有关人工智能的几个认识问题

李国杰

2021.06

开场白

- 1999年我担任中科院计算所所长前曾去问候科学院的老领导张劲夫，事后他给朱镕基总理写信。

  松下问童子，言师采药去。

  只在此山中，云深不知处。

- 30多年前人工智能第二波高潮时，我也算是“弄潮儿”之一。但我早已不在人工智能科研和产业化的第一线。现在只能讲几点“过来人”的观感与认识，仅供大家参考。不对之处，请批评指正。

- 准备本报告时受到潘云鹤、张钹、李德毅、高文、王飞跃、王坚、鄂维南、金观涛、卜东波、米歇尔、泰格马克、李飞飞等诸多国内外学者文章报告的启发，在此一并表示感谢。

目录

- 人工智能技术究竟已发展到什么水平？
- GPT-3和AlphaFold是否代表了人工智能的发展方向？
- 通用性和可解释性是不是当前最重要的研究方向？
- 符号主义与连接主义融合的前景如何？
- 新一代人工智能到底“新”在哪里？
- 发展人工智能仍要坚持“顶天立地”战略
- 发展人工智能应有更理性的态度

人工智能技术究竟已发展到什么水平？

斯坦福大学2021年人工智能指数报告

- AI系统可以合成高质量的文本、语音和图像，甚至人类都很难辨别真伪。过去十年计算机视觉研究取得了巨大进展，已实现产业化。
- 自然语言处理(NLP)近年来进展较快。得益于NLP的快速发展，已经出现了语言能力显著提升的人工智能系统，这些系统已经开始产生了有意义的经济影响。
- DeepMind 的 AlphaFold 应用深度学习技术在蛋白质折叠生物学挑战中获得重大突破。药物发现是 2020 年私人 AI 投资额最大的一个项目，超过 138 亿美元。
- AI 最近取得的进展和突破为企业提供了大量利益和机遇，从自动化提高生产率、使用算法为消费者定制产品到大规模分析数据等等。 然而，企业必须注意采取措施来降低使用 AI 的风险。

2020年AI研究10大成就
智源人工智能研究院评选

1、OpenAI发布全球规模**最大的预训练语言模型GPT-3**

2、DeepMind的**AlphaFold2** 破解蛋白质结构预测难题

3、深度势能**分子动力学研究**获得**戈登·贝尔奖**

（深度学习用于基础科学研究，将第一性原理精度分子动力学模拟规模扩展到1亿原子，计算效率相比此前人类最好水平提升1000倍以上）

4、DeepMind等用深度神经网络**求解薛定谔方程**促进量子化学发展

5、美国贝勒医学院通过动态颅内电刺激实现高效率“**视皮层打印**”，（斯坦福团队 脑中想的文字可在屏幕上显示，每分钟90字符）

2020年AI研究10大成就（续）

智源人工智能研究院评选

6、清华大学首次提出**类脑计算完备性**概念及计算系统层次结构，

7、北京大学首次实现**基于相变存储器的神经网络高速训练系统**

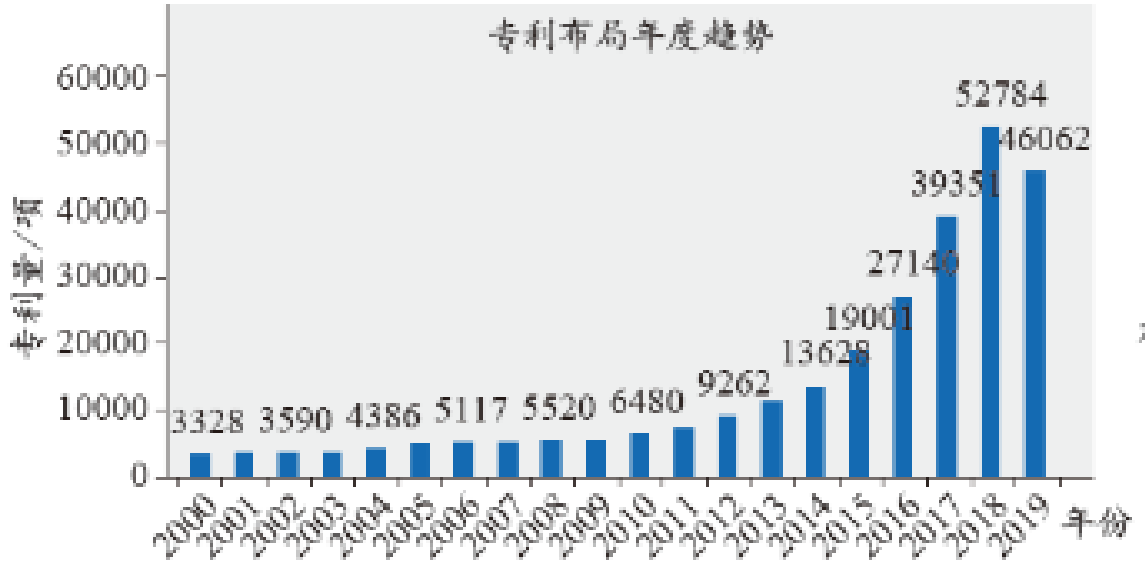
8、**MIT仅用19个类脑神经元实现控制自动驾驶汽车**

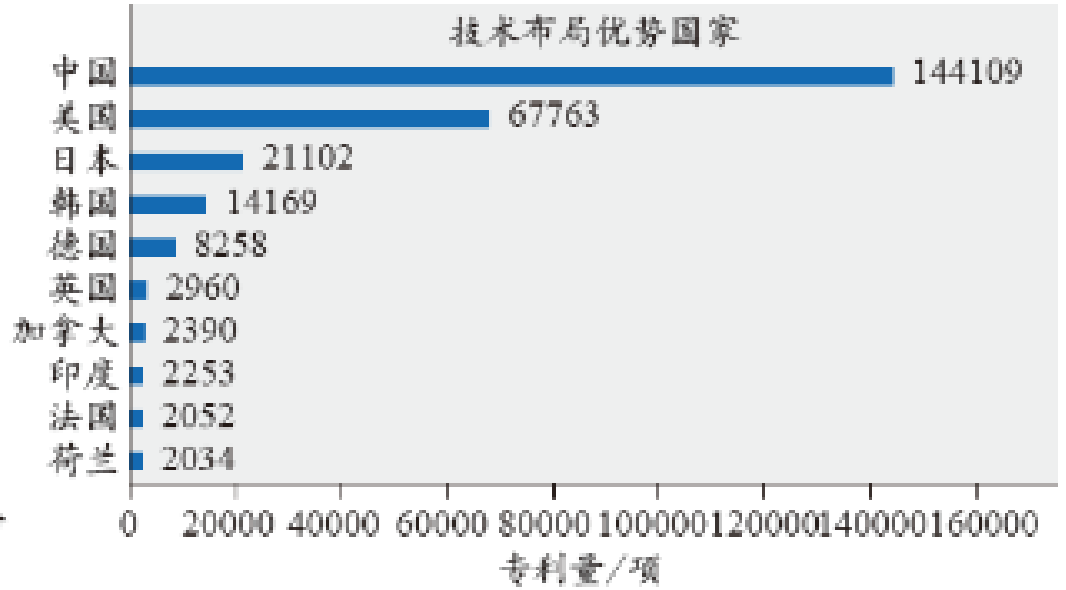
9、Google与FaceBook团队分别提出全新无监督表征学习算法

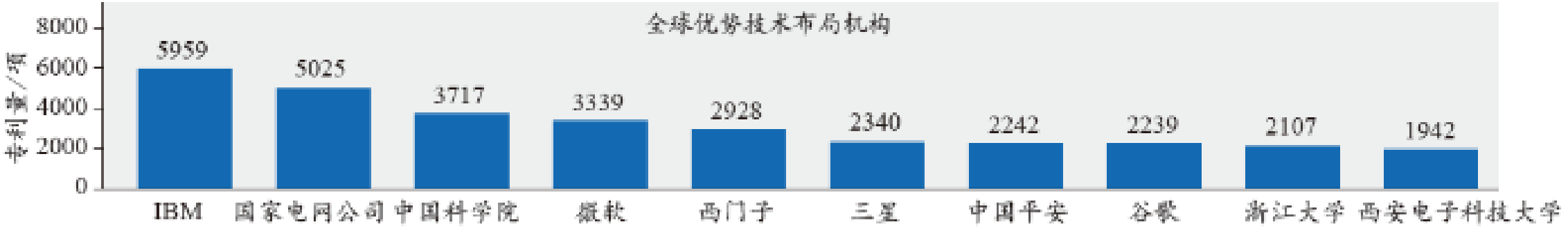
10、康奈尔大学提出无偏公平排序模型可缓解检索排名的马太效应问题

- 智源人工智能研究院评选的成就还不够全面。算法“自造”，即AI 自动生成学习算法应该是2020年的重大进展。DeepMind推出的LPG （Learned Policy Gradient）拥有元学习模型，具有学会学习的能力。李飞飞等提出了一种新型计算框架——深度进化强化学习，能够在环境、形态和控制这三种复杂度维度下同时规模化创建具身智能体。这是算法背后的算法，可以发现“要预测的内容”，从而形成其特有的值函数。谷歌大脑部门的工程师在2020年展示了AutoML，在没有人类干预的情况下产生自己独特的神经网络。

全球机器学习专利态势


专利布局年度趋势
专利量/项
60000
50000
40000
30000
20000
10000
0
3328
3590
4386
5117
5520
6480
9262
13628
19001
27140
39351
52784
46062
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
年份


技术布局优势国家
中国
美国
日本
韩国
德国
英国
加拿大
印度
法国
荷兰
144109
67763
21102
14169
8258
2960
2390
2253
2052
2034
0
20000
40000
60000
80000
100000
120000
140000
160000
专利量/项


全球优势技术布局机构
专利量/项
8000
6000
4000
2000
0
5959
5025
3717
3339
2928
2340
2242
2239
2107
1942
IBM
国家电网公司
中国科学院
微软
西门子
三星
中国平安
谷歌
浙江大学
西安电子科技大学

机器学习全球 TOP10 机构

机构	专利申请量	专利申请量 全球排名	有效专利量	有效专利占比	
IBM	5959	1	2349	39.42%	
微软	3339	4	1569	46.99%	
中国科学院	3717	3	1216	32.71%	
国家电网公司	5025	2	1020	20.30%	
谷歌	2239	8	997	44.53%	
西门子	2928	5	840	28.69%	
西安电子科大	1942	10	699	35.99%	
三星	2340	6	521	22.26%	
浙江大学	2017	9	516	24.49%	
中国平安	2242	7	63	2.81%	

要充分肯定深度学习技术的历史性作用

- 人工神经网络研究从1943年算起，已有近80年历史。2012年卷积神经网络等深度学习技术的崛起推动了人工智能技术的大规模产业化应用，这是人工智能发展史上里程碑式的事件，应高度肯定深度学习技术的历史性作用。
- 数学家鄂维南指出：通过对高维函数的拟合，数学上可以推导出两层神经网络和 梯度下降法。物理学家泰格马克指出：深度神经网络的形成符合重要的物理学原理。
- 加州大学伯克利分校马毅教授最近发文指出，从数据压缩(和群不变性)的角度提供了对深度(卷积)网络的完全“白盒”解释，展示了深层架构、线性算子等所有参数都可以从最大化速率缩减(具有群不变性)的原则推导出来。深度学习技术在计算机视觉和自然语言理解等领域取得成功有深层次的原因。

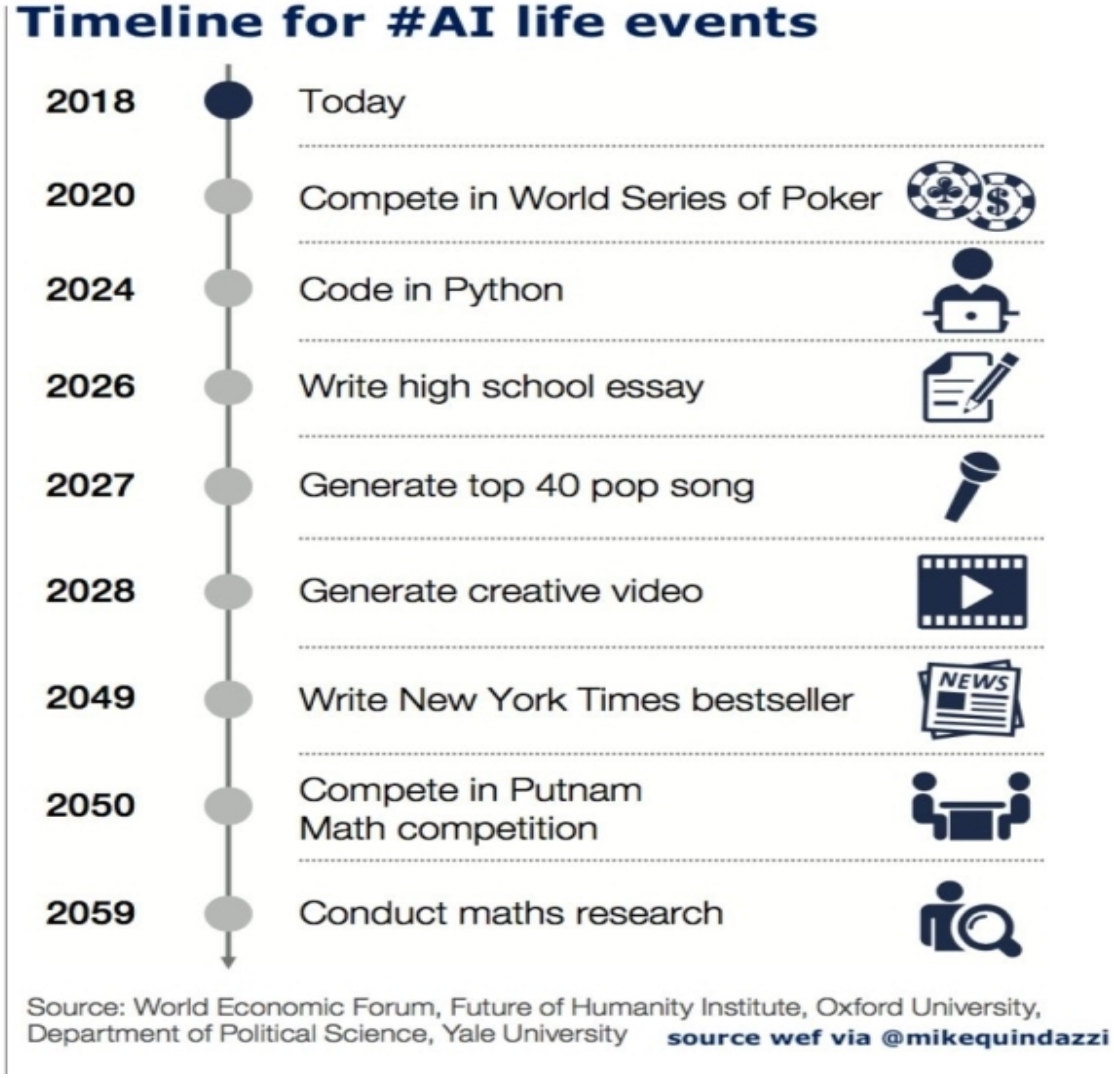
人工智能应用效果举例

- **大渡河水电公司**。根据存水、发电、电价和需求各种不同的情况综合进行优化调度。这三个水电站每年增加的发电量1.2亿千瓦。2018年四川的洪水非常厉害，他们利用这个模式，不但安全地运行，而且发电量大大增加。
- **海康威视。**里面已经装了智能芯片，不仅可以告诉你video，而且还可以告诉你很多数字，比如说车牌号码等，追查交通事故时非常方便。
- **大疆公司。** VR眼镜背后有一个头旋转的传感器，智慧摄像机随着头而转动。
- **盲人眼镜**。两个摄像头，一个管视觉，一个测距离，用声音告诉盲人，比导盲犬使用更方便
- **自主智能装备。** 比机器人更快的是无人系统。快递公司智能仓库自动分捡

目前人工智能最热门的研究方向

- 不可解释的深度机器学习支持了今天人工智能应用的繁荣
  - MIT通过深度学习对2300多种抗生素进行训练，研发出解除耐药性的新抗生素
  - Google的用AlphaFold做蛋白质折叠预测，研制出抗疫药物。
  - 英国基于机器学习研发人工智能控制的吸气式高超声速六代战斗机(暴风雨)，最大速度5马赫，是歼-20速度的2倍多，2035年前服役。
- 正在兴起的4个热门研究方向：
  - 可解释的机器学习（DARPA 的XAI项目）
  - 脉冲神经网络和神经拟态计算机
  - 基于知识图谱的大规模语义网络 （GPT-3）
  - 脑机接口（用于帕金森病，意念写字等）

世界经济论坛发布的人工智能技术的时间表



- 2020年 世界扑克比赛取胜
- 2024年 用Python编程
- 2026年 写高中水平作文
- 2027年 TOP 40 流行歌曲
- 2028年 有创意的视频
- 2049年 写纽约时代畅销书
- 2049年 在普特南数学竞赛中获胜
- 2059年 开展数学研究

人工智能研究的目标不应过分追求某项能力超过人，应更注重智能化的适应性和鲁棒性

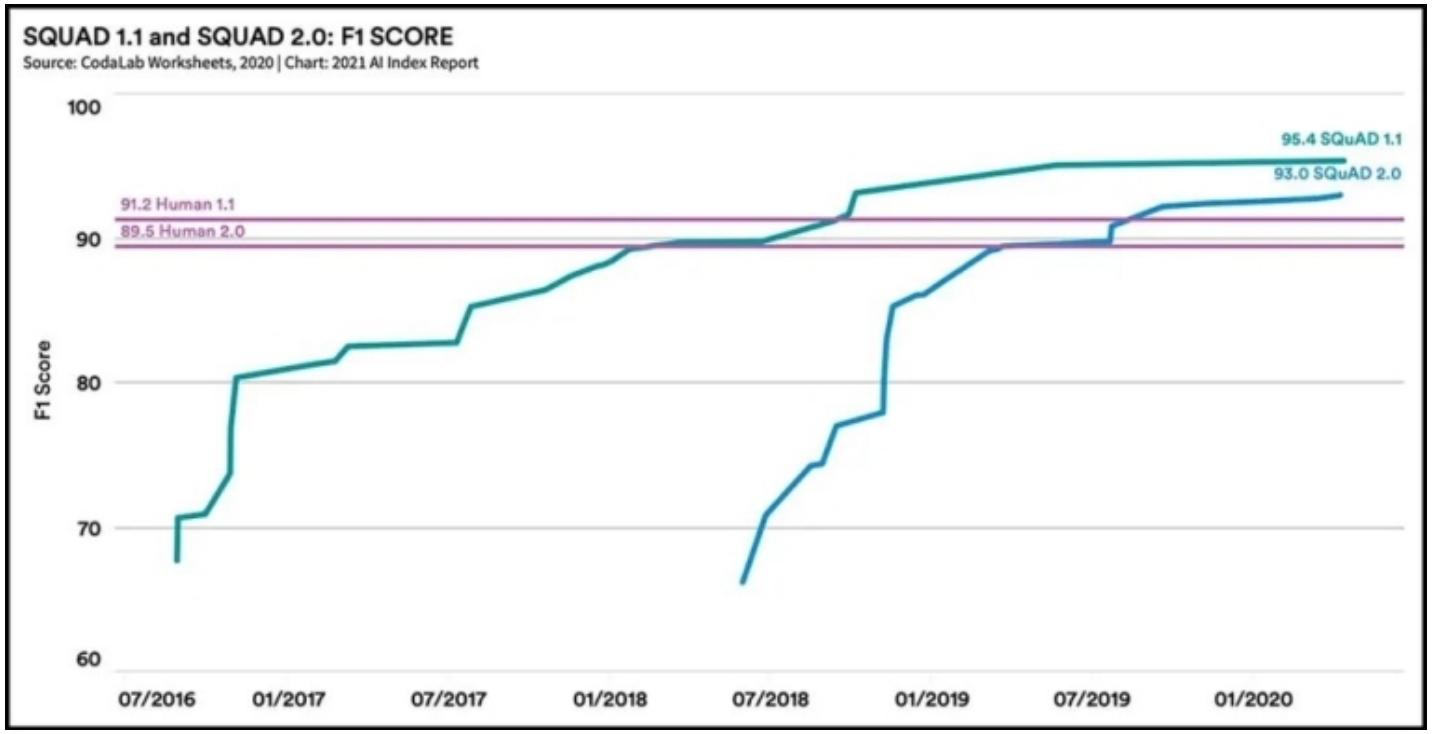
值得高度重视的两项人工智能技术

- 我认为，人工智能的目标主要是两个：一是计算机的智能化，二是人造物理世界（主要是工业世界）的智能化。
- 计算机的智能化离不开听说读写能力，核心是自然语言理解。广泛认为这是人工智能最困难和最具标志性的任务。近几年统计机器翻译采取了“不懂装懂”的模式，对“理解”绕道而行，在语义层面实际上后撤了一大步。
- 人造物理世界的智能化除了计算机视觉等感知技术外，最重要的应该是机器的自动（辅助）编程技术，或者说机器控制程序的自动进化技术。早年遗传编程曾红火过一阵。相对于深度学习，国内学者对机器自动编程几乎无动于衷。这应该是我国人工智能发展布局上的一大失误。
- 上世纪80年末到90年代初，在863计划支持下，国内软件界的许多学者都投入“软件自动化”课题研究，但后来不了了之。近年来中科院计算所卜东波等人开始做计算机辅助算法设计研究，自动学习出“贪心算法”，让神经网络自动生成可读的Prolog 代码。

自然语言理解取得较大进展

- SQuAD（Stanford Question Answering Dataset）是行业内公认的机器阅读理解领域的顶级水平测试，相当于机器阅读理解领域的 ImageNet。包含十万个问题的大规模机器阅读理解数据集，选取超过 500 篇的维基百科文章。机器在阅读数据集内的文章后，通过回答与文章内容相关的问题来测试。
- 2018年，斯坦福大学著名的机器阅读理解赛事SQuAD，阿里巴巴曾凭借82.440的精准率打破了世界纪录，超越了人类82.304的平均得分。2018年11月，谷歌发布的BERT模型，在机器阅读理解顶级水平测试SQuAD1.1中，两个衡量指标上超越人类，在11种不同NLP测试中创出最佳成绩。
- 在微软、谷歌等公司都参加的全球自然语言理解GLUE顶级赛事中，2019年百度研制的自然语言理解平台文心 ERNIE以90.1的成绩获得第一，率先突破90分大关，超越人类平均水平（87.1分）。2020年阿里巴巴达摩院以90.3分的成绩夺冠。2021年，百度以 90.9 的分数再登榜首。

在SQuAD 2.0比赛中
机器阅读理解全部指标超越人类平均



- SQuAD 2.0中不仅包含十万个问题-答案对，还有超过五万个由人类众包者对抗性地设计的难以回答的问题。

目前计算机“理解”自然语言的实际水平

- 以下三句话代表理解自然语言的三种不同水平：
  - 中国男子足球队正在和韩国队进行比赛 → 一群男性正在进行体育活动
  - 百米成绩，张三15秒23，李四14秒56 → 张三比李四跑得慢
  - 把3GB的文件读入Java字节数组 → JVM可能发生内存溢出
- 第一段推理是对左侧句子的简单概括。第二段推理涉及“慢”这个概念和时间、距离的关系。第三段推理要求理解Java的机制：Java数组的索引类型是32位整数，因此字节数组最多储存约2GB数据。整个推理过程衍生出了更多的概念和关系。
- 目前NLP技术能解决第一类推理，可以在特定范围内解决第二类推理，基本无法解决第三类推理。第三类推理中常识和领域知识界限已模糊。

威诺格拉德常识测试

- 2016年7月开始的“威诺格拉德模式挑战赛”（Winograd Schema Challenge，WSC）是图灵测试的变种，要求人工智能回答关于语句理解的**常识性问题**，主要测试模糊指代，由加拿大多伦多大学的计算机科学家赫克托·莱维斯克发起。这次竞赛的结果显示，人工智能最好的结果也只是48%（人类随机选择的情况下，该问题答对的概率是45%）。
- 测试题举例如下：
  - 市议员们拒绝示威者的游行许可，因为他们害怕暴力
  - Babar 不知怎样才能得到新衣裳。幸运的是，一个一直喜爱他的阔老头立刻看出来了，他想要一身漂亮的套装。因为他喜欢带给别人开心，他就为他拿出了他的钱包。
- 2020年GPT-3的WSC达到88.3%，BERT 达到84.6%。一篇提出WINOGRANDE测试（WSC的变种）的论文荣获了 AAAI 2020 最佳论文奖。WINOGRANDE 是一个有 44000 个问题的大规模数据集，在规模和难度上比WSC数据集更大，说明我们现在高估了模型的常识推理的能力。

机器自动编程

- 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，因为只有程序员自己才知道他想要什么。高水平的程序员是难以被机器取代的工作。
- Intel与MIT等共同推出了新的机器编程系统，称为机器推断代码相似系统性（machine inferred code similarity，MISIM），目标是使每个人都能表达自己的意图创建软件，目前识别相似代码段的精确度已提高40倍。
- 机器编程主要有两种方法，一种是形式化方法，一种是随机方法，目前业界更重视随机的方法。在有些案例中，英特尔已做到开发软件的时间减少到千分之一，三年开发的软件，借助机器编程只需要花费一天就可以完成。
- 计算机编程出现两种对立的趋势。一方面，计算资源越来越异构，需要了解硬件并最大限度使用硬件的专家级程序员；另一方面，软件开发人员越来越青睐于使用更抽象的语言，以提高工作效率，导致硬件难以发挥出它本身的性能。我们对程序员的要求过高，当前开发软件的方式难以持续发展。

IBM 放弃Watson智慧医疗系统的启示

- 2011年美国智力竞猜节目《危险边缘》中，IBM的Watson计算机击败两位最优秀的选手后，IBM 将医疗作为人工智能科研转化的核心，启动了Watson Health项目。烧掉数百亿美元，年营收只有10亿美元，十年内也没有通过FDA的审批，仍然无法有效应用于临床。最近IBM打算出售Watson业务。
- 高质量数据稀缺（最多的肺癌也仅有635例），存在认知偏见，缺乏有效的逻辑推理，缺乏质控和本土化优化，价格昂贵（4500元/例）。 目前更像是“AI图书馆员”，不会创造新的知识，不能提供医生想不到的方案。
- Watson不成功的案例说明，对正确性和安全要求很高、环境复杂的行业，AI应用还有很长的路要走。应该根据具体的应用场景定制需求，先用成熟的AI技术解决某些痛点问题，而不是强行将AI算法嵌入现有业务流程之中。

人工智能是处在夏天还是正在进入寒冬？

- 人工智能已经经历过两次严冬，会不会再度经历一个冬天？之前人工智能遭遇寒冬，是因为当时的技术还不能创造较大的经济价值。第三波人工智能的兴起不是来自学术界，而是企业界的驱动。本质上不是人工智能突破了什么关键技术，而是数字化的普及产生了智能化的需求。只要智能化的需求旺盛，学术界不像前两次那样盲目乐观，人工智能就不会马上进入冬天。
- 与其他技术一样，近几年人工智能也经历了过度炒作的泡沫期（夏天？），只要学术界和企业界能理性地反思，扎扎实实地在在合适的场景推广合适的智能技术，人工智能可能会度过一段较长时间的秋天。如果人工智能技术在10年之内在引领数字经济发展还达不到预期的效果，可能又会降温一段时间。
- 技术波浪式的发展是平常事，技术发展也有其自身内在的规律，我们只能顺势而为，做技术发展的促进派。过高的期望、不切实际的承诺只会加速寒冬的到来。

GPT-3 和 AlphaFold
是否代表了人工智能的发展方向？

机器学习的标志性成果：GPT-3

- 2015年马斯克等硅谷科技大亨共同创建非营利组织OpenAI，后来获得微软10亿美元的投资，转变为以盈利为目标。2020年5月OpenAI发布了无监督的转化语言模型GPT-3（Generative Pre-trained Transformer，生成性预训练变换器）。 这个模型包含的1750亿个参数，训练数据量达到了45TB（1万亿单词量），在语义搜索、文本生成、内容理解、机器翻译等方面取得重大突破。GPT-3最大的价值是无监督下的自我学习能力，只要用户提供少于10个培训示例，证明了通过扩大规模可以实现性能提升。GPT-3对外已开放API，但GPT-3仍然存在巨大的知识空白，距离真正的商业化还有一定的距离。
- 2021年一开始，OpenAI在GPT-3方向上又实现重要突破，它可以成功跨界，按照文字描述、生成对应图片。新的AI叫做DALL·E，它是被重新训练过的120亿参数版的GPT-3，参数只有GPT-3 的1/14。如输入命令：“一颗白菜穿着芭蕾舞裙在遛狗”，输出结果是：

GPT-3的局限性

- GPT-3需要巨大的计算量。Microsoft和OpenAI合力开发了一款超级计算机，专门用于GPT-3模型训练，这款超级计算机拥有超过 285000 个 CPU 内核、10000 个 GPU 和 400Gbps 的网络连接。在TOP500 超级计算机中可排名 第5位。Open AI 训练1750亿参数要花费**1200万美元。**
- 一味增加参数，只是表面上“懂”了许多事情。GPT-3还缺乏常识，有时会犯愚蠢的错误。例如， 如果问：“太阳有 几只眼睛？”时，GPT-3会给出回答：“太阳有一只眼睛”。
- 运用巨大的算力不是人工智能发展的唯一方向。探索人脑的奥妙机理，实现小数据学习、举一反三的迁移学习也是重要的研究方向。大脑的功耗只有20W，实现低能耗的智能系统是更重要的努力方向。

深度学习参数的指数型增长

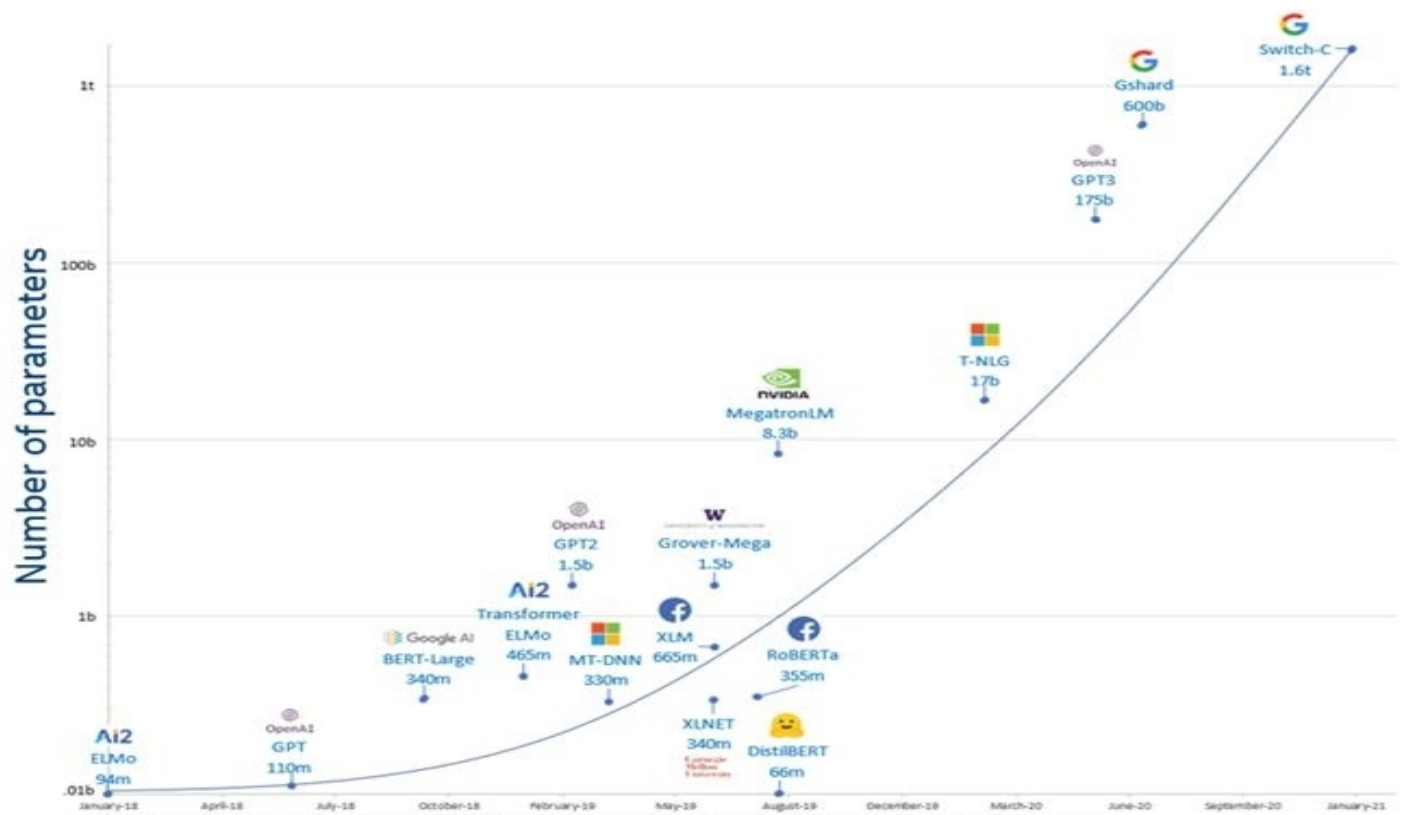


Figure 1: Exponential growth of number of parameters in DL models

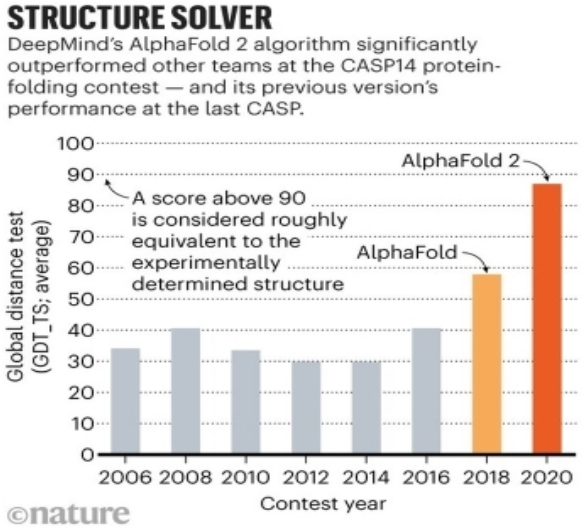
- 2018年9月
  GPT 1.1亿
- 2020年6月
  GPT-3 1750亿
- 2020年8月
  Gshard 6000亿
- 2021年1月
  Switch-C 16000亿

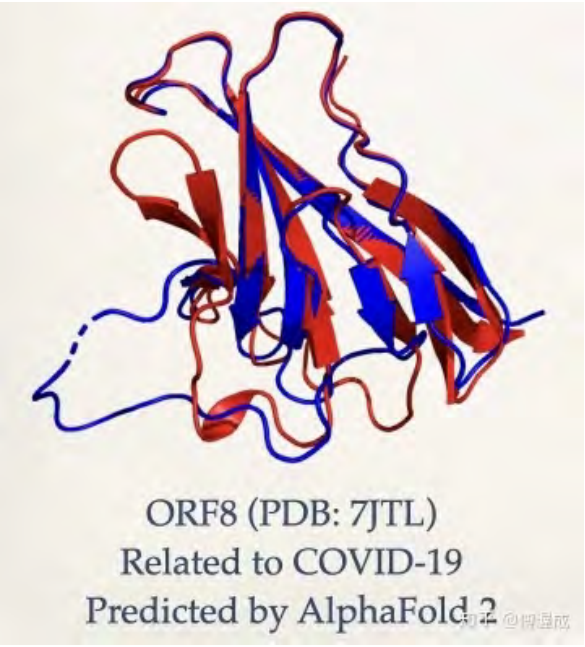
减少深度学习对数据的依赖性

- 减少深度学习对数据的依赖性，已经成为AI研究最重要的探索方向之一。深度学习并不只是监督学习，也不只是神经网络。深度学习是将参数化的模块组装到计算图中构建AI系统。它的优势在于不需要对系统进行直接编程，只需要定义架构并调整参数。不过其中需要调整的参数可能多达数十亿之巨。
- 自监督学习类似无监督学习，都是从没有明确标签的数据中进行学习。但无监督学习侧重于学习数据的内在关系和结构，而自监督学习算法是通过揭示数据各部分之间的关系，从数据中生成标签。
- AlphaZero 采用的强化学习还极少在实践场景中得到应用。在工业界，目前也极少应用大规模Transformer，因为工业模型太庞大，数据太复杂，算力跟不上。2020年所有顶会的最佳论文，没有一篇能反映出通用机器学习的进步。目前各种测试集上的结果可表明新方法性能有提升，但忽略了这些算法在实际问题上的适用范围。

AlphaFold 蛋白质结构预测

- 谷歌旗下DeepMind公司开发的人工智能程序AlphaFold2，在2020年度蛋白质结构预测大赛CASP14中，对大部分蛋白质实现了原子精度的结构预测，取得了92.4分的高分。实现了蛋白质结构预测领域史无前例的巨大进步，这是人工智能技术的标志性应用成果。（计算所卜东波团队的CopulaNet预测精度达到70分， 超过AlphaFold 50分，低于AlphaFold2 ）
- AlphaFold是用人工智能方法做基础科学研究的成功探索，为人工智能研究开辟的新研究方向。
- DeepMind 最近公布了AlohaFold 生成的六种可能与新冠病毒有关的蛋白质结构预测结果，为新冠肺炎的治疗方案提供一个假设生成平台，可加速疫苗研制。 AlohaFold 将是新药研制和发现传统药物新功能的得力工具。





算法的巨大进步和局限性

- 人工智能的巨大进展体现为算法的改进。投入AI算法研究可以比硬件研发收益更高。自2012年以来，训练一个人工智能模型在基准测试ImageNet图像分类任务中达到同等的分类效果，所需的算力每16个月就会减少1/2。与2012年相比，训练神经网络达到AlexNet的水平所需的算力会已减少到1/44。
- 人工智能研究的重大难题不仅仅是好的学习算法而是研究机器行为，包括单机行为、机群行为、人—机行为三个层次的机器行为方式和准则。
- 算法的目标往往定义为效用函数，关键是要保证效用函数可以准确表达一个对社会有利的行为目标和对不利行为的限制，从而实现符合伦理的行为。人工智能的挑战重要的不是机器能做多少事，而是知道机器做的对不对！
- 人工智能的应用严重依赖算力，没有智能加速芯片和智能超算系统，算法就发挥不出预想的作用。

从“大数据、小任务”到“小数据、大任务”

- AlphaFold在方法学和概念理论上没有大的创新，还是黑匣子，只是用的资源比较多，工程能力比较强，是应用层面的突破和创新。
- 目前人工智能算法主要特点是实现“大数据、小任务”。希望在少量数据的情况下，任务不是那么具体的时候，取得“小数据、大任务”的成功。
- 李飞飞表示：我仍然认为我们的 AI 时代是牛顿物理学之前的时代。我们还在学习现象学和工程学。总有一天，我们会开始理解智能的原则。
- 杰夫·辛顿、李飞飞和吴恩达等AI领域的领军人物，都在呼吁重启AI，因为深度学习尚未被证明具有可扩展性。培养基于小数据的抽象能力和概述能力，是人类认知的核心。

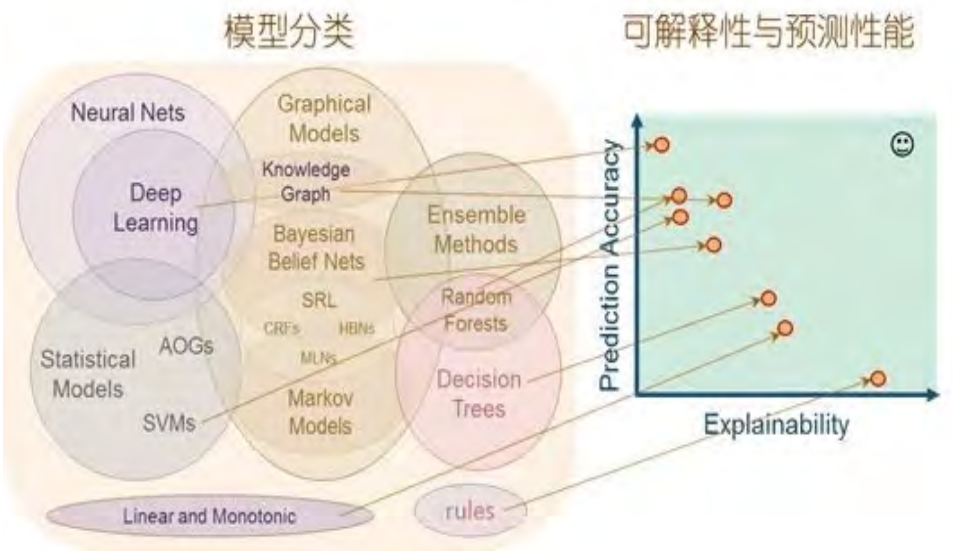
人工智能研究值得重视的几个方向

- 目前的机器学习是针对特定的、预先编程好的某个目的，只能说是算法、模型的优化，如何实现人造智能体的自动“进化”是体现智能的关键，要关注“用人工智能创造出更好的人工智能”，进化中的鲍德温效应值得重视。
- 一堆狭义智能永远不会堆砌成通用人工智能，“信息整合”是人脑涌现智慧的关键，要高度重视提高机器学习通用性的“迁移学习”。
- 人工智能的伦理、法律和监控技术研究是当务之急，智能技术的应用和对智能技术合理合规性的监控必须同步进行。
- 从基础研究的角度看，人工智能正在面临着一场研究范式的转换，基于图灵可计算概念的“认知可计算主义”研究纲领的局限已经显现。需要探索非图灵机（超越图灵机“计算”概念）来解决人工智能更深层的问题。

通用性和可解释性是不是当前最重要的研究方向？

需要加强人工智能系统可解释性的研究

- 可解释性是AI在关键领域得到应用的前提，不可解释的预测和决策在安全、金融、政府、医疗等关键部门不敢采用。但可解释性和预测精度是相互矛盾的两个维度，如同高性能和低功耗，通用性和高效率，难以得兼。
- 2017年，DARPA发起XAI项目，从可解释的机器学习系统、人机交互技术以及可解释的心理学理论三个方面，全面开展可解释性 AI 系统的研究。
- 可解释性也是分层次的，最严格的可解释性是数学。要求像数学一样从几条公理出发可能会扼杀人工智能研究。



鄂维南：从数学家的观点看，人工神经网络就是一种应对维数灾难的数学工具，与对人脑的模仿无关。

人工智能技术研究目前不必追求全领域通用

- 通用人工智能是 AI 研究的终极目标。这个目标何时能实现？在对全球23位顶尖AI学者的一次调查中，最乐观的专家给出的时间为2029年，最悲观的专家认为要到2200年。平均来看，时间点为2099年。我个人的看法这个时间取决于对“通用”的定义，如果要通过洞察人脑的奥秘实现所谓强人工智能，恐怕要到22世纪以后。
- 通过获取足够多的背景知识，让机器具有更丰富的常识，可以逐步提高人工智能系统的通用性。但近期内不必将追求像人脑一样的通用性作为研究目标。在一个领域内有足够的通用性就有很宽广的应用前景。对于人工智能研究，比最大程度的通用性更紧迫的研究目标包括：针对具体应用的应对复杂环境的鲁棒性和自适应性，智能系统的可解释性和安全性等
- 强人工智能追求的是智能纵向的深度，通用人工智能追求的是智能横向的宽度。纵向智能、横向智能都没有尽头。新一代人工智能既要通用，又要有专门领域或者多个专门领域的强智能。

通用性和可解释性不应是AI研究的首要目标

- 美国的AI研究计划将提高通用性和可解释性置于优先考虑的位置。但智能应用是多种多样的，不同的应用对通用性和可解释性的要求大不相同。即使是对安全性要求很高的应用，需要解释到什么程度也是有区别的。
- 人类智能本身也是一个黑箱，相比人类的大脑的不可解释，人工神经网络也许能解释更多决策的过程。至于深度神经网络的输出究竟如何形成的，用现有的知识解释每个参数的作用无济于事，需要创立新的理论才能做出解释，还需要等待下一个图灵一样的科学家。
- 实践是检验真理的标准，解释性弱的技术也会延续发展，例如中医。
- 对于人工智能，人们最担心的是不知道它什么时候会出现错误，比可解释性更重要的是人工智能的防错技术，要有科学根据地将出错率降到可接受的范围，特别是解决攻击性出错的问题。

要实现可解释性必须牺牲准确性吗？

- 一般规律中，模型的复杂度和准确性往往是正相关的关系，而越高的复杂度也意味着模型越可能无法解释。信任黑盒模型意味着你不仅要信任模型的方程式，而且也要信任它所基于的可能有偏见的整个数据库。
- 许多人认为：“必须牺牲一些可解释性才能获得准确性的模型”。这种观点未必正确。刑事司法系统已经反复证明，利用黑盒模型的复杂性预测未来的逮捕情况，其准确性远不及基于年龄和犯罪记录的简单预测模型。
- 可以考虑先训练出庞大、精确的、上百层的深度神经网络，再将深度神经网络压缩成较浅的神经网络，在保持它的准确率的同时提高可解释性。
- 大多数机器学习模型的设计没有可解释的约束条件，只是为了在静态数据集上为准确的预测变量而设计，有些应用可考虑增加可解释性的约束。

AI芯片要权衡芯片性能和通用性

- 每款新的AI芯片问世，往往只讲性能多高，不会细说芯片的适用范围。实际上通用性较强的芯片在某些单项性能上会低于专用芯片。但市场卖得多的还是通用性较强的芯片。
- 在AI领域，目前销量最高的还是NVIDIA的GPU芯片，GPU的垄断性主要是靠它的生态系统。 CUDA生态系统包括工具、库、应用程序和合作伙伴，CUDA库支持线性代数、图像和视频处理、深度学习和图形分析等应用。
- 不同的AI芯片推出的性能采用不同的测试标准，不能直接比较，在人工智能芯片和机器学习领域推出一套公认的性能测试Benchmark十分必要。
- 一款芯片要兼顾性能和通用性，选择哪些功能，舍弃哪些功能是一门大学问，中国芯片公司需要向Intel和NVIDIA 学习。

AI（训练）芯片比较

厂商	型号	峰值性能	功耗	能效	时间
Intel	NNP-T	110 Tops（FP16）	250W	0.44Tops/W	2019.04
AMD	M150(GPU)	26.5Tops (FP16)	300W	0.09Tops/W	2018.11
Nvidia	V100	312Tops (FP16)	400W	0.78Tops/W	2020.05
Google	TPU v3	90Tops (FP16)	200W	0.45Tops/W	2018.05
Google	TPU v4	待发布（2.7X）	待发布	待发布	2020.07
Grapecore	IPU GC200	250Tops (FP16)	300W	0.83Tops/W	2020.07
华为	Ascend 910	256Tops（FP16）	310W	0.83Tops/W	2019.08
寒武纪	MLU 270	128 TOPS（INT8）	70W	1.82TOPS/W	2019.06
寒武纪	MLU 290	1024 TOPS（INT4）	350W	2.93TOPS/W	2021.02

●Cerebras Systems公司2019年推出晶圆级芯片（WSE）面积422 平方厘米（英伟达GPU的56.7倍），拥有1.2 万亿个晶体管，40万个核心，片上内存18G字节，内存带宽19PByte/s，fabric带宽100Pbit/s。。

Cerebras Systems 的第二代晶圆级芯片

2.6万亿晶体管、85万个核心， 100%的良率，
7nm工艺，15KW功耗，2021年3季度推出，
每台CS-1系统 250万美元，
一台CS-1可以提供超过1000个高性能GPU的性能

Cerebras Wafer Scale			
AnandTech	Wafer Scale Engine Genl	Wafer Scale Engine Gen2	Increase
AI Cores	400,000	850,000	2.13x
Manufacturing	TSMC 16nm	TSMC 7nm	-
Launch Date	August 2019	Q3 2021	-
Die Size	46225 mm^2	46225 mm^2	-
Transistors	1200 billion	2600 billion	2.17x
(Density)	25.96 mTr/mm^2	56.246 mTr/mm^2	2.17x
On-board SRAM	18 GB	40 GB	2.22x
Memory Bandwidth	9 PB/s	20 PB/s	2.22x
Fabric Bandwidth	100 Pb/s	220 Pb/s	2.22x
Cost	$2 million+	*arm+leg*	?

- CS-1的功率是谷歌TPU v3芯片1/5，体积是它的1/30，速度是 TPU v3的3倍.
- 晶圆内部互联能否软件定义可重构?
- 内存过小，存在延迟或带宽瓶颈。 比GPU更专用的AI芯片，既不是基于x86，也不是基于Linux。

符号主义与连接主义融合的前景如何？

“符号主义”在人工智能发展中的历史作用

- 能否使用符号是人和动物的本质区别。早在19世纪，德国哲学家恩斯特·卡西勒指出人是会使用符号的动物。从动物向智人进化的分界线是发明符号和使用符号，这是人类大脑统演化中的“涌现”行为。
- Newell和Simon 提出的物理符号系统假设：对于一般智能而言，具备物理符号系统是一个充分必要的条件。所谓必要，就是任何表现出智能的系统都可以经过分析被证明是一个物理符号系统；所谓充分，就是任何足够大的物理符号系统都可以通过组织而表现出智能。
- 物理符号系统假设是符号主义的理论基础，符号主义确实为人工智能发展做出了历史性的贡献。但这一假设没有得到人工智能界的公认。人脑是不是“物理符号系统”至今也说不清楚。智能系统涉及信号、亚符号和符号的处理和转化，符号处理应该不是智能处理的全部内容。

自然选择了生命的杂乱而非逻辑的严谨

- 牛顿只用简单的3个定律就能描述各种物体的运动；麦克斯韦只用4个定律就能解释所有的电磁活动。但在认知领域，没办法找到一个明确的公式去描述人的智能。经过几十亿年的进化，我们的大脑最终形成了很多不同的应对环境的方法。自然选择了生命的杂乱而非逻辑的严谨。
- 物理符号系统模拟人类智能的方式：先把问题形式化了，再从中归纳出算法。但不是所有问题都能被形式化，有的问题就算被形式化了，也找不到对应的算法。
- 仅仅通过反馈学习的神经网络不可能具有创造和使用符号的能力，这就是主体的自由。如果没有自由的主体，只是对外来刺激做出反应，根本不会发明符号并用它来表达对象。使用语言的背后正是主体的出现和主体的自由，但对于推动这一巨变的机制，现在仍一无所知。

机器学习的局限性

- 现在机器学习的基本思想是，把学习的目的表达为一个效用函数，学习的过程就是用，通过大量数据的训练来优化这个效用函数，拟合出一个模型。所以，现在的机器学习实际上是一个优化问题。
- 下棋的目标是单一的，就是赢这盘棋，下棋的效用函数相对简单。但是人做的很多决策，并不是优化一个目标，还有很多因素要考虑，所以表达人的学习目标的效用函数非常复杂。
- 目前机器学习的局限性：没有全局的抽象能力；没有运用知识的能力；缺乏常识；无法理解感知的内容，缺乏可解释性；容易受到“对抗性扰动"攻击。
- 机器学习离不开Hinton教授提出的**反向传播**，但他断然宣称要放弃反向传播。未来机器学习还有很长的路要走。

机器学习与逻辑推理的结合

- 在人脑的认知系统中存在两个系统：System 1 （快系统）和 System 2（慢系统）。System 1 是一个直觉系统，可以通过人对相关信息的直觉匹配寻找答案；而 System 2 是一个分析系统，要通过一定的逻辑推理才能找到答案。从 System 1 到 System 2 的认知是深度学习未来发展的重要的方向
- 历史已证明，只靠基于人类知识和特定规则创建人工智能，往往会失败。逻辑推理要与机器学习相结合。要真正实现基于知识的推理，需要万亿级的常识知识库支持。四五十年前费根鲍姆做过的事情，也许现在要重做一遍，但是现在可以实现更大规模的常识知识图谱，用大规模的常识知识图谱来支撑深度学习，有可能实现更通用的 AI。
- 人工智能已经两起两落，第一波的主流是逻辑推理，第二波是基于人工神经网络的机器学习。普遍认为，未来的第三波将是机器学习与逻辑推理的有机融合，追求更加通用、更鲁棒、更具有可解释性的人工智能。

知识驱动和数据驱动结合的尝试

- 清华大学张钹团队提出第三代人工智能，试图充分地利用知识、数据、算法和算力四个因素，采用自编码神经网络等新技术，争取解决不完全信息、不确定和动态变化环境下的智能问题，解决随机应变、举一反三的难题。
- 知识驱动和数据驱动相结合，需要将离散的符号语义空间与连续的向量数据空间无缝地连接起来。一条出路是通过脑启发计算，将连续数据空间的向量提升到离散的语义空间，实现从感性到理性的提升。另一条出路是将符号嵌入到连续向量空间并保持语义不变性。
- 基于数据驱动的机器学习不能举一反三，鲁棒性差（易受噪声的影响），不能在学习过程中不断提高学习能力。目前机器学习与人类学习的机制和方法存在巨大的差异。为了提高机器学习的水平，需要与心理学、认知科学和神经科学结合，借鉴人类的学习机制，在学习中融入常识和推理，改变目前机器学习的方法与机制。

连接主义和符号主义结合十分困难

- 希望两种有互补性的技术结合起来是人们习惯的想法，但连接主义和符号主义的结合比常人想象的要困难的多。符号主义到连接主义再到可解释的连接主义是否定之否定螺旋式上升。简单地用符号逻辑解释深度神经网络，可能是走回头路。
- 人脑的智慧是感知到认知的“涌现”，之所以用涌现（emerge）这个词，是因为它过于复杂而无法用公式或任何确定的方式来表达。但理解从低层次的感知到高层次的逻辑推理，必须明白“涌现”如何发生。简单地互相借用另一个层次的某些思路或方法，难以实现真正的结合
- 涌现的特征是混沌性，理解它如何形成可能已超越目前人类的智力。人工智能的下一步发展要更加解放思想，跳出现有的符号主义和链接主义的框框，从神经科学、生物科学、人文科学等更广泛的领域获取灵感。

打通“不可言传”的两种概念体系

- 连接主义和符号主义的结合的困难还在于，深度神经网络中隐结点上发生的事情是不可言传的，因为隐结点可能并不表达我们使用的任何概念或概念组合，可能只有把认知过程分解成远比我们的概念体系细得多的碎片，再按另一种方式重新组合才能得到一点语义的蛛丝马迹。
- ‘不可言传’的隐结点的行为，说不定正是神经网络通过训练对世界所做的另一种概念化。符号主义是做显式推理，与连接主义的概念体系相差很远。每一套概念体系都无法言传另一体系中很容易言传的事情。中医所使用的理论概念，如脏、腑、经络等术语，在西医看来就象是深度神经网络中的‘隐结点’，所以中西医的结合十分困难。
- 是否能够通过增加网络层数或输入更多的训练数据就能实现真正的理解，尚未可知。

神经符号主义

- 1980年以来，很多有远见的人工智能学者就试图实现神经网络和符号智能的结合，这个方向被称为神经符号主义（Neural-symbolism）。
- 1990年，Towell 等人便提出了 KBANN（基于知识的人工神经网络），采用已有的经验知识去构建人工神经网络的结构和网络中的连接权重。
- 1999 年Garcez 等人提出了 CILP 系统，他们将背景知识转化为命题逻辑，并基于此构建前向人工神经网络。2001 年Garcez 等人 提出了一种在训练好的神经网络中抽取逻辑知识的方法，可以增强神经网络的可解释性。
- 2006 年Richardson 等人在 对一阶符号逻辑和概率图模型结合的方式进行了探索，提出了马尔科夫逻辑网络。
- 受当时机器学习技术和自然语言处理技术的制约，这些探索并不能充分利用神经网络的优势，因而没有取得更进一步的成功。

生物学计算

- 有人倡导的“生物学计算”是一种新的模式。它指的是人工培养的神经元在一个适宜的营养基中进行生物学意义的生长，根据构建系统的计算要求完成定型。有学者预言，以自然为基础的半人工智能40年之内可接近人脑水平。
- “生物学计算”既不同于作为抽象神经元计算机模型的人工神经网络，也不同于用于解决复杂问题的利用DNA分子的化学性质的DNA计算机，指的是人工培养的神经元，可以进行真正生物学上类似人脑那样的操作。
- 人脑是经过数百万年的演化逐步形成的。从这个角度来讲，在依靠人类设计之外，智能模型是否也能通过演化过程去自动发现最佳的模型结构？传统的遗传算法是一种基础的演化计算模型，我们能否借鉴自然系统中的智能行为，将其形式化为可计算的智能范式？

某些人文学者对人工智能的看法

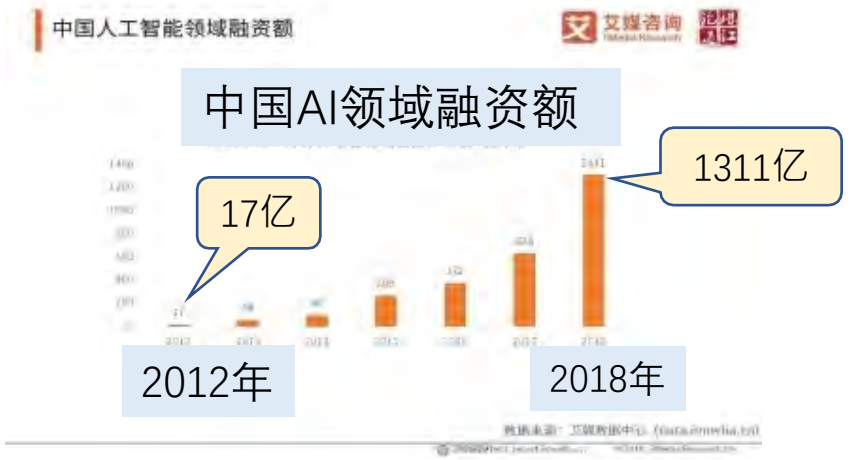
- 当前正在发生的这场人工智能革命，实质是一场“退回到原点的运动”。所有这一切都和人文精神的沦丧有关，特别是对人工智能的发展历史的忽略。今日忽略对行为主义的批评、回到控制论刚兴起时状态的“人工智能革命”，是某种对智能认识的大倒退。
- 因为神经网络模型对应的是生物本能，而不是人类智能。深度神经网络研究对应的是仿生学。所谓深度学习，其理论基础是行为主义和连接主义，这只是生物学意义上“自然选择”导致物种形成的过程。
- 数学符号系统运行规则背后还存在以下四个层面，它们才涉及智能的本质。第一，主体是自由的，可以给出符号并用符号系统指涉对象。第二，人用语言传递知识，组织社会，产生社会行动。第三，人会意识到自己有自由意志，创造一个应然世界；第四，应然世界的演变会进一步放大主体的自由，创造出更为复杂的符号系统。人工智能的符号主义学派只着眼第四个层面上的一小点（即数学推理），只是冰山之一角，当然不能把握智能。

新一代人工智能
到底“新”在哪里？

人工智能技术发展的分期

- 学术界多数学者将人工智能的发展分成三个阶段，从1956年开始前30年是符号主义主导，上世纪80年代以后进入多种研究纲领竞争的第二阶段。1986年Rumelhart和 McClelland两卷本PDP专著出版和反向传播提出，标志者联结主义研究纲领复兴，行为主义也开始出现。2006年以后深度学习为代表的机器学习开始成为主流，从21世纪第二个10年开始，AI进入高速发展的第三阶段。

- 三阶段的分期是从学术界的小圈子看人工智能，站在更高的高度，从全社会的宏观视角看人工智能的发展，前60年AI的发展总的来讲还是在“象牙塔”内发展。2015年全球AI市场规模只有200多亿美元（国内约200亿人民币）。2012年国内AI领域融资只有17亿元，2020增长到1311亿元，AI企业数也翻了两番。因此2016年工程院的报告称人工智能进入了AI2.0时代，这是指信息化的发展促使AI成为数字经济的领头雁。



工程院人工智能2.0建议的要点

- 人工智能2.0是基于重大变化的信息新环境和发展新目标的新一代人工智能。其中，信息新环境是指：互联网与移动终端的普及、传感网的渗透、大数据的涌现和网上社区的兴起等等。新目标是指：智能城市、智能经济、智能制造、智能医疗、智能家居、智能驾驶等从宏观到微观的智能化新需求。
- 人工智能2.0已显露大量新特征：一是大数据上的深度学习＋自我博弈进化技术。二是基于网络的群体智能已经萌芽，实现任务分配的众包模式；较复杂支持工作流模式的群智；最复杂的协同求解问题的生态系统模式。三是人机融合技术导向混合智能，生物智能系统与机器智能系统的紧密耦合。四是跨媒体智能已经兴起。在语言、视觉、图形和听觉之间语义贯通，实现联想、推理、概括等智能的重要关键。五是自主智能装备涌现，类人或类动物的机器人，往往不如对机械进行智能化和自主化升级来得高效。

新一代人工智能“新”在何处？

- 我们的世界原来是两元空间：人类社会—物理世界。近半个世纪以来，以计算机为代表的信息力量迅速壮大，形成了人类—计算机—物理世界三元空间。从另一个角度看，计算机诞生以后出现人—机二元世界，物联网的发展产生大量来自物理世界的数据和信息，大数据时代实际上是人-机-物三元世界。
- 进入新的三元空间以后；空间变化形成信息流的新变化；新的信息流会生成认知的新变化。这是人工智能走向2.0的本质原因。所谓的人工智能走向2.0，就是说在新空间的互动中间，可以产生出大量新的人工智能需求和新的技术。
- 智能城市、智能医疗、智能交通、无人驾驶、智能游戏、智能制造等等，和传统人工智能研究的对象不一样。，传统人工智能是模拟一个人的智能行为，AI2.0要求模拟人类社会和物理世界。

新一代人工智能
“新”在融入科学和经济的大舞台

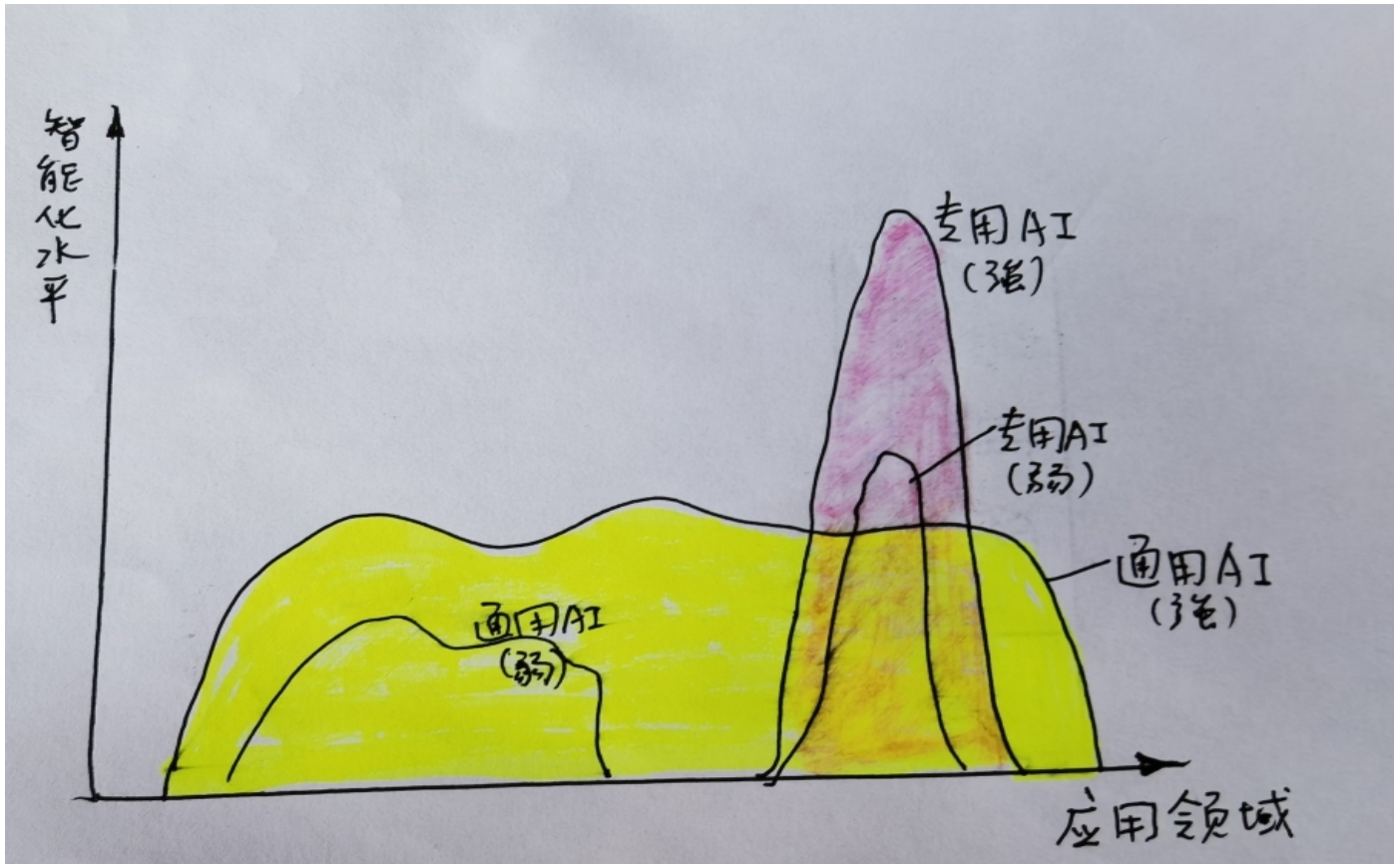
- 早期的AI研究大多在“八皇后”之类的“Toy Problem”上兜圈子，我的朋友顾钧首次求解“百万皇后”问题，被认为是人工智能研究的一次思想解放。现在AI可以作诗、画画、写小说，但这还是AI的小舞台。人工智能研究的知识表示、学习优化、问题求解都是通用的技术，可以融入数理化、天地生，尤其是与大数据结合，可以解决许多原来解决不了的问题。新一代AI就是要走出小舞台，奔向科学研究和数字经济的大舞台。
- 不要太在意哪里是人工智能自己的“一亩三分地”，去掉AI头上神圣的光环，放下AI高贵的身段，在生物学、社会科学、医学、工业制造等诸多领域甘当配角，让AI成为众多行业工攻坚克难的得力工具，引领各行各业向高端发展，这就是新一代人工智能的内在的含义。

发展人工智能仍要坚持“顶天立地”战略

究竟什么是强人工智能？

- 强人工智能有各种各样的定义：
  - ◆机器本身具有思维能力，有自我意识
  - ◆能完成任何人类可以做到的智力任务，即通用人工智能（AGI）
  - ◆超越人类的智能
  - ◆类脑智能（真正模拟人脑机制的人工智能）
  - ◆不用计算机实现的智能，非计算智能等等
- 大部分学者讲的强人工智能是指通用人工智能（Artificial General Intelligence），并不强调有自我意识的人工智能，认为计算机没有意识但它有智能，我们可以继续做没有意识但有智能的机器。
- 对意识问题的讨论是哲学家的事情，科技人员更应关注人工智能技术的探索及其合法、合理与合情的应用。

通用性和智能化水平是两个维度



- 通用性等价于强人工智能的看法已在学术界流行，但这并不是一种科学合理的分类。
- 通用性和智能化水平是两个维度，通用性有强弱，智能化水平也有强弱。与通用相对的是专用。我们要在扩展应用领域和提高智能化水平两个维度发力。

关于计算机不可能有意识的几种观点

- “广义歌德尔定理”。智能分为算法智能（AI）、语言智能（LI）和想象智能（II）三个层面，算法智能无法超越语言智能，语言智能又无法超越想象智能。计算机的智能只能是AI，无法达到人类所具有的LI和II层面，计算机不能真正理解语言和想象等相关的活动。（爱因斯：“智能的真正标识不是知识，而是想象。”）
- 卡普坦尼1997年证明，符号逻辑方法不能完全描述意识现象。
- 彭罗斯《皇帝新脑》：目前我们的物理理论，甚至包括量子力学还无法刻画意识的规律，需要一种建立在微观理论基础上的新的量子力学理论。
- 霍兰1998年《涌现》：我们目前还没有理论和模型能够清楚地表现意识的自涌现的现象。
- 必须意识到意识的解释不能排除意识。正因如此，在人类智能研究中，必须发展出一种将科学与哲学结合在一起的新方法即意识解释的递归方案。

人脑也许不具备了解自己的能力

- 当你想描述上百万件同时发生的事件时，语言就不是合适的工具了。有些学者认为：人类的大脑并不具备了解自己的能力。没有一台机器能够输出比它本身更精密的东西。与制造汽车的工程技术相比，汽车本身的复杂程度不值一提。同理，人类大脑所能做的事与其自身构造相比同样微不足道，
- 大脑可能过于复杂，人类无法在通俗的意义上理解它，而描述它可能是个更好的方向。但当你想描述上百万件同时发生的事件时，语言可能不是合适的工具了。
- 几十年前神经科学家就已经得到了完整的秀丽隐杆线虫神经链联接图，但还没有完全理解这个仅由300个神经元组成的生物。

以“4E”为标识的“新认知观”

- 上世纪末，一些认知科学家提出以“4E”为标识的所谓“新认知观”，即主张认知在基本方面并非是计算—表征的，而是具身的（embodied）、嵌入的（embedded）、延展的（extended）和生成的（enactive）。后来又加上“S” ，即情境(situated)认知，形成4E+S 理论模型。
- 其核心观点是：我们的心灵不仅存在于大脑之中，还可以存在于大脑之外。心灵是分布于大脑、身体和世界中的一系列复杂的状态、过程和行为。
- “4E+ S”理论曾被看作是一场新的认知科学革命，但目前看来，“4E+S”理论相对于标准认知科学来说，尚没有形成一场哥白尼式的革命，但值得我们重视。

对人脑应有足够的敬畏

- 认知科学的大量实验事实表明，认知的基本单元不是计算的符号，不是比特，而是一种整体性的“组块”（ chunk ）， Kahneman 提出注意选择的基本单元是整体性的“知觉物体”。
- 目前基于计算概念的理论，不能满意地解释诸多基本认知现象，如“大范围首先”、“不变性的直接知觉”、“认知偏向”、“意识涌现”等。认知和计算的关系只能通过实验来回答，认知科学本质上是实验科学。
- 对上百亿年宇宙演化形成的极为精巧的人脑应有足够的敬畏，破解人脑的奥秘可能需要几百年甚至更长的时间。短期内不要指望神经科学对人工智能的发展起到关键推动作用。
- 通用智能是人脑最本质的特征，在对人脑机制缺乏了解的情况下，短期内对实现像人脑一样通用的人工智能也不要抱过高的期望。要根据应用需求扩展人工智能的通用性。人类也不需要像人一样通用的产品。

发展人工智能仍要坚持“顶天立地”战略

- 1991年，在第一届全国人工智能与智能计算机学术会议上，我代表863-306专家组做特邀报告，提出了发展智能计算机的“顶天立地”战略：一方面要努力突破传统计算机甚至图灵机的限制，探索关于智能机的新概念、新理论与新方法；另一方面要充分挖掘传统计算机的潜力，在目前计算机主流技术基础上实现计算机智能化。
- 我认为，30年后的今天发展人工智能技术仍要坚持“顶天立地”战略，采取“弱人工智能” 和“强人工智能”两条腿走路的方针。 “强人工智能”还处在基础研究阶段，要解放思想，争取“广种奇收”，近10年内不要对其产业效益提什么要求；要毫不犹豫地大力发展和推广“弱人工智能”技术，以计算机和控制设备（系统）的智能化为重点，将人工智能技术融入数字经济和智慧社会之中。

发展人工智能应有更理性的态度

30年前Simon对人工智能的判断

- 1990年我率领国家智能计算机研究开发中心代表团访问CMU，拜访了图灵奖得主司马贺（Simon）和Newell教授。
- 我曾问司马贺教授，未来10年人工智能有望取得的最大突破（Biggest breakthrough）是什么？他肯定地回答：未来10年人工智能不会有根本性的突破，但可能有上千个小的进展。
- 我认为，Simon教授的判断至今仍有参考价值。人工智能要获得根本性的重大突破还要付出艰苦的努力。

吸取上世纪80年代的历史教训

- 智能应用比常规应用需要更快的计算速度、更大的存储容量和更高的数据传输能力。这些要求将趋使计算机系统结构产生重大创新。这些要求与计算机主流技术的发展是一致的。若干年后适合智能应用的芯片与系统结构将成为计算机的主流技术。
- 上世纪80年代，许多人鼓吹计算机将按照人工智能的要求从第四代发展到所谓的第五代。可是，计算机的发展恰恰走了一条与此相反的路。大部分所谓的智能计算机（LISP机、PROLOG机）、智能软件和智能工具并入了计算机主流。
- 今天的形势不同于上世纪80年代，但历史的教训值得吸取。我们要重视智能应用的特殊要求，但不能忽视通用的计算机主流技术的巨大包容能力。

如果AI是答案，那么要解决的问题是什么？

- 上世纪80年代，日本人研制Prolog计算机即“第五代计算机”时，美国的学者质问他们；“如果Prolog机是解决方案，那么要解决的问题是什么？”正是因为日本人没有想明白第五代计算机要解决什么问题，当时预想的智能计算机应用（如法院用计算机判案等）还不是社会的真实需求，第五代计算机的研制以失败告终。
- 今天我们不能重犯日本人当年的错误，一定要时刻问自己：“如果AI是答案，那么，要解决的问题是什么？” 研究一项智能技术，在某个单项智能水平上超过人不是发展人工智能产业的目的。也不能把人工智能当做“锤子”，把所有要解决的问题都看成“钉子”。人工智能不是万能药，一定要了解目前人工智能技术究竟能解决那些实际问题。不要把实验性成果当成可普遍推广的产业技术。

行百里者半九十，在最后10%上下大功夫

- 在机器翻译等领域，深度学习做得非常好，可以达到90%以上的准确率。问题是最后的10%的提升，可能需要完全不同的方法。在自动驾驶系统中，机器做出的决策必须是非常准确，容不得一丝马虎，这样才能确保乘客的安全。所以对于最后的1%的提升，甚至最后万分之一的提升，需要做大量的科研攻关，也许最终还是要人机结合。
- 每种技术途径都有性能的极限，最高水平的科研人员有眼光判断技术途径的极限。二流的科研人员要么半途而废，要么盲目地浪费精力做无用功。计算复杂性理论就是研究问题和算法的上下界，要高度重视各种复杂性研究。
- 工程技术和产业技术问题都要考虑成本。不计成本和不惜一切代价是不可取的。如果没有新突破，将ImageNet的误识率降到1%，需要投入几千亿美元。

智能技术是蛋糕上的奶酪，要重视信息技术整体的作用

- 智能技术是蛋糕上的奶酪，主要的价值在“蛋糕”，不能只顾奶酪不顾蛋糕。智能技术对其他产业的作用如同蜜蜂传粉对农业的作用，不能只关注“蜂蜜”的价值。
- 大数据和人工智能（数据智能）的巨大驱动作用本质上是整个信息技术的作用。信息技术酝酿了几十年，现在是见效的时候了。智慧城市建设不完全是人工智能问题，而是建设什么样基础设施的问题。
- 数据智能技术的兴起得益于计算能力的提升、存储成本降低和网络通信技术的普及，不能只见树木不见森林，要重视信息技术整体的作用！
- 智能技术是整个系统技术的组成部分，要重视系统技术的研究，用系统技术弥补器件技术的不足。

不要将替代人工当成发展智能技术的唯一目标

- 发展人工智能的目标是创造有益的智能，而不是漫无目标的智能。
- 人工智能的效果不局限于自动化，不能将替代人工当成发展智能技术的唯一目标。要不要用智能技术替代现在的人工要做全面分析，应考虑整体成本和就业、稳定等社会问题。
- 韩国是工业机器人使用密度最高的国家，考虑大量使用机器人可能增加失业，韩国政府决定对投资工业自动化设备的企业取消税收减免，变相征收“机器人税”。这一政策可能不利于发展机器人技术，但反映出政府不能不顾一切地支持发展替代人工的自动化技术。
- 对在发展智能技术的过程中可能失业的工作人员要未雨绸缪，有计划地做新职业培训，对任意解雇劳动者的企业要有适当的约束措施。

从DARPA投资的下一代AI项目看AI布局

（转引自网络Peter Highnam博士的报告 仅此致谢）

- DARPA的下一代AI计划部署了90项应用AI，27项高级AI和18项前沿探索

  （比例为9：3：2）

- 此布局以AI应用为主，兼顾高端技术和前沿探索，是一种综合考虑近、中、远期需求的全面布局，中的我们借鉴。

建立自主可控的AI开源平台

- 不能只关注技术本身，还要关注比技术更广泛的生态系统。竞争可能会在新旧生态系统之间发生，而非在技术之间发生。一个新技术替代老技术，不能只看技术是否足够先进。替代速度取决于新技术生态系统的挑战能在多长时间内解决。
- 现在大家做AI研发大都基于国外大企业的开发平台，如谷歌、Facebook、亚马逊、微软等，这些开源程序都放在GitHub中（现在是微软下面的托管平台），按照美国法律，GitHub要受美国法律管辖，存在断供的风险。
- 我国一定要建立自主可控的AI开源平台。企业牵头的AI开放创新平台是一条路，但企业的开发工具不一定开源，数据和模型不一定能共享。国家科研机构和大学要花更多精力打造培育人工智能开源软件和开源开发工具。

提倡百家争鸣的自由探索精神

- 现在并不知道如何实现通用人工智能，想要真正取得进展，应当容忍人工智能基础研究中方法论的发散性，不应人为切断任何一种探索途径。
- 国家新一代人工智能规划中的“五智”（大数据智能、跨媒体智能、群体智能、混合增强智能和自主智能）源于工程院的AI2.0建议，“五智”是指现在较流行的技术途径，未来的人工智能不一定局限于“五智”。
- 图灵奖得主Bengio创建的蒙特利尔学习算法研究所（MILA）现在是全球AI研究的重镇，每年进行关于AI的的辩论会。辩论主题包括“AI架构与挑战”、“神经科学与心理学带来的洞见”、“构建可信任的 AI”等
- 我国学术界习惯于随大流，辩论风气不浓。中国计算机学会启智会也开始对AI的局限和未来发展趋势展开激烈辩论，希望国内人工智能界在会议和期刊中多开展辩论，弘扬百家争鸣的自由探索精神。

AI应用技术和监管技术要齐头并进

- 科技本身的发展和科技监管技术必须齐头并进，不能等风险不可收拾时才想到要监管。一定要把科技关进伦理和法律的笼子里。金融科技（FinTech）兴起后，马上就有金融监管科技（ RegTech ）。人工智能领域也必须大力发展AI-RegTech。“自主”的AI武器系统中可能需要安装检查“合理性”的自监控系统。
- 越是先进的技术越需要监管。社会各界十分担心人工智能的风险，要像监管核武器一样加强人工智能技术的监管。AI伦理和AI监管是我国明显的短板，应立即加强有关布局和规划。
- 未来的智慧社会中，“智警”和有关智能业务的“法务工程师”应成为重要的从业人员，其数量可能会多于普通的刑警和法官。从现在起，就要着手培养“智警”和熟悉智能业务的“法务工程师”。

请批评指正！