改变了我们的成功衡量标准，大大减少了技术挑战，至少是早期人工智能研究人员所接受的，同时保持了我们利用所获得的商业结果的能力（我将在后面更详细地讨论最后一点）。

有趣的是，以上这些都不是AI本身的一个重大技术突破或里程碑，比如在80年代末建立概率作为常识推理的基础，或者五十多年前引入神经网络。（6）这些因素的组合对现实世界产生了深远的影响。各种人工智能技术（与神经网络结合）的应用和成功部署，这些技术已经投入了很长时间。

# 我不同意

在准备这篇文章的过程中，我在不同的背景下分享了上述观点。受众范围从人工智能、计算机科学，到对人工智能感兴趣的法律和公共政策研究人员。我发现令人吃惊的是对这次讨论的极大兴趣和对我所说的话的安慰，如果不是普遍同意的话。不过，我确实收到了一些“我不同意”的回答。它们都集中在与优化函数有关的最新进展上，这是神经网络成功训练的关键（例如，随机梯度下降、辍学等结果）。反对意见源于没有将这些命名为突破或里程碑（在人工智能中）。我的回答是：这些都属于我上面列出的“拟合函数的更复杂的统计技术”的范畴。接下来的问题是：与传统的人工智能技术相比，它们是统计技术有关系吗？答：承认和欣赏科学研究和进步并不重要，但重要的是要解释刚刚发生的事情，更重要的是，预测下一步可能发生的事情。

想想坐在你旁边的一个受过教育的人，人工智能研究人员，在飞机上（我经常听到）。他们发现你在做人工智能，然后问：是什么发展促成了目前人工智能的进步？你告诉他们基于功能的故事，并列出三个促成因素。他们很可能会被打动，也会在智力上得到满足。然而，如果答案是“我们刚刚发现了一种新的思维理论”，你可能不会感到惊讶，如果他们也结束了担心天网即将到来的混乱我们的生活。公众对人工智能进步及其未来的看法非常重要。对刚刚发生的事情缺乏科学、准确和大胆的观点，助长了目前的错误认识和相关的恐惧，因此，给人们留下了很大的想象空间。

这并不是说，只有一种新的思维理论，或是这种规模的进步，才能证明围绕人工智能的一些当前和合理的担忧是合理的。事实上，即使是有限的人工智能技术也能吸引可能带来各种风险的自主系统。

然而，这些担忧对我们的社会来说并不新鲜，例如，当自动驾驶仪被引入航空航天行业时的安全考虑，以及ATM机引入银行业时的工作考虑。因此，这里的标题应该是“自动化”而不是“人工智能”，因为后者只是一种碰巧用来改进和加速自动化的技术。因此，为了解决这些问题，我们应该把重点转向政策和监管方面的考虑，以应对我们社会正在开始的新的自动化水平，而不是害怕人工智能

# 5移动球门柱

现在让我谈谈目前事态发展的第三个原因，那就是我们衡量成功的方式发生了变化。这个特殊的原因是相当重要的，但基本上没有被注意到，特别是年轻的研究人员。

以机器翻译为例，它在人工智能早期受到了极大的关注。代表和推理方法被认为在这项任务上失败了，机器学习方法现在是最先进的（超越了基于函数的方法）。在人工智能的早期，衡量成功与否的标准是系统的准确率与100%之间的差距，而智能是主要的驱动应用程序（翻译不正确可能导致政治危机）。翻译系统所接受的测试将其性能与人类的翻译能力（例如，将一个句子从英语翻译成俄语，然后再翻译回英语）。今天，机器翻译几乎唯一的应用是在网络上。在这里，成功的有效衡量标准是系统的精确度离0%有多远。如果我在看一个用法语写的页面，一种我不会说的语言，我对任何能让我感觉到页面所说内容的翻译感到满意。事实上，机器翻译界理所当然地称之为主旨翻译。它可以很好地处理数据中经常出现的典型句子，但在新颖的文本中却可能失败。这仍然是非常有价值的，但它所对应的任务与早期人工智能研究人员所处理的任务有很大不同。如果接受早期翻译研究者所采用的测试，当前的翻译系统将会失败。此外，如果将这些系统集成到一个旨在模仿人类行为或能力的机器人中，这些系统将不适合，也不被认为是成功的。

最近推出的“艾伦人工智能科学挑战赛”将这一观察放在了更进一步的角度。这里的任务是理解一段陈述中学阶段科学问题的段落，然后回答一个选择题。有很多关于深度学习在不同语境下应用于文本的成功报道。但这些系统能否通过艾伦人工智能科学测试，达到合格的中学生水平？

对语音识别系统也有类似的观察。目前，这些系统的主要应用是用户界面，如自动化技术支持和软件系统的指挥（如车辆上的电话和导航系统）。你试过这些吗？他们经常失败（试着说一些不太典型的话，或者如果你有口音的话，尽量不要掩饰自己的口音）。但是当这些系统出现故障时，它们会把你送回给人工操作员，或者强迫你用经典的方法来命令软件。这还有用吗？对。它是否符合人工智能领域的期望？不。

在视觉应用方面，关于对图像进行简单的改变有时会破坏神经网络正确识别物体的能力的报告。

对图像中的对象进行某些变换或变形，可以保持人类对它们的识别能力，也会破坏网络对它们的识别能力。虽然这并没有达到早期人工智能研究人员，甚至是当代视觉研究人员的期望，但我们仍然设法从这些技术中获益，比如在智能相机自动对焦过程中识别人脸（人们通常不会变形他们的脸，但如果真的变形，那就倒霉了，这是一个没有聚焦的图像）；在网上搜索中查找包含猫的图片（如果你最终得到的是一只狗，这是可以的），并在一个自动驾驶汽车的摄像头拍摄的图像中定位周围的车辆（这些系统容易出错的漏洞在其范围和如何在政策或监管层面上处理都是一个有争议的问题）。

这些观察结果的意义来自于它们对我们预测未来的能力和我们决定投资什么研究的影响。特别是，成功地解决这些由受限商业应用程序驱动的选定任务，是否证明了对世界末日情景的担忧是合理的？这种成功是否证明了我们已经找到了一种将常识推理形式化的新方法的说法？它是否证明了我们现在可以理解语言或言语，或在人类水平上理解视觉的说法？我自己的答案是否定的。这是否证明了目前对各种机器学习和人工智能方法的态度不平衡？如果你为一家对这些应用程序感兴趣的公司工作，那么答案也许是肯定的。然而，如果你更广泛地关注科学研究，那么答案可能是否定的。

总而言之，刚刚发生在人工智能领域的事情与一个突破相去甚远，而这正是人们担心世界末日的原因所在。刚刚发生的事情是人工智能技术成功地应用于一类特定的商业应用，这在很大程度上得益于姐妹领域的发展，以及比最初寻求的更宽松的性能要求。换句话说，作为对我今天看到的标题的回应，比如“人工智能已经到来”，“我没有看到人工智能的到来”，真正出现的是大量可以从改进的人工智能技术中获益的应用程序，这些应用程序仍然达不到人工智能的雄心壮志，但已经足够商业化。这本身就是积极的，直到我们开始把它和其他东西混淆。

让我在结束这一节时强调两点。第一是再次强调先前的观察，即虽然目前的人工智能技术仍然相当有限，但它可能对自动化和社会产生重大影响（例如，工作和安全）。这反过来又要求在技术层面、政策层面和监管层面进行深刻的处理。我想强调的第二点是，“移动球门柱”是一项有利的、积极的发展，我们需要准确地意识到并更好地理解其含义。这里的一个关键发现/暗示是，一些认知任务可以在一个合理的范围内被模拟出来，而不需要像最初相信和寻求的那样理解或形式化这些认知任务（例如在一些语音和视觉应用中）。也就是说，我们在这些应用中成功地避开了某些技术挑战，而不是直接解决它们。这一观察并不是要贬低目前的成功，而是要突出它的性质，并为以下问题奠定基础：我们能朝着这个方向走多远？我稍后再讨论这个问题。

6人类水平还是动物水平？

现在让我首先谈谈引发本文标题的想法。我认为，把人类水平的智力归因于目前由神经网络完成的任务是值得怀疑的，因为这些任务几乎没有提升到许多动物所具备的能力水平。朱迪亚·珀尔（JudeaPearl）已经提到，鹰和蛇的视觉系统超过了我们今天所能建造的。猫的导航能力远远优于现有的自动导航系统，包括自动驾驶汽车。狗能识别人类的语言并作出反应，非洲灰鹦鹉能发出模仿人类语言的声音，达到令人印象深刻的程度。然而，这些动物都没有人类特有的认知能力和智力。

我对这些话的反应之一是：我不知道有什么动物会玩围棋！这是指AlphaGo系统，该系统去年曾因击败世界冠军而成为新闻，被广泛认为是一个深层神经网络，甚至被一些人工智能研究人员所认识。

事实上，我们还不知道动物能玩像围棋这样复杂的游戏。但是AlphaGo并不是一个神经网络，因为它的体系结构是建立在人工智能技术的基础上的，这些人工智能技术已经投入使用至少50年了。[10]这包括两人游戏的极小极大技术、随机搜索、自玩学习、使用评估函数来切断极大极小搜索树，以及强化学习，除了神经网络。因此，围棋玩家可以被看作是一个映射板配置的函数（输入）对于一个动作（输出），AlphaGo播放器不是使用基于函数的方法构建的（即通过从输入输出对中学习单个函数），但是只有它的一些组件是这样构建的。在这里所要解决的问题是如何分配一个计算量较小的存储量，而不是一个关于收集足够多的竞争性数据的函数。如果是这样的话，这将是非常有趣的，但是我们还不知道答案（我还应该注意到，AlphaGo系统是一个很好的例子，说明了今天通过集成基于模型和基于函数的方法可以实现的目标）。

# 7推门槛

当然，我们不能排除纯粹从数据或类似的复杂函数中构造一个有竞争力的围棋函数的可能性，即使我们现在还没有达到目前的阈值。但这就引出了一个问题：如果这是一个门槛的问题，鉴于目前的成功，为什么不把我们所有的注意力集中在进一步提高这些门槛上呢？虽然这项建议有其优点，目前似乎已被关键行业采纳，但它确实面临着来自学术和政策考虑的挑战。接下来，我将讨论学术方面的考虑，而将政策考虑留给后面的章节。

从学术的角度来看，人工智能的历史告诉我们要非常谨慎，因为我们以前也见过类似的现象。我们这些在这方面有足够长时间的人可以回忆起20世纪80年代专家系统的时代，那时，我们发现了使用规则来构建函数的方法，这些规则是通过当时所谓的“知识工程”会议设计出来的。通过这一过程产生的功能，称为专家系统或知识系统，据说在某些情况下，特别是在医疗诊断方面，其性能超过了人类专家。[11]当时创造了“知识就是力量”等术语，象征着一种兴高采烈的状态，类似于“深度学习”成为今天的象征。这个时代之后的一段时间被称为人工智能冬天，因为我们最终可以划分出一类应用程序，这些应用程序屈服于没有实现人工智能目标的系统。

虽然目前神经网络进展的衍生工具令人印象深刻，但它的持续时间还不够长，不足以让我们充分了解下面的问题。如果将基于功能的方法应用于比已有目标更广泛的新应用程序，特别是那些要求更严格的成功衡量标准的应用程序，其效果如何？这个问题有两部分。第一个问题是关于一类认知任务，这些任务的相应功能足够简单，能够有效地进行评估（如在神经网络中），并且其估计值在我们期望在10到20年内达到的当前阈值或阈值范围内？问题的第二部分暗示了这样一个事实：这些功能只是认知任务的近似值（即，它们并不总是正确的）。当目标定位于比当前目标应用程序要求更严格的成功衡量标准的认知任务时，这些近似值有多合适或可接受？

# 8成功的力量

在我评论政策考虑之前，让我强调一下科学史上反复出现的一个相关现象，人工智能也不例外。我把这叫做欺负者-*成功*一种现象，指的是将一个社区压制成主要追求目前成功的东西，而牺牲了追求更成功或更长期需要的东西

回溯历史，在专家系统时代提倡的一些观点今天可以被安全地描述为在科学上是荒谬的（例如，见脚注10）。然而，由于当时专家系统的成功感，这些观点对科学对话和指导的过程起着主导作用，导致了一个被成功社区欺负的现象。[13]我在常识推理从基于逻辑的方法向基于概率的方法的过渡过程中（1980年代后期）看到了类似的现象。当时流行的观点，比如“人们不会用概率来推理”，我相信这是有价值的，当概率方法开始解决超过10年逻辑方法的常识推理问题时，这些争论就完全被压制了。当符号逻辑几乎从人工智能课程中消失时，被成功欺负的群体在这个案例中做出了更为深远的选择。因此，那些被视为世界符号逻辑代表和推理中心的部门几乎没有开设任何逻辑课程。现在我们要付出代价了。我们没有意识到概率推理一开始就把数字归因于布尔命题，逻辑是概率推理的核心，除了最简单的形式外，我们现在得出的结论是，我们需要将概率归因于更复杂的布尔命题，甚至一阶句子。由此产生的框架被称为一阶概率模型或关系概率模型，需要符号逻辑的技巧来改进这些形式。唯一的问题是，这项技能几乎已经从人工智能社区中消失了，这导致了与早期人工智能研究人员所能完成的相比显得幼稚的治疗方法

这种现象不能归咎于任何一方。对成功者来说，过度高兴有时也会夸大成功，这是很自然的。人们预计，工业界将以重新定义就业市场和影响研究生学术兴趣的方式来开发成功。同样可以理解的是，学术界的其他成员可能会为了维持其生存而合作（赢得资助、获得论文、吸引学生）。虽然这些行为中的每一种在局部看来都是合理的，但它们的结合有时会对科学研究有害，从而导致全球范围内的非理性。除了提高人们对这种反复出现的现象的认识之外，政府和学术层面的决策者对减轻其负面影响负有特别的责任。学术界的资深成员也有责任把当前的发展放在历史的角度，让初级研究人员能够追求他们真正的学术兴趣，而不是仅仅屈服于当前的成功

# 9政策考虑

现在让我来谈谈政策上的关切，即把我们所有的注意力集中在功能上，而不是模型上。这里最大的担忧源于所学函数缺乏可解释性，以及它们的结果缺乏可解释性。如果一个医疗诊断系统建议病人做手术，病人和他们的医生需要知道原因。如果一辆自动驾驶汽车撞死了人，我们需要知道原因。如果一个面部识别系统导致了一个错误的逮捕，我们还需要知道原因。如果语音命令无意中关闭了发电系统，这也需要解释。回答为什么？问题是分配责任和责任的核心，而这恰好是法律制度的核心。模型可以用来回答这些问题，但函数不能。这让我想到了两个与我的观点相关的反应，因为第一个将帮助我进一步解释这种特殊的差异。

这是一个关于为什么一个功能不符合模型的问题。考虑一个工程系统，它允许我们把空气吹入气球，然后将位于气球顶部的杠杆升起。这个系统的输入是我们吹出的空气量（），而输出是控制杆的位置（）。我们可以学习一个函数，通过收集-对，然后估计函数=f（X），来捕捉这个系统的行为。虽然这个功能可能是我们在某些应用程序中所需要的全部功能，但它不能作为一个模型，因为它不能捕获系统机制。对机构进行建模对于解释是必不可少的（例如，为什么操纵杆位置的变化不是吹气量的线性函数？）更一般的因果推理（例如，气球被捏了怎么办？）。*十是的十是的是的*

在即将出版的《为什么之书》中，朱迪亚·珀尔进一步解释了（因果）模型和函数之间的区别，尽管他没有明确使用“功能”一词。在第一章中，他写道：“只有一种方式，一个思维实体（计算机或人类）可以计算出在多种情况下会发生什么，包括一些它以前从未经历过的场景。然后他给出了一个导航系统的例子，这个导航系统要么基于（1）地图（模型）的推理，要么（2）咨询一个只给出到达目的地（功能）的左右转弯列表的GPS系统。剩下的讨论集中在模型可以做什么，而不是函数。Pearl的论点特别集中于模型如何处理新的场景，例如遇到使函数建议无效的路障，同时指出在函数中编码这种偶然事件的组合不可能，因为它必须有一个有限的大小。[16]

有趣的是，“为什么书”的第一章从四万年前的进化论开始讨论，同时认为回答“如果呢？”？“为什么？“问题使我们成为“独一无二的人类”

# 10认知功能理论

我对基于模型和基于功能的观点的第二个反应是在西蒙斯研究所（Simons institute）的一次深度学习研讨会上（2017年3月）。研讨会的题目实际上是“表征学习”，这一术语越来越多地被深度学习研究者使用。如果你听过关于深度学习的演讲，你会注意到让这些系统工作的一个关键组成部分就是找到神经网络的正确结构。此外，这些体系结构根据任务的不同而有所不同，它们的一些组件有时被描述为正在做一些可以在直观级别上描述的事情。这里的反应是，深度学习不是学习函数（黑匣子），而是学习表示（因为体系结构不是任意的，而是由给定任务驱动的）。我不这么看。构建神经网络的结构是“功能工程”而不是“表示学习”，特别是因为结构由于与输入输出对的一致性而受到惩罚和奖励。函数工程的结果相当于限制了可以使用参数估计技术学习的函数类。这类似于限制在修复概率图形模型的拓扑之后可以学习的分布类。表征学习的实践就是识别适合特定任务的函数类的练习。

在这方面，我认为最需要的是认知功能理论。认知功能是一种捕捉与认知相关的关系的功能，例如将音频信号映射到单词，或将单词映射到某种意义。除了研究认知功能的可学习性和可接近性外，似乎还需要一份认知功能的目录，以及对其表征复杂性（表示这些功能所需的建筑的大小和性质）的研究。对于布尔函数，我们有一个很深的理论。特别是，研究人员已经根据以不同形式表示它们所需的空间（例如，CNFs、DNFs、OBDDs）对各种函数进行了编目。我们需要的是类似于实值函数的东西，这些函数是用来捕捉认知行为的。在某种意义上，我们已经有了一些关于这种理论的线索例如，研究人员似乎知道什么体系结构（又名函数类）对某些对象识别任务更有效。这需要理论上的和坚实的基础。该理论还将包括使用深度学习社区使用的估计技术（尤其是梯度下降）对函数类的可学习性的结果。有趣的是，这些结果在上面提到的表征学习研讨会（题为“深度学习的失败”的演讲）上展示了，会上展示了非常简单的函数，可以击败当前的估计技术。更有趣的是，一些人在旁听讨论中否定了这些结果的重要性，理由是所确定的功能没有实际意义（阅读“这些不是认知功能”或“我们通过学习函数的近似值已经取得了很大的进展”）。事实上，如果我有发言权的话，我会把深度学习的领域重新命名为学习近似认知功能的领域。

# 11总结性思考

本文的出发点是我和其他人对当前在人工智能今天被陷害和感知。如果没有对近期成就的原因和影响进行学术讨论，也没有对所取得的成果有一个正确的看法，就可能误导年轻一代的研究人员或在学术、工业和政府各级错误分配资源，从而阻碍进一步的进展。还有一种说法是误导公众，他们对人工智能及其含义产生了极大的兴趣。目前公众对人工智能奇点（又称超级智能）的负面讨论，只能归咎于缺乏准确的框架和对最近进展的描述。由于几乎每个人都对新的发展过于兴奋或不知所措，大量的学术讨论和思考都不见了。

我有幸在20世纪80年代中后期开始了我在人工智能领域的研究生涯，当时正值该领域的一次重大危机（这段时期的特点是无能力而非能力）。当我坐在斯坦福大学的课堂上，目睹了人工智能研究人员如何被人类常规执行的一些简单任务严重挑战时，我感到很沮丧。现在回想起来，我现在意识到这样的危机是如何促进科学发现的，因为它们促进了学术思考，增强了研究人员的能力，并为深刻的科学贡献创造了基础。[18]另一方面，我今天被提醒到，成就的时代是如何通过改变学术兴趣来减缓科学进步的，资源和脑力太过注重开发刚刚发现的东西，而牺牲了对这些发现的理解，并为它们的实际应用被限制或耗尽的时刻做好准备。这种准备有很多方面。对于深度学习社区来说，最重要的可能是从“看我们还能做什么”模式过渡到“看看你还能做什么”模式。这不仅是一个邀请，以接触和授权更广泛的人工智能社区。这也是一个挑战，因为这种转变不仅是一种态度的作用，而且是一种能力，以使社区外的人能够理解、描述和利用进展。更广泛的人工智能社区承担着同样的准备责任。这个社区也被邀请和挑战，以确定基本的方式，功能可以调整为一个有益的建设和学习模型，并提供环境，以促进这一识别的人工智能研究人员的光谱。

现在，我将以这个额外的思考作为结束。我在2016年11月写了这篇文章的初稿。当时，很多同事都给出了肯定的反馈，其中一位同事警告我，文章中出现了一种消极的语气。因此，我将草案搁置了几个月，同时继续在各种情况下口头交流其内容，并据此进行了修订。现在发布它的决定是由两个事件引发的：加州大学洛杉矶分校法学院组织的研讨会上对这些想法的讨论，以及与人工智能以外的同事（架构、编程语言、网络和理论）的其他讨论。这些讨论揭示了人们对这一主题的浓厚兴趣，并使我得出结论：我应该寻求的最重要的目标是“开始讨论”。我可能在某些方面犯了错误，我可能没有给予应有的信任，我可能错过了演变中的部分场景。我只希望我在这里分享的想法能引发一场讨论，希望社区的集体智慧能纠正我可能犯的错误。

# 致谢

我从同事们那里得到的讨论和反馈让我受益匪浅，他们的意见对于形成本文所表达的思想至关重要。

[[1]](" \l "_ftnref1" \o ")提交给ACM通信。

关于这篇文章的演讲可以在https://www.youtube.com/watch？v=UTZCWIC Do

[[2]](" \l "_ftnref2" \o ")关于因果推理的数学讲座，包括对机器学习和科学逻辑的思考，https://www.youtube.com/watch？v=zHjdd--W6o4。

[[3]](" \l "_ftnref3" \o ")斯蒂芬·霍金说，“全人工智能的发展可能意味着人类的终结”，埃隆·马斯克说人工智能“可能比核武器更危险”

[[4]](" \l "_ftnref4" \o ")有些人将其称为曲线拟合方法。虽然这一术语强调了对函数（“曲线”）的有效评估，并抓住了基于函数的方法的精神，但它低估了当今（深层）神经网络编码的复杂而丰富的函数结构。

[[5]](" \l "_ftnref5" \o ")机器学习包括基于函数的方法，但有足够大的跨度与基于模型的方法重叠。例如，人们可以学习模型的参数甚至结构，但这可能仍然需要复杂的推理才能从模型中获得答案（即，既不要求也不保证有效的评估）。

[[6]](" \l "_ftnref6" \o ")这里值得一提的是，神经网络的研究自20世纪40年代早期出现以来，经历了许多转折，例如，早期关于感知器（一种浅层神经网络）的局限性的结果几乎在60年代停止了这方面的研究，直到20世纪70年代第一个多层（深层）网络投入使用时，才更为积极地恢复使用。

[[7]](" \l "_ftnref7" \o ")mikewellman告诉我，深度学习的成功在很多方面都令人惊讶，它应该提醒我们，人工智能的进步可能会突然到来，我们应该做好准备。

[[8]](" \l "_ftnref8" \o ") The recent White House initiative, “Preparing for the Future of AI,” particularly raised awareness about the safety of AI systems, a term that was mentioned dozens of times in the final report of this initiative. We now hear about efforts such as “safe reinforcement learning” and “verified AI systems,” which highlight how safety considerations are starting to impact AI technology.

[[9]](" \l "_ftnref9" \o ") Eric Horvitz brought up the idea of subjecting certain AI systems to trials just as we do to approve drugs. I think that proper labeling of certain AI systems should also be considered, again as we do with drugs. For example, some have suggested that the term “self driving car” is perhaps responsible for the misuse of this technology by some drivers, who are expecting more from this technology than is currently warranted.

[[10]](" \l "_ftnref10" \o ") Oren Etzioni laid out this argument in a talk he gave at UCLA in March, 2016.

[[11]](" \l "_ftnref11" \o ") One academic outcome of the expert system era was the construction of a dedicated Masters degree at Stanford University, called the “Masters in AI,” which was separate from the M.S. in computer science degree and had significantly looser course requirements. It was a 2-year program, with the second year dedicated to building an expert system (I was a member of the very last class that graduated from this program before it was terminated). I personally recall that one of the justifications for this program was that classical computer science techniques can be harmful to the thinking needed for effectively building expert systems!

[[12]](" \l "_ftnref12" \o ") One cannot here but bring into attention the long time it took neural networks to have the impact they are now having.

[[13]](" \l "_ftnref13" \o ") A colleague couldn&apos;t but joke that the broad machine learning community is being bullied now by the success of its deep learning sub-community just as the broader AI community has been bullied by the success of its machine learning sub-community.

[[14]](" \l "_ftnref14" \o ") Early AI researchers left us with a wealth of results on first-order deduction (e.g., resolution and unification) but basically left untouched the subject of first-order model counting, which is at the heart of current needs for probabilistic reasoning and machine learning.

[[15]](" \l "_ftnref15" \o ") I made these remarks at a dinner table that included a young machine learning researcher. His reaction: “I feel much better now.” Apparently, this young researcher was subjected to this phenomena by supportvector-machine (SVM) researchers during his PhD work, when SVMs were at their peak and considered to be “it” at the time. Another young vision researcher—when pressed on whether deep learning can address our ambitions in vision research—told me, “The reality is that you cannot publish a vision paper today in a top conference if it does not contain a deep learning component, which is kind of depressing.”

[[16]](" \l "_ftnref16" \o ") This is similar to responses given to Searle&apos;s Chinese Room Argument (1980), which comes down to having a function (of bounded size) that emulates competence in speaking Chinese. Searle basically missed the (impossible) combinatoric assumption underlying his argument.

[[17]](" \l "_ftnref17" \o ") Dan Roth alerted me to another perspective on functions versus models based on the distinction between discriminative versus generative learning. In particular, when learning a function (discriminatively), all one can usually do is estimate the value of its output variables. However, a learned (generative) model can be used to answer more queries through reasoning. Hector Geffner made a related observation that functions are perhaps more suitable when the “goals” are fixed.

[[18]](" \l "_ftnref18" \o ") The seminal work of Judea Pearl on probabilistic approaches to commonsense reasoning is one example outcome of the mentioned crisis.