

关于 AI 的畅想

谈到人工智能，人们的第一印象大多数来源于有关人工智能的书籍、电影、纪录片。《终结者》是笔者最早接触的一部人工智能影片。作为经典的“人工智能威胁论”影片，《终结者》讲述的是在未来超人工智能脱离人类控制、接管世界统治的背景下，超人工智能为剿灭人类反抗组织，派遣机器人杀手回到二十年前杀死反抗组织领袖的科幻故事。

在影片中，人类与人工智能的矛盾被表现的淋漓尽致——脱离控制、无所不能、攻击人类……无不反映着大众对超人工智能未知的恐惧。这类畅想并不易让人感到诧异——比著名的芯片摩尔定律更迅猛，人工智能进步的速度之快足以令人产生焦虑。人工智能威胁论与英国物理学家史蒂芬·霍金的观点是吻合的，即“人工智能有可能是人类历史上最后的事件”。

与人工智能威胁论相对，在另一部分人的畅想中，人工智能则是人类的得力助手——《星际穿越》中刻画了三个智能机器人，以其中的 Tars 为例，在科幻的故事背景下，它不仅操控飞船，还在尽力保护人类的生命和利益——即使掉进黑洞也收集整理了重要数据，使得人类获得了在重力方面的重大突破。关于人工智能有益的畅想也有其忠实支持者——谷歌等互联网公司的主要技术人员大部分都认同人工智能有益的观点。

就电影本身而言，以上两部影片对人工智能的演绎均存在一些片面——《终结者》通过华丽的特效，将人工智能刻画成最便于演绎的机器人的形象——天崩地裂、靛光掣目、迷雾扑朔、外形与人类完全一致的终结者在机械传动轴的运作声中缓缓站起；《星际穿越》虽然避免了为人工智能赋予“人类外形”，但终究还是以机器人的形式进行了表现。这在一定程度上使得部分人形成了刻板偏见——人工智能等价于机器人。

实际上，机器人仅仅是人工智能的一种形式。同样的，电影也仅仅是人们对人工智能的一种畅想方式——漫漫历史长河中从不缺乏仰望星空的人。

未曾止步的畅想

在人类文明的初始阶段，远远早于“人工智能”这个词汇被正式提出之时，出于对神权的敬仰和对于未知的想象，中外均已出现了类 AI 的思想。我国春秋时期名著《列子·汤问》中的《偃师献伎》有记载——偃师向王进献一具木偶，虽然木偶看上去在被偃师控制，但实际上有自己的意识。与《列子》几乎同期，古罗马诗人奥维德在《变形记》中也有相似的思想体现：塞浦路斯国王皮格马利翁以自己想象中的理想女性为模板，亲手制作了一个象牙雕塑并起名为伽拉忒亚，随后国王便爱上了她，爱神维纳斯出于对国王的同情，赋予这件雕塑生命。不难看出，在人类文明的萌芽阶段，AI 的概念仅仅停留在想象中——即以形而上的方式将生命赋予非生物。

数千年白驹过隙，1900 年的画家们畅想 100 年后的世界，创作了一系列展示在烟盒、明信片等上的插画——小到家庭扫地机器人，大到建筑机器人^[2]。此时，人们对于类 AI 的畅想已经不局限于“赋予生命”，更多在关注其实用价值。



时间来到 1950 年，艾伦·图灵发表了题为《机器能思考吗？》的论文并在其中提出了

图灵测试和机器思维的概念，将现代意义上对人工智能的畅想以论著的形式表现了出来。六年后，达特茅斯会议如期召开，约翰·麦肯锡将世界志同道合的科学家聚集在一起集思广益。会议期间，人工智能的概念被暂时统一——“人工智能就是要让机器的行为看起来就像是人所表现出的智能行为一样。”

自此，人们对于人工智能的畅想已不再是空想，而更倾向于付诸实践——即有想法便会尽可能去实现。

19世纪60年代，由于人工神经网络的突破，人们开始憧憬神经网络可以解决大量的算法函数逼近和逻辑策略问题。人们对人工智能的未来大为看好，人工智能进入了第一个蓬勃发展的时期。但随着1969年马文·明斯基出版的《感知机》提出了人工神经网络的局限（即不能解决高阶谓词问题），人们对于人工智能的乐观期望遭到了严重打击，人工智能的研究迎来第一次低谷。

低谷时期的人工智能技术发展很艰难，但是科学家们一直保留着有想法便要尽力付诸实践的品质。

上世纪八十年代，一类名为“专家系统”的AI程序成功模拟了人的知识分析技能，与当时人们想象中的完美人工智能（能够与人交流、翻译文字、分析图像甚至像人一样推理）恰好方向一致，人们眼中的人工智能再一次变得无所不能。但正如约翰·麦肯锡的名言：“一旦它开始发挥作用，就不会再有人叫它人工智能了”，在1987年，随着用户的期待不断提高，技术的发展难以满足用户的要求，Lisp机器市场崩溃，人工智能领域声誉再次下降，第二次凛冬降临。

在这次人工智能寒冬中，科学家们将思想的洪流回溯到人工神经网络，尝试着去突破明斯基所预言的人工神经网络的局限。

2006年，杰弗里·辛顿提出了神经网络深度学习算法，显著提高了神经网络的能力，也丰富了人工智能的概念——人工智能应当具备学习的能力。自此，人工智能重获人们信赖，迎来了第三次高峰。时至今日，人工智能成为了人们心中最为前沿、最具潜力的“新兴”行业。

畅想是发展源动力

与人工智能发展的整个历史背景相类似，每项人工智能的成就都经历了“畅想—尝试—成功—继续畅想”的过程。其中，自动驾驶技术和面部识别技术的发展过程最为典型。

1926年12月8日的清晨，一切都和往常一样，人们洗漱、进餐、赶赴工厂。车流停滞，一位穿着绅士的先生从伸进车窗的报童小手中买下了《密尔沃基日报前哨》，只见那新闻标题赫然：“幻影汽车将自主探索城市”。绅士只是笑笑，心想：“真是新颖的观点啊”，随即踩下离合、推上排挡，驱车离开。

当“自动驾驶”这个划时代的概念出现在一封地方性的日报上时，其重大意义被局限在人们无趣时光的消遣、茶余饭后的谈资之中。在那个计算机尚未出现的年代，自动驾驶汽车被设计成无线电遥控驾驶的模式——驾驶员通常会坐在不远处控制“幻影汽车”，这虽然与当代理念大相径庭，但对于那个年代来说，出现如此前瞻的概念已经是难能可贵的了。

随着1946年电子计算机的诞生，在1950年前后渐渐出现了汽车自动化的实验。经历无数次测试，第一台真正意义上的自动化汽车生产于上世纪八十年代。此后实验的进程越来越快，2012年美国内华达州机动车辆管理局为谷歌自动驾驶汽车颁发了合法车牌；2018年底，Waymo公司将自动驾驶汽车推入市场，正式开始其商业化——在经历了百年漫长的研发过程后，自动驾驶终于冠以人工智能的名义，在润物无声中渐入人们的生活。自此，自动驾驶与每个人息息相关。

当代自动驾驶与人工智能的联系紧密。自动驾驶汽车一方面具备准确探测外界的信息的能力，可以在各类传感器的帮助下，收集车辆运行的数据，通过中央处理器处理并迅速做出调整；另一方面，与众多人工智能领域专家的观点一致，自动驾驶汽车具备深度学习能力。在对海量数据的分析与处理后，汽车将收集到的图像、波等直观信息转化为对其有价值的信息。经过深度学习，汽车识别道路、行人、障碍物的效率和准确性均得到了明显提高。

即使自动驾驶技术已经达到了超人预料水平，但科学家们依然没有停下对其可优化性的畅想。他们认为：“自动驾驶依然有很大的提升空间。除了检测和避开行人之外，自动驾驶汽车还应当可以通过推断精神状态而更准确地预测行人和其他司机的行为（例如通过判断他们是否在注意车流来增减车速）。这种类型的心理推理，以及其他类型的基于模型的因果和物理推理，对于缺乏相关培训数据的新颖驾驶环境很有价值^[1]”。

比起尚未占据主流的自动驾驶技术，面部辨识技术已经在真正意义上深入了大众的生活，从手机的面容解锁到支付宝的“刷脸”支付：从住宅小区门口的“刷脸”门禁到全国范围的嫌犯排查。可以说，人工智能对于人脸的识别已经成为了每个人生活不可或缺的一部分。

虽然面部辨识的突然兴起发生在最近五年，但其历史可以追溯到1888年，《自然》杂志发表了法兰西斯·高尔顿的一篇代表性的论文，介绍如何使用人的面部特征进行识别。自此，拥有理论依据的科学家们产生了对自动识别人脸技术的思考。

有畅想，科学家们便去尝试。近现代传统的面部辨识技术大致分为五个方法：基于模板匹配的方法；基于几何特征的方法；基于代数特征的方法；基于人工神经网络的方法；基于稀疏表示的方法。随着人工智能的不断发展，面部辨识技术也产生了基于深度学习的方法。相较于传统，新的方法基于对大量人脸数据集的学习，大大提高了识别的速度与精确度。

技术总是在不断发展的，面部辨识技术亦然引发了人们的进一步期待——模糊图片中人像的识别、图片识别之后的重构都是亟待解决的问题。

或许我们不难发现，人工智能的发展总是趋向于摒弃原有的程序算法，使计算机自主学习处理数据、面部识别技术的实现，展现了弱人工智能向强人工智能的转化，处处体现着人工智能的美学——高效、便利、精准、强大。

但对于笔者，当下的技术已经足以使得我们的生活效率和生活质量显著提高——就仅面部识别技术而言，在生活节奏不断加快的今天，有什么能比不用扫码就可以买到一罐廉价的饮料更能让人欣慰的呢？

未来需要更多的畅想

人工智能作为当下最热门的行业，自然吸引了大批的关注。难以想象人工智能将会有多么广阔的未来，所以在这个时代，所有人都有权利继续畅想下去。笔者与不同年龄阶段、不同知识储备程度的朋友们深入座谈交流、收集调查问卷，同时一并阅读了部分文献，得到了许多有价值的观点。

对于大部分少年儿童而言，由于受到科幻小说、影视作品的影响，他们对于人工智能的理解是停留在想象阶段的，人工智能对于他们而言更多是想象中的那些可以成为朋友的机器人——类似于日本动漫中“哆啦A梦”的形象。

作为青年学生，在接触了有关程序设计和人工智能的课程之后，相较于孩童执着AI的科幻成分，同学们更多地关注了自己畅想的可实现性。对“你想象中的未来人工智能”问题搜集的有效问卷样本进行分析后，大致可以将观点划分为三大类——形成自主思维、大幅提升服务业质量、未知环境探索。

形成自主思维的观点中，大部分人认为，未来的人工智能应该可以辅助解决更多高难度的数理分析、战略统筹决策问题、产生审美、情感等抽象类功能、基于一定的数据分析预测尚未发生的事件；由于当下人工智能虽然功能多样化，但质量并不能达到令所有人满意的程度，所以有关提升服务业质量的样本问卷中，“精准医疗”、“工作助理”、“生活管家”、“智能家居”的高智能化被提到的频率最高，“感同身受的虚拟现实”、“人机结合交互”等也是值得思考的方向；对于未知环境的探索方面，人们对火灾、地震等的救援很感兴趣，也有人提出了以人工智能作为太空、深海探索先锋的观点。除此三大类观点外，“人机界面完善”、“机器人趋向小型化”也是有参考意义的畅想。

同时，问卷中对“人工智能有益有害”的思考方面，74.29%的人认为人工智能不会失控；对“未来AI专一化和集成化”的讨论中，57.14%的人认为专一化（即机器仅可以高质量地实现单一功能）相较于集成化（即质量较低地实现多种功能）更具优势。

科学家们在著作中也对当下并非尽善尽美的AI提出了很多优化性的畅想。

人们比机器更善于解决一系列困难的问题，包括场景理解、概念学习、语言学习、语音识别、语言理解等。而创造力、常识和通用推理等人类的认知能力对于机器仍然难以理解。未来几年有望在自动驾驶汽车、遗传学、医药、药物设计和机器人等领域推出更多人工智能应用^[1]。

在当前备受关注的神经网络方面，未来几代的神经网络将看起来与目前的最先进的神经网络大相径庭。他们可能被赋予直观的物理、因果推理、心智理论和其他能力^[3]。

业界权威人士在其著作中总是一副严肃的面孔，为此，笔者采访了清华大学自动化系主任张涛教授，希望可以用一种轻松的方式表现他对人工智能的畅想。

《终结者》等电影中人工智能突破了人类权限的限制，接管了世界的统治。那

么请问您对此类观点有何看法呢？

人类肯定是希望人工智能可以帮助人类做一些事情，而不是让它危害人类。将来如果出现这种危害人类的人工智能之后，大家肯定有办法把它控制住，这是矛和盾的关系。

您在机器人的领域当中做出了很高的成就，那么请问在当下，人工智能与机器人

是如何有机结合的呢？人工智能是如何被应用的呢？

其实机器人就是人工智能的一部分，人工智能是有三大学派的，其中一个学派就是行为学派，行为学派就是机器人，就是用机器来模拟人的行为，这就是人工智能。机器人在现实中应用太多了，因为它被设计出来就是要代替人从事很多危险的、恶劣环境下的工作，或者从事人做不到的事情，它功能和应用太多了。

阿西莫夫在小说中对人工智能机器人有三大原则，现实编译人工智能机器人的

自我学习功能时，是否有这三大原则的约束呢？

比如说伤害人，技术上机器当下是无法自我学习到的，我们现在还是弱人工智能，你让它有攻击人类的行为，那肯定是人教的。这个问题就像法律和道德一样，法律由公安机关来执行，道德是靠大家自觉遵守。三原则现在只是道德道义上的，暂时没有什么约束的

方法，但至今大家还都认可这些原则。

人工智能应当有一定的深度学习能力 ,那么这种深度学习是通过什么途径进行实现的呢？学习的效果是什么样的呢？

人工智能实际上它主要包含了四大部分内容，学习只是其中一部分。人工智能还包括推理、搜索、知识表达这三大部分。深度学习就是用人工神经网络来学习，是模拟人的神经系统——突触、神经元等等，但它只是模拟了神经系统的结构，实际人工智能的学习跟人的神经网络的工作原理是不一样的，它的方式是基于大数据。以前的神经网络可能就是几个节点，一层两层的规模，现在由于计算能力增强了，可以有成百上千个节点、成千上万层深度，于是它就被定义为数据学习。人工智能通过学习以后，把学习中得到的特点特征全都存在了神经网络的函数、参数里，下次再输入相同情况的时候，它就可以迅速得到跟学习相似的结果。

人工智能这个概念已经提出了七十年左右，但近些年人工智能突然经历第三次兴起，您认为是偶然现象还是大势所趋？

我觉得它就是一个不断的在发展、在进步的过程，每一个科学门类都是这样起起伏伏的，这是很自然的事情。如果有新的突破了就成为潮流，一直没有什么突破大家就不关注了。这次人工智能热主要是由 AlphaGo 打败了围棋世界冠军引起的，它兴起就是由于深度学习算法的提出——深度学习能够很好的解决这个棋局，后来又解决了很多计算机视觉等等的问题。

您觉得人工智能会不会迎来它的这种第三次寒冬呢？

这个绝对有可能，如果这个行业持续没有什么新发展的话，可能公众关注的就会变少，大家转而去关注别的了，这是有可能的，但当下我对人工智能这个领域还是非常看好的，觉得很有前景。不好说第三次寒冬会因什么而起，可能只有科幻小说家会猜这个。

人工智能今后在人文社科方面大概会有什么建设性的发展？

现在人工智能在社科类下也很重视啊，比如说现在有智能法庭，就是让机器帮法官判案，还有人工智能作曲等等，结合挺多的。人工智能本身和各行各业都可能存在交叉，都可能找到会交叉的点。

您（或者清华大学）在人工智能方向上是否取得了一些有意思的突破呢？

咱们系的老师一直都在努力，包括计算机视觉、机器学习等等都有很多的成果，老师们做的工作都很不错。至于突破的话，比较有意思的是带类脑芯片的自行车，那个就是咱们学校做的。

当下的人工智能向您期待的方向发展的瓶颈有什么？

我主要做机器人这一方面，瓶颈有很多。比如说现在想做机器人，但是拿什么来做机器人？这就遇到很多的问题。简单的说，像人的四肢现在做不出来，机器很难做到像人那么灵活、那么柔韧、有肌肉、有力量。这牵扯到各种各样的问题——涉及到能源、涉及到材料，这不仅是人工智能的问题，更是医学的问题、制造业的问题。

您对人工智能还有什么畅想？

人工智能未来发展肯定越来越强的。现在确实很难预见人工智能后面会有什么样的突破，但是任何的突破都有可能。我觉得无论是具有更高形态的实体、突破算法限制形成思维、人与人工智能融合等等，我们都可以大胆地、尽情地去想象。人工智能从它的定义来讲就是要模拟人类的智能，初衷就是希望能够用机器来帮助人，所以我觉得各种各样的可能都会出现。

更多的畅想该何去何从

《终结者》在反派机器人被液压机碾碎的结局中落下帷幕，仿佛表明人机最终难免沦为仇敌，但似乎这种不切实际的幻想并不能对社会做出实质性的贡献。事实上，在《终结者》中更有很多渐渐被付诸实践的畅想，这当中最主要的就是学习能力的体现——终结者刚刚降临时并不能与人类进行语言交流，但它依靠后期的机器学习得到了与他人交流的能力。这类畅想对于当下这个已经诞生过“AlphaGo”等深度学习机器的时代并不是新鲜之事，但对于1984年而言，那个时代的畅想真的在最后被付诸了实践。对于《星际穿越》而言，Tars试图实现的“幽默”功能也正是当下人工智能领域竭力解决“机器情感”难题。

所有畅想都应该归于实践——皮格马利翁的伽拉忒亚对应着如今的“AI女友”、偃师的木偶变成了当下“自动机器人”、扫地机器人以全新面貌进入千家万户、建筑机器人在水泥森林中崭露头角……历史长河中从不缺乏仰望星空的浪漫主义者，更不缺乏埋头苦干的实干主义者——无论是跨越千年还是遥隔万里，美好的畅想终究要实现。

正像张教授所言，当下的人工智能还在路上、未来的人工智能无限可能。畅想之瀚海无限，实践之天路同样长远，“俱往矣，数风流人物，还看今朝”，笔者愿寄希望于未来——征途不止，畅想无极！

##参考文献

<div id="refer-anchor-1"></div>

- [1] Brenden M. Lake et al. Building machines that learn and think like people[J]. Behavioral and Brain Sciences, 2016, 40

<div id="refer-anchor-2"></div>

- [2] [100年前插画师幻想的未来世界是什么样的？这些插画告诉你](https://www.sohu.com/a/453821546_99947083)

<div id="refer-anchor-3"></div>

- [3] Jing Pei et al. Towards artificial general intelligence with hybrid Tianjic chip architecture[J]. Nature: International weekly journal of science, 2019, 572(7767) : 106-111.

