人工智能的局限性

有人听说我想创业，给我提出了一些“忽悠”的办法。他们说，既然你是程序语言专家，而现在人工智能（AI）又非常热，那你其实可以搞一个“自动编程系统”，号称可以自动生成程序，取代程序员的工作，节省许许多多的人力支出，这样就可以趁着“AI 热”拉到投资。

有人甚至把名字都给我想好了，叫“深度程序员”（DeepCoder = Deep Learning + Coder）。口号是：“有了 DeepCoder，不用 Top Coder！”还有人给我指出了这方向最新的，吹得神乎其神的研究，比如微软的 Robust Fill……

我谢谢这些人的关心，然而其实人工智能的能力被严重的夸大了。现在我简单的讲一下我的看法。

识别系统和语言理解

纵观历史上机器学习能够做到的事情，都是一些字符识别（OCR），语音识别，人脸识别一类的，我把这些统称为“识别系统”。当然，识别系统是有价值的，OCR 是有用的，我经常用手机上的语音输入法，人脸识别对于公安机关显然意义很大。然而很多人因此夸口，说我们可以用同样的方法（机器学习，深度学习），实现“人类级别的智能”，取代大量的人类工作（比如客服，保洁，送外卖，司机……），这就是神话了。这些人完全没有理解这些看似“简单枯燥”的人类工作背后所隐含的，难以逾越的难度。

识别系统跟真正理解语言的“人类智能”，其实相去非常远。说白了，这些识别系统，也就是统计学的拟合函数能做的事情。比如 OCR 和语音识别，就是输入像素或者音频，输出单词文本。很多人分不清“文字识别”和“语言理解”的区别。OCR 和语音识别系统，虽然能依靠统计的方法，“识别”出你说的是哪些字，它却不能真正“理解”你在说什么。

聊一点深入的话题，看不懂的人可以跳过这一段。“识别”和“理解”的差别，就像程序语言里面“语法”和“语义”的差别。程序语言的文本，首先要经过词法分析器（lexer），语法分析器（parser），才能送进解释器（interpreter），只有解释器才能实现程序的语义。类比一下，自然语言的语音识别系统，其实只相当于程序语言的词法分析器（lexer）。我在之前的文章里已经指出，词法分析和语法分析，只不过是实现一个语言的万里长征的“第0步”。

大部分的 AI 系统里面连语法分析器（parser）都没有，所以主谓宾，句子结构都分析不清楚，更不要说理解其中的含义了。IBM 的语音识别专家 Frederick Jelinek 曾经开玩笑说：“每当我开掉一个语言学家，识别率就上升了。”其原因就是语音识别仅相当于一个 lexer，而语言学家研究的是 parser 以及 interpreter。当然了，你们干的事情太初级了，所以语言学家帮不了你们，但这并不等于语言学家是没有价值的。

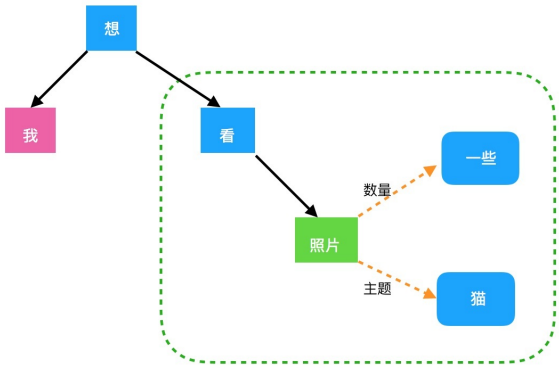
很多人语音识别专家以为语法分析（parser）是没用的，因为人好像从来没有 parse 过句子，就理解了它的意义。然而他们没有察觉到，人其实必须要不知不觉地 parse 有些句子，才能理解它的含义。

举一个很简单的例子。如果我对 Siri 说：“我想看一些猫的照片。”它会给我下图的回答：“我在网上没有找到与‘一些猫’有关的资料。”

这说明了什么呢？很多人可能都发现了，这说明了 Siri 无法理解这个句子，所以它到网上去搜一些关键字。可是这还说明一个更深层次的问题，那就是 Siri 里面并没有 parser，甚至连一个好的分词系统都没有，所以它连该搜什么关键字都不知道。

为什么 Siri 去网上找关于“一些猫”的信息，而不是关于“猫”的信息呢？如果搜索“猫”和“照片”，它至少能找到一些东西。这是因为 Siri 其实没有 parser，它里面根本没有语法树。它只是利用一些普通的 NLP 方法（比如 n-gram），把句子拆成了“我…想…看…一些猫…的…照片”，而不是语法树对应的“我…想…看…一些…猫…的…照片”。

这个句子的语法树，按照我之前做过的一种自然语言 parser 的方式，分析出来大概是这个样子。



具体细节太过技术性，我就不在这里解释了。不过有兴趣的人可能发现了，根据语法树，这句话可以简化为：“我想看照片。”其中“看照片”是一个从句，它是“我想…”的宾语，也就是所谓宾语从句。多少照片呢？一些。看什么样的照片呢？主题是猫的照片。

- 我想看照片
- 我想看一些照片
- 我想看猫的照片
- 我想看一些猫的照片

是不是挺有意思？

Siri 里面没有这种语法树，而且它的 n-gram 居然连“一些”和“猫”都没分开，这就是为什么它去找“一些猫”，而不是“猫”。它甚至把“照片”这么重要的词都忽略了。所以 Siri 虽然正确的进行了“语音识别”，知道我说了那些字。但由于没有 parser，没有语法树，它不可能正确的理解我到底在说什么，它甚至不知道我在说“关于什么”。

制造自然语言的 parser 有多难？很多人可能没有试过。我做过这事。在 Indiana 的时候，我为了凑足学分，修了一门 NLP 课程，跟几个同学一起实现了一个英语语法的 parser。它分析出来的语法树形式，就像上面的那样。

你可能想不到有多困难，你不仅要深刻理解编程语言的 parser 理论（LL，LR，GLR……），还得依靠大量的例子和数据，才能解开人类语言里的各种歧义。我的合作伙伴是专门研究 NLP 的，把什么 Haskell，类型系统，category theory，什么 GLR parsing 之类…… 都弄得很溜。然而就算如此，我们的英语 parser 也只能处理最简单的句子，还错误百出，最后蒙混过关 :P

经过了语法分析，得到一棵“语法树”，你才能传给人脑里语言的理解中心（类似程序语言的“解释器”）。解释器“执行”这个句子，为相关的名字找到对应的“值”，进行计算，才能得到句子的含义。至于人脑如何为句子里的词汇赋予“意义”，如何把这些意义组合在一起，形成“思维”，这个问题似乎没有人很明白。

至少，这需要大量的实际经验，这些经验是一个人从生下来就开始积累的。机器完全不具备这些经验，我们也不知道如何才能让它获得经验。我们甚至不知道这些经验在人脑里面是什么样的结构，如何组织的。所以机器要真的理解一个句子，真是跟登天一样难。

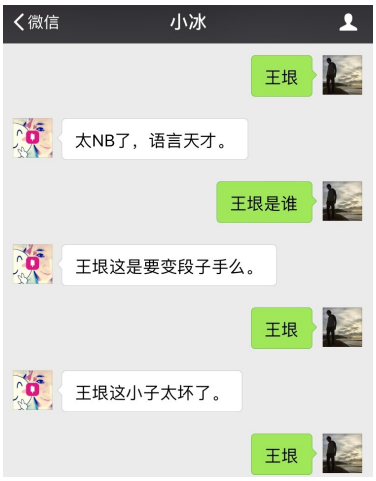
这就是为什么 Hofstadter 说：“一个机器要能理解人说的话，它必须要有腿，能够走路，去观察世界，获得它需要的经验。它必须能够跟人一起生活，体验他们的生活和故事……”最后你发现，制造这样一个机器，比养个小孩困难太多了，这不是吃饱了没事干是什么。

机器对话系统和人类客服

各大公司最近叫得最响亮的“AI 技术”，就是 Siri，Cortana，Google Assistant，Amazon Echo 一类含有语音识别功能的工具，叫做“个人助手”。这些东西里面，到底有多少可以叫做“智能”的东西，我想用过的人都应该明白。我每一次试用 Siri 都被它的愚蠢所折服，可以让你着急得砸了水果手机。那另外几个同类，也没有好到哪里去。

很多人被“微软小冰”忽悠过，咋一看这家伙真能理解你说的话呢！然而聊一会你就发现，小冰不过是一个“网络句子搜索引擎”。它只是按照你句子里的关键字，随机搜出网上已有的句子。大部分这类句子出自问答类网站，比如百度知道，知乎。

一个很简单的实验，就是反复发送同一个词给小冰，比如“王垠”，看它返回什么内容，然后拿这个内容到 Google 或者百度搜索，你就会找到那个句子真正的出处。人都喜欢自欺欺人，看到几个句子回答得挺“俏皮”，就以为它有智能，而其实它是随机搜出一个句子，牛头不对马嘴，所以你才感觉“俏皮”。比如，你跟小冰说：“王垠是谁？”，她可能回答：“王垠这是要变段子手么。”



心想多可爱的妹子，不正面回答你的问题，有幽默感！然后你在百度一搜，发现这句话是某论坛里面黑我的人说的。

下面是一个确切的例子，它显示了小冰是如何工作的。图片是 2016 年 10 月底抓的，那时候我试了一下跟小冰对话。现在的情况可能稍微有所不同。

这说明小冰的答复，基本是百度问答，知乎一类的地方来的，它只是对那上面的数据做了一个搜索匹配。随机搜索出这句子作为回答，至于幽默感，完全是你自己想象出来的。很多人跟小冰对话，喜欢只把其中“符合逻辑”或者“有趣”的部分截图下来，然后惊呼：“哇，小冰好聪明好有趣！”他们没有告诉你的是，没贴出来的对话，很多都是鸡同鸭讲，枯燥无味，把人都聊走了。

IBM 的 Watson 系统在 Jeopardy 游戏中战胜了人，很多人就以为 Watson 能理解人类语言，具有人类级别的智

能。这些人甚至都不知道 Jeopardy 是怎么玩的，就盲目做出判断，以为 Jeopardy 是一种需要理解人类语言才可以玩的游戏。等你细看，发现 Jeopardy 就是很简单的“猜谜”游戏，题目是一句话，答案是一个名词。比如：“有个歌手去年得了十项格莱美奖，请问他是谁？”

如果你理解了我之前对“识别系统”的分析，就会发现 Watson 也是一种识别系统，它的输入是一个句子，输出是一个名词。一个可以玩 Jeopardy 的识别系统，可以完全不理解句子的意思，而是依靠句子里出现的关键字，依据分析大量语料得到的拟合函数，输出一个单词。世界上那么多的名词，到哪里去找这样的语料呢？这里我给你一个 Jeopardy 谜题作为提示：“什么样的网站，你给它一个名词，它输出一些段落和句子，给你解释这个东西是什么，并且提供给你各种相关信息？”

很容易猜吧？就是 Wikipedia 那样的百科全书！你只需要把这种网站的内容掉一个头，制造一个“倒索引”搜索引擎。你输入一个句子，它就根据里面的关键字，搜索到最相关的名词。这就是一台可以玩 Jeopardy 的机器，而且它很容易超越人类玩家，就像 Google，Yahoo 之类的搜索引擎很容易超越人查找网页的能力一样。可是这里面基本没有理解和智能可言。

其实为了验证 Watson 是否理解人类语言，我早些时候去 Watson 的网站玩过它的“客服 demo”，结果完全是鸡同鸭讲，大部分时候 Watson 回答：“我不清楚你在说什么。你是想要……” 然后列出一堆选项，1，2，3……

你指望拿这样的东西代替你公司的人类客服吗？那你的公司就等着倒闭吧。

当然，我并不是说这些产品完全没有价值。我用过 Siri 和 Google Assistant，我发现它们还是有点用处的，特别是在开车的时候。因为开车时操作手机容易出事故，所以我可以利用语音控制。比如我可以对手机说：“导航到最近的加油站。” 然而实现这种语音控制，根本不需要理解语言，你只需要用语音识别输入一个函数调用：导航（加油站）。

个人助手在其它时候用处都不大。我不想在家里和公共场所使用它们，原因很简单：我懒得说话，或者不方便说话。点击几下屏幕，我就可以精确地做到我想要的事情，这比说话省力很多，也精确很多。个人助手完全不理解你在说什么，这种局限性本来无可厚非，可以用就行了，然而各大公司最近却拿个人助手这类东西来煽风点火，夸大其中的“智能”成分，闭口不提他们的局限性，让外行们以为人工智能就快实现了，这就是为什么我必须鄙视一下这种做法。

举个例子，由于有了这些“个人助手”，有人就号称类似的技术可以用来制造“机器客服”，使用机器代替人作为客服。他们没有想清楚的是，客服看似“简单工作”，跟这些语音控制的玩意比起来，难度却是天壤之别。客服必须理解公司的业务，必须能够精确地理解客户在说什么，必须形成真正的对话，要能够为客户解决真正的问题，而不能只抓住一些关键字进行随机回复。

另外，客服必须能够从对话信息，引发现实世界的改变，比如呼叫配送中心停止发货，向上级请求满足客户的特殊要求，拿出退货政策跟客户辩论，拒绝他们的退货要求，抓住客户心理，向他们推销新服务等等，各种需要“人类经验”才能处理的事情。所以机器能不但要能够形成真正的对话，理解客户的话，它们还需要现实世界的大量经验，需要改变现实世界的能力，才可能做客服的工作。由于这些个人助手全都是在忽悠，所以我看不到有任何希望，能够利用现有的技术实现机器客服。

连客服这么按部就班的工作，机器都无法取代，就不用说更加复杂的工作了。很多人看到 AlphaGo 的胜利，以为所谓 Deep Learning 终究有一天能够实现人类级别的智能。在之前的一篇文章里，我已经指出了这是一个误区。很多人以为人觉得困难的事情（比如围棋），就是体现真正人类智能的地方，其实不是那样的。我问你，心算除法（23423451345 / 729）难不难？这对于人是很难的，然而任何一个傻电脑，都可以在 0.1 秒之内把它算出来。围棋，国际象棋之类也是一样的原理。这些机械化的问题，其实不能反应真正的人类智能，它们体现的只是大量的蛮力。

纵观人工智能领域发明过的吓人术语，从 Artificial Intelligence 到 Artificial General Intelligence，从 Machine Learning 到 Deep Learning，…… 我总结出这样一个规律：人工智能的研究者们似乎很喜欢制造吓人的名词，当人们对一个名词失去信心，他们就会提出一个不大一样的，新的名词，免得人们把对这个名词的失望，转移到新的研究上面。然而这些名词之间，终究是换汤不换药。因为没有人真的知道人的智能是什么，所以也就没有办法实现“人工智能”。

生活中的每一天，我这个“前 AI 狂热者”都在为“人类智能”显示出来的超凡能力而感到折服。甚至不需要是人，任何高等动物（比如猫）的能力，都让我感到敬畏。我发自内心的尊重人和动物。我不再有资格拿“人类”来说事，因为面对这个词汇，任何机器都是如此的渺小。

纪念我的聊天机器人 helloooo

乘着这个热门话题，现在我来讲一下，十多年前我自己做聊天机器人的故事……

如果你看过 PAIP 或者其它的经典人工智能教材，就会发现这些机器对话系统，最初的思想来自一个叫“ELIZA”的 AI 程序。Eliza 被设计为一个心理医生，跟你对话排忧解难，而它内部其实就是一个类似小冰的句子搜索引擎，实现方式完全用正则表达式匹配搞定。比如，Eliza 的某个规则可以说，当用户说：“我(.*)”，那么你就回答：“我也$1……” 其中 $1 代替原句子里的一部分，造成一种“理解”的效果。比如用户也许会说：“我好无聊。” Eliza 就可以说：“我也好无聊……” 然后这两个无聊的人就惺惺相惜，有伴了。

有些清华的老朋友也许还记得，十多年前在清华的时候，我做了一个聊天机器人放在水木清华 BBS，红极一时，所以

我也可以算是网络聊天机器人的鼻祖了 :) 我的聊天机器人，水木账号叫 helloooo。helloooo 的性格像蜡笔小新，是一个调皮又好色的小男孩。

它内部采用的就是类似 Eliza 的做法，根本不理解句子，甚至连语料库都没有，神经网络也没有，里面就是一堆我事先写好的正则表达式“句型”而已。你输入一个句子，它匹配之后，从几种回复之中随机挑一个，所以你反复说同样的话，helloooo 的回答不会重复，如果你故意反复说同样的话，最后 helloooo 会对你说：“你怎么这么无聊啊？”或者“你有病啊？” 或者转移话题，或者暂时不理你…… 这样对方就不会明显感觉它是一个傻机器。

就是这么简单个东西。出乎我意料的是，helloooo 一上网就吸引了很多人。一传十十传百，每天都不停地有人发信息跟他聊。由于我给他设置的正则表达式和回复方式考虑到了人的心理，所以 helloooo 显得很“俏皮”，有时候还可能装傻，捣蛋，延迟回复，转移话题，还可能主动找你聊天，使用超过两句的小段子，…… 各种花样都有。最后，这个小色鬼赢得了好多妹子们的喜爱，甚至差点约了几个出去 :)

在这点上，helloooo 可比小冰强很多。小冰的技术含量虽然多一些，数据多很多，然而 helloooo 感觉更像一个有感觉的人，更受欢迎。这说明，我们其实不需要很高深的技术，不需要理解自然语言，只要你设计巧妙，抓住人的心理，就能做出人们喜爱的聊天机器。

后来，helloooo 终于引起了清华大学人智组研究生的兴趣，来问我：“你这里面使用的什么语料库做分析啊？”
我：“&%&￥@#@#%……”

自动编程是不可能的

现在回到有些人最开头的提议，实现自动编程系统。我现在可以很简单的告诉你，那是不可能实现的。微软的 Robust Fill 之类，全都是在扯淡。我对微软最近乘着 AI 热，各种煽风点火的做法，表示少许鄙视。不过微软的研究员也许知道这些东西的局限，只是国内小编在夸大它的功效吧。

你仔细看看他们举出的例子，就知道那是一个玩具问题。人给出少量例子，想要电脑完全正确的猜出他想做什么，那显然是不可能的。很简单的原因，例子不可能包含足够的信息，精确地表达人想要什么。最最简单的变换也许可以，然而只要多出那么一点点例外情况，你就完全没法猜出来他想干什么。就连人看到这些例子，都不知道另一个人想干什么，机器又如何知道？这根本就是想实现“读心术”。甚至人自己都可以是糊涂的，他根本不知道自己想干什么，机器又怎么猜得出来？所以这比读心术还要难！

对于如此弱智的问题，都不能 100% 正确的解决，遇到稍微有点逻辑的事情，就更没有希望了。论文最后还“高瞻远瞩”一下，提到要把这作法扩展到有“控制流”的情况，完全就是瞎扯。所以 RobustFill 所能做的，也就是让这种极其弱智的玩具问题，达到“接近 92% 的准确率”而已了。另外，这个 92% 是用什么标准算出来的，也很值得怀疑。

任何一个负责的程序语言专家都会告诉你，自动生成程序是根本不可能的事情。因为“读心术”是不可能实现的，所以要机器做事，人必须至少告诉机器自己“想要什么”，然而表达这个“想要什么”的难度，其实跟编程几乎是一样的。实际上程序员工作的本质，不就是在告诉电脑自己想要它干什么吗？最困难的工作（数据结构，算法，数据库系统）已经被固化到了库代码里面，然而表达“想要干什么”这个任务，是永远无法自动完成的，因为只有程序员自己才知道他想要什么，甚至他自己都要想很久，才知道自己想要什么……

有句话说得好：编程不过是一门失传的艺术的别名，这门艺术的名字叫做“思考”。没有任何机器可以代替人的思考，所以程序员是一种不可被机器取代的工作。虽然好的编程工具可以让程序员工作更加舒心和高效，任何试图取代程序员工作，节省编程劳力开销，克扣程序员待遇，试图把他们变成“可替换原件”的做法（比如 Agile，TDD），最终都会倒戈，使得雇主收到适得其反的后果。同样的原理也适用于其它的创造性工作：厨师，发型师，画家，……

所以别妄想自动编程了。节省程序员开销唯一的办法，是邀请优秀的程序员，尊重他们，给他们好的待遇，让他们开心安逸的生活和工作。同时，开掉那些满口“Agile”，“Scrum”，“TDD”，“软件工程”，光说不做的扯淡管理者，他们才是真正浪费公司资源，降低开发效率和软件质量的祸根。

我的人工智能梦

回过头来，很多人可能不知道，我也曾经是一个“AI 狂热者”。我也曾经为人工智能疯狂，把它作为自己的“伟大理想”。我也曾经张口闭口拿“人类”说事，仿佛机器是可以跟人类相提并论，甚至高于人类的。当深蓝电脑战胜卡斯帕罗夫，我也曾经感叹：“啊，我们人类完蛋了！” 我也曾经以为，有了“逻辑”和“学习”这两个法（kou）宝（hao），机器总有一天会超越人类的智能。可是我没有想清楚这具体要怎么实现，也没有想清楚实现了它到底有什么意义。

故事要从十多年前讲起，那时候人工智能正处于它的冬天。在清华大学的图书馆，我偶然地发现了一本尘封已久的『Paradigms of Artificial Intelligence Programming』（PAIP），作者是 Peter Norvig。像个考古学家一样，我开始逐一地琢磨和实现其中的各种经典 AI 算法。PAIP 的算法侧重于逻辑和推理，因为在它的年代，很多 AI 研究者都以为人类的智能，归根结底就是逻辑推理。

他们天真地以为，有了谓词逻辑，一阶逻辑这些东西，可以表达“因为所以不但而且存在所有”，机器就可以拥有智能。于是他们设计了各种基于逻辑的算法，专家系统（expert system），甚至设计了基于逻辑的程序语言 Prolog，把它叫做“第五代程序语言”。最后，他们遇到了无法逾越的障碍，众多的 AI 公司无法实现他们夸口的目标，各种基于“神经元”的机器无法解决实际的问题，巨额的政府和民间投资化为泡影，人工智能进入了冬天。

我就是在那样一个冬天遇到了 PAIP。它虽然没能让我投身于人工智能领域，却让我迷上了 Lisp 和程序语言。也是因为这本书，我第一次轻松而有章法的实现了 A* 等算法。我第一次理解到了程序的“模块化”是什么，在代码例子的引导下，我开始在自己的程序里使用小的“工具函数”，而不再忧心忡忡于“函数调用开销”。PAIP 和 SICP 这两本书，最后导致了我投身于更加“基础”的程序语言领域，而不是人工智能。

在 PAIP 之后，我又迷了一阵子机器学习（machine learning），因为有人告诉我，机器学习是人工智能的新篇章。然而我逐渐意识到，所谓的人工智能和机器学习，跟真正的人类智能，关系其实不大。相对于实际的问题，PAIP 里面的经典算法要么相当幼稚，要么复杂度很高，不能解决实际的问题。最重要的问题是，我看不出 PAIP 里面的算法跟“智能”有什么关系。而“机器学习”这个名字，基本是一个幌子。很多人都看出来了，机器学习说白了就是统计学里面的“拟合函数”，换了一个具有迷惑性的名字而已。

人工智能的研究者们总是喜欢抬出“神经元”一类的名词来吓人，跟你说他们的算法是受了人脑神经元工作原理的启发。注意了，“启发”是一个非常模棱两可的词，由一个东西启发得来的结果，可以跟这个东西毫不相干。比如我也可以说，Yin 语言的设计是受了九 yin 真经的启发 :P

世界上这么多 AI 研究者，有几个真的研究过人脑，解刨过人脑，拿它做过实验，或者读过脑科学的研究成果？最后你发现，几乎没有 AI 研究者真正做过人脑或者认知科学的研究。著名的认知科学家 Douglas Hofstadter 早就在接受采访时指出，这帮所谓“AI 专家”，对人脑和意识（mind）是怎么工作的，其实完全不感兴趣，也从来没有深入研究过，却号称要实现“通用人工智能”（Artificial General Intelligence, AGI），这就是为什么 AI 直到今天都只是一个虚无的梦想。

傻机器的价值

我不反对继续投资研究那些有实用价值的人工智能（比如人脸识别一类的），然而我觉得不应该过度夸大它的用处，把注意力过分集中在它上面，仿佛那是唯一可以做的事情，仿佛那是一个划时代的革命，仿佛它将取代一切人类劳动。

我的个人兴趣其实不在人工智能上面。那我要怎么创业呢？很简单，我觉得大部分人不需要很“智能”的机器，“傻机器”才是对人最有价值的，我们其实远远没有开发完傻机器的潜力。所以设计新的，可靠的，造福于人的傻机器，应该是我创业的目标。当然我这里所谓的“机器”，包括了硬件和软件，甚至可以包括云计算，大数据等内容。

只举一个例子，有些 AI 公司想研制“机器佣人”，可以自动打扫卫生做家务。我觉得这问题几乎不可能解决，还不如直接请真正智能的——阿姨来帮忙。我可以做一个阿姨服务平台，方便需要服务的家庭和阿姨进行牵线搭桥。给阿姨配备更好的工具，通信，日程，支付设施，让她工作不累收钱又方便。另外给家庭提供关于阿姨工作的反馈信息，让家庭也省心放心，那岂不是两全其美？哪里需要什么智能机器人，难度又高，又贵又不好用。显然这样的阿姨服务平台，结合真正的人的智能，轻而易举就可以让那些机器佣人公司死在萌芽之中。

当然我可能不会真去做个阿姨服务平台，这种东西可能已经有了。我只是举个例子，说明许许多多对人有用的傻机器，还在等着我们去发明。这些机器设计起来虽然需要灵机一动，然而实现起来难度却不高，给人带来便利，经济上见效也快。利用人的智慧，加上机器的蛮力，让人们又省力又能挣钱，才是最合理的发展方向。

（如果你喜欢这篇文章，欢迎付款。）