智能极限：后人类的未来

作者：十五言.从科幻到科技 著

目　录

智能极限：后人类的未来

十五言.从科幻到科技　著

出版社：青苹果数据中心

湖南省青苹果数据中心有限公司

注册时间：1992年8月13日

注册地址：中国湖南省长沙市开福区青竹湖大道399号

互联网出版许可证：新出网证（湘）字013号

电子邮箱：GA@egreenapple.com

网　　址：www.egreenapple.com

青苹果数据中心为作者和相关机构提供数字出版服务。

本书电子版如有错讹，敬请指正，我们会及时更新版本。

接触未来

主编：岳川

机器能思考吗？奇点临近是危言耸听？人类是否可以突破肉体的极限，与机械合而为一？面对末日，我们是该自生自灭，还是飞向太空？

针对这些问题，本书中收录了十五言社区的十篇文章，将以不同的视角为我们提供可能的答案。作者们从科幻出发，根据已知或可预见的科技，结合哲学思辨，试图向我们展现一个智能达到极限的世界，包括人工智能与人类智能本身，以及智能意识如何走向永生。最后，本书还展现了两个未来——地球末日和太空前景。

科幻作品如何描绘人工智能和后人类的未来，每位作者都有不同的解答。他们的观点彼此呼应，对概念的解读互相交叉，引用的经典作品时常不谋而合，思考的方向却不尽相同。

当你看完《奇点之后的世界》，好奇人工智能是怎么诞生的，那么思故渊的第二篇文章《P与NP——计算机科学的圣杯》，将为你彻底解答这个数学问题，并介绍科幻作品所展现的赛博化世界——超级AI觉醒。

《机器中的幽灵》将从笛卡尔心身二元论出发，穿插以科幻梗“机器中的幽灵”，讲述近现代哲学家们对心身二元论的批判，以及对机器思维的思考，并最终指向图灵的提问“机器能思考吗？”

张小星的《放开那只机械姬！》除了探讨人工智能的意识外，还将讨论前几篇文章未涉及的伦理问题。梦见乌鸦的《机器人进化修炼指南》则向我们一一展现了经典科幻电影中的机器人。

让我们将目光转向人类本身，戴桃疆的《基于赛博格的后人类主义》和子非鱼的《后人类时代》探讨人类机械化的可能性，解读后人类主义与后人类的几种可能形式。《永生》提出了人类亘古不变的愿景，不断追求智能极限的我们，最终能否上传意识，直抵永恒？

或许，人类来不及走向进化的终极，末日便不期而至。艾守义的《Who Killed the World？》以简洁优美的文笔，讲末日成因，抒科幻情怀。文章透露出的对文明轮回的观点，让我想起刘宇昆的短篇小说《河图洛书》对不同生命形式追求智慧之书的想象，他在开头写道：“每个文明都有自己独特的方式传承智慧，令其穿越岁月。他们令想法变得可感可知的方式尘封在时间中，像舷墙抵抗着不可遏止的时间洪流。”

刘慈欣曾说人类有两种未来，内向的未来和外向的未来。机械化和赛博化是内向的未来，漫游太空的奥德赛征程是外向的未来。选择外向未来的人类，又该如何打造太空中的家园？Ent的《流浪家园：如何打造太空方舟》，从封闭体系、开放体系和多体系平衡三个方面讨论了这个问题，并再次提出刘慈欣的观点：“决定宇宙图景的未必是物理学，很可能是社会学。”

这本电子书由果壳网旗下十五言社区打造，征稿活动得到科幻星云网的支持。感谢编辑组戴一、思故渊、王江山、语月和刘三尺的辛苦审阅，以及十五言AI们的全程监督。最后也感谢各位作者投稿，让我们得以瞥见科幻作品中未来世界的一隅。

奇点之后的世界

——技术爆炸与严厉的赛博

作者：思故渊

科幻浪潮的兴起从来都不是单独发生的，它总和当时的科学的发展所密切相关。从这个意义而言，正如大刘所说，科幻作家的想象力其实并不如科学家，而科幻的繁茂的枝叶也正是从科学坚实的根基上生长起来的。五六十年代的科幻黄金时代，也正是科学技术发展的黄金时代；而进入新世纪，最新的科幻浪潮无疑是“后奇点”（Post Singularity）科幻。

“技术奇点”的概念是基于人类科学进步的一个自然观察：总的来说，人类科学进步的速度是越来越快的。人类从学会使用工具到进入农业时代花了数十万年的时间；从农业时代到蒸汽机用了上万年时间；而蒸汽机到电气时代，不过200年光景；而电气时代到目前的信息时代，也就不到一百年。这说明人类的科技进步速度是一个加速度，或者按照某些理论，是呈指数发展的，那么可能在不久的未来，人类科技的进步将会快到这样一个程度，之后技术将会变得与之前完全不同，就跟数学上的概念“奇点”一样，这之后的世界将完全无法预测。

技术奇点将以一个什么样的形式到来呢？从目前来看，人类的科技发展的主要方向有三个：生物、能源和信息。这三个领域中的任何一个产生了突破，都会导致奇点的到来。在生物领域，人类如果能够彻底的搞清楚生物的秘密，突破进化的限制，最终将生物世界和人类自己塑造成任意的形状，那无疑是技术奇点的产生；在能源领域，人类如果最终发现了核聚变或者比这要更加先进的，近乎无限能源的产生办法，那么科幻迷们常常爱说的“星辰大海”的时代就会到来，人类会突破地球的重力井，扩展到星际之间，那也是奇点的一个产生方式；而在信息科学领域，人类最终制造出了人工智能，它的智力必然会远远超过人类自身，技术奇点也会到来。这个过程还有另外一个称呼，叫做“超人巨变”。

计算机的性能是一条指数发展的曲线

从目前的科学发展来看，生物或者能源科技的突破需要基本理论的革命性突破，就如同当年相对论和量子力学的产生或者DNA的发现那样的突破，而这样的突破目前来说还没有任何先兆。相比于生物或者能源，信息科学领域是最有可能产生奇点的，因为最能够支持“奇点”理论的就是计算机科学的进步。著名的摩尔定律：“每十八个月芯片的晶体管数量会增加一倍，价格降低一半”就是一个标准的指数定律，而现实中计算机性能的发展也很好的符合了摩尔定律的预测。在未来学家Ray Kurzweil的那本《奇点临近》里，作者就举出了很多的指标说明我们现在处于指数发展的一个平缓期；到了一个临界点之后，指数发展会变成近乎于无穷大的斜率，技术发生了爆炸，人类社会就进入了“奇点”时代。

“技术奇点”理论的兴起是八十年代，当然，它算不上一个严肃的科学理论，只能算作对未来的某种估计。在科学哲学上，“奇点”的理论基础来自于科学的“范式转换”说。

对科学理论的发展路径最广为人知的理论是“可证伪”说；这个理论是波普尔所推广开来的，在这个理论的视野里，科学是一个不断进化的过程；后人不断的提出新的证据证伪前人的科学理论，并且将新的证据化归为新的理论；然后新的理论再被更新的、更精细的观察数据所证伪，这就是科学前进的步伐。

而著名的科学哲学论者托马斯·库恩在他的《科学革命的结构》中提出了一个更接近于科学实在发展路径的理论：“范式转换”（Paradigm Shift）。科学并非是一个缓慢进步的过程，而是科学革命，从一个范式飞跃向另一个范式的过程，这两个范式之间不是兼容的，并不存在一个“循序渐进慢慢改进”的科学。一段时间之内，只会存在一个统一的科学研究的框架，亦即范式，而当时的科学则在范式内进行常规问题解决。从古希腊的力学到经典牛顿物理学，从牛顿物理学到爱因斯坦的相对论，再到量子力学，都是范式转换；生活在旧范式里的人是绝无可能理解新范式的。推而广之，农业社会到工业社会再到信息社会，也是范式转换的过程，一个生活在农业社会中的平民是理解不了工业社会的。而奇点的到来就是一

个新的范式转换。我们这样的生活在信息社会的平凡人类是无法理解后奇点的社会的。

“技术奇点”概念的普及本身就与科幻密不可分，给予这个理论的普及很大帮助的就是一个数学家兼科幻小说家——弗诺·文奇，他的一系列科幻小说有力地普及了这个概念。实际上，连“奇点”这个词的使用，都是文奇开始的。而他的科幻小说也基本上是围绕着这个概念进行。其中最杰出的就是“界区”系列：《天渊》和《深渊上的火》，还有前年才出版的《天空的孩子》。

《深渊上的火》里的三界区设定

“界区”系列基于一个初看上去非常玄幻的宇宙设定：银河系被分为了三个部分。处在银河系内层的爬行界，在这个界区之内，任何物质的速度都无法超过光速，地球就处于这样一个界区之内；处在外层的飞跃界，在这个界区之内，物质可以超过光速，智慧生命可以超光速旅行，但是这一界区之内的智慧生物最多只会有跟人类差不多的智力；再就是处在银河系悬臂之外的超限界，这一界限内存在的都是超级智能，人类的智力水平对他们来说就像蚂蚁之于人，这些超级智能有一个统一的称呼叫做“天人”，对于普通的智慧生物，他们就跟神一样。

为何会出现这种界区的分野，小说里没有说明，但是暗示很可能是人工的，上古时期的超级人工智能为了保护普通智慧生物所设下的分类。而为什么只有超限界才会出现超级人工智能，书中并没有明确说明；有一种猜测，认为很可能是因为超限界的数学规律改变了，P=NP。

“界区”的设定起初会让很多科幻迷不太容易接受；因为世界设定本身是奇幻小说所擅长的内容。但是“界区”的设定更多的，是在暗示人类发展的阶段。所谓的“爬行界”，就是人类在没有发展出信息科技的时代；而“飞跃界”，则是人类发展出信息科技之后，已经大大扩张了生存空间的时代；最终的“超限界”，就是奇点之后的时代了，超级智能的出现迅速地将世界改造成了现在的我们完全无法想象的状态。

克拉克三定律中最著名的一条无疑是：“足够先进的技术与魔法无异。”而文奇所设定的超限界非常符合这样的概念。超级智能所发展的技术对于普通的人类等级的智慧生物而言，与魔法毫无差别。所以很有趣的一点是文奇的界区设定的适应范围极广，很多科幻都可以与这个设定无缝兼容。比方说日本漫画家贰瓶勉所画的《Blame!》，就可以视作典型的“后奇点”类型，设定为超限界里的一个世界所开发的自动工厂失控之后的故事。

漫画《Blame!》。其实是一个典型的后奇点故事，发生在一个遥远的未来，无论是地球还是人类都变成了我们不能理解的样子

弗诺·文奇还写过一些非智能爆炸的“后奇点”小说，最典型的就是《彩虹尽头》和相同世界观的《费尔蒙特中学的流星岁月》。极度发达的信息技术让整个世界分化的千奇百怪。其中一个很有趣的技术是这样的：需要将图书馆里的纸质书电子化，最快的方法是什么？将书切成碎片，然后用鼓风机吹到一个封闭管道里。管道四周布满了高速照相机，可以将书的碎片飞舞的场景全部照下来，然后通过图像识别将碎片拼接起来，这样就能将纸书电子化了。看到这里，如果有生物学基础的读者肯定会想到当年DNA测定所使用的“霰弹枪法”，这可以说是直接的推广。这个方法快速、有效，但是是毁伤性的，于是《彩虹尽头》的故事，就从抗议这种纸书电子化技术开始。

实际上，生物技术产生爆炸之后的“后奇点”科幻小说也有一些，最著名的应该是格雷格·贝尔的《血音乐》和《达尔文电波》。《血音乐》就非常典型，生物技术上的突飞猛进导致了整个世界乃至宇宙的巨变。

真正意义上的人工智能的出现的前景现在还不清楚，在“奇点大学”的预测里，这个时间节点是2060年前后。也就是说，这篇文章的大多数读者都很可能能够活着看到奇点的来临。

而奇点的到来对于人类是不是个好消息呢？很不幸，大多数的未来学家都对此持悲观态度。超级智能产生之后，“增进人类的福祉”这一项很有可能并不包括在它的目标里。就算是包括了，人类自身对于所谓“人类的福祉”究竟是怎样的一个范畴都会有争议，就更别提让超级智能来决定如何增进人类的福祉了。在它“优化”世界的这个过程中，人类很可能会像遇到了先进欧洲文明的美洲文明一样被消灭掉，只是这个过程会更快，更无情（几个小时到几周时间）；而人类在这个过程中甚至都不会像《终结者》里的那样，还有能力组织一支反抗军；在这个过程中人类连一丝一毫的机会都不会有。

Transcendence这部电影大概是第一部描写“超人巨变”过程的电影。当然，最不科学的是，人类赢了

关于为什么人类没有赢的可能性，一位AI理论研究者Eliezer Yudkowsky曾经提出了一个非常有趣的思想实验，叫做“AI盒子实验”（AI-Box Experiment）。

这个实验的假设是这样的：假如我们制造出了一个超级人工智能，我们将它关在一台电脑里，切断对外界的一切联系，只留下一个文字窗口和外面的人联系，我们能不能用这种方法来确保它不被放出来，直到我们确认了它是安全的为止？

Eliezer Yudkowsky就做了五次这样的实验。在实验中，他扮演AI，另外一个人扮演“看门人”，规则如下：

AI方必须说服看门人让AI出来，完全通过一个文字交流窗口，而且必须是看门人完全自愿的（套出看门人的密码或者指纹等方式不算在内）；这个过程有两个小时的限制，但是交流必须达到一个最低量级，看门人不能够完全拒绝交流直到两个小时耗完；同时，AI方不能够用真实世界的手段来说服看门人（比方说“让我赢了我给你20美元”）。

可以看出，文字交流窗口，完全自愿的同意，以及两个小时的限制，这对于AI方的限制极大。但是在五次试验当中，Eliezer Yudkowsky仍然成功的三次说服了“看门人”，赢得了比赛。需要注意的是，他赢了前两次，而后面三次两方是在赌博，输的一方也会输钱，所以看门人一方有可能会因为这个原因而非理性判断来拒绝释放AI。

这个实验其实相当有力的支持了Eliezer Yudkowsky的论断：一个比人类更加聪明的人工智能是不可能被限制住的。普通人对于超级智能的想象往往是类似“雨人”的形象，有超级强的计算能力却缺乏心智；实则不然。很自然的，超级的智能肯定会有超级的策略水平，人类不可能在耍阴谋诡计或者战略战术上胜过超级智能。

那么更加自然的推论就是，人类在与超级智能的斗争中不可能赢。奇点之后的世界对于人类而言，很可能并不是什么友好的地方。

比起设定不错但是故事老套的《超验骇客》（Transcendence）之外，好莱坞更加严肃而深刻的奇点影片应该是电视剧《疑犯追踪》（Person of Interest）。剧中的主角Harold开发除了一套系统The Machine，用来大规模的监视人类社会来侦测恐怖分子，但是这套系统看到了一切，它的智能足以推断出人类社会中发生所有的事件，于是故事就此展开。

The Machine sees everything

这部电视剧非常深刻的讨论了超级智能与人类的关系。剧中的世界与现实的世界没有多大差别，但是实际上已经平顺的进入了后奇点时代，两个“神”（也就是两个超级人工智能）已经接管了世界，人类对此无能为力。而神与神之间的战争由微妙的人类之间的互动模式展开，这部剧正好讨论了之前所说的主题：人工智能是否会照顾到人类的福祉？开发出The Machine的Harold经历了非常艰苦的努力才使得他的Machine对人类充满善意，而另外一个人工智能“撒玛利亚人”（Samaritan，显然是来自“善良的撒玛利亚

人”的典故）并不关心人类，这意味着，编剧（克里斯托弗·诺兰的弟弟，乔纳森·诺兰）对于人工智能对人类的态度同样是悲观的。

“后奇点”科幻浪潮里，除了弗诺·文奇以外，还有一个非常杰出的科幻作家，就是查尔斯·斯特罗斯。基本上他的所有小说都是关于奇点的世界的小说，但是在这些小说里，奇点发生的原因个个不同，有很多非常脑洞大开的设定。

在经典短篇《抗体》（Antibodies）里，数学家证明了P=NP，于是AI觉醒，奇点到来；经常会认为人类可能会是在机器人或者核武器或者基因灾难中毁灭，但是在数学规律里毁灭，仍然是个非常新鲜的设定。在另外一个短篇小说《导弹差距》（Missile Gap）里，人类实际上是外星人的实验动物，将整个地球表面摊开之后传送到一个太阳系那么大的平面星体上……于是所有的卫星和洲际导弹统统作废。

查尔斯·斯特罗斯最先锋的“后奇点”科幻小说，应该是他的《渐加速》（Accelarando）。在这篇小说里，人类所创造出来的超级智能最终拆解了太阳系的内行星，用这些行星的原材料做成了一台超级计算机——这个概念叫做Matrioshka Brain。它是由戴森球发展而来的：利用恒星能量的最优模式是将恒星整个包起来，不让任何能量泄露。那么，Matrioshka Brain就是一台这样的将恒星包裹起来的超级计算机。更高阶的推论则如下：凡是能够产生文明的智慧生物，发展的终极都是产生这样一台超级计算机，这些超级智能们通过人类目前科学尚不知晓的方式互相通讯，而不会泄露任何电磁信号。这也是费米悖论的一种可能的解答。

艺术家笔下的Matrioshka Brain。虽然从轨道动力学角度来讲，其几乎不可能是一个完整的球体

“后奇点”科幻小说与80年代以来兴起的“赛博朋克”类型实际上有相当大的差别。这个类型的确是“赛博”的，而不是“朋克”的。赛博朋克的类型元命题是在网络的时代，单枪匹马的个人反抗系统的故事，这也是“朋克”的真意（“赛博朋克”的源流需要另一篇文章来讲述）；而“后奇点”科幻关注的是庞大的系统，和信息科技的进步本身息息相关，所以，称其为（小者语）“严厉的赛博”更为合适。

人工智能界也有“强AI”和“弱AI”的假说：“弱AI”假说认为，真正的智能需要一种特殊的结构，也就是人脑的结构才能够产生；也就是说，“智能”这种软件需要搭配特殊的硬件（神经网络等）才能够运行；而“强AI”假说认为，无论是什么样的结构，一旦系统的复杂度达到了一个临界点，智能就会自发的涌现出来。至于现在的计算机科学界对于“人工智能”的探索，仍然无法判断那种假设是正确的。但是机器学习（Machine Learning）领域告诉我们，实际上只要有足够多的样本，简单的线性定律就能够收敛到足够有意义的结论，简直就是一种巫术（Witchcraft）。这种技术不由得让人担心，可能人的智慧比我们想象的要浅薄得多。

周穆王西狩于昆仑山，遇到了一个大师工匠偃师，制造出了和真人一模一样能够跳舞唱歌甚至能够向宠姬抛媚眼的假人。偃师制造的有可能是人类有史以来的第一个人工智能。制造人是神的工作，人自己来做，是僭越。当年，各地的人们集合起来要造巴别塔直通天堂，上帝觉得这么做根本就和自己把他们当初放在地上的目的完全冲突了，“变乱他们的语言”，于是巴别塔垮掉了。人类制造超级智能的努力最终是否会成功，现在并不知道。但是，制造超越我们自己的智能，做上帝的工作，或者干脆说，制造出神本身，是人类一直以来不断追求的目标；我们这一代人在临死之前，可能真的能够看到一位新的神灵的诞生，但是这对于我们以及我们的子孙后代的福祉到底有什么影响，那就真的只有神才能回答。

P与NP——计算机科学的圣杯

P?=NP

作者：思故渊

我们所喜欢的科幻，大多数是关于物理学和生物学的。说到物理学，我们有相对论量子力学薛定谔的猫；说到生物学，我们有基因工程DNA进化论。就连数学，我们也有哥德巴赫猜想心理史学。但是关于计算机——大家除了知道赛博朋克或者人工智能这些概念之外其实并不知道计算机科学到底是做什么的学科。这样说来，计算机科学其实是科幻里提得最多，但是大家最不了解的一门科学。当然，这与计算机科学自身也有关系，所谓Computer Science实在是一大堆互相关系不大的分类的集合，就好比有人把建筑设计结构力学室内装潢电器维修全部统合起来称作“造楼学”一样。这篇文章里，我就给大家介绍一个计算机科学里最基本的，可以称之为“圣杯”的问题：P?=NP。

P与NP其实是一个数学问题，但是它是跟计算机紧密联合在一起的。我得先从最基本的东西开始介绍，就是计算机之父——图灵和诺依曼、罗素等人在上个世纪初所做的数学研究。所谓计算机，就是可以用来计算的机器；那么在制造出这种机器之前，我们得先定义，什么是“可计算”的，这就是计算机科学里的基础理论：可计算理论。图灵提出了计算机的通用模型所谓“图灵机”，并且定义了什么是可以计算的：任何能够在图灵机里完成的操作都是所谓“可计算”的。而著名的“停机问题”就是一个“不可计算”的操作的例子。具体的停机问题的内容其实在概念上与罗素悖论是一样的，即一个只给“不给自己理发的人”理发的理发师是否应该给自己理发一样，计算机要判断自己会在哪个问题上死掉，这就是停机问题。更加完整和详细的解释大家可以去看维基百科，或者任何一本集合论或者计算理论的教科书。

计算理论不只是判断什么是可计算的，进一步的，它还要研究“如何计算”。“如何计算”归根结底，就是如何更容易的计算。就跟有些动物比其他动物更平等一样，有些计算比其他计算更容易，这就涉及到计算复杂度的问题。对于计算机来说，最基本的操作只有四个：加减乘除。乘就是连加，会比加难一点点，但是不会难太多。但是除就要比乘更难，这个大家小学的时候做算数就知道。所以，加减乘叫做短操作，除叫做长操作，在这里，计算并不是平权的。

完成一个特定计算的一系列操作被称为算法。算法的核心价值就在于需要多长时间来完成它，这个评价标准就是算法的时间复杂度。时间复杂度一般用O(f(n))来表示。也就是说，对于一个大小为n的输入，需要一个关于n的函数来描述算法的运行时间。比方说，给一个n的数组，要求输出数组中的一个元素，那么，只需要访问数组的那个元素就够了，这个算法就不依赖于数组的大小，它是常数时间的，即O(1)；如果需要找出数组中最小的那个数，那么就需要对数组中的所有元素做遍历，这是线性时间的，即O(n)；最简单的数组排序冒泡算法，它的复杂度是O(n^2)。可以看出这些算法的时间复杂度都可以用n的冥来表示，比方说n^2+n之类，这些算法就称之为多项式时间算法，算法所解决的问题也就是可以在多项式时间解决的问题，称之为P问题（Polynomial）。

但是有些问题并不能在多项式时间解决，解决它的算法的时间复杂度可能是O(2^n)，最浅显的例子，是这样的：给出一大堆正负整数的集合，找出任意两个整数：它们的和等于0。这是个看上去非常简单的问题，但它就是没办法在多项式时间解决！这一类问题，就称之为NP问题（Nondeterministic Polynomial）。NP问题与P问题的差别不光是能否在多项式解决，还有一个关键在于NP问题是可以在多项式时间验证的。这就回到了之前所说的计算难易不平等的问题上来。将两个素数乘起来得到一个数，这很容易，但是它的逆操作，将一个数分解为两个素数的乘积就很难；但是如果给出两个素数，来验证是不是这个数的因子，又很容易。上面所提到的这个NP问题，也是这样：虽然找出两个整数满足和为0的条件没有P时间内的算法，但是给出两个整数来验证是否满足条件却是一眼就能看出的——所以，这个问题不能在多项式时间得到答案，但是可以在多项式时间得到验证。上述素数乘积的问题，就是现代密码学的基础。

那么问题来了，挖掘机技术哪家——不对，我们能否找到一个算法，把NP问题在P时间内解决？这就是计算机理论的圣杯——P?=NP.

这是个到目前为止都没有得到证明或者证否的问题。这是一个本质问题，即有些计算是不是本质上比其他计算更容易，或者更难——人们不知道NP是不是等于P。假如哪一天，人类真的证明了P=NP了呢？乐观的预测Matrix67给出了一个：假如P=NP，世界将会怎样？不过第一个遭殃的就是现在的银行系统，所有的密码统统失效了。

但这只是乐观的预测。还有一个悲观的预测：这将是人类的末日。

斯特罗斯在他的短篇小说《抗体》里就描述了这个景象。在这篇小说里，一个数学家证明了P=NP，真正的AI出现。它的智力迅速超越了人类的一切可能性，奇点降临，人类灭亡。我们同样可以用上述那个整数集合的问题来说明：假如说，有一台电脑存放着全世界所有人的脸；目标则是，找出任意两个长得相似的人。这个问题就绝对是NP难的（大家也可以看出这跟整数集合问题的相似之处）。假如P=NP，那么也就是说，计算机可以迅速地把全世界所有人按照“长像相似”或者任何一种分类方式进行分类，它能够看到一切联系，也就意味着，计算机全知全能。

从某种意义来说，这倒更可能是P=NP之后的场景。从这个意义来说，这将会是对人类影响最大的一个数学问题也不为过。不过，到目前为止仍然没有看见任何P=NP的可能性，而且绝大多数计算机科学家和数学家都认为P!=NP，所以，安啦……

再就是，就算有人证明出了P=NP，可能这只是一个存在性证明而非构造性证明，即只是证明存在这种算法而算法到底是什么也不清楚；或者找到了算法，但是算法的复杂度是O(n^10000000000000000)这样，没有现实意义。

除了《抗体》是直接描述P=NP的后果如何，科幻中提到这个的惊人地少。美版福尔摩斯《基本演绎法》中有一集讲述一个证明了P=NP的数学家被杀，原因仅仅是因为“这个算法会攻破银行的安全系统”，这比起这个问题被证明所真正会引起的震动，仅仅是九牛一毛。

弗诺·文奇的“界区”系列《深渊上的火》也跟这个问题有关。在这部科幻经典里，银河被分为三个部分：速度不能超越光速的爬行界，速度可以超越光速的飞跃界，和居住于其上的全是超级智能的超限界。爬行界和飞跃界住的都是智力相当的生物，两边的差别很容易理解，就是一个能超光速另外一个不能；而飞跃界和超限界的差别就不是那么容易理解的。为什么，超限界可以诞生神一样的超级智能“天人”而飞跃界不能，并且天人不能进入下界？书里没有明确的说明。可以猜想，其中的一个隐藏设定就是，超限界的数学规律是不一样的，在那里，P=NP，所以才能够出现超级智能。

我一直很喜欢的美剧《疑犯追踪》（Person Of Interest）就是科幻界最近很火的“后奇点”（Post-Singularity）题材，刚才提到的《抗体》和《深渊上的火》都是这个题材的，讲述技术达到奇点之后的故事，而这类科幻的核心创意都是人工智能，人工智能超越人类。POI里主角开发出了监视系统“机器”（Machine），它就是一个能够知晓一切事情的超级人工智能，新时代的神灵。最近两季的主线就是两位神灵的战争：两个人工智能，“机器”与“撒玛利亚人”之间的争斗。可以合理地猜测，POI的最终结局将是，一位数学家证明了P=NP，于是所有人都围绕着他开始行动。

P=NP，作为计算机科学中的圣杯问题，从某种意义上，关乎人类的未来：有科学家就认为这个问题是真正的人工智能能否出现的关键要素。或者可以这样表述：如果P!=NP，那么奇点就永远不会到来。数学规律是否能够决定未来人类的命运？让我们拭目以待。

机器中的幽灵

——从《疑犯追踪》和《真实的人类》说起

作者：岳川

"Ghost in the Machine."

《疑犯追踪》第一季第十集截图

《真实的人类》第一季第三集截图

如果单从犯罪题材来看，科幻剧《疑犯追踪》（Person of Interest）的故事已足够精彩。它由骨灰级科幻爱好者乔纳森·诺兰和J.J.艾布拉姆斯打造，讲述了一个智能机器与人共存的故事。人工智能之父和他的搭档们依靠“The Machine”的帮助惩恶扬善。而它又不局限于描绘人对机器的掌控，乔纳森·诺兰着重探讨的是奇点时代人工智能的进化。消灭罪犯的故事包裹着奇点的科幻内核。乔纳森在第三季开始，逐渐表达了对机器和秩序如何共存的忧虑，引入另一个不受约束的人工智能“撒玛利亚人”，孤立无援的机器“The Machine”开始向“人”进化。

由瑞典电视剧改编的英美合拍剧《真实的人类》（Humans）正在AMC和英国电视4台热播，背景设定在近未来，也是一部描述奇点临近的“低科技科幻”（Lo-Fi Si-Fi）。该剧探讨的是已经具备人类形态的人工智能和人的界限逐渐模糊，大量机器人从事人类的大部分工作，而与此同时人性被赋予在机器身上。

这两部热门电视剧都用细节展示了亚瑟·库斯勒（Arthur Koestler）的《机器中的幽灵》（The Ghost in the Machine），《疑犯追踪》甚至给了它两次特写。这本哲学心理学著作首次出版于1967年，1990年由企鹅出版社再版。需要插一句的是，库斯勒并非纯粹的哲学和心理学死宅，他本人的生活相当精彩。二十世纪三、四十年代，作为记者的他报道过西班牙内战和二战，战后他开始写政治小说和非虚构作品，曾获三次诺贝尔文学奖提名，此外他还积极参与政治和社会活动。

有关机器人和人工智能的科幻，或多或少会提及这本经典著作，《太空堡垒卡拉狄加：卡布里卡》、《星际之门：亚特兰蒂斯》和《4400》的某一集都曾以此命名，《X档案》系列曾以“Ghost in the Machine.”命名人工智能计算机。赛博朋克系列《攻壳机动队》（Ghost in the Shell）亦以此为灵感，作者士郎正宗正是借用了“Ghost”这个概念。由阿西莫夫小说改编的电影《我，机器人》（I, Robot），以及1985年上映的英国反乌托邦科幻电影《妙想天开》（Brazil），也曾引用过“机器中的幽灵”，前者与《真实的人类》一样探讨了机器人三大定律约束下智能机器人的生存困境，后者则和《疑犯追踪》一样表达出对监控的担忧。

The Police乐队专辑《Ghost in the Machine》封面

"Cogito, ergo sum."

牛津大学哲学教授吉尔伯特·赖尔（Gilbert Ryle）曾用比喻“机器中的幽灵”来贬低笛卡尔二元论。赖尔是日常语言哲学牛津学派创始人，沿袭维特根斯坦的方法，是心理学行为主义代表人之一。在他1949年出版的首部著作《心的概念》（The Concept of Mind）中，他批驳了以笛卡尔为代表的心身二元论，这个“官方”学说被17至18世纪的西方哲学家普遍接受，他们认为自然是一个复杂的机器，人性则是被注入“灵魂”的更小的机器，并以此来解释智慧、自发性和其他人类特性。

笛卡尔所处的时代已经有水力驱动的机器，对机器作为物质客体的思考，对心灵的探寻，以及对神学的虔诚和对力学（机械定律）的笃信，促成了他的二元论，他认为心和物是本质上不同的独立实体。他总