重新审度人工智能

秦颖，何纯阳 译

北京外国语大学

公众对于人工智能千篇一律的看法是，我们正在制造愈来愈聪明的机器，总有一天这些机器的能力会超过人类，窃取我们的工作，乃至摆脱人类的控制、消灭全人类。这一错误的观点，尽管不是所有的人工智能研究者都持有，但它很有可能阻碍甚至破坏人工智能的实际应用，并对公共政策产生不利的影响。历史的进程和当前的发展都证明，更为合理的架构是，人工智能仅仅是长期以来自动化任务的延续。自动化任务至少可以追溯到工业革命之初。仔细分析人工智能这种毁天灭地的巨大能力，有利于我们评估这一重要技术的利与弊，也有助于削弱夸大其词和受挫感所造成的不断自我毁灭。自出现以来，过分夸大和令人失望的事一直充斥着人工智能领域。

问题的核心在于那些外行却令人尊重的公众人物，还有部分业内人士对于夸大其辞的新闻报道采取了容忍甚至认可的态度。这些报道称，人工智能的每一个进展都出乎人们意料、超乎人们想像，已经接近（超越）一般人类的智能水平，散布“机器人”会掌控世界的恐慌。那些新闻标题部分往往鼓吹引人注目的工程成就，暗示着人工智能是对独一无二、至高无上的人类的挑衅。如果电脑可以炒股和开车，那么是否很快就会比最佳销售员更加出色，取代法官，赢得奥斯卡奖和格莱美奖，为自己购买和开发高端房地产？“他们”会怎么想“我们”？

显然，并不存在所谓的“他们”。“他们”不过是源于好莱坞大片的拟人修辞，不过是在人工智能技术的公开展示中加入了无足轻重的类人特征以增强效果，比如自然人的声音、面部表情和人类情感的模拟。上述所有技术在人-机界面的应用中意义重大，但主要用于愚弄或误导公众就不对了。

通过引发不恰当的问题，以及隐含的弦外之音，使得对人工智能重大成果包装的现象大行其道，这样只会有害无益。IBM 的 Watson（沃森）是否真的会为在 Jeopardy! 中胜出而高兴？表面上看，的确如此。和人类选手的比拼不仅使得电视节目火爆，也促使观众对这一重要成果的实际意义进行过度的解读。机器没有思想，也罕有证据表明未来机器会有。

近来涌现的突出成果源于越来越多的工具和技术得到应用，这些工具和技术使我们能够更好地利用计算能力、存储空间和大规模可用数据库。这自然是计算机科学的伟大之处，但并不足以证明超级智能的发展能够在任何可能的任务上胜过人类。虽然一些新的工具——最突出的是机器学习——能够广泛地用于看起来与技术无关的领域，实际上它们经常依赖于领域问题的某些共同属性，例如大量数字化的样本。快速交易算法、在视频中跟踪对象，以及预测传染病的传播都有赖于采取技术手段在有噪声的实时数据流中发现微妙的模式，因此尽管任务迥异，使用的工具却类似。

当然，我们正用机器来执行各种需要人类智能来解决的实际任务，但这并不意味着计算机真的具有智能，只能说明这是解决问题的另一种方式。人类和电脑都能下棋，但我们尚不清楚人类和电脑下棋的方式是否一样。机器翻译近来所取得的进展令人瞩目，但机器更多依赖的是从大量相似文本中获取统计意义上的关联性，而非在理解自然语言上取得了根本性的进展。

原本需要人类付出努力的工作，既有体力工作也有智力工作，机器使之变得自动化了，而且机器经常能比人做得更好，成本更低，或者兼而有之，否则机器也就失去了作用。工厂的自动化取代了从钣金工人到咖啡品尝师等无数具备高超技能、训练有素的工人。过去专门由人类“计算师”解决的算术问题现在由十分廉价的工具完成，这些工具在贸易展览会上常常被当作促销的小礼品而被堆在一边。过去，制作动画中灰姑娘的头发需要一批艺术家的协作，而现在，CGI 技术就能再现长发公主飘逸的头发。这些进步并不是对人类能力的藐视或者挑衅，相反，机器使我们脱身而出，能够执行更多更重要的任务。

Machines don't have minds, and there is precious little evidence to suggest they ever will.

机器没有思想，也罕有证据表明未来机器会有。

一些专家警告称，整体上，电脑（尤其是人工智能）会导致广泛的失业。当机器能够承担现在几乎所有的工作时，我们靠什么生存？从历史角度看，这个想法是有问题的。劳动力市场因自动化程度的提高而不断发展。200 年前，美国 90% 以上的劳动力在农场工作。现在，2% 的劳动力就能以低廉的成本生产出更多的食物。然而，任何人并未因此而失业。事实上，如今就业的人数比以往都多，而且大多数人认为，比起辛苦劳作的祖先来说，他们的工作缴税更少，报酬更多。这是因为自动化使得社会更加富裕，从而产生对各种新产品和服务的需求，最终扩大了对工人的需求。技

术不断使很多行业消失，但我们的经济最终会以新的、截然不同的行业取而代之。毫无疑问，人工智能的最新进展可能使如今的许多乃至大多数工作得以实现自动化，但我们没有理由认为创造就业的历史模式会停止。

上面说的都是好的方面。坏的方面是，技术驱动劳动力市场转型可能需要相当长的时间，会给流离失所的工人造成很多困难。如果人工智能加速了自动化的步伐，许多人预测，这种快速转型可能会导致社会大的动荡。

哪些工作最容易受到威胁？为了回答这个问题，我们需要了解，实际上并不是让工作实现了自动化，我们是让任务实现了自动化。因此，工人是否会被取而代之或能否提高生产效率，要取决于所执行任务的性质。如果工人的工作是重复性的、具有确定的流程和明确的目标，那么他们要保住工作确实有风险。但如果是纷繁杂乱的活动，在混乱的或不断变化的环境中具有挑战性的工作，或是需要流露真实情感的工作，那么风险就要低得多。

那么未来的工作是什么样的？很多人把工作视为一种交易，但很多职业要依赖于与其他人建立信任或融洽的关系。如果您只是在周末提取一点现金，那么 ATM 机和柜员的作用是一样的。但是，如果您希望投资者能够帮助您建立新的业务，相信在任何时候您都不会选择一台机器去做这件事。

这并不是断言说机器永远不会拥有感知能力或是表达情感的能力。事实上，情感计算方面正在迅速发展。问题是用户如何看待这些功能。这些功能如果只是简单地被理解为沟通的辅助方式，很可能被广泛接受。但是，如果被认为是表达假同情或减少合理诉求，那么很可能会招致不信任和拒绝——正如任何人都说，他们愿意拿着电话听着循环播放的录音等待，以说明他们的电话很重要。没有人想要一个机器人牧师接受他们的忏悔，或者一个机器承办人在他们失去亲人时帮助祷告。

有些工作涉及到技能演示，或是传达他人所需的安慰。除了猎奇之外，有谁会想看自动驾驶的赛车比赛？或是在你喝酒时，一个机器调酒师来询问你一天过得怎么样？很多职业需要更多的社会技能，而且随着我们可支配收入的增加，对这些职业的需求只会增长不会减少。原则上，凡事都无法阻挡人们成为报酬丰厚的各种专业人才，如工匠、设计师、个人购物者配送员、演员、看护人、网络游戏玩家、会员、策展人和顾问。就像在几十年前几乎没有这样的工作一样，未来也可能出现超乎我们想像的职业。

So the robots are certainly coming, but not in the way most people think.

所以机器人肯定会出现，只是不一定是大多数人所想的方式。

所以机器人肯定会出现，只是不一定是大多数人所想的方式。有人担心机器人会替代我们，会造反，会统治我们，持有这种观点的人只能说是被误导了。有些人担心，随着超级智能机器与我们的生活密不可分，会分享我们的物理空间和公共空间，使得我们无法面对困难。如何确保这些机器尊重我们不言自明的社会习俗是一个严峻的挑战。是否应该允许机器人与您一起排队、将钱放在停车计时器中延长您的停车时间、使用拥挤的人行道交货、承诺购买、

签订合同、代您投票、替您占公车上的座位？哲学家们会关注更多显而易见的、重要的伦理问题，例如自动驾驶的汽车是否应该不顾您的生命去拯救两个行人？现实问题可能涉及的面更广。多数人工智能研究人员自然而然地专注于解决直接的问题，但在未来的几十年中，关系到他们的工作是否会被人们广泛接受的因素很可能是系统能否遵守我们的社会、文化习俗。

科幻小说讲的都是机器人的故事，但从工程的角度来看，这些都是设计问题，而不是由一些假定的自然普遍秩序造成的不可预知的后果。良好的产品，包括日益自动化的机器和应用，都不会乱套，除非我们设计得不好。如果 HAL 9000 杀死机组人员是为了避免自己被停用，那也是因为它的设计师没有正确地设置目标的优先级。

为了应对这些挑战，我们需要为自主能力越来越强的系统制定工程标准，或许我们可以借鉴土木工程中的潜在危险的概念。例如，系统可以包含预期作业场的模型（即“标准操作环境”或 SOE），并且当系统偏离界限时，输入明确的“安全模式”。我们需要研究人们如何合理地调整自己追求的目标，以满足他人的利益和权利。系统在部署前应通过认证考试，相当于汽车做碰撞测试。最后，我们需要有关基本伦理的程序概念来指导在意外发生时的行为。这并不是说机器必须要有道德，只是要求它们在特定情况下按照道德“行事”。我们是如何在人类生命、动物生命、私有财产、自我保护中排出优先次序的？什么情况下才可以打破法律的束缚？

当计算机在有限的、确定的领域运行时，这些问题都无足轻重。但如果我们希望人工智能系统得到广泛的信任和利用，就应对该领域的目的、目标和潜能进行仔细的重新评估，至少我们的评估结果应当获得大众的普遍认可。显而易见，人工智能的公关问题可能对自身不利。我们需要减少新闻界夸大其词的说法，避免公众狂热情绪的蔓延，着力于为人类世界制造出有教养的机器。

致谢

本工作受到校级教改研究项目“计算机专业大学生综合素养研究”的资助。

（原文刊登在美国计算机学会通讯（Communications of the ACM），2017 年第 1 期（第 60 卷），36-38 页。作者 Jerry Kaplan（jerrykaplan@stanford.edu），斯坦福大学的计算机科学客座讲师、斯坦福大学法律信息学中心研究员。他的新作是《人工智能：每个人都需要知道的东西》（牛津大学出版社，2016））。

秦颖

博士，北京外国语大学计算机系副教授。主要研究方向为计算语言学。

何纯阳

北京外国语大学高级翻译学院硕士研究生。主要研究方向为机助翻译。