人工智能的思考

宇宙就是一座黑暗森林，每个文明都是带枪的猎人，像幽灵般潜行于林间，轻轻拨开挡路的树枝，竭力不让脚步发出一点儿声音，连呼吸都必须小心翼翼：他必须小心，因为林中到处都有与他一样潜行的猎人，如果他发现了别的生命，能做的只有一件事：开枪消灭。在这片森林中，他人就是地狱，就是永恒的威胁，任何暴露自己存在的生命都将很快被消灭，这就是宇宙文明的图景，这就是对费米悖论的解释。

一旦被发现，能生存下来的是只有一方，或者都不能生存。

——《三体·黑暗森林法则》

《银河系漫游指南》中，外星人造了一台超级计算机，问了它一个终极的问题:“嘿，生命、宇宙和一切答案的是什么？”计算机用了很长时间计算说：“42”。外星人又很好奇的问：“你这答案是什么意思啊？”计算机回答说：“这个问题的答案就是 42 啊……”

人工智能日益炒热，在许多年前我们会恐惧机器人会统治人类，而现在盛誉“Alpha Go”创造的神迹象。已经有人期待着去问“她”关于宇宙、人类的终极问题，希望能从人类制造的全知全能的“神”口中获得自己想要的答案。

神学与机器之灵

人类总是热衷于造神，宗教、科学皆是如此。对于人工智能的幻想从蒸汽朋克时代就已经开始，当手术刀第一次切开人体大脑表面的灰质时，就会想着用二极管和电流来代替突触和神经元；而当计算机更新换代，从埃尼亚克更新到超级计算机，集成电路进入超微小时代，电脑的计算速度早就把人脑远远甩开。科学在进步，心理学、哲学变成了最后一块处女墙，守护着人类最后的尊严。

沃卓斯基姐弟的《黑客帝国》其实是赛博朋克般残忍的后乌托邦故事，智能圈养人类从而完成自我升级。2016 年三月“阿尔法狗”击败李世石，2017 年初“MASTER”席卷棋坛，人工智能的研究也成功的从“仿生派”过度到了“学习派”，AI 不再需要像个人一样作出决策，而是选择最好的那个决策。

自动驾驶带来便利的同时会让一部分人失去工作，高盛投入大量资金在人工智能领域，同时裁员力度一年大过一年，股票期货交易已经有一部分智能交易员开始参与。蒸汽时代工业革命让人类在体力上被机器击败，现代人工智能可以在智力方面击败人类吗？这难免又落入了一个讨论了一万年的话题——机器人可以替代人类吗?

本质上，这是一种人性对自我超越的渴望和恐惧。乍得沙赫人于 700 万年前行走于非洲时，就会挥舞手中的大棒击退群狼、狩猎野兽，机器作为“体力”的替代早就成为我们身体的一部分，这就是非洲大陆上的赛博朋克。凭借着不错体力在进化论中存活下去的人类“体力”上早已屈服，这种屈服已经轮到“智力”。医学科技的发达让义肢和器官替代出现，这一种自我超越的触手从“体力”伸向“智力”。 笛卡尔时代的思想家就在思考人类和机器谁是社会的主宰，而人工智能的日益普及讲这种忧虑直接抛给了每一个人。

人对人的控制无处不在，鲜血涂脸的巫术、宗教、处刑用的断头台，工业时代签订契约，信息时代舆论操纵，控制的手段科幻小说都不及。于内心深处，控制他人却怕被他人控制，这里的他人可以是一切客观存在，自然、客体甚至是趋势。

即使身体被局限，起码内心是自由的。现代医学却告诉我们仍然是受制于神经结构的凡人，思想也会深受局限，就如同动物无法运算高等数学一般，思维有限而且脆弱。

在自我意识和抽象思维能力的共同作用下，理性思维诞生突如其来。

人类意识到自己不再是动物，因为我们已经无法像动物般满足于本能带来的生存，开始追问更多的问题；但也不是神神——因为深植于内心的动物本能作为早已跟不上社会发展的自然进化产物，却对我们的思维产生最根本的影响。

所以人的全部悲剧，就在于理性对自身的绝望。

即使在学会了控制本能之后，整个神经系统的基本结构也依然让我们无法如神一般全知全能。汉语中将人类思维与超然存在都用同一个汉字——“神”的深意是，要么找到神，要么造神，要么成为神。

现代科学证明找到神的这一步失败了，对自由和智慧的追求一直在内心无法阻挡。从雕刻在石碑上的汉谟拉比法典到超级计算机诞生出来的人工智能，理性驱使我们一步步抵达本身想到达的地方——用机器之灵创造新的神学。

哥德尔：上帝创造了不完美

任何自然科学研究都没有比人类彻底认识自己、了解自己、找出解决自身面临的问题更为迫切。在人类最为关注的四大自然科学领域——物质的结构、宇宙的起源、生命的本质和智能的产生中，“智能是如何产生的？”，是最具有挑战性也是最为困难的课题。

1900 年希尔伯特在巴黎数学家大会上提出 21 世纪亟待解决的 23 个问题，其中第二个问题是建立整个数学的无矛盾性。哥德尔寻此方案解决希尔伯特第二问题，希望先建立算术理论的无矛盾性，然后再建立相对于算术，实数理论的相对无矛盾性，却得到了相反的结果：他证明，即使限制在算术这样狭小的数学范围内，要想证明关于它的形式系统的无矛盾性都是不可能的。换句话说，任何丰富到足以展开初等数论的形式系统，如果它是一致的，它就是不完全的，在这个系统中至少有一个真的数学命题不可证，虽然它是真的，但它不是系统中的定理。

在《无穷与心》这部书中，作者记录了他在参观罗马教堂时的情形，教堂外面有一个巨大的石头圆盘。圆盘上刻着一张毛乎乎的长满络腮胡子的男子的脸。他的窄缝状的嘴巴被刻在大约腰部位置。按照民间传说，上帝有令，任何把手插进这张嘴里并且说一句假话的人，绝不能再把手抽回来。鲁克说，他来到教堂，把手插进那张嘴里，并且说了一句“我绝不可能把手再抽回来”。无须说，鲁克根毛无损地离开了罗马。实际上，鲁克说的是一个自指句。这个故事阐明了为什么永远不可能造出能够捕获所有可能的真实世界真理的“万能真理机器”的逻辑基础。

哥德尔是在形式系统 S 中构造了一个命题，这个命题断定自己在 S 中是不可证的，进而指出，这个命题和它的否定都不是 S 的定理，（显然这个命题是真的）即这个命题在 S 中是不可判定的，从而给出不完全性结果的。

那么，能不能添加更强的公理扩充这个系统 S1 到更大的系统 S2 呢？哥德尔说，不行，因为，还有新的真数学命题在新扩充的系统 S2 中是不可证的，继续扩张，情形依然如此，…… 实际上，除非你把这种扩张过程持续到超穷，否则这种系统连最简单的算术真理都不能穷尽。看来，可证数学命题和数学真理之间永远隔着一个超穷距离，仅使用有穷方法甚至都没有希望逼近它。正如哥德尔所说，数学不仅是不完全的，还是不可完全的。 另一个看似更让人吃惊的结论是，如果一个形式系统是一致的（不含矛盾的），不可能在该系统内部证明系统的不矛盾性。 这就是著名的哥德尔第一和第二不完全性定理。

哥德尔定理与数学家的期望相去甚远，在定理发现之后，数学家不得不重新调整自己的思维方式：因为，一方面人们期望数学形式系统应当囊括所有数学真理，一方面又分明知道总有数学真理不可证；一方面经验和直觉告诉人们数学是不含矛盾的，一方面理性又教导人们数学不能证明它自身的一致性。著名数学家外耳当时曾感慨说，“上帝是存在的，因为数学无疑是一致的；魔鬼也是存在的，因为我们不能证明这种一致性。”以生动的语言道出了当时处于两难境遇的数学家的困惑。这个宇宙给了我们一种选择，就人类认知而言，我们要么拥有一种正确的但却是极不完整的小书，要么拥有一本正确缺乏内在和谐的大书，我们可以选择完整也可以选择和谐，但鱼和熊掌不可得兼。

哥德尔定理似乎表明，在机器模拟人的智能方面必定存在着某种不能超越的极限，或者说计算机永远不能做人所能做的一切。

人工智能的极限

1936 年图灵发表重要文章《论可计算数》指出，“我们将假定需要计数的心的状态数是有穷的。这是因为，如果我们承认心的状态有无穷多，它们中的某些状态就会由于‘任意接近’而被混淆”。图灵的这段话曾被看作“人类心智活动不可能超越任何机械程序” 的一个论证。1950 年图灵在《计算机器与智能》中指出，我们不能因为一台机器不能参加选美大赛而责备它，就像我们不能因为一个人没有飞机跑得快就责备他一样，机器也能够思维。这篇文章还隐含着“人心等价于一台计算机”的论断。图灵的观点对当时刚刚兴起的人工智能方案无疑是一强有力的声援，也自然引起了一场大争论。

1961 年美国哲学家鲁卡斯在《哲学》杂志上以极其激烈的言辞首先撰文《心、机器、哥德尔》，试图用哥德尔定理直接证明“人心超过计算机”的结论：“依我看，哥德尔定理证明了机械论是错误的，因为，无论我们造出多么复杂的机器，只要它是机器，就将对应于一个形式系统，就能找到一个在该系统内不可证的公式而使之受到哥德尔构造不可判定命题的程序的打击，机器不能把这个公式作为定理推导出来，但是人心却能看出它是真的。因此这台机器不是心的一个恰当模型。这就是著名的鲁卡斯论证。随后，另一位美国哲学家怀特利在接下来的《哲学》杂志上发表了强有力的批驳文章《心、机器、哥德尔——回应鲁卡斯》，遂引起许多人卷入并长达几十年的争论。

1963 年，美国哲学家、认知科学家、现象学家德莱弗斯出版了《计算机不能做什么？——人工理性批判》，1982 年和 1986 年又相继出版了《胡塞尔、意向性和认知科学》和

《心灵优于机器：人的直觉在计算机时代的力量》，批判了强人工智能的观点，反对把人仅仅看成一种抽象的推理机器，与机器不同，人有具有识别、总和以及直觉的能力，这些能力植根于一些与计算机程序的计算理性截然不同的过程中，直觉智能的力量使人能够理解、言说以及巧妙地调整我们与外部环境的关系。并说胡塞尔是认知科学的先驱。

完全可以在精确的意义上说，计算机具有人类理解故事和解答相关问题的能力。在塞尔看来，计算机的理解力与汽车和计算器的理解力没有什么不同，计算机与人类的心智相比，其理解力不仅是不完全的，而且可以说完全是一个空白。当然，对塞尔来说，重要的不是要论证“计算机不能思维”，而是要回答“正确的输入输出加上正确的计算本身是否足以保证思维的存在？”他认为,“如果我们所说的机器是指一个具有某种功能的物理系统，或者只从计算的角度讲，大脑就是一台计算机”，然而在他看来,心的本质并非如此。计算机程序纯粹是按照语法规则来定义的，而语法本身不足以担保心的意向性和语义的呈现，程序的运行只具有在机器运行时产生下一步形式化的能力，只有那些使用计算机并给计算机一定输入同时还能解释输出的人才具有意向性。意向性是人心的功能，心的本质绝不能被程序化，也就是说，心的本质不是算法的。

大脑的许多行为是“突现”的，即这种行为并不存在于像一个个神经元那样的部分之中，只有很多神经元的复杂相互作用才能完成如此神奇的工作。因此，从神经元的角度考虑问题，考察他们的内部成分以及他们之间复杂的、出人意料的相互作用的方式，这才是研究意识问题的本质。许多哲学家和心理学家认为目前从神经元水平考虑意识问题的时机尚不成熟。然而事实恰恰与此相反。仅仅用黑箱方法去描述脑如何工作，特别是用日常语言或数字化编程计算机语言来表达，这种尝试为时尚早。脑的语言是基于神经元之上的。“要了解脑，你必须了解神经元，特别是巨大数目的神经元是如何并行工作的。”

人脑是一个丰富的相互关联的信息的载体，它的许多内容是连续变化的，然而机器却能使我们通过内省得到非常有限的体验，其他机器却不具有这个特性。那么我们将来能否造出这样的机器，如果可能，它们看上去是否会有意识？克里克相信，最终是可以实现的，也可能存在着我们几乎永远不能克服的技术障碍。短时期之内，我们所能构造的机器就其能力讲，与人脑相比很可能很简单，因为在理解了产生意识的机制以前，我们不大可能设计一个恰当形式的人造机器，也不能得出关于意识的正确结论。

值得注意的一点是，哥德尔第二不完全性定理的一种形式是说，任何恰当的定理证明机器，或者定理证明程序，如果它是一致的，那么它不能证明表达它自身一致性的命题是定理。哥德尔说，一方面，人心不能将他的全部数学直觉形式化，如果人心把他的某些数学直觉形式化了，这件事本身便要产生新的直觉知识（如该系统的一致性）。一方面，不排除存在一台定理证明机器确实等价于数学直觉，但重要的在于，假定有这样的机器 M，由不完全性定理，我们不可能证明 M 确实能做到这点。

当我们应用哥德尔定理试图严格地作出“人心胜过计算机”的论证时，其中包含着一个令人难以察觉的漏洞：问题的核心并不在于是否存在能捕获人类直觉的定理证明机器，而恰恰在于，即使存在这样一台机器，也不能证明它确实做到了这一步。恰如哥德尔所说：“不完全性结果并不排除存在事实上等价于数学直觉的定理证明机器。但是定理蕴涵着，在这种情况下，或者我们不能确切知道这台机器的详情，或者不能确切知道它是否会准确无误地工作。”

展望未来：生存还是毁灭？

我们无法确定“心不是计算机”的结论为真，而且，认知的本质为何？计算的含义应当是什么？人工智能是否存在某种不可逾越的逻辑极限？大脑和心的功能究竟为何？心是否有物质载体？这些问题需要更深刻的科学的进展。 我认为，现在一个更值得思考的问题是，我们以上的讨论都是建立在图灵意义上的可计算概念基础上的，目前人工智能领域也完全是在图灵意义上可计算概念基础上产生的“认知可计算主义”的纲领指导下工作。我们最初是从希尔伯特元数学方案开始考虑问题的，是想用有穷手段，用能行的方法建立一个没有内在矛盾的形式系统囊括所有的数学真理，哥德尔告诉我们，这样做不可能，但是我们仍然在追求一种严格一致的算法来模拟人的智能。即使不论用一个形式系统表达图灵机的方式不唯一，我们也应当考虑到，对于模拟人类智能的计算机，完全可以采用某种新型的形式系统，采用包含非古典逻辑的具有动态性质的形式系统。但是，同样不容忽视的一个问题是，这种形式系统至少应当保证紧致性定理成立，应当在原始递归的范围之内，这样一来，哥德尔不完全性定理就自然成立，因此仍然没有超出哥德尔所言的逻辑极限范围。

我们对世界的理解来自对我们经验规律性的发现和学习。规律或因果律有两种，一是可精确重复的或可以预期的，二是统计的。所有人工物都是按机械规则(精确重复性)被设

计和制造的，我们尚无法制造一种东西，它的原理和行为指向是遵守统计性规律的。生命和智能服从统计性规律，所以许多理论称人是机器是极其荒谬的。量子效应虽然遵守统计性规律，但对于量子计算机我们尚未克服界面的困惑。

因此，解决人类智能的极限和人工智能的极限问题，除了与哥德尔定理有关外，还需要对大脑和计算机更精细的模型作更大胆的的研究，而且还需要将学习、问题求解、对策理论与实数论、逼近论、概率论和几何学知识结合在一起，探索其如何对问题的解起实质性作用。