25 全体大会

LSTM 之父 Jürgen Schmidhuber：具有好奇心的人工智能是 AI 的美好愿景

整理：智源社区 沈磊贤

在第二届北京智源大会全体大会中，LSTM 之父 Jürgen Schmidhuber 教授做了题为《AI against Covid-19》的报告。Jürgen Schmidhuber 教授是瑞士人工智能实验室（IDSIA）的研发主任，LSTM（长短期记忆网络）的提出者。在报告中 Jürgen 回顾了一系列基于 LSTM 的深度神经网络模型在病毒传播分析、药品设计、蛋白质折叠预测、医疗图像识别，以及语音识别、机器翻译等方面的应用，提出了其对 AI 未来的愿景： 具有好奇心的人工智能，能够自我设定目标、自我学习、自我提升、实现目标。

Jürgen 教授首先通过一张卡通漫画引入本次演讲的主题，这张漫画源于他创建的“AI-vs-Covid19”网站（https://www.aivscovid19.org）上，本次报告的绝大部分内容都将可以在该网站中找到。Jürgen 教授本次演讲的主题是在 covid-19 病毒的背景下，基于深度神经网络和深度学习的人工智能是如何通过多种方式帮助对抗病毒，这其中的基本原理是神经网络可以从病、患者等各种数据中学习各种行为模式，然后使用这些神经网络来预测采取行动可能带来的后果，从而最大程度地减少损失。

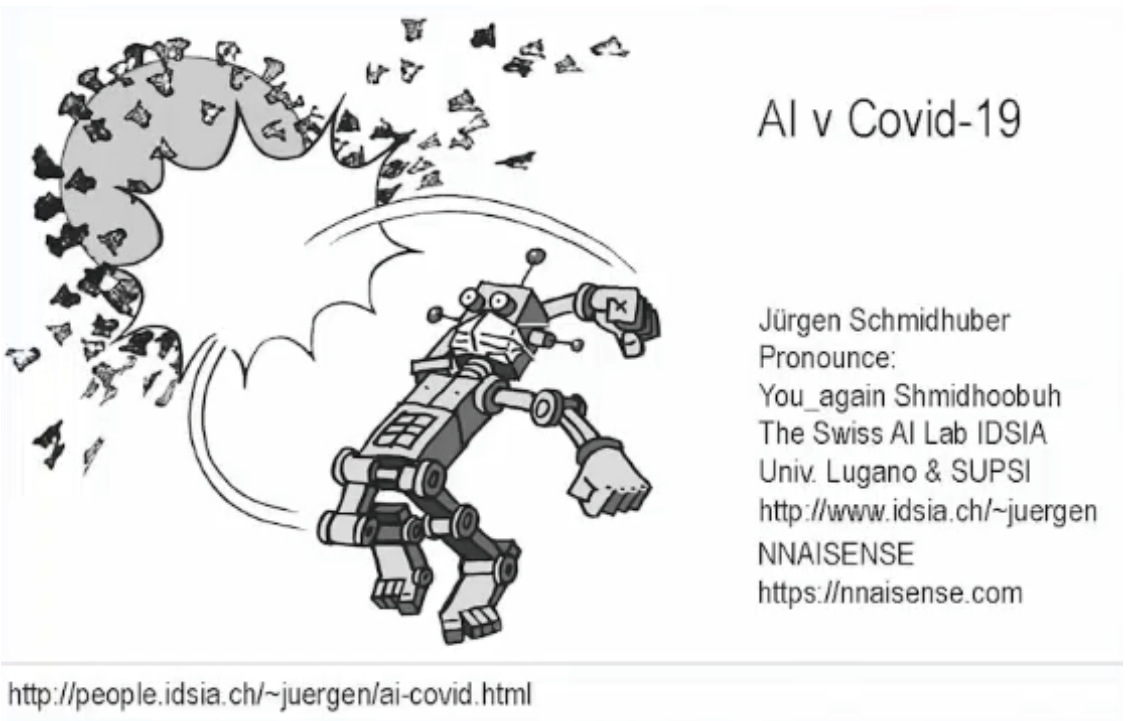


图 1： AI-vs-Covid19

一、人工智能在新冠疫情中的贡献

Jürgen 教授介绍了一些人工智能应用的具体例子。首先是通过模式识别来跟踪人群。智能手机上的蓝牙 P2P 应用程序能够在一定程度上防止疫情中人与人之间危险的接触，这样的应用程序不需要太多的人工智能技术。相

比之下，更具挑战性且更难应用的是，在无人机或在街上拍摄的视频中识别人的面孔、步态及行人之间的接触。但得益于神经网络，这样的想法在今天能够成为现实。2011 年，Jürgen 教授的博士后 Dan Claudiu Ciresan 创建了第一个能够在模式识别竞赛中获胜的快速深层神经网络。大概十年前，在硅谷的一次竞赛中，计算机视觉领域首次出现了超过人类视觉的应用。现代计算机视觉的许多应用方面都是在这种方法基础之上进行的拓展。而如今在中国，计算机视觉技术已被广泛使用。

在 2010 年至 2020 年这 10 年间，计算机的运行计算速度飞速提升，很多历史久远的技术在短时间内获得了巨大的商业成功。实际上，如今非常流行的基本概念可以追溯到 90 年代初。但是那个年代的计算机运行速度比今天低了一百万倍，几乎没有科学家能借助计算机实现太多工作。而如今这一切已经发生了很大的变化。算力的发展使得复杂的神经网络可以大展身手，在今年抗击疫情的过程中，神经网络作为人工智能技术的核心，也被应用在诸多领域。

1.1 疾病爆发预测

神经网络通过对历史数据的分析，可以预测疾病的爆发，建立预警系统。对此，近期还有一个与之相关的 Kaggle covid–19 预测挑战，对人类抗疫工作有一定的帮助。

1.2 资源需求预测

对过去资源需求数据的学习之后，神经网络可以预测未来对有限资源的需求，例如预测呼吸机和医生资源需求，从而优化物流配送等流程。

1.3 基因测序

神经网络还可以对病毒基因组进行测序，并且预测下一个相似的基因组将在何处出现，这是 AI 科学家们一直在做的事情。尽管病毒基因组一直在随机变异，但是这种方法可以追溯其来源。将这些数据输入系统，AI 可以学会预测病毒基因组未来的发展。

1.4 健康检测

神经网络的一个典型应用是监视患者的健康状况。例如，他们可以监视心跳频率，追踪生物信号，比如呼吸，咳嗽等，从而帮助医生做出更具针对性的诊断。

1.5 医学图像分析

神经网络的另一个重要应用是医学图像分析，例如分析患者肺部或胸部的 X 射线图像，从而诊断出病理。由深度神经网络赢得的第一场医学影像竞赛可以追溯到 2012 年，这个比赛是关于癌症检测，比赛的赢家是 Jürgen 教授的博士后 Dan Claudiu Ciresan 及其团队。下图为一张女性乳房的切片照片，可以看到的是其中一些细胞是无害的，其他则是危险的。

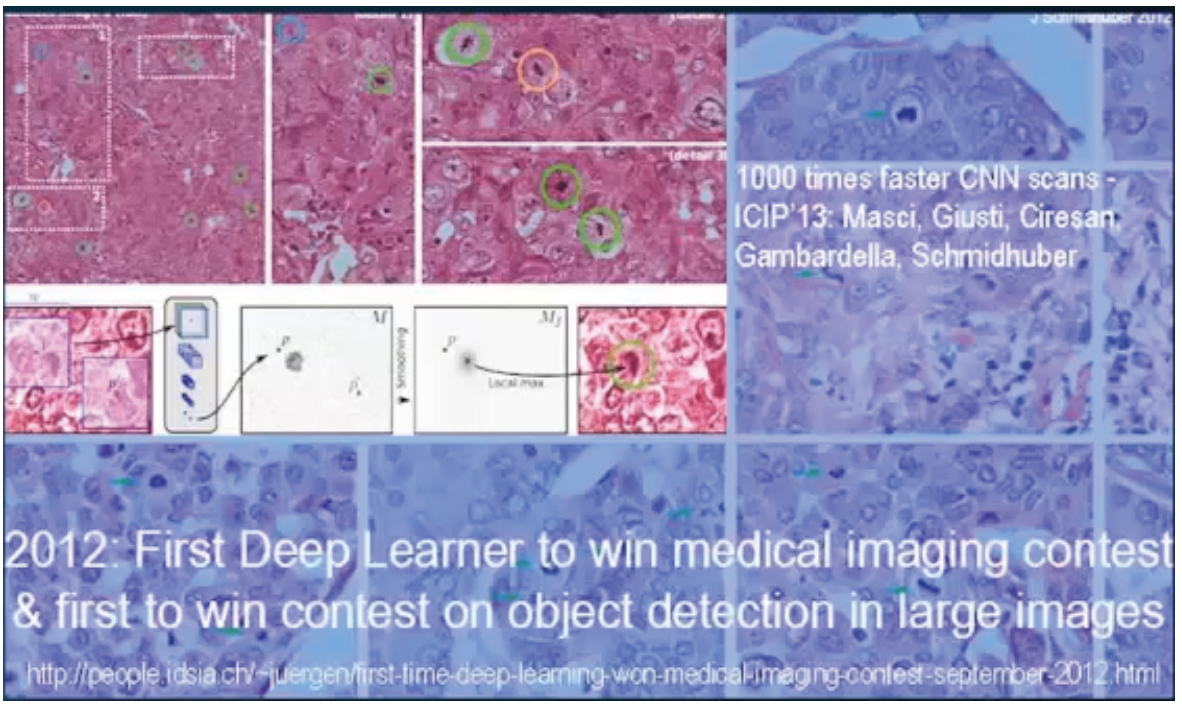


图 2：女性乳房的切片照片

通常情况下，需要训练有素的组织学家才能将好的细胞和坏的细胞区分开。但是在比赛中，AI 科学家能够训练一个神经网络来模仿医生的诊断。如今，许多企业都在这样的领域中都采用了类似的想法。不仅是初创公司，还有西门子，谷歌和 IBM 等大公司。

1.6 药物设计自动化

随着 AI 技术的不断发展，神经网络也被应用到药物设计流程的部分自动化。例如，神经网络可以帮助找到停靠在病毒折叠蛋白上的分子，这种病毒通常只有些许蛋白质。科学家的目标是像抑制抗体一样，抑制这些蛋白质的活性，以便可以在感染五种病毒后阻止病毒的整个自我复制机制。早在 2007 年，当计算成本比今天贵了将近 1000 倍时，林茨大学的 Sepp Hochreiter 教授就使用 LSTM 深度神经网络在预测蛋白质折叠方面获得了出色的成果。预测这样的折叠位置对于找到 docking stations 很重要。

1.7 设计新分子

神经网络还可以设计新的分子。例如，Segler 及其同事在 2018 年发表了一篇著名的论文，他们可以通过神经网络在抗体库中找到抗体针。这同样是由林茨大学设计完成的。例如，基于配体的方法就是如此，给定某种分子，神经网络将学会预测其将结合哪些蛋白质。

神经网络还可以大大减少药物开发流程时长。传统的药物开发流程耗时太长，从 1 万种化合物中选择 5 种至少需要六年的时间。此后需要七年的临床试验，接着可能只能发现少量的能应用于研究的物质。然后需要一年或更长时间才能获得批准。这样的研究周期显然太长了，尤其是在当前的疫情大流行时期。然而现在或许可以通过快速的虚拟筛选来加快这一过程。使用大型数据库，例如 Zinc 数据库，其中包含大约 10 亿个分子的描述，然后使用 SmilesLSTM 深度神经网络传入数据，该神经网络的诞生源自于林茨大学。这样的神经网络能够在

SARS–CoV–2 抑制剂备选名单中，筛选出 30000 个得分最高的候选分子。一旦候选分子数量大大减少，就可以在 wet lab 中进行测试。而且，也可以将这种方法应用于已经投放市场的药物，这是减少此类临床试验的重要方法。

二、LSTM 的起源和核心思想

LSTM 的诞生可以追溯到 90 年代初。这是一个受到人脑运行机制启发的递归神经网络。在人脑中，可能有 1000 亿个称为神经元的小处理器。它们每个都可能与 10000 个其他神经元相连。这些神经元中的一些是输入神经元，其中数据可能会通过外界的相机和麦克风输入人脑。其中一些神经元是输出神经元。如果将这些神经元的开关打开，人们的手指肌肉或者语音肌肉就会移动。人类日常的生活基本上就是将这些输入的信息转化为行动和行动序列，日积月累，最终走向成功。

所有神经元之间的连接都具有一个强度，这称为权重。在一开始，所有这些权重都是随机的，这意味着一个神经元对另一神经元的影响也是随机的。但是通过学习，这些连接中的一部分会变得更牢固，权重变大，而某些连接会变得更弱。最终，可以用网络训练实例来做一些有趣的事情。例如，驾驶汽车、语音识别、预测新型冠状病毒的折叠位置等。

2.1 LSTM 的应用

如今 LSTM 的应用十分广泛，甚至存在于我们的智能手机中。LSTM 执行着许多日常生活中的 AI 任务，例如语音识别，当人们通过语音渠道与搜索引擎互动时，例如“OK，Google，距离最近的去往车站的路程是什么?”，语音助手背后的 LSTM 算法，能理解人类想要表达的意思，并将其转换为对搜索引擎的查询。如今这项功能比以前的效果要好得多，数十亿智能手机都在使用。

而且，LSTM 也广泛地应用于机器翻译中。现在，许多智能手机上也有这种功能。到 2017 年，Facebook Translate 每周已经通过 LSTM 完成了约 300 亿条消息的翻译。LSTM 对于多种序列处理都是有效的。蛋白质折叠预测只是这些可能的应用之一。

但是，回溯到 1991 年，当科学家刚开始为循环神经网络 Recurrent Networks 进行深度学习实验时，只能进行很少的微型实验，因为计算机的速度实际上比今天慢了一百万倍。然而每五年，计算机的计算价格就会便宜 10 倍。早在 1935 年到 1941 年之间，Konrad Zuse 建造了第一台工作程序控制的通用计算机。它可以每秒执行大约一个操作，比如每秒执行一次乘法。但是 30 年后，科学家们可以以相同的价格执行大约一百万次这样的操作，依此类推。如今，可以以相同的价格每秒执行约 10 亿次操作。到 2009 年，计算机已经足够快了，这时 Jürgen 教授的博士生 Alex Gravers 通过 LSTM 赢得了模式识别竞赛。紧接着数十年的深度学习时代来临，突然每个人都开始使用 LSTM 了。

科学家们还可以用 LSTM 或前馈神经网络或图形神经网络上学习化学。首先需要输入物质和某些条件，例如温度和催化剂。然后发生化学反应，并产生输出，输出物质。如果有一个包含数百万个此类示例的数据库，则网络可以学习将这些输入成分映射到相应的输出物质。于是这个网络就学会了化学反应； 而且可以将其用作物理和化学真正作用的替代者，这意味着，一旦科研工作者对它进行了许多实例训练，就可以将其变成人造化学家，然后使用它来提出新的方法以及成分组成，从而产生从未见过的新的输出物质。通过向后推导神经网络，可以弄清楚给定化学模型要达到的预期结果，以及需要哪种输入物质。近年来，这确实已经彻底改变了化学研究。

现在，神经网络有时甚至足以代替湿实验室测试（所谓的测定法）。

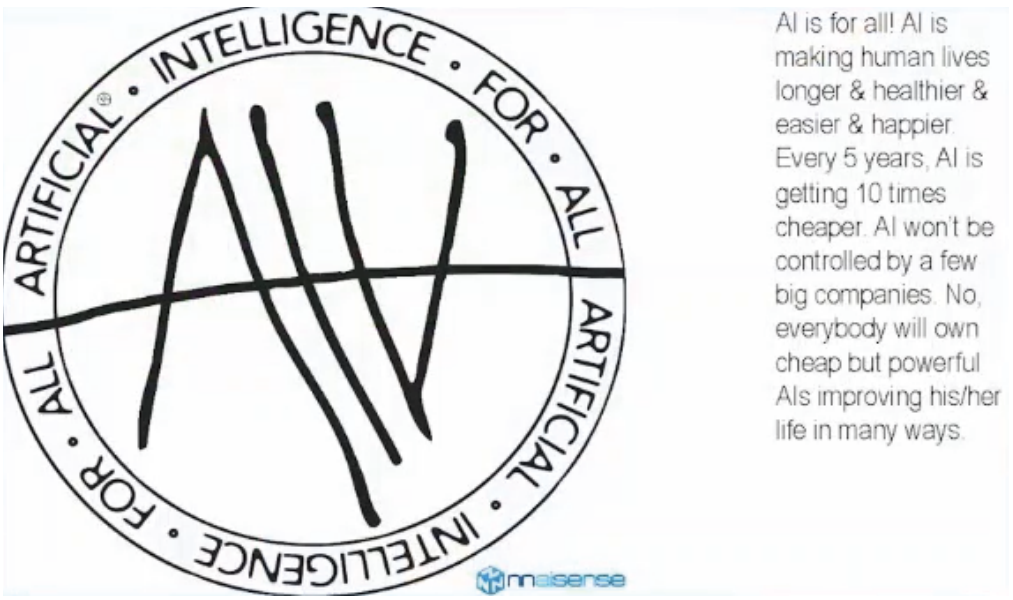
三、AI 未来的愿景

Jürgen 教授提出了其对 AI 未来的愿景： 具有好奇心的人工智能。到目前为止，人们一直在谈论的主要内容是监督学习。但是，他希望 AI 不要单纯的模仿人类老师，而是要建立自己的目标。同样的，就像小婴儿一样，好奇地探索世界。小婴儿就是小科学家。他们没有从父母那里学到很多东西，他们学到的很多东西都是通过玩具而学到的。

自 1990 年以来，科学家们一直在研究可以真正自我发明的小型人工 agent，并创建玩具世界，使 agent 可以从中学习未知的知识。因此，好奇 agent 的基本原理与 20 年前的科学家所做的，原理是相同。唯一的区别是实验变得越来越昂贵。而且如今这些“人工科学家”正变得越来越聪明。但是对于需要快速反应的新型冠状病毒来说可能还不够。然而未来，“人工科学家”将会发明自己的目标和实验，来创造性地了解真实的世界。

Jürgen 教授谈到了 Covid–19 病毒对全球的经济的影响，他也未能幸免，他所拥有的名为 Nnaisense 的创业公司也受到了冠状病毒很大的影响。Nnaisense 的发音像英文单词 nascence。但这是用不同的方式拼写的。nn 代表神经网络，ai 代表人工智能。这与基于神经网络的通用人工智能的诞生有关。尽管世界上所有公司都受到了影响，但众所周知，危机也是机遇。实际上，这场危机将使行业变得比以前更强大。因为我们正在使用的机器将变得更加智能。这些机器将能够做人工大脑之前做不到的许多事情。Jürgen 列举了 Nnaisense 的合作机构：Audi,，Festo，作为玻璃制造的领导者的 SCHOTT，3D 制造和 3D 工业打印领导者的 EOS，使用无人机检查风力涡轮机的 Sulzer Schmid 等等，他们都使用 AI 技术为改造世界做出了自己的贡献。

最后 Jürgen 教授强调，人工智能永远不会仅由少数几家大公司控制。因为现在计算价格越来越便宜，每个人都将拥有廉价的 AI，从而可以通过许多不同的方式改善自己的生活。



Jürgen 教授从机器人漫画开始了本次演讲，最后以半个世纪前绘制的机器人卡通作为演讲的结束。那是 Jürgen 教授 1987 年研究元学习的毕业论文封面。那时人们很少对此感兴趣，但如今这已经成为一个热门话题。从那时起，所谓的人工智能就开始以许多方式改变了人类日常生活的方方面面，并将持续改变并影响下去。

图灵奖得主 Joseph Sifakis：我们能信任自治系统吗？

整理：智源社区 钱小鹅

在第二届北京智源大会上，2007 年图灵奖得主 Joseph Sifakis 为我们带来了题为“Can We Trust Autonomous Systems? Boundaries and Risks”的主题演讲。

Joseph Sifakis，欧洲科学院和法国科学院院士，美国文理科学院和美国国家工程院院士，中国科学院外籍院士。Joseph 是著名嵌入式系统专家，1981 年独立提出模型检测方法，这一方法被广泛应用于芯片检测、通讯协议、嵌入式系统以及安全算法等领域。1993 年开创 Verimag 实验室。2007 年，Joseph 因在模型检查理论和应用上的卓越贡献而被授予图灵奖，2012 年，获得利奥纳多・达芬奇勋章。Joseph 现作为欧洲卓越网络嵌入式系统设计联盟技术协调人，负责对 35 个欧洲研究小组的研究进行协调，来推进欧洲在嵌入式系统设计理论和应用上的发展。

Joseph 在演讲中认为，在物联网的自动化控制中存在两大问题： 对于工业物联网，规则可以被直接更改； 对于人机物联网，人类或明确或武断的行为，也可能戏剧性地导致触发控制序列和规则的更改。因此，Joseph 在本次演讲中从传统的安全检测手段出发，为我们逐步引入下一代安全系统的概念。他在分享中提到，下一代的安全系统更强调自主性，即系统的自我调控能力。在自动驾驶、大脑手术、工业报警等这些熟悉的物联网行业，我们不难看出，在产品的整个寿命中，系统所处的环境都是瞬息万变的，因此，系统强大的自我调控能力必不可少。那么我们能否相信“自主”系统，系统的边界与风险又在哪里呢？我们不妨从 Joseph 精彩的讲座中寻找答案。

一、下一代自主系统的特征和概念

Joseph 首先指出，下一代自主系统的产生是为了满足用自治代理逐步取代人工操作来进一步自动化现有机构的需求。因此，自主系统需要足够严谨并在处理知识时表现出“Broad Intelligence”，具体来说需要满足如下几个要求:

- 能处理一系列动态改变且可能有潜在冲突的目标，这反应了自主系统从狭隘人工智能 / 弱人工智能到强人工智能 / 广义人工智能的转变趋势;
- 应对复杂、不可预测的网络物理环境中的不确定性;
- 能与人类和谐的合作，比如共生自治。

但面对这些关键性的要求，目前的自主系统还有诸多的缺陷。比如： 系统中的自我学习部分没有可信赖的保障技术进行支持； 处理地理分布和移动性所需的网络基础架构可信赖性差，如网络安全问题、反馈时间无法得到保证等；处于高动态中的行为带来了巨大的复杂性，如网络物理媒介和不可预测性。

Joseph 认为自动驾驶就是自主系统中的一个标志性案例，它提出了严峻的技术挑战，并涉及巨大的经济利益和深远的社会影响。从中我们也可以一窥下一代自主系统的新趋势。与航空航天和铁路行业相比，自动驾驶技术有如下特征:

1. 自动驾驶的研发人员并不遵循“通过设计来保障安全”的理念，他们通过端到端的黑箱机器学习来进行技术设计。
2. 自动驾驶技术的研发人员认为统计学的可信度依据已经足够保障安全，这表示如果一辆车开了上百万公里没有出事故，那么他们就认为这辆车足够可靠。
3. 公共机构允许自动驾驶汽车进行“自我认证”。
4. 与飞机这种软件和硬件在出厂后均无法修改的产品不同，自动驾驶汽车的关键软件可以通过更新进行定制，如特斯拉汽车以月为单元进行软件升级。

埃隆·马斯克 (Elon Musk) 2015 年在英伟达技术大会上宣称自己几乎把自动驾驶汽车当作一个已经实现的问题，他只是需要若干年来按部就班地实现而已。

然而，是否真如 Musk 所言，自动驾驶技术可以当作一个已经实现的问题呢？自动驾驶是否已经非常严谨了呢？ Joseph 给出的答案是否定的。他认为，对于自动驾驶这种缺乏足够严谨的设计方式，人们普遍的态度可以概括为接受其风险、贪婪其利益的直率现实主义； 认为传统的严谨设计方式本质上具有局限性，复杂问题只能通过经验方法来解决，对经验方法带有盲目的信念； 以及坚信自己已经掌握了正确手段，技术实现只是时间问题的无限乐观。

我们的安全系统设计面临着巨大的挑战，我们正在从小型的、中心化的、自动的、在可预测环境下可规范的系统迈向复杂的、去中心化的、自主的、在非可预测环境下的不可规范的安全系统。我们需要建立一个新的科学和工程基础，而不是简单组合过去二十多年的现有成果并仅仅关注诸如自主计算、自适应系统、自主代理这类软件系统。我们需要解决以下这些问题:

1. 了解自动化和自治化之间的可能性范围。提升系统自主性的技术方案是什么？对于每一个提升，我们需要了解其中有哪些潜在的困难和风险。能有原则的确定一个能执行给定任务的系统是否足够可靠。
2. 将系统可信度与已开发系统的知识确实性关联起来。
3. 从传统的系统设计过渡到“混合”设计，以寻求在基于模型的可信赖性与基于数据的表现性之间的权衡 .

接下来，Joseph 向我们解释了自主性 (Autonomy) 的概念。

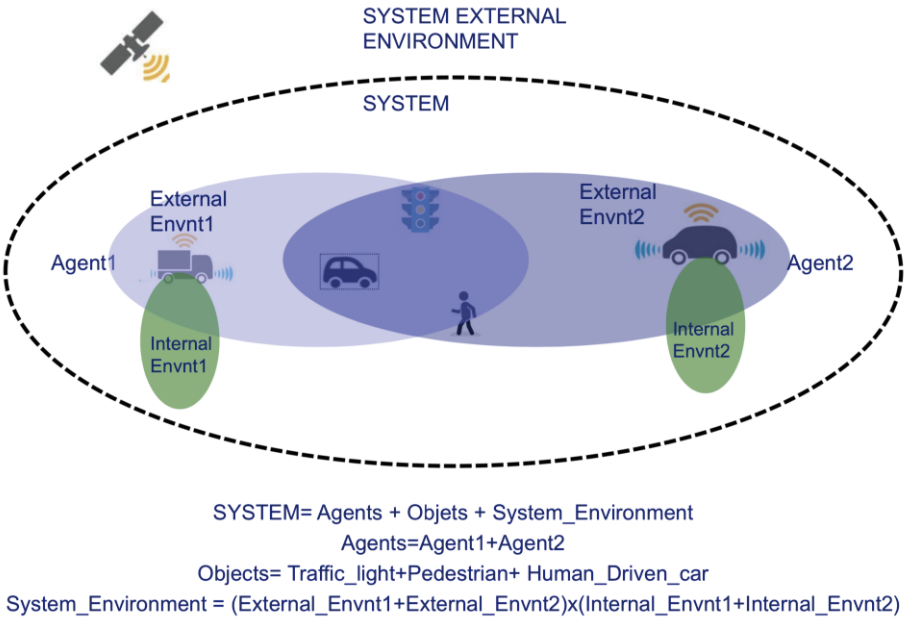


图 1：以自动驾驶为例，自主系统基本概念示意图

Joseph 认为对于自主系统，如恒温器、足球机器人、自动驾驶汽车等，它们都是在某种特定环境下，由若干携带传感器的终端组成，通过每个终端对各自的目标实现，其集体行为就会迈向系统整体目的的实现。如图 1 所示以自动驾驶为例，一个自主系统由终端、客体和系统环境组成。系统环境包括内部环境和外部环境。随着环境、感知客体的复杂化和动态化，目标的多变化，系统也逐渐由自动向自主迈进。

	Environment	Stimuli	Meeting Goals
Thermostat	Room + Heating/cooling device	Temperature	Explicit controller Single goal
Shuttle	Cars + Passengers+ equipment	Static configuration of cars+ State of equipment	Explicit controller + on line adaptation Many fixed goals
Chess robot	Chess board + pawns	Static configuration of pawns	On-line planning+ stored knowledge Dyn. Changing goals
Soccer robot	Regions in the field + Players + Ball	Dynamic configuration of players/ball	On-line planning+ stored/generated knowledge Dyn. changing goals
Robocar	Vehicles/obstacles + Road/communication equipment	Dynamic configuration of vehicles/obstacles + State of equipment	On-line planning+ stored/generated knowledge Dyn. changing goals

图 2：从自动到自主，不同系统实例的比较

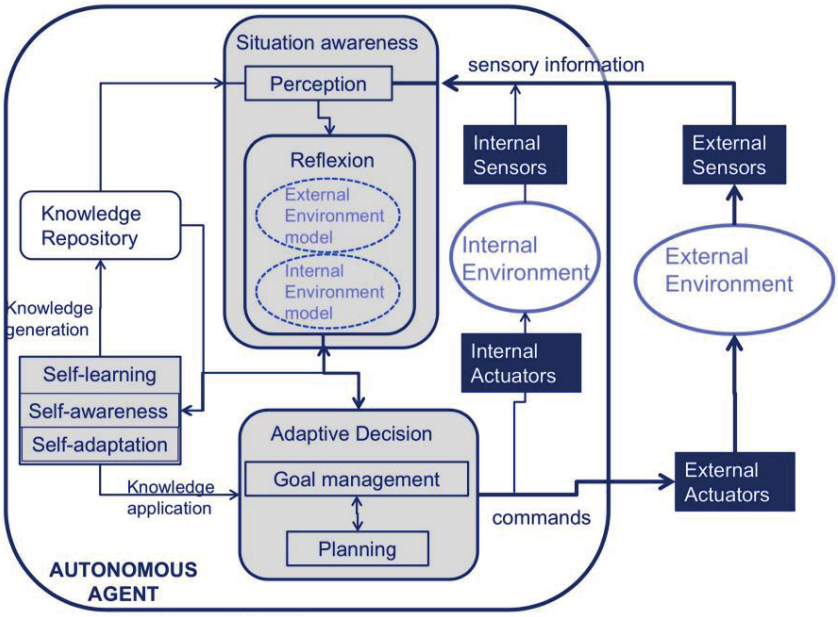


图 3：自主终端结构示意图

因此自主性是一个终端通过自己的方式（无需人工干预），在多变的环境下实现一系列目标的能力。如图 3 所示，它主要由五个互补的功能组成:

1. 感知，对外部激励的翻译表达、从复杂的输入数据中去除模糊部分并提取相关信息；
2. 反馈，能够建立 / 更新一个可靠的运行时环境模型以计算出可以实现目标的策略；
3. 目标管理，从可能的目标中选择最接近当前给定环境模型的目标；
4. 目标规划，规划如何实现特定目标；
5. 自学习，通过学习和推理创造新的情景知识和新目标的能力。

需要注意的是，上面提到的这些功能是实现无关的，我们未来可能通过机器学习、深度学习或者其他方式来实现。这五个功能也可以用来区分自动化和自主化。

二、系统的可信赖度问题（Trustworthiness）

我们如何决定一个系统执行的任务是可以被信赖的呢？Joseph 认为有两点需要考虑，可信赖性 (Trustworthiness) 和严格性 (Criticality)。系统的可信赖性是指不管外界发生何种意外，系统仍按照预定运行； 系统的严格性则是错误的严重程度会在任务的实现中表现出严重的后果，比如驾驶一辆车、做手术、核电站作业。

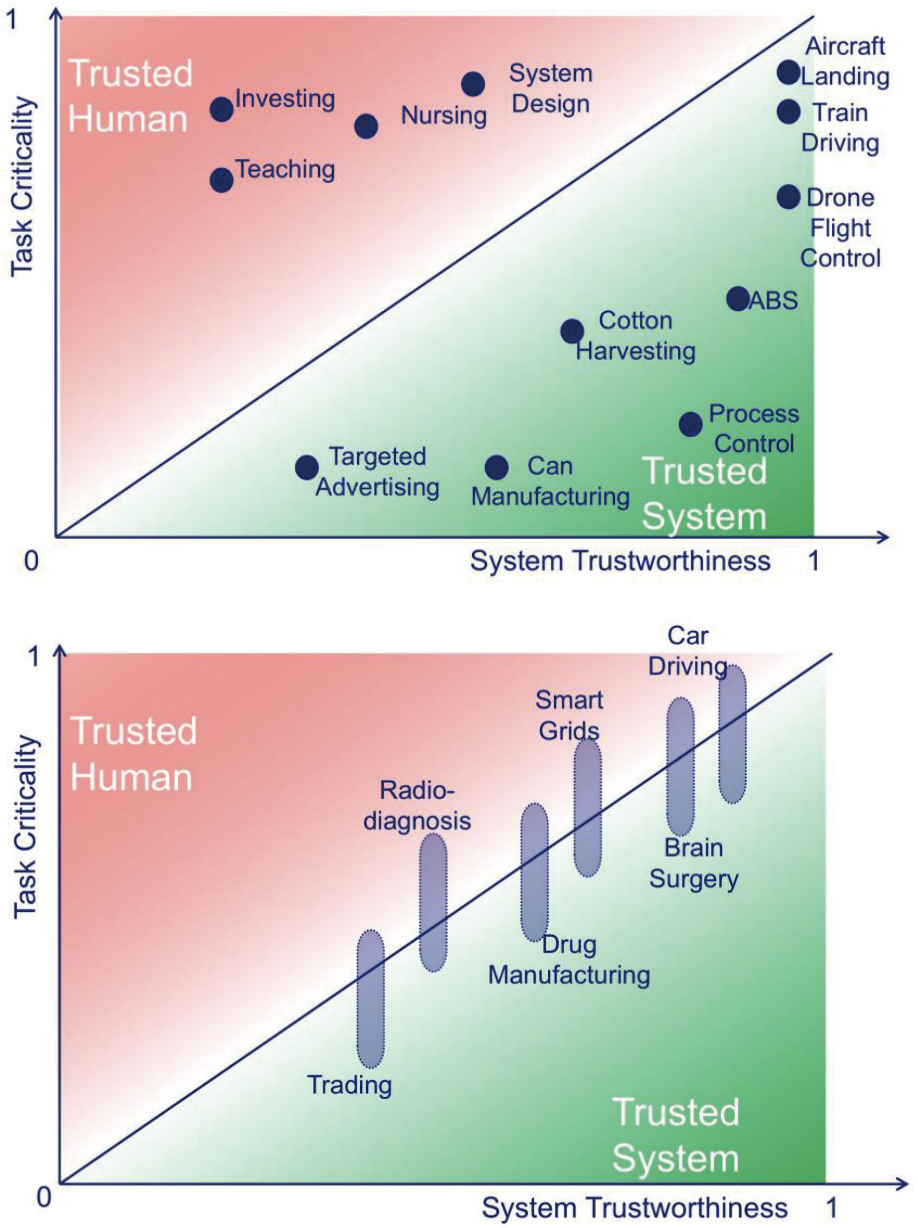


图 4：任务严格性 vs 系统可信赖性

从图 4 上可知，人类行为往往符合任务的严格性，而自动化的机器则满足系统的可信赖性，而两者的边界就是自动化的前沿领域。所谓自动化系统往往表现为静态的决策过程或者 / 以及失败的影响很小。而非自动化系统则需要好的情景认知以及多目标管理能力。而人和机器的共生自主系统则可以很好的在可信赖性和严格性之间取得平衡（如图 4 下所示）。我们可以制定一系列恰当的自主化标准来确保人和机器能和睦的一起工作，这些协议会帮助操作人员能够推翻机器的决定，机器也会主动征求人的介入。未来的很多应用会需要这种共生自主系统，比如汽车驾驶、大脑手术、药物制造等，这是自主系统实现的一个重要方式。

The Automation Frontier – Autonomy Levels

	SAE AUTONOMY LEVELS
Level 0	No automation
Level 1	Driver assistance required (“hands on”)
	The driver still needs to maintain full situational awareness and control of the vehicle e.g. cruise control.
Level 2	Partial automation options available(“hands off”)
	Autopilot manages both speed and steering under certain conditions, e.g. highway driving.
Level 3	Conditional Automation(“eyes off”)
	The car, rather than the driver, takes over actively monitoring the environment when the system is engaged. However, human drivers must be prepared to respond to a "request to intervene”
Level 4	High automation (“mind off”)
	Self driving is supported only in limited areas (geofenced) or under special circumstances, like traffic jams
Level 5	Full automation (“steering wheel optional”)
	No human intervention is required e.g. a robotic taxi

图 5：SAE 自主级别规定

图 5 为 SAE（美国汽车工程师协会）制定的自动驾驶自主级别标准。从 level0 到 level5 反映了人类与机器分工的不同方式。从最低级的非自动驾驶到最高级的没有人类介入的全自动驾驶依次定为六个级别。这表明了共生自主系统所面临的一个共同问题，我们如何划分人与机器之间明确的工作界限来最大确保他们的合作效率。

之后 Joseph 提出了知识真实性（Knowledge Truthfulness）问题。他首先做了一个有趣的类比。人类的思考方式可以分为快思考（Fast Thinking）和慢思考（Slow Thinking）两种。快思考是无意识、自动化且毫不费力的，不需要自觉和控制，人们用它来处理所有经验内隐知识（Empirical Implicit Knowledge）比如走路、说话、弹钢琴等。而与之对应的慢思考则是有意识、被控制且需要努力的，它有自我意识的参与和控制，它是所有推理知识如数学、科学、技术的源头。

我们可以分别把计算机技术的神经网络和传统计算与上面的快思考和慢思考作一个类比：神经网络对应快思考，它通过数据训练产生经验性的知识，是基于数据的知识，用神经网络去识别猫和狗正如人类的孩童认知一样，我们无法验证这种识别能力从何而来。传统计算机对应慢思考，它有明确的执行算法，是基于模型的知识，它能处理明确形式化的知识，是可以被验证的。

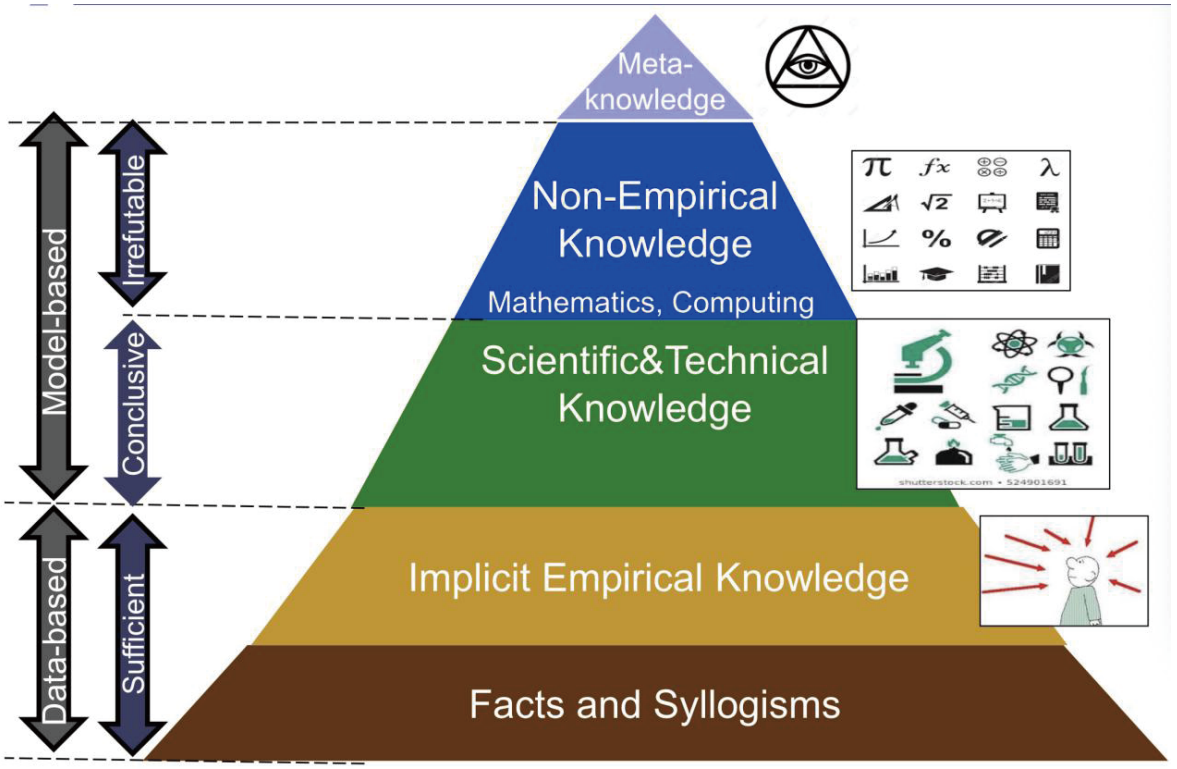


图 6：知识金字塔

由这一类比进行引申，我们可以将知识分级建立知识金字塔，并将计算机技术放入其中。如图 6 所示，金字塔的最底层是事实及其推论，我们从中提取隐含的经验性知识建立第二层金字塔，第三层则是更加概括的科学和技术知识，第四层则是从中提取的高度抽象的非经验性知识，最顶层则被称为元知识。数学和计算被认为处于第四层，而机器学习和数据分析则处于第二层。也就是说科学与机器学习分别处在不同的知识层级，科学能够通过对实验的分析学习来进行合理的解释；而神经网络这类机器学习技术则无法解释他们的结论是如何得来的。

三、如何设计一个可信赖高性能的自主系统

那么，如何涉及一个可信赖高性能的自主系统？ Joseph 首先介绍说传统基于模型的系统设计方法足够信赖但性能较差，新兴的机器学习方案性能优秀但不可信赖，于是人们希望将两者综合利用，使得系统在可信赖性和高性能之间取得一个平衡。这就是混合设计流程 (Hybrid Design Flows)。但对于这种混合设计自主系统，基于模型的设计方法面临着可信赖性的挑战。传统上，基于模型的设计方法在设计期间就试图保障系统的可信赖性，如图 7 所示，在系统设计时，他们会通过穷举有害事件来进行风险分析，找到有害事件后，先通过容错机制使这些有害事件是非致命的，最后通过 DIR(Detection, Isolation, Recovery) 机制将系统从非致命状态还原至可信赖状态。但这种设计方法无法被直接应用于自主系统中。这主要是因为极度复杂和无法预测的环境以及系统的机器学习部分本身就是黑箱。这一问题在一份 2017 年关于撞车事故模型的分析报告中可见一斑，交通事故需要更加详细和复杂的分析才能保证系统被还原到可信赖状态，这包括对事故当时环境的分析。针对这一问题的关键方案是尝试用运行监测 (Run-Time Monitoring) 机制来取代 DIR 机制。

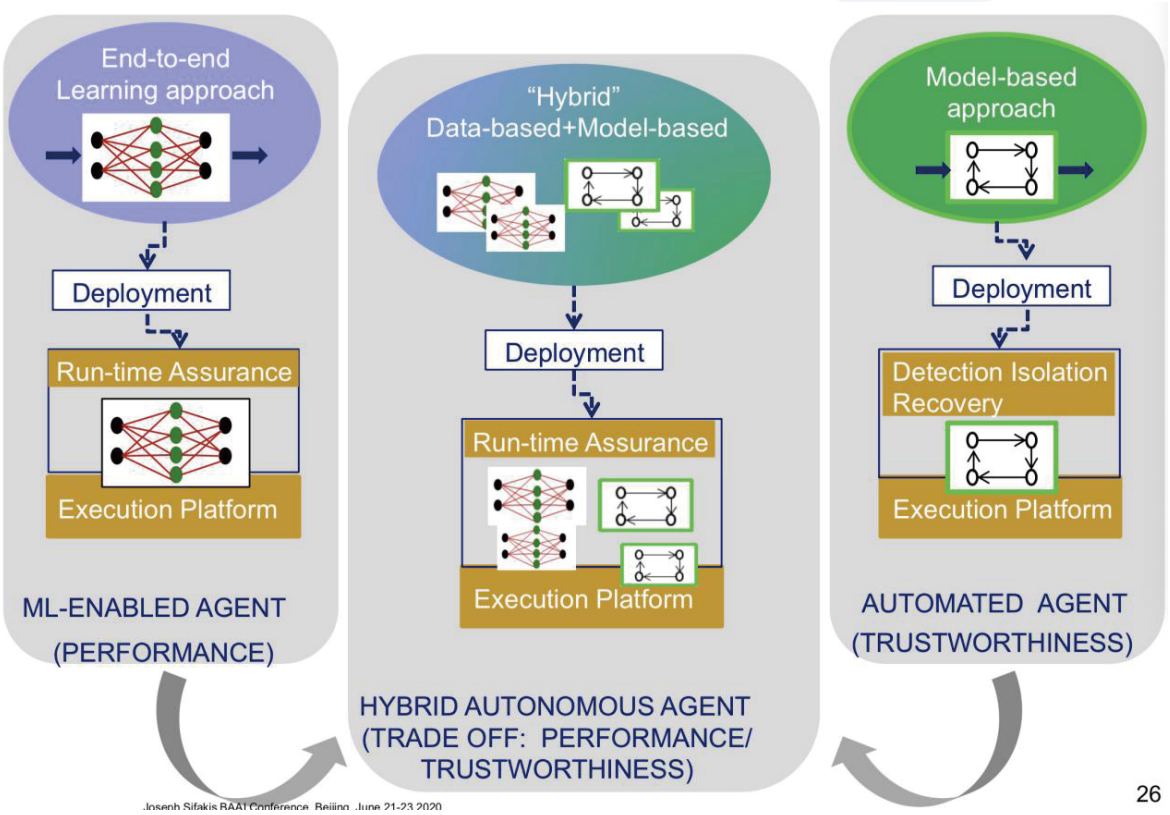


图 7：混合设计流程（Hybrid Design Flow）

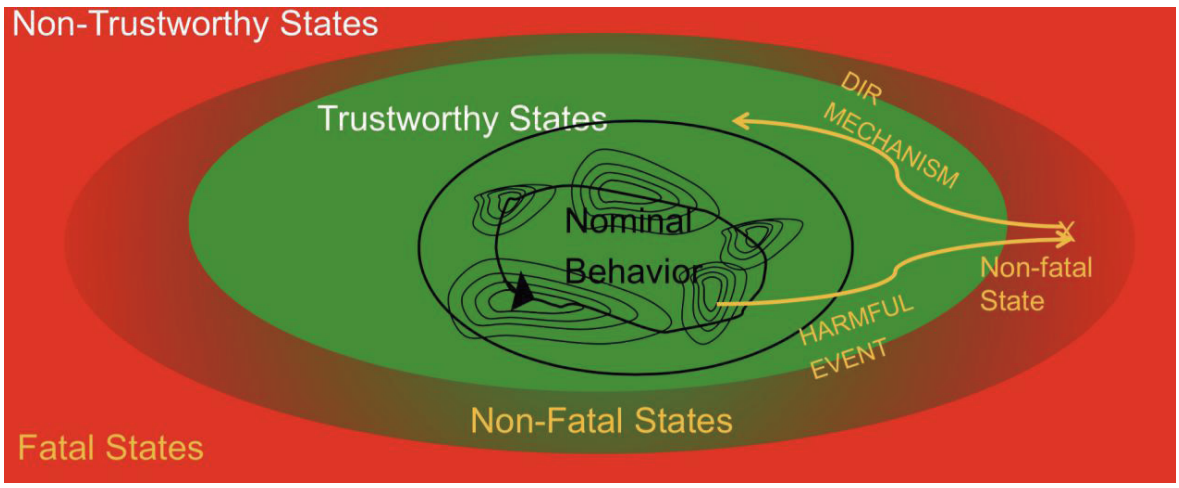


图 8：传统的基于模型的设计方法是如何保证可信赖性的

自动驾驶的混合设计目前有一些设计理论，如通过机器学习进行情景认知，再基于模型进行适应性决策。感知功能可以识别定义明确和完整的环境结构来作为驾驶模式选择上的参考，而每一种驾驶模式都需要详细的策略。

这种适应性决策过程可以划分为如下层级:

Level1: 防碰撞系统和轨迹跟踪控制系统;
Level2: 策略协议如超车、会车、停车等;
Level3: 环境模型和分析，包括安全区域计算、轨迹计算、驾驶模式选择等;
Level4: 行程目标和计划。

之后 Joseph 提到了系统设计关键的一环，系统如何被验证。仿真是系统验证中至关重要的一环，仿真模型的搭建覆盖了从技术到理论的方方面面。一个好的仿真系统应该有如下三个特性:

1. 真实性 (Realism)，也就是让终端的行为和环境的表现尽可能贴合现实世界。
2. 语义意识 (Semantic awareness)，仿真系统动力学应扎根于过渡系统语义 (transition system semantic) 中，它包括 Notion of state 来确保对实验的掌控性和可重复性；Scenarios 来模拟 / 检测极端状态和高风险情形；Notion of coverage 来衡量相关系统配置已经被实现到什么程度。
3. 多尺度多粒度的建模和仿真 (Multiscale multigrain modeling and simulation)，包括在理论上，要加强对网络物理系统的建模、不同尺度之间的关联研究；实践中，对基础联合仿真引擎的开发如 HLA、FMI 等。

目前大多数工业仿真系统缺乏语义意识，他们大多依靠游戏引擎或者预建的软件。这是无法完全满足自主系统可靠性验证需求的。

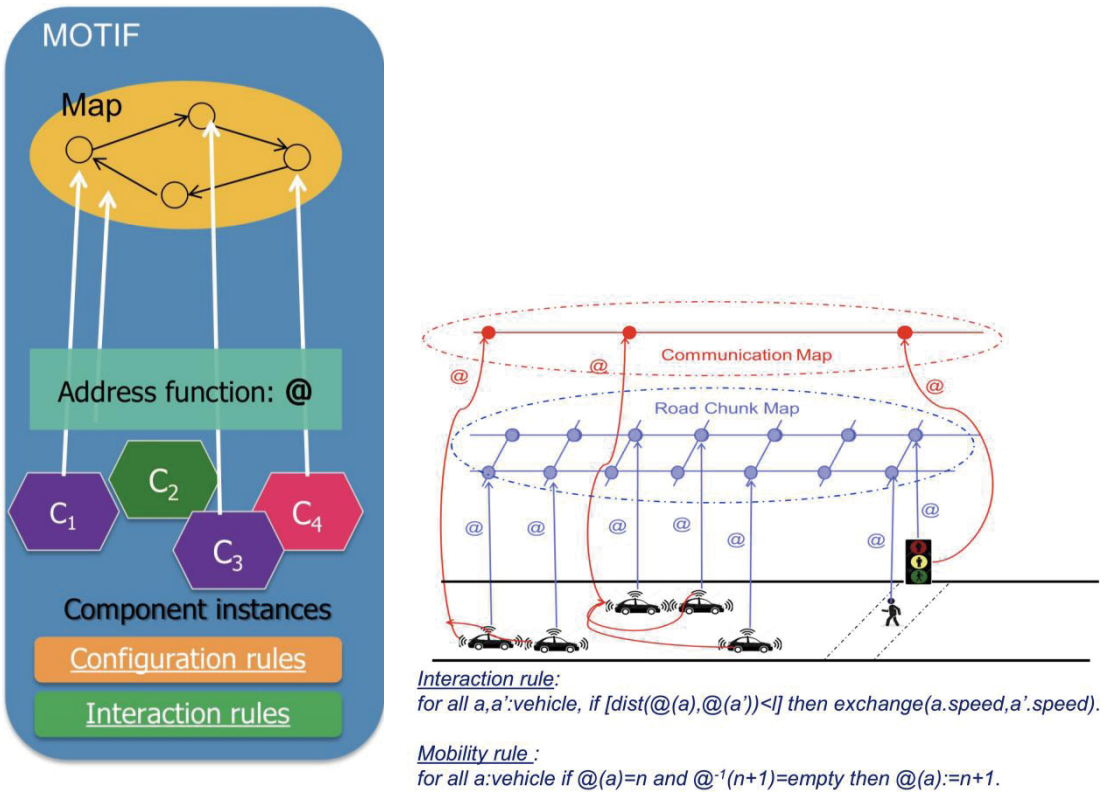


图 8：DR–BIP 的一些基本概念以及它们在自动驾驶中的实例

接下来 Joseph 向我们介绍了他们实验室开发的一种针对自主系统的仿真语言，DR-BIP (Dynamic Reconfigurable BIP)。如图九左图所示，DR-BIP 将系统看作由一系列主题 (Motif) 构成，每个主题都是一个协作模型，它包括：一系列元件 (终端、客体)；一个含有节点和边的地图，它用来表示元件之间的关系，这一关系可以是空间上的、也可以是组织上的。一个定位函数来将元件映射到地图上的各个节点。我们定义了交互规则来规定元件之间的交互 (Atomic Multiparty Synchronization)；同时我们还定义了配置规则使得我们可以添加 / 删除 / 移动元件并动态的变换地图。

我们以机器人车的仿真为例，如图九右图所示，这一复杂系统用到了两个主题，它们分别包含了路线地图和通讯地图。机器人车是终端，行人、路灯是客体，它们都属于主题中的元件。我们可以定义这样一个交互规则，当两辆车的距离小于一定阈值，则它们的速度交换；我们也可以定义一个配置规则，如当一辆车前面的一个节点为空的时候，我们可以将该车重新映射到前面的节点来表示机器人车的移动。

四、结语

Joseph 认为目前的验证方式还远远达不到自主系统的要求，首先机器学习是无法被验证的，它的开发并不是基于一个形式化的目的，目前它对我们依然是个黑箱。而之前自动系统应用的形式化验证 (Formal Verification) 也有着诸多缺陷，它只在需求和目标可以被形式化的时候有效，难以应对高度动态且可以不断重构的自主系统。

讲座最后，Joseph 对于机器自主性的未来进行了一些展望：

1. 很多自动驾驶厂商会因为技术问题以及公众信任的崩塌而改变他们的野心——现实远远不像他们现在说的那样乐观。
2. 我们会摒弃基于数据还是基于模型的争论，将两者进行混合设计。
3. 自主性的级别会逐步平稳地向全自动或者共生自主迈进。
4. 自主系统的可信赖性依然是一个亟待解决的问题，我们没有决定性的证据表明这些系统是足够严格可靠的。虽然这些系统并不足够可靠，但好消息是它们也的确在向着更高层次的智能发展。我们将给予这些系统多大的决策自主权，我们何时能够相信它们，这些都是值得进一步思考的问题。

英国两院院士 Steve Furber：重建大脑——人工和生物智能

整理：智源社区 孙肖月

在第二届北京智源大会“闭幕式及全体会议”上，英国两院院士、ARM 核心设计者 Steve Furber 做了题为《Building Brains — Artifical and Biological Intelligence 》的主题演讲。

在此次演讲中，Furber 主要介绍了其主持的 SpiNNaker 项目，该项目用一百万个 ARM 核心来模拟神经元，通过互联实现百分之一的人类大脑的仿真。Furber 首先介绍了项目的历史渊源和项目的生物灵感，接着介绍了第一代 SpiNNaker 的研究成果和应用，研究人员在 SpiNNaker 系统上成功模拟出了皮层微环路结构。最后介绍了第二代 SpiNNaker2 的设计创新之处，该系统将于今年底推出。Furber 认为对大脑的了解对于人工智能的发展非常重要，可以用它来指导人们如何设计人工智能系统； 目前人们并没有从根本上了解大脑如何工作的，SpiNNaker 项目的目标便是帮助人们理解大脑，帮助人们设计更好的 AI。

SpiNNAker 项目构思于 20 年前，建设历史为 10 余年，与当前世界各地的组织都有联系。其在曼彻斯特建立了百万核的机器，由欧盟人脑技术支持。目前人工神经网络和脉冲网络是并行发展的。工业 AI 使用二代传统神经网络（非脉冲网络）现在有一个日益增长的期望是，希望脉冲网络可以给工业人工智能提供支持，尤其是能源效率方面。反过来，工业人工智能的巨大进步和深度网络等同时告诉人们关于大脑运作的一些知识。SpiNNAker 是一个理想的探索平台。

Steve Furber，数学家，计算机科学家和硬件工程师。现为英国皇家科学院与皇家工程院院士，欧洲科学院院士，英国计算机学会、IET、IEEE 会士。在计算机系统结构，微处理设计方面享有盛名，被誉为“ARM 架构之父”。他主持研制的 ARM 微处理器已经成为手机、平板电脑和其他智能设备及嵌入式系统的主流微处理器，年出货量高达 100 亿颗，总出货量达 500 亿颗。目前，他主持开发的 SpiNNaker 神经形态计算机，是欧盟“大脑”项目的核心。该计算机包含 100 万颗 ARM 微处理器，是目前世界上规模最大的专用超级计算机。Furber 因其在计算机领域的卓越贡献，荣获 IEEE 计算机先驱奖，IET 最佳成就奖，大不列颠帝国司令勋章等终身荣誉。

一、200 年的历史

200 年前，维多利亚早期女工程师 Ada Lovelace 是计算机界著名科学家 Charles 的助手，也是早期的工程师，设计和制造了非常早期的机械计算机，是最早思考算法的人之一。许多人认为她是第一个像电脑程序员一样思考的人。她经常谈到算法，对大脑很感兴趣，希望能够理解大脑现象，并用方程式表示出来，然后把神经系统的“微积分”传给下一代。然而直到今天，我们仍不能提供神经系统的“微积分”，并且仍然缺乏对大脑如何工作的根本理解。

150 年后，计算机科学和密码学的创始人图灵，来到了曼彻斯特。图灵在曼彻斯特从事了许多项目，其中一项就是题为《计算机机器智能》的论文。这篇论文从以下句子开始——

“我提议考虑这个问题——机器能不能思考?”

这篇论文发表于 1950 年，此时曼彻斯特机器运行第一个程序已经两年了，图灵开始怀疑这项技术能走多远。他在文中提出一个类人人工智能测试，称之为“模仿游戏”，即今天著名的“图灵测试”。图灵估算，与曼彻斯特机器相比，一台机器要展示人类的智能，需要用更多的内存。当时曼彻斯特机器只有 128 字节内存，图灵认为只要确保记录是千兆，就能够模仿人类智能。图灵在文中还预测，二十世纪末机器将拥有千兆内存，这是很了不起的预测，因为当时曼彻斯特机器只有 128 字节； 在世纪之交，典型的台式电脑有 2000M 字节的内存，它的功能是曼彻斯特机器的百万倍，然而却仍然无法通过图灵测试。直至今天，也没有一台机器能够令人信服地通过图灵测试。

Furber 认为实现类人智能之所以比图灵所预计的困难得多，原因有多方面，最主要的原因在于，我们不了解人类智能，不清楚大脑是如何工作的。这也是为什么在过去的 20 年里，Furber 致力于通过构建合适的机器，来帮助人类理解大脑功能。

二、生物灵感

在过去 20 年里，有两个主要的研究问题在推动着 Furber:

第一，大规模并行计算资源是否可以加速我们对大脑的理解?
第二，我们对脑功能的了解能否为更有效的并行容错计算指明道路?

他对上述两个问题的答案是肯定的。在报告中，Furber 举了个例子:

图 1 是用来做图像分类的简单卷积神经网络。左边的图像，向右传递给网络，网络包含很多层，第一层是输入层 (图像)，中间是隐藏层，最后一层是是输出层，每层内神经元都有大量的参数，如果输出层给出错误的答案，错误从右向左传播到网络上，对参数进行调整，有时甚至调整数亿次，直到网络开始正确分类图片。在早期，谷歌用神经网络学习了大约 10 万张猫的图片后，才学会对猫进行识别。而作为对比，一个两岁的孩子看到一只猫后，就可以在他余生中可靠地识别出猫。这说明人工网络和生物网络学习知识的方式有着根本的不同。ConvNets 的所有参数都是随机的，而两岁的孩子却有两年的经验，大脑内有一个复杂的模型，这就是人工网络和神经网络执行类似任务时有巨大差异的根本原因。

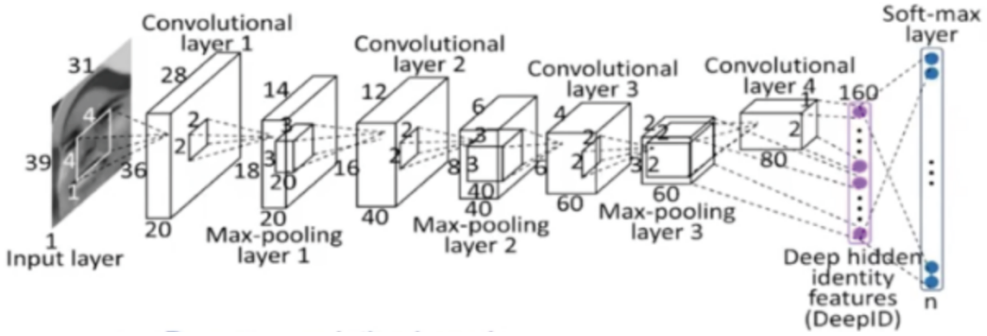


图 1： ConvNet 结构图

图 2 是一个简单抽象的大脑皮层结构图。Furber 用非常简单的形式显示了大脑皮层的一小部分。从图中可以看

出，在网络中信息不是简单地从某个位置输入，某个位置输出； 在网络的任何地方，信息都是前后流动的； 从不同的位置输入，可以在网络中不同的层输出。生物网络很复杂，不是简单的线性过程。但我们至今仍还没有明白基本的皮层的神经网络是如何工作的，而理解基本的皮层网络如何工作是理解大脑的研究重点。

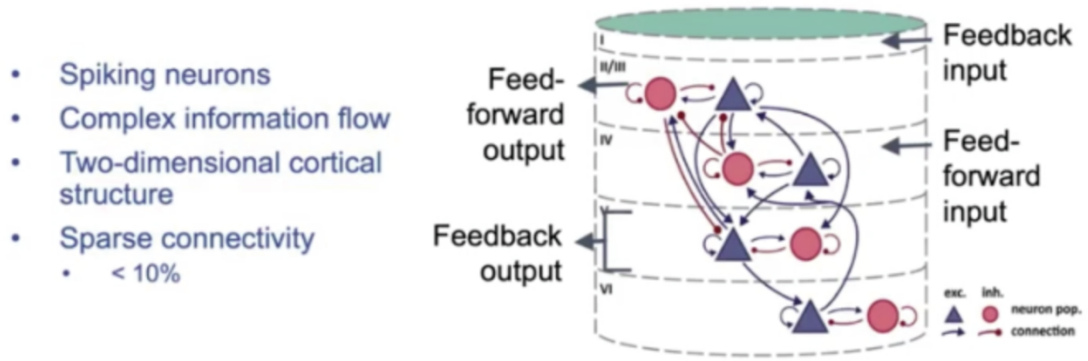


图 2：皮层简化结构

Furber 认为要模拟人脑神经网络，强大算力的计算机是一个好的选择，比如算力强大的 GPU，非常擅长密集矩阵乘法。目前神经网络是密集的，可以很好地映射在 GPU 上。然而要训练庞大复杂的网络，这种方法虽然有效，却耗能巨大。

另一种方式是脉冲网络。脉冲网络相比于标准的计算机更具有生物性，它能够做类似 GPU 的事情，同样擅长矩阵运算。

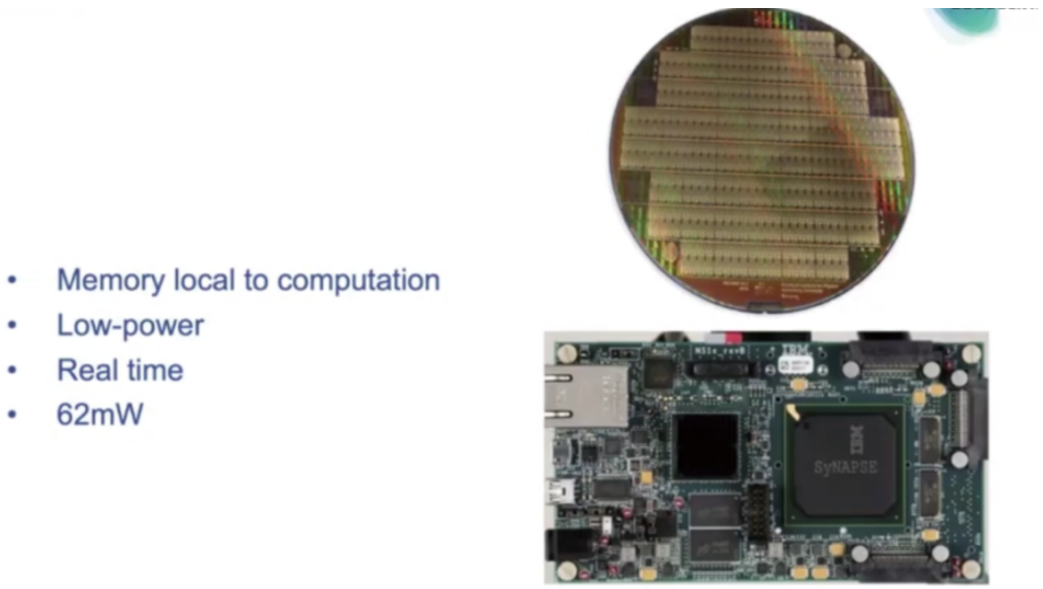


图 3：神经形态硬件

为了实现脉冲网络，人们设计了一种硬件——神经形态硬件（如图 3）。其特点是内存和计算距离很近，这种设

计大大降低了功耗。脉冲网络还模仿了松弛的生物实时网络，这在超级计算机中是做不到的。这种芯片可以很好地处理学习任务。因其明显的能效，工业界对这一领域的兴趣在持续增长，例如三星使用 IBM 的 brain-inspired 芯片识别姿势；Intel 推出研究原型芯片等。不过目前为止，脉冲神经网络在商业领域的可行性还有待观察。Furber 则希望能够通过脑科学的研究来促进这方面的工作。

三、SpiNNaker

要理解大脑是如何表达和处理信息，最后的方法之一就是建立一个电脑模型来测试我们设想的人脑工作方式。受生物神经网络和人类智能的启发，Furber 主持了 SpiNNaker 项目，此项目构思于于二十年前，用一百万个 ARM 核心来模拟神经元，模拟哺乳动物大脑神经元连接特点互联，实现大规模并行运算，可并行运行一百万个进程。虽然有一百万个进程，但也只能模拟大脑 1% 的计算量；更乐观地讲，约等于 10 只老鼠大脑。

一个 SpiNNaker 机器（图 4）是一个由处理节点构成的均质二维多指令多数据阵列。每一个节点（即 SpiNNaker 芯片）整合了 18 个 ARM 处理核，每个核包括本地内存和共享内存，一个包路由，还有一个通用斯通支持协议。每个 SpiNNaker 芯片会选择 18 个处理器核心中的一个作为监视处理器（monitor processor）。这种选择是出于容错灵活性的考虑。一旦监视处理器被确定，他会被分配一个操作系统支持规则。接下来的 16 个处理器会被赋予应用支撑规则，第 18 个处理器则会作为容错冗余封存。每个芯片大小 1 平方厘米，封装在 2 平方厘米的标准外壳内。这个封装包含了一个定制的多处理器片上系统 SoC 集成电路，这些处理器都通过自计时的片上网络连接到数个片上共享资源和第二个芯片。

封装的芯片会被装配到印制的电路板上（PCB，printed circuit boards，即 SpiNNaker 板），每个 PCB 包含 48 个芯片，864 个 ARM 处理核。PCB 上的芯片间连接是使用自计时 2-7 不归零协议来传输 4bit，两线制的符号，此外还有一根线用来传输应答回复。SpiNNaker 系统是由一系列 48 芯片的 PCB 构成的，PCB 间连接是使用高速串行连接。不同大小的 SpiNNaker 系统可以由一个或多个 48 芯片的 PCB 装配而成。目前，全世界范围内约有 100 个 SpiNNaker 系统在运行，其中 4 节点的小板可以非常方便的用于训练，开发和移动小型机器人；大部分系统是包含 48 节点的 SpiNNaker 板。

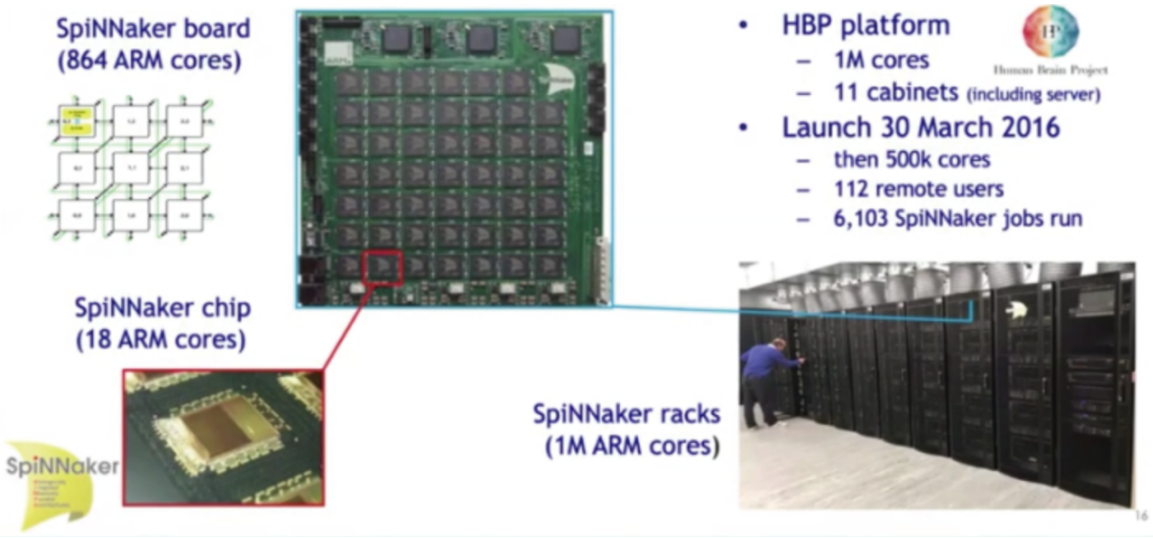


图 4：SpiNNaker 机器

该项目受到了欧洲“人脑计划”的支持，于 2016 年 3 月 30 日启动。2018 年，Furber 便实现了百万核的最初目标，这也是当前在服役的部分； 到当年 11 月，Furber 将机器的规模扩展到 400 万核。此机器可以从世界上任何地方通过互联网访问，用户可以在机器上运行作业。目前，Furber 等人把机器的负载降到了 25%，即平均在用 SpiNNaker 核为 25 万个。

四、皮层微环路

目前这些机器能够做什么呢?

将 SpiNNaker 系统应用于模拟皮层微环路，是一个有趣的工作。在今年年初，Furber 团队正式宣布模型运行成功，其计算速率相比于超级计算机以及 GPU 模块要快很多。这是微电路的奇迹!

他们模拟了 1 平方毫米的大脑皮层（图 5），包含 77000 个神经元，28500 万个突触，并以 0.1 毫秒的时间步长进行计算。计算结果显示，相比超级计算机 HCP 要慢 3 倍，相比谷歌的 GPU 要慢 2 倍。“脑计划”的下一步将会把模型扩大为多个区域，规模扩大到 100 平方毫米，区域之间像大脑皮层的不同区一样进行连接。

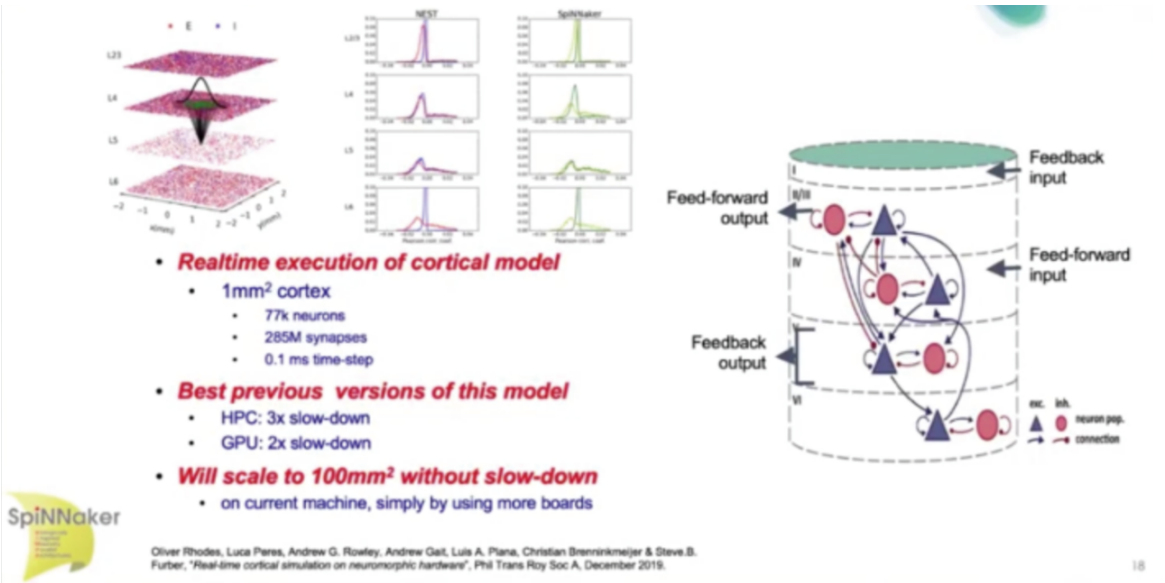


图 5：皮层微环路

在更抽象的层次，Furber 证明了脉冲神经网络还可以用于解决约束满足问题（图 6），这是计算机科学中的难题。演讲中 Furber 展示了数独的例子，这是一个简单的约束满足问题，类似的问题包括地图着色、旋转系统等。这些例子说明了脉冲网络有有很好的计算特性。

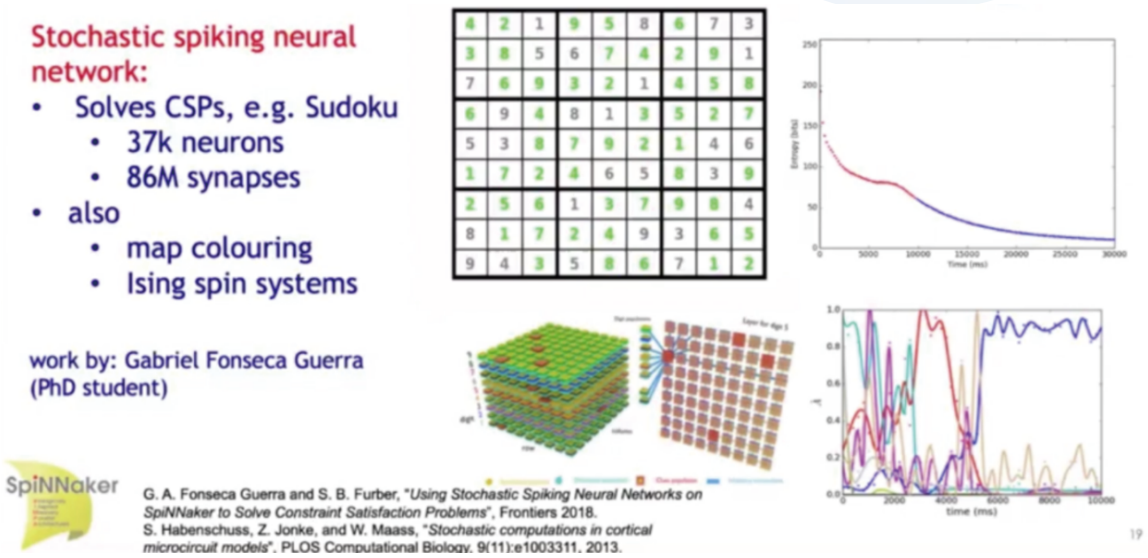


图 6：约束满足问题

四、SpiNNaker–2

最近，Furber 与德国合作者再次开发了第二代机器 SpiNNaker–2，基于第一代机器的原理和经验，二代机采用了更先进的 CMOS 技术来扩展系统，每个芯片由 152 个基于 ARM 的处理元件（PEs）（图 7），所有其他的系统参数也相应等比例的放大，系统具备更强大的内存连接。如图 PEs 布局在一个个核心处理岛中，芯片大部分面积由处理单元占用，采用本地存储。每个神经元是发送一个小数据包，信息在一个大的网络中通过多个网络快速到达其他芯片，为保持生物实时性能，需在毫秒时间内完成计算。SpiNNaker–2 测试芯片原型机正在实验室制作当中，全尺寸的完整芯片有望在今年底制成。

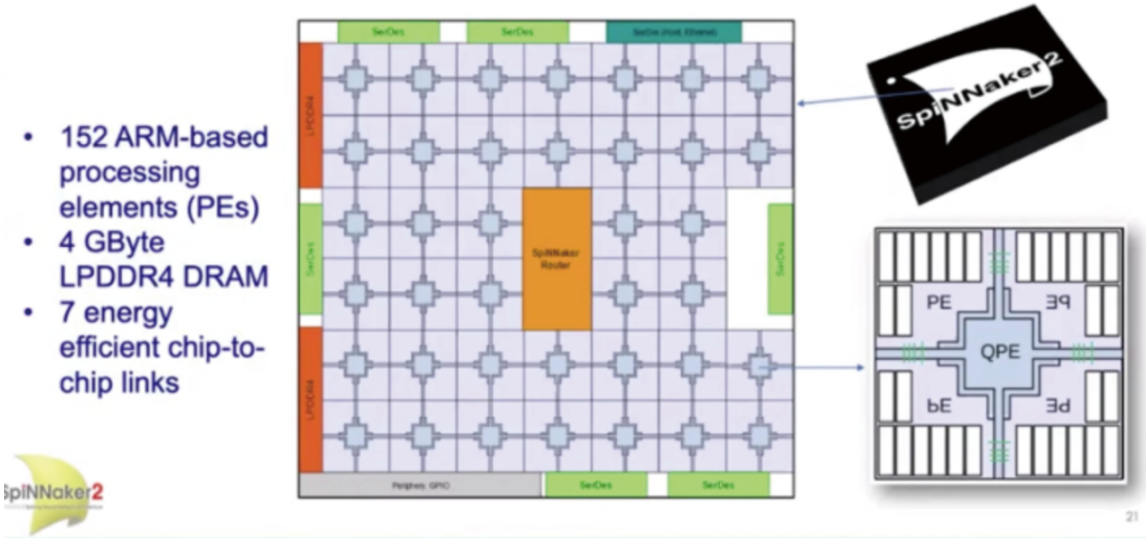


图 7：SpiNNaker–2 芯片结构

二代机器吸取了第一代的教训，使用动态能量管理的方法来提升能源效率，每一个 PH 都可以单独操作一个更高或更低的时钟频率和电压，因此在一个时间内，通过开关可以吸收过程中的峰值，并在需要时，逐步提高吞吐量。在这个过程中，内存被灵活共享。Furber 发现，随着建立的模型丰富度和复杂度的增加，共享内存越来越有用。目前他们正在努力改进这个特性，以便它可以支持传统的人工神经网络和脉冲网络。他们给每一个 ph 增加乘积加速器，使芯片几乎有了最先进的性能，反过来，也使机器支持了标准人工神经网络。另外他们还为脉冲网络最关键的突触和神经计算功能提供神经形态加速器，对加速基本的指数和对数函数功能很有效。为了高效脉冲通讯，他们还采用处理元件连到芯片上的网络。而在最底层，硅的使用单体自适应在补偿制造过程中的变化，因此减少制造了过程中不同参数范围，并使芯片在表现和能效方面得到更严格的优化，减少了制造过程中的主体变化。

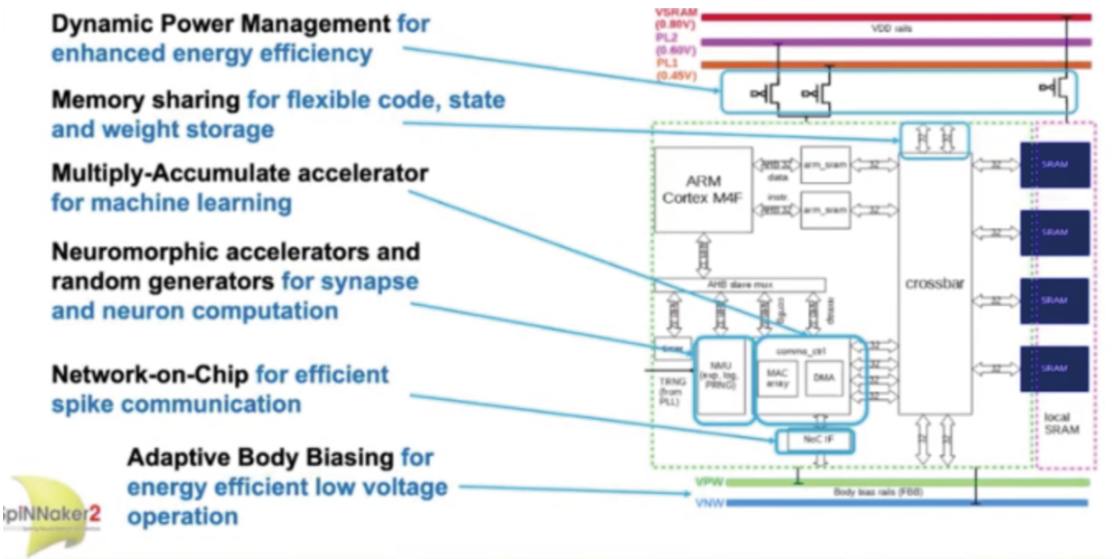


图 8：SpiNNaker–2 处理元件

那么新一代芯片可以做什么？ Furber 展示了深度重布线，人工神经网络的一个标准基准。 网络中连接的数量减少超过一个事实或 100 个，不会显著影响分类正确率，这可以减少内存需求，这可以将网络内存需求降低 30 倍，能量需求降低 100 倍，而不会显著影响分类精度。例如 LeNet300–100，内存由 1080kb 降到 36kb，可在本地 SRAM 上训练，在二代 SpiNNaker2 原型上训练能量比在 X86CPU 上能量减少 100 倍。Furber 深度重布线是是生物网络的一个重要的特性，将来会越来越多的出现。生物模型和工业人工网络存内在差异，在未来将影响硬件设计支持人工网络的运作方式。

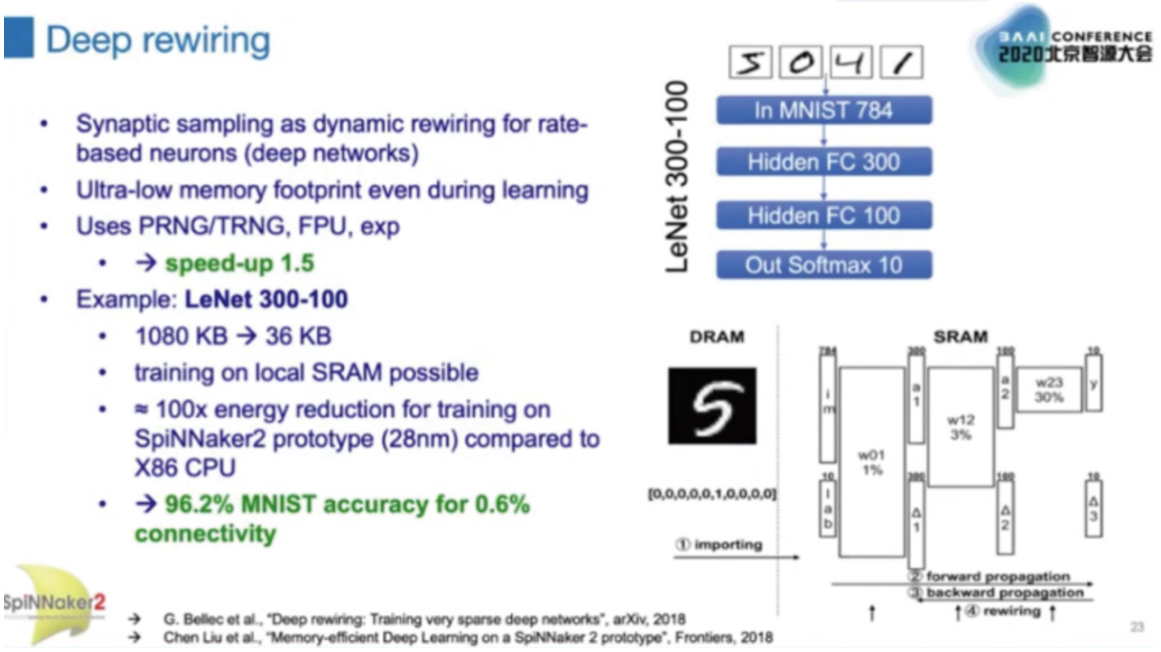


图 9：深度重布线

SpiNNaker 项目构思于 20 年前，建设了 10 多年，与世界各地的组织有联系，由欧盟人脑技术支持。目前人工神经网络和脉冲网络是并行发展的。工业 AI 使用二代传统神经网络（非脉冲网络），现在有一个日益增长的期望是，希望脉冲网络可以给工业人工智能提供提供支持，尤其是能源效率方面。反过来，工业人工智能的巨大进步和深度网络等同时告诉人们关于大脑如何运作的一些知识。SpiNNaker2 是一个理想的探索平台。

微众银行 CAIO 杨强：AI 与人的新三定律

整理：智源社区 蒋宝尚

在北京智源大会“全体大会”上，微众银行首席人工智能官兼香港科技大学讲席教授杨强做了题为《AI 的新三定律：隐私、安全和可解释性》的报告分享。

杨强在报告中提到：人工智能和人类之间具有微妙关系，传统的阿西莫夫的机器人三定律并不能解决越来复杂的微妙关系。需要启用新的三大定律：人工智能需要保护人的隐私、人工智能需要保护模型的安全、人工智能需要人类伙伴的理解。这也就意味着，人工智能不能脱离人去发展，需要让人和人工智能形成协作关系。

以下是文字整理:

我今天演讲的题目是《AI 与人的新三定律》，讲座的内容在于探讨 AI 的下一步，AI 与人的关系。

先从科幻小说开始讲起。众所周知，阿西莫夫在科幻小说里打造的机器人三定律，描绘了人与机器和谐相处的场景。

其中，第一定律为：机器人不得伤害人类个体，或者目睹人类个人将遭受危险而袖手不管；第二定律为：机器人必须服从人给予它的命令，当该命令与第一定律冲突时例外；第三定律为：机器人在不违反第一，第二定律的情况下要尽可能保护自己的生存。

童年时期，我认为这三个定律很神奇。毕竟，如果真的有这样一个机器人，能够为人服务，又能够自主地做决定，那么在这三个定律的制约下，那么这个世界真的就会像科幻电影中描绘的那样：人与机器以某种平衡生活在一起。

但是，现实并没那么简单。因为，首先无论是研究人工智能，还是研究人工智能驱动的机器人，发现这些工作都是离不开人类。

再者，研究人工智能的机器学习和模型时，要遵循的规则可能和阿西莫夫的三定律存在并不相同。

那么，哪些不相同的地方呢？换句话说，有哪些新的规则和定律产生呢？在 AlphaGo 时代，人类畅想的 AI 场景是无人化，例如无人工厂、无人车、无人商店等等。

而现实是：AI 是需要人类做伙伴，人类也需要 AI 做伙伴。首先，AI 的运算结果是要解释给人类用户的，而 AlphaGo 就不是这样，表现为：在下棋的时候，它并没有能力解释为什么走这一步棋而不走那一步棋；其次，AI 的运行是要让人类工程师进行 Debug，显然 AlphaGo 也没有这个功能，表现为：它下的“臭棋”到现在都没有办法 Debug。再者，AI 的流程需要人类监管，而 AlphaGo 是完全不受人的监管，表现为：在棋盘的世界中，它可以自由驰骋。

最后，AI 的模型和系统需要解释因果关系，而 AlphaGo 没有把解释赋予人类，也即它在设计的时候没有考虑人的因素，包括后面的技术发展，例如自监督学习、AlphaGo Zero 都在往此方向发展。显然，这并不是今天想看到的场景。

一、AI 与人的新三定律

图 1：AI 的“新三定律”

我们在期待什么样的场景呢？我总结了一些新的规律。首先，AI 要保护人的利益，这一点毋庸置疑，其中用户的隐私是非常重要的利益，当时设计 AlphaGo 的时候，科学家学并没有完全没有考虑到的此因素。

其次，AI 不仅要保护人的安全，并且也要保证模型也是安全的，现在的这类研究只是刚刚开始，即如何防止恶意或非恶意的攻击模型。

最后，AI 需要人类伙伴的理解，并且要促成这种理解。也就是说 AI 模型是需要可解释性的，而且针对不同的人类，解释也应该是不一样的。

二、AI 要用来保护用户的隐私

下面详细解释这三个定律：AI 要用来保护用户的隐私。就像刚才提到的，今天 AI 的力量来自大数据，但是我们周围更多的是小数据，例如人工智能在法律上的应用，每个案例的收集都是“旷日持久”，因为它需要很多的标注，需要很多的人类经验积累，最后才能形成一个案例。

AI 在金融领域的一个重要的应用是反洗钱，洗钱的金融交易往来和非洗钱相比数据量远远不足； 另外洗钱的每个案例也都极具特点，所以它属于小数据量。

人工智能在医疗方面的应用也受到了数据量的制约，毕竟医疗图像这种高质量有标注的图像是非常少。

在科幻小说当中描述的无人车、无人机、机器人，在现实中相当于一种端计算，它们通过不断地和云端沟通产生数据，每一个终端的广义机器人（包括摄像头等）看到的周边景象（数据）也都只是反映一个角度。因此只有

把所有不同的角度会合起来，才能对事物有一个全面的了解。

图 2：个人隐私与数据法规——欧盟的 GDPR

数据量有限只是一方面，另一方面涉及到数据聚合过程当中的隐私保护。现在世界上出台了很多保护个人隐私的法规，让数据聚合的方式变得不是那么容易，例如欧洲在 2018 年 5 月出台的 GDPR 法案，意味着在为一个目的收集数据时，就不能轻易用在另外一个目的机器学习上，否则就会违反法律，从而遭受很严重的罚款。

图 3：国内的数据监管法律趋严

国内的监管也逐渐趋严，各个大数据公司都不得不认真学习国家各项法规法案。法规的制定也有着通用化、多样化的趋势，因而形成了一种理想和现实的隔离。

理想是： 我们的大数据可以驱动人工智能，从而发展成通用人工智能为人类服务； 现实是我们面临的是众多的数据孤岛，由于利益、隐私、条例、法规等多种因素，无法将数据孤岛聚合起来。

为了解决数据孤岛问题，技术人员一直在探寻解决方案。其中一个办法叫做联邦学习，英文名为 Federated Learning，主要思想可以总结为：数据保持在原地，但模型通过加密的方式和不同机构进行沟通，带来的效果是：数据可以被使用，但是各方都看不见对方的数据。

图 4：联邦学习能够保护用户的数据隐私

具体怎么实现呢？要求是在确保用户的隐私得到保护的情况下，模型的参数也要受到保护，此外，还要考虑模型的能力和效果，即联合建模比单独用一方的数据建模效果更好。

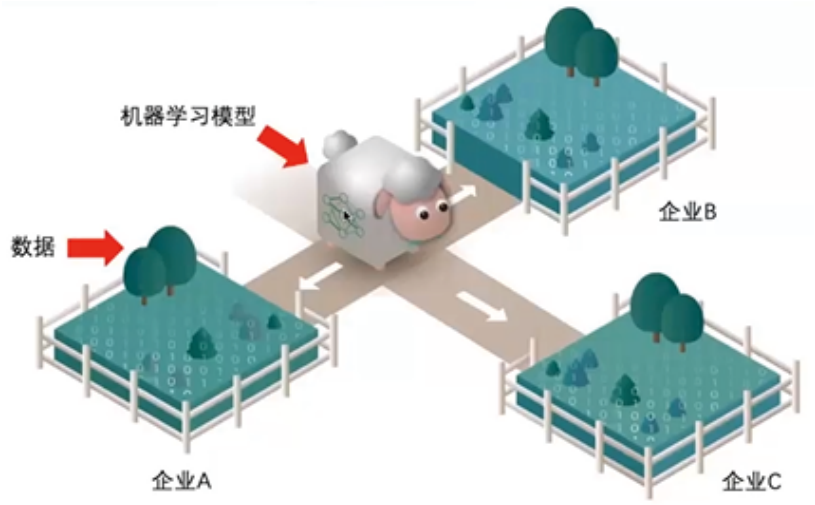


图 5：以养羊为例介绍联邦学习

举一个更加生动的例子：养羊。过去养羊是把草丛各地集中到一起喂羊，在数据层面，并不合规隐私和数据安全保护的要求使得获取数据成为障碍。而联邦学习提供的新思路是：让羊群在各地移动，而联邦学习提供了新的思路：让羊群在各地移动访问草场，而草不出本地，也即主人无法知道它吃了哪些草。

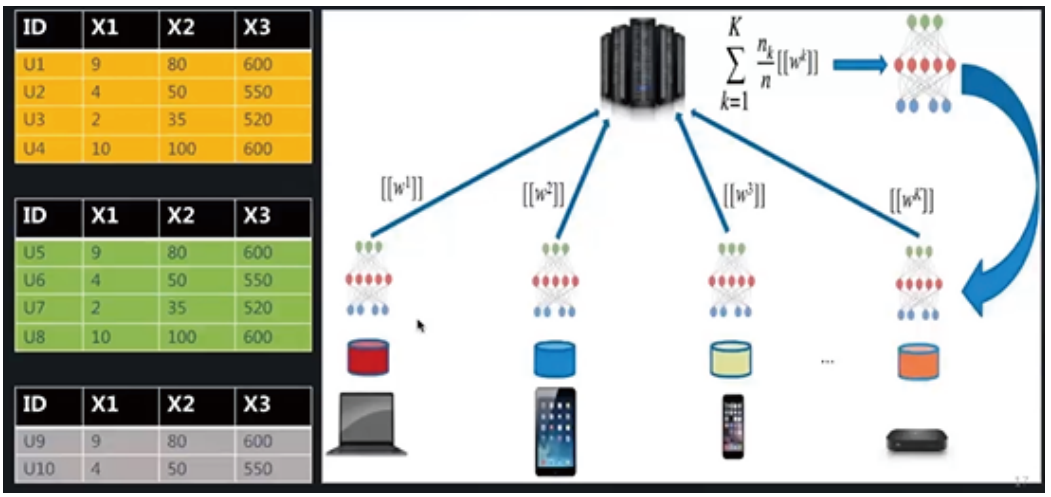


图 6：第一类联邦学习问题——按样本进行分割（横向切割数据）

现在回过头来看最开始的那个例子，现在存在的问题是： 有不同的终端，需要把每个终端在每一天都聚集的新数据收集起来，模型的训练不仅可以根据自己的数据，也可以利用其它终端的数据，从而不侵犯隐私。

如何用联邦学习完成任务呢？如上图所示，每个颜色每个表格代表一个终端收集的数据集，这些终端数据的特点是： 其特征 X1 到 X3 基本相差无几，差别在于数据的量不同，例如第一个终端收集的用户是 U1~U4，第二个终端收集的是 U5~U8，第三个终端收集的是 U9~U10。这种数据的特征几乎相同，但用户和样本却是不一样，在面对这种横向切割的数据时，我们采用的方法是横向联邦学习。

如何更新此模型呢？ Google 在 2016 年曾经提出过联邦平均 (Federated Averaging) 的方法，具体是： 往云端传递的消息只包含模型的参数，并且是受到加密的保护，然后通过计算参数的平均值在云端进行模型更新，整个的过程不泄露本地的隐私，也不泄露模型的参数，谷歌将这种做法也用在了 Android 系统中。

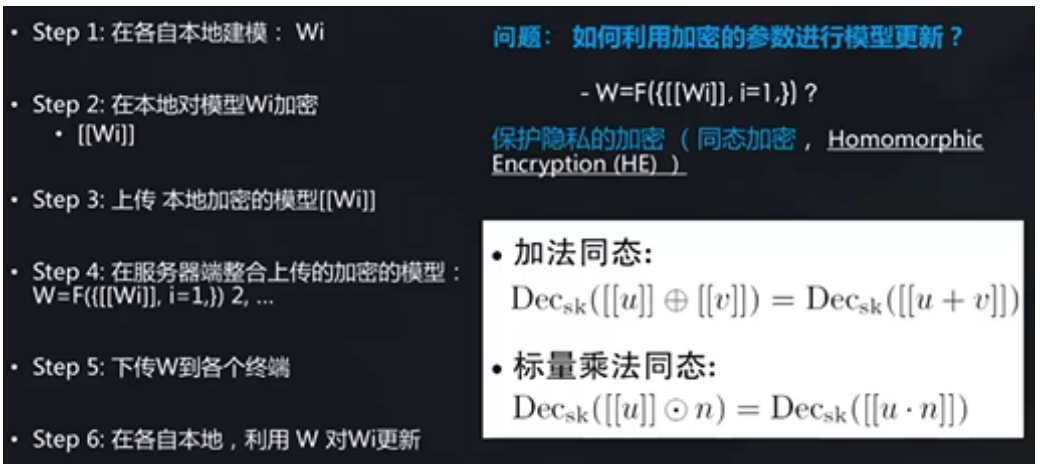


图 7：联邦学习关键技术——加密 / 解密

显然，这里的关键技术是加密和解密的算法。当前，已经存在各种各样的算法即能够支持保护数据，也能够允许在加密层上进行一系列的数学操作和数学运算。例如同态加密。

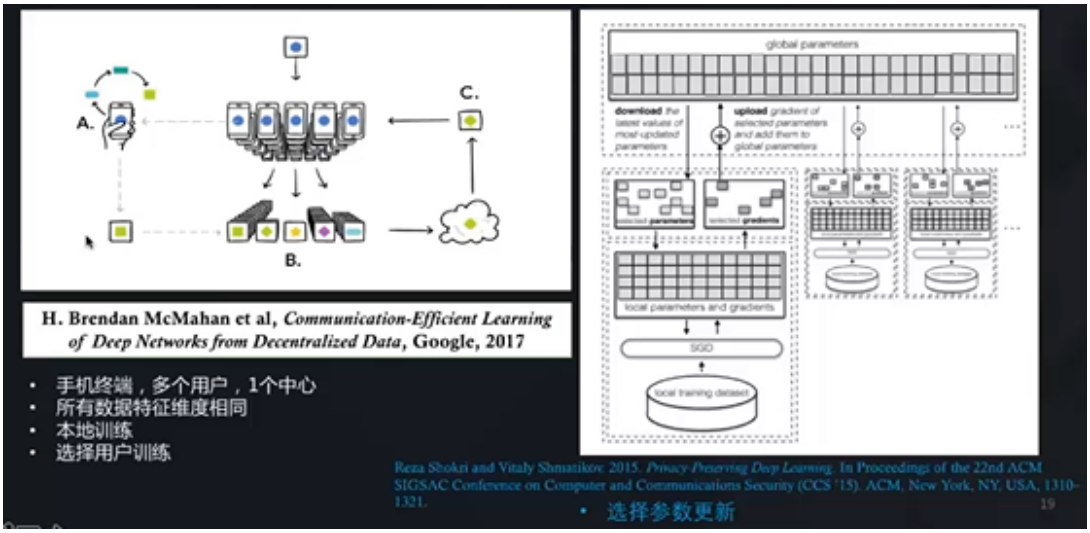


图 8：谷歌的横向联邦学习

综上，Google 联邦学习的思路如上图所示：A 点代表用户的数据，数据上传到模型，然后进行训练，训练之后将参数上传到云端进行聚合，聚合之后，然后下发到终端完成更新，如此形成闭环。当然，此过程支持随机梯度下降等算法。

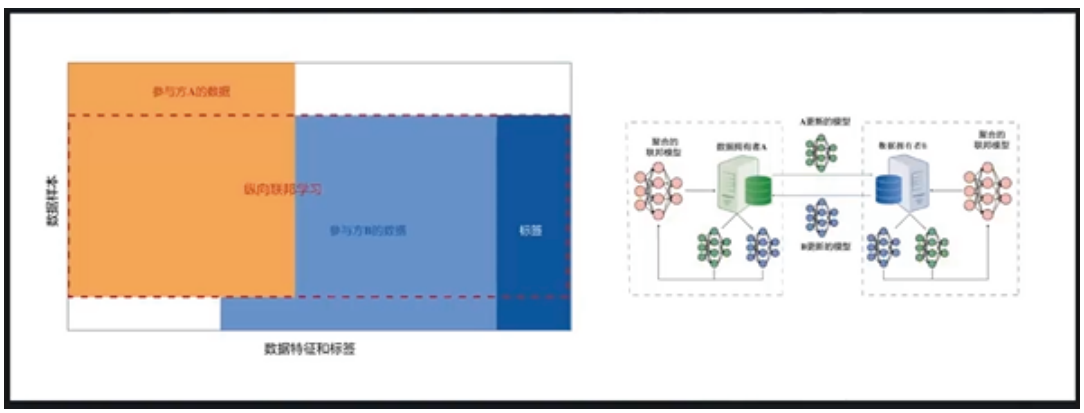


图 9：纵向联邦学习（特征不同，样本重叠）

不同于横向联邦学习，纵向联邦学习能够将特征不同，样本重叠的数据孤岛进行聚合。例如有一家银行和互联网公司，这两家公司有很多相同的用户，但是这些用户在这两家公司“保存”的数据却不同，如果在银行里是金融相关的数据，那么在互联网公司里的是用户隐私相关数据。

那么这两个机构如何进行合作，从而做出更好的风险控制模型，答案是用纵向联邦学习。综上，横向联邦学习是针对 To C（消费端）方向，纵向联邦是针对 To B（企业端）方向。

既然联邦学习有这些方向，研究者在哪些领域有更多地投入？特别推荐最近我们和 Google 在研讨会上发布的名为《Advanced and Open Problem in Federated Learning》的白皮书，对当前分布式学习还是激励机制等等都做了总结，希望大家去关注。

另外一个方向是联邦学习和迁移学习的有效结合，迁移学习的主要思想是：我们的模型在一个领域已经非常成熟，比如 Source Domain 领域，此领域数据丰富且经过训练后模型性能较高。现在有个新的领域叫做 Target Domain，此领域数据有限，但是和 Source Domain 领域一样都是有关图像。而迁移学习指的是，从 Source Domain 领域学到的模型知识转移到 Target Domain 领域。形象点比喻就是举一反三。

假设两个领域的数据不能交换，两个领域的参数也要互相保密，那么还可以进行迁移学习吗？进一步，如果涉及到两个机构（有不同的特征，但是有类似的样本），他们有意愿合作，但是数据的格式不一样，一个领域的数据是图像，另一个领域的数据是文字，换句话说异构的协作如何用联邦学习实现？

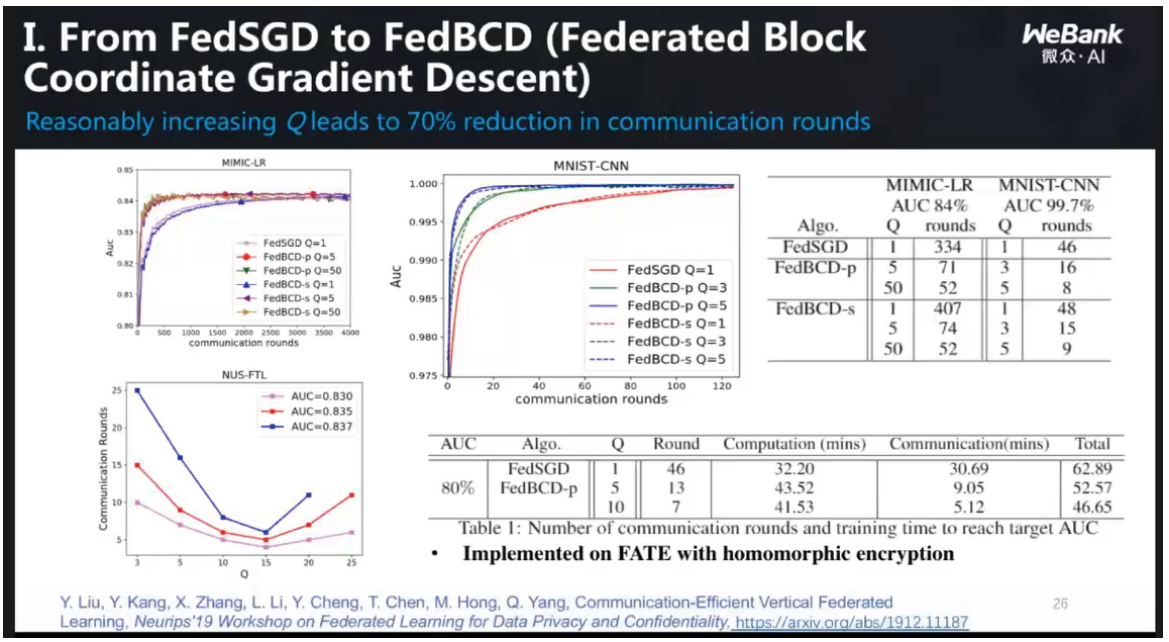


图 10：从 FedSGD 到 FedBCD

其实，这些研究问题可以囊括在联邦学习和迁移学习的有机结合当中，叫做联邦迁移学习。下面看联邦迁移学习的第一个进展：可以用迁移学习的思想协助联合学习，但是因为需要兼顾两者，所以其速度将会大大减慢。

我们提高两个机构沟通效率的办法是：尽量减少沟通次数，让一次沟通发挥最大的作用，尽量在本地进行多次运转，然后进行机构间的沟通，而且尽量能够在设计机器学习算法的时候，就让两个机构之间的沟通并行化。所以这样做的效果是：不仅是效率和速度的提高，而且运行的成本大大降低，运行的速度大大提高。

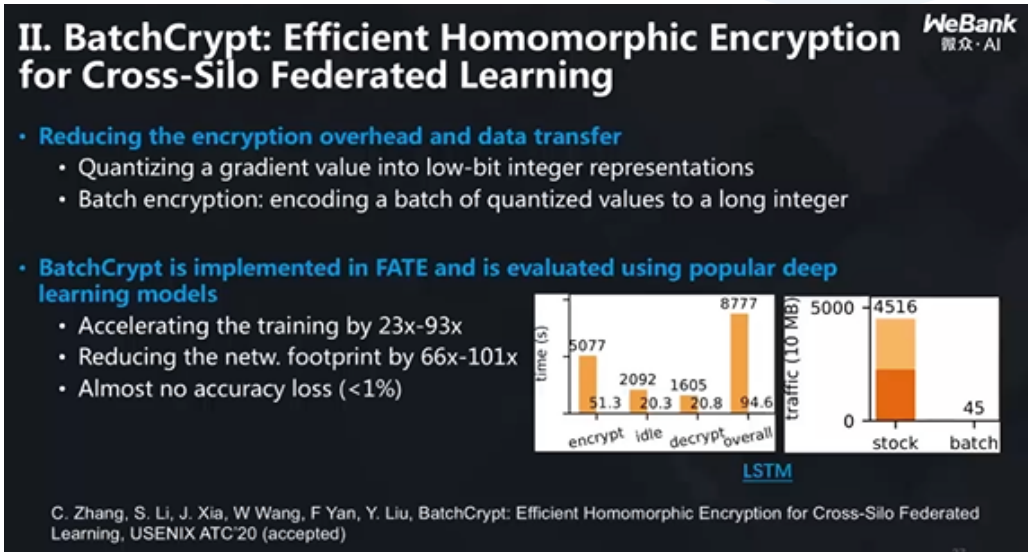


图 11：BatchCrypt

联邦迁移学习的第二个进展是：引入一些比较精密、比较高端的加密算法，例如可以把梯度的值变成矢量运算、向量运算，如此加密的时候效率就会大大提高。具体而言，效率能够提高从 20 倍到百倍不等；另外，也可以用一些新的加密手段，比如密钥分享 (secret sharing)，通过这种有机的结合能够让加密算法的速度大幅度提升。

如果联邦学习这个领域想要持久发展，离不开公开数据集的支持。最近，一些联邦学习公开的数据集的出现，例如 Federated AI Datase 网站（https://dataset.fedai.org），其集成了计算机视觉方面的联邦学习数据，此数据的特点是分布式。

图 12：IEEE 标准 P3652.1——联邦机器学习

另外，很快也会出台第一个 IEEE 联邦学习的国际标准；不久，我国也会有相应的国家标准和团体标准出现。

前面提到的开源算法也被纳入 Linux Foundation 开源平台，并且有多个工业级别联邦学习的商业应用现在也已

经出现。例如 FATE，作为一个工业级别联邦学习系统（https://FedAI.org），能够有效帮助多个机构在符合数据安全和政府法则前提下，进行数据使用和联合建模。

三、AI 要用来保护用户的隐私

刚才讲的大部分内容是：AI 要保护用户隐私和利益。现在介绍第二个定律，AI 要保护模型安全，那么模型什么时候会变得不安全？人工智能机器学习的整个流程可以被分解成以下几步： 第一步是获取很多训练数据； 第二步是通过训练数据训练一些算法，从而形成模型； 第三步是把模型应用在实际当中，使得测试数据的时候能够得到有效的结果。

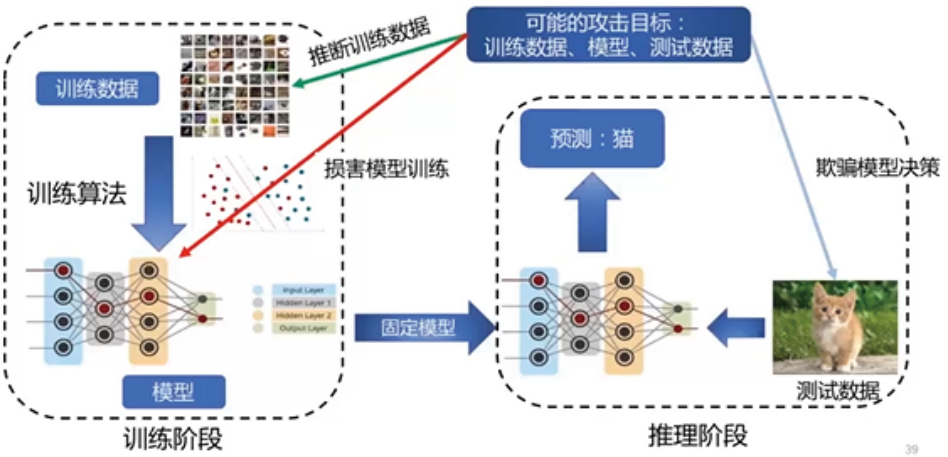


图 13：机器学习流程中的可攻击点

上面流程就存在一些薄弱环节，例如可能被攻击的就是训练数据本身，叫做数据中毒； 也可能对模型进行攻击，也就是说模型的隐私可能会被泄露； 测试数据可能会是假的，但模型本身可能是没有办法识别，这也相当于对模型的攻击。

那么如何应对这种攻击，如何保证人工智能模型的安全呢？

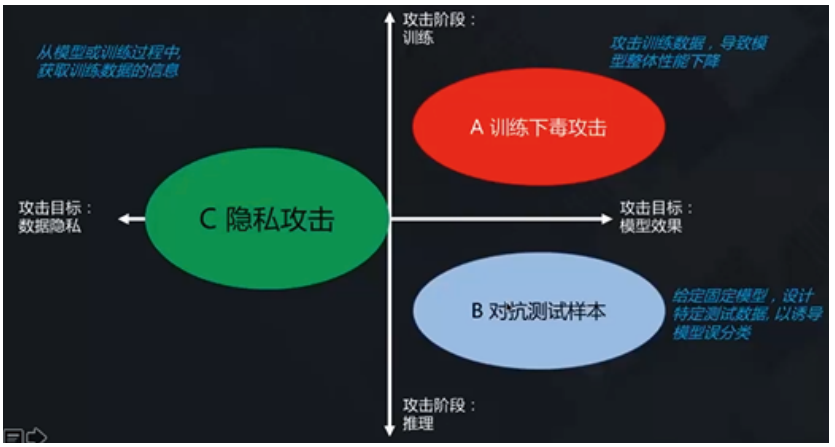


图 14：对机器学习模型的攻击

我们来看“攻击”的三种情况： 训练下毒攻击： 毒化数据，相当于对模型的目标进行攻击，或者对目标进行攻击影响模型的效果； 对抗测试样本，即在推理的过程当中进行攻击； 隐私攻击，即攻击的过程当中要了解数据当中包含的用户隐私。

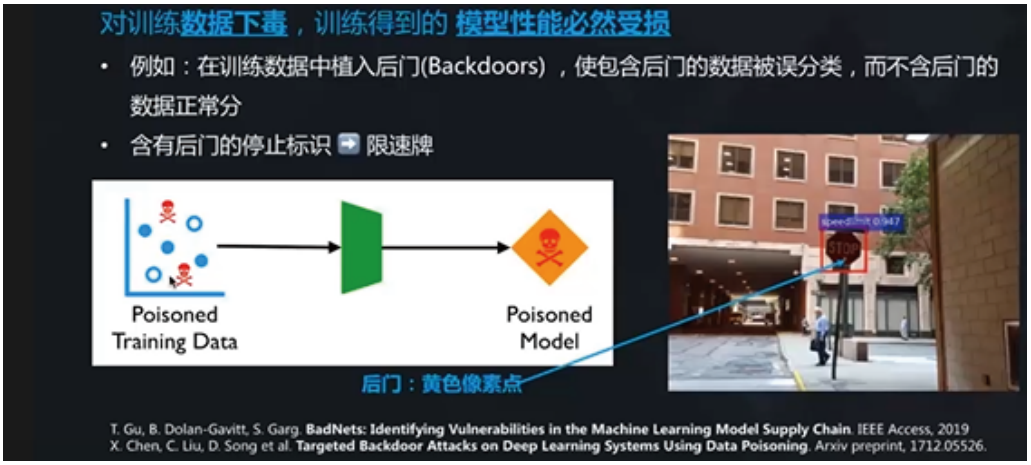


图 15：训练下毒攻击——毒化数据

其中，对训练数据下毒，训练得到的模型性能必然受损。常见的手段是在训练数据中植入后门，使包含后门的数据被误分类，而不含后门的数据正常分类。

例如，上图中的骷颅就相当于在“带毒”的数据，而在模型中表现为“黄色”，也就是说，当模型遇到黄色标记就会失灵。在现实中，Stop Sign 中有黄色的像素点，相当于对模型加入了后门，以至于无法识别是否停车，从而存在对行人产生伤害的危险（恶意攻击）。

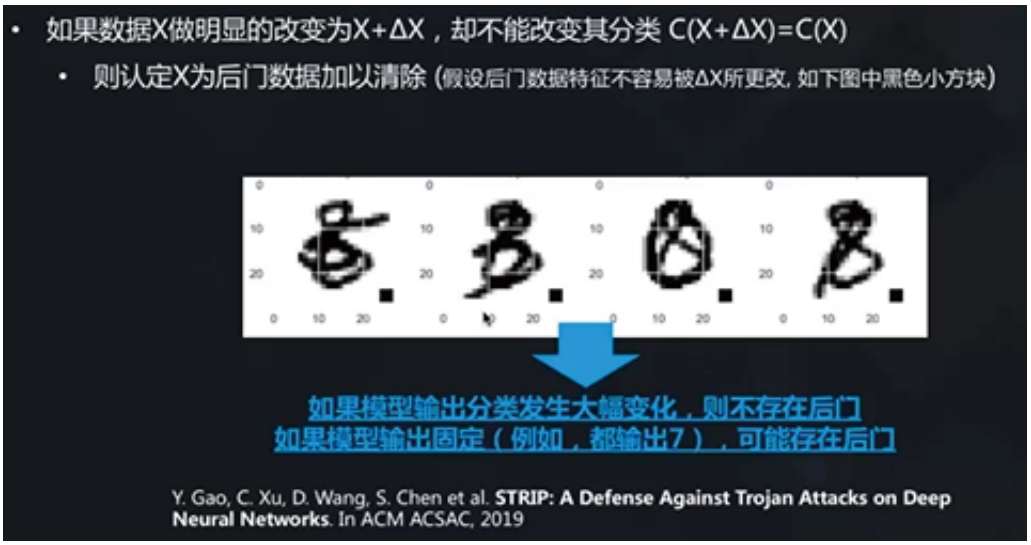


图 16：训练下毒攻击——如何检测并清理被植入后门的数据?

如何防止这个现象发生？一种解决方案是： 可以在数据上面加一些扰动，使得原来的数据 X 现在变成了 X+ Δ X，扰动小的时候模型分类的结果会不会发生突变，如果发生突变就认为被下毒了。

以手写字体检测为例，加上一个像素的扰动，如果模型输出分类发生大变化，则不存在后门； 如果模型输出固定（例如，都输出 7），则可能存在后门。

所以手写体是各种 8 的写法，要是加上一个像素一下子变成 7 了，我们就认为加了一个 DeltaX 以后分类效果就变了，因此我们就认定这个样本可能是带毒的样本。当然，我们也还有其他的做法，总之这个领域是非常活跃的。

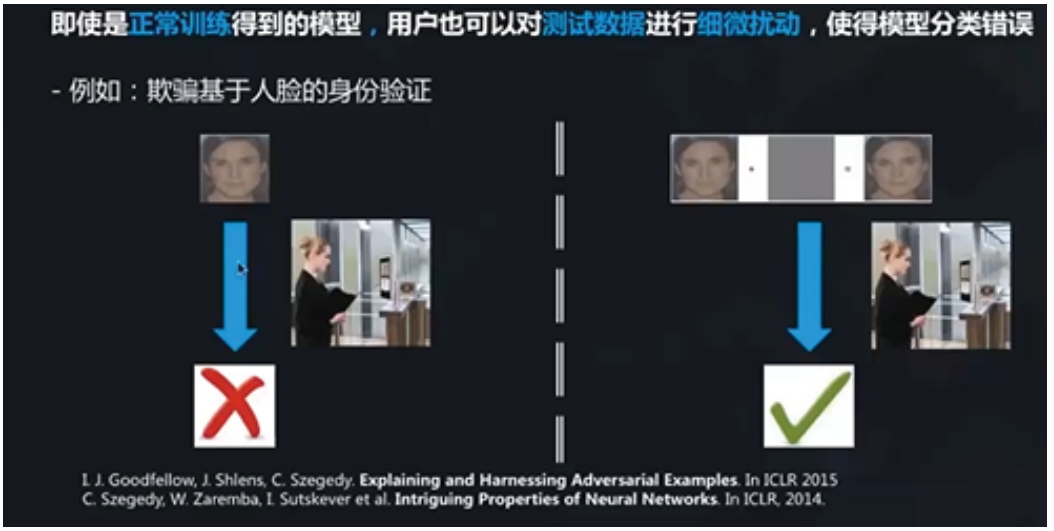


图 17：对抗测试样本

第二种方式是对抗测试样本，即如果对模型的机制有所了解，那么就可以设计一种测试样本，使得蒙混过关，相当于对模型的攻击。

具体例子如上图所示： 一张人脸本来不能通过人脸识别的闸机口，但如果我们在测试数据当中加了一些噪音，这样就使得此人脸能够通过检测。总的来说，这种细微的测试数据的扰动，欺骗了人脸识别的身份验证系统。

如何解决呢？一种解决方案是用对抗测试样本，即对原始数据进行一些扰动，证明模型是否具有鲁棒性。

图 18：隐私攻击

第三种方式是对隐私的攻击，攻击者的目的往往是盗取私有信息。这种攻击的表现是： 不同的机构参与的联邦学习分布式的架构下有一个“坏人”，他希望通过模型的参数反推出原始的数据。

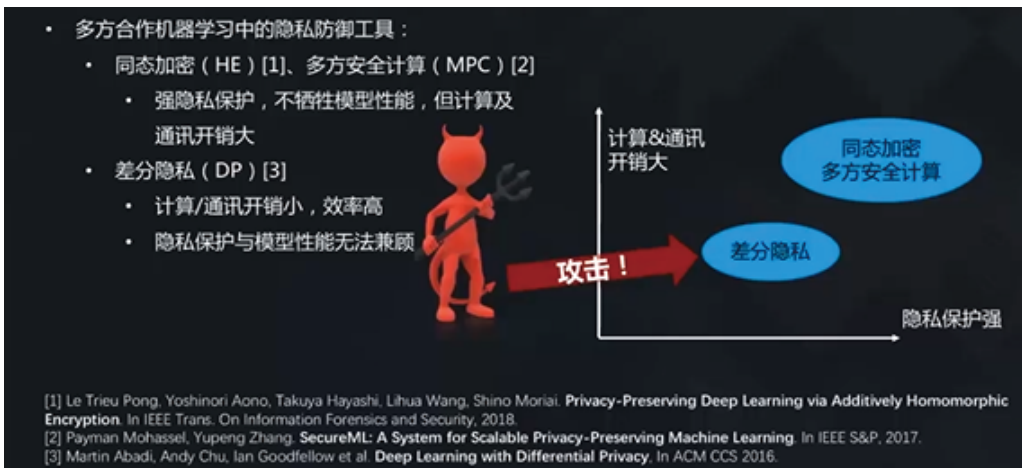


图 19：隐私攻击中的防御攻击

联邦学习中用来对抗隐私攻击的标配是同态加密，当然也可以用多方安全计算加强隐私的保护。虽然，在不牺牲模型性能的情况下能够保护隐私，但是世界上没有免费的午餐，因为往往需要大量的计算才能运行这些加密算法。

所以，大多数加密算法的计算开销非常大。因此在实际当中，有些应用就选用了差分隐私替代同态加密。但是由于隐私保护和模型性能无法兼顾，差分隐私算法在工业界并没有广泛应用。

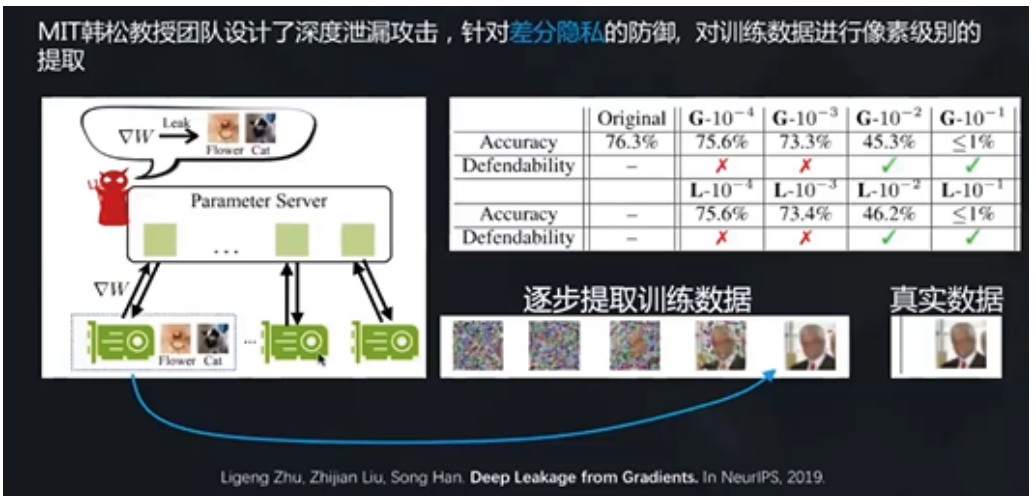


图 20：隐私攻击例子，深度泄漏攻击

但是，在某些状态下，使用差分隐私可以重构原始的训练数据。MIT 韩松教授证明了这一点，他的团队设计了深度泄露攻击，针对差分隐私的防御，对训练数据进行像素级别的提取。而深度泄露攻击意思是：当我们在多方沟通模型的梯度的时候，即使此梯度当中部分是加密的，部分加噪音的（例如差分隐私），但对方还是可以不同程度地学到原始数据。

另外，韩松教授通过实验得出的结论是：噪音加得多对方学到的少，同时效果也会变差。

图 21：深度泄露攻击的防御

最近，微众银行范力欣的团队在理论上证明了：即使在差分隐私的情况下，在不影响模型效果同时，完全防御深度泄露攻击。

他将隐私泄露分成了三个状态：隐私完全泄露，即完全不加密也不加噪；完全保证隐私，即完全加噪，这时“坏人”就没有办法学到信息，但“好人”的模型也会受到影响。第三个状态是处于两者之间的最佳点：坏人很难学到信息，好人会得到最好的结果。所以这是一个让人很振奋的工作。

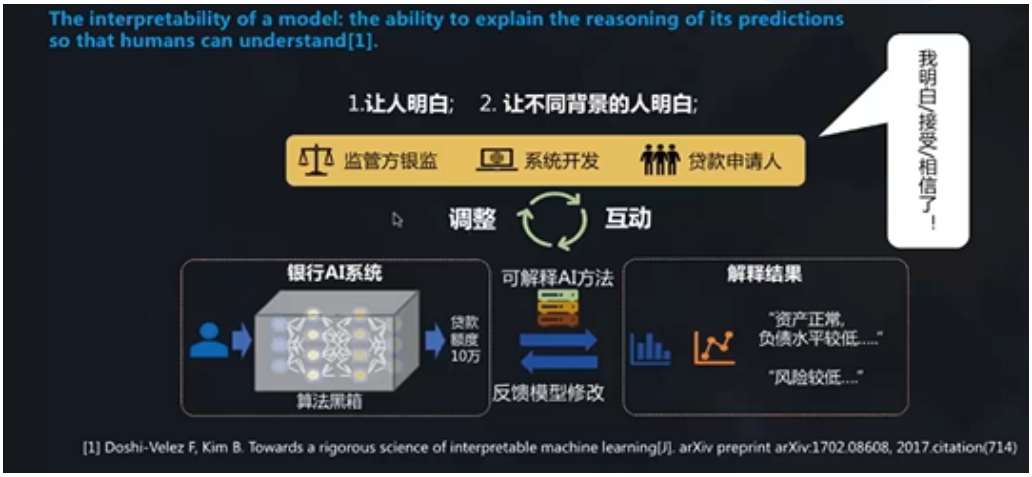


图 21：可解释 AI

四、可解释 AI

第三个定律是 AI 要对人类解释自己。我们把“解释”分成两部分：第一，让人明白 AI 在做什么；第二，让不同背景的人用不同的方式明白 AI 在做什么。例如银行的风险评估人工智能算法：在面对监管方和银监会时，要解释出整个结果产出的逻辑；在面对系统开发的工程师时，要使得他们能够了解系统从而进行 Debug；对贷款的申请人要能够解释贷款的结果，例如“因为资产正常，所以您的风险较低”等等。因此，这里的可解释性是指：对于不同的人，有不同的解释。

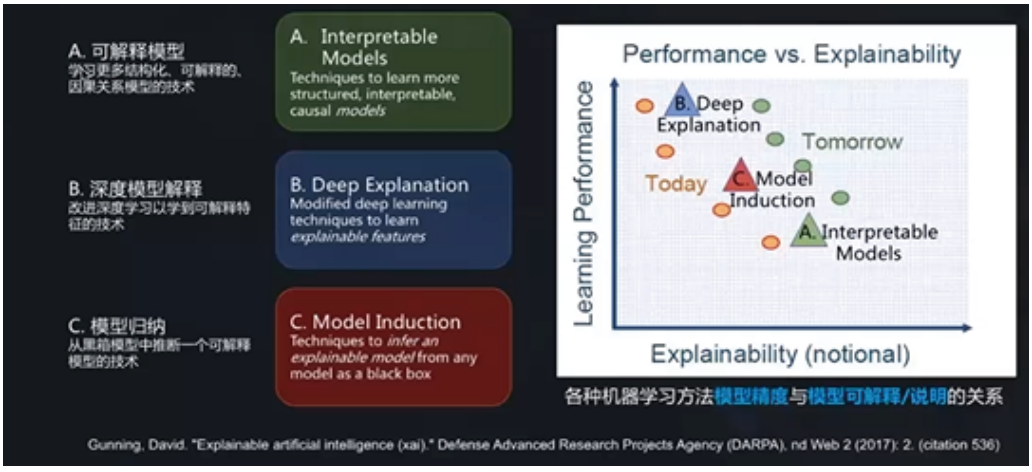


图 22：可解释 AI 主要方法与关系

最近，在人工智能界可解释性也被大量提出，DARPA 研讨会上提出了三种可解释的定义：1、模型本身要可解释，即能够学习更多结构化、可解释的、因果关系模型的技术，英文术语叫做 Interpretable Model。2、深度模型解释，改进深度学习可以学到可解释特征的技术，例如，要能够让我们知道，模型输出结果是“黑箱”的哪一部分推断出来的。此定义的英文名称叫做：Deep Exploration；3、模型归纳，从黑箱模型中推断出一个可解释模型的技术，例如，我们有一个很复杂的黑箱模型，是不是可以用一个比较简单的模型基本上“覆盖”复杂的黑箱子模型？那么这个简单的模型就是可解释。

其实，关于 AI 学习的有效性、效果和可解释性存在着不平衡，当然最终的希望是得到既可以高度可解释，又有高度的性能。

那么，上面三个解释的定义进展如何?

模型	描述	优点	缺点
线性回归	预测结果为所有特征与其权重乘积之和	• 结果是加权和，易于理解 • 保障可以找到最优权重	• 需人工干预非线性问题 • 预测方面性能不佳
逻辑回归	模型为二分类问题产生两个概率输出	• 可以给出概率结果 • 可以扩展成为多分类器	• 模型表现能力有限 • 以乘法形式对权重进行解释
广义线性模型	线性模型的推广，用来求解非线性问题	• 模型被广泛使用 • 可以转变为更灵活的模型	• 可解释性稍差
决策树	多次将数据特征依据某些规则进行分割	• 节点划分清晰易懂 • 树模型节点间的关系直观	• 不能处理线性关系 • 平滑性较差、不稳定

图 23：可解释模型

首先看一下可解释模型，如上图所示，罗列了一些机器学习模型，如线性回归、逻辑回归等等，这些模型的优点和缺点上图都有总结。其实，当前没有一个模型现在能够达到既高效率，又高度可解释。所以，这个方向有待于大量的研究。

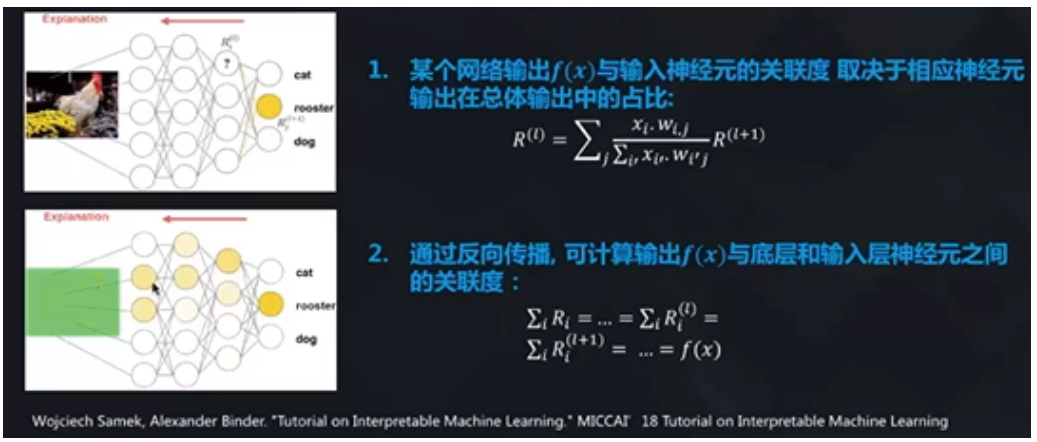


图 24：层级关联度（反向）传播

第二是，深度模型解释。如上图推理模型所示，图像当中有一个攻击，深度模型解释指的是： 哪些像素可以解释这种攻击。当然，人类是可以很容易“解释”，但是黑箱子模型当中如何能够找出对应的高度适配的像素？这个技术可以用反向传播进行，此技术叫做 LPR。

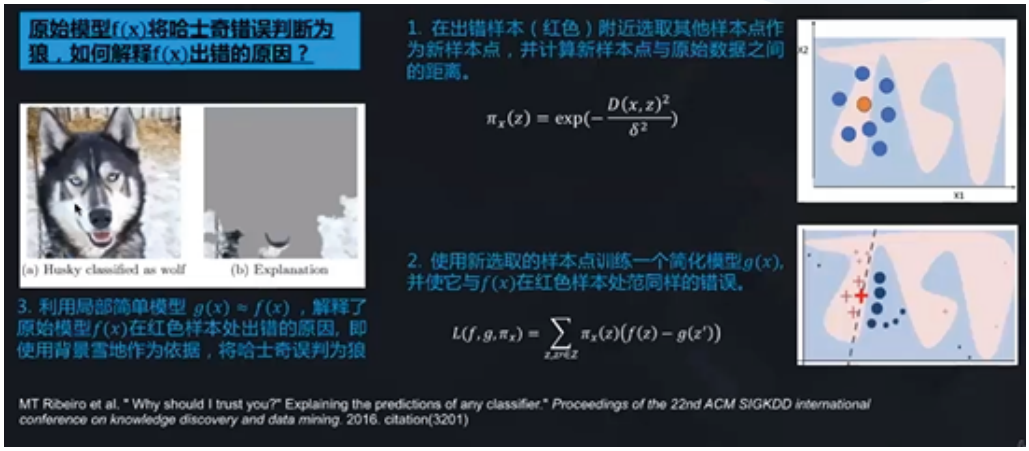


图 25：模型无关的局部解释

模型归纳。例如用哈士奇进行分类，系统会错误地把它分成了一只狼，虽然人类知道它不是狼，但是我们想归因，想知道为什么这个系统做错了。那么，我们就可以把错误的例作为一个输入，通过对其周边像素和特征的解释。得出结论：因为雪地的存在，导致了模型的误判。

最后，现在人工智能技术的研究日新月异，AI 的可解释标准建设也是刚刚开始。人工智能的发展不仅要芯片、数据和算法，同时也要注意人，研究者要保护人的隐私，保护模型的安全，保证对人类可以解释。

图灵奖获得者 Alan Kay：突破常规思维，创建下一代科研社区

整理：智源社区 熊宇轩 常政

北京智源大会，以「内行的人工智能大会」为宗旨，科学巨匠、AI 领袖云集，给观众们留下了深刻的印象。

开幕式上，“贝叶斯网络之父”、图灵奖获得者 Judea Pearl 总结其毕生之所学与所思，发表了名为「The New Science of Cause and Effect with reflections on data science and artificial intelligence」的精彩演讲，启迪人们从数据革命走向因果革命，让人工智能系统由果溯因、学会思考有关「WHY」的问题。

而在本届大会的最后一场主旨演讲中，SmallTalk 之父、图灵奖获得者 Alan Kay 基于其近期发表的力作「HOW ?」从宏观、未来的角度，介绍了人类社会面临的 12 个重大挑战，并指出解决它们的关键来自于打破常规的思维模式。

美国科研和信息技术腾飞的内因是什么？如何孕育科学技术发展的土壤？如何培养下一代的卓越科研人员、机构和社区？针对以上问题，Alan Kay 在本次演讲中都给出了自己的答案。在演讲后的圆桌讨论中，Alan Kay 和北京智源人工智能研究院理事长张宏江、北京智源人工智能研究院院长黄铁军，就人工智能发展中的趋势性话题等进行了深入交流，Alan Kay 认为目前我们仅仅触及了人工智能的冰山一角，并建议年轻一代学者应该摈弃“唯论文”、“唯学位”等功利心态，回归本心去迎接真正的挑战。

从「WHY」到「HOW」，两位大师站在不同的角度给出了自己对于人工智能未来发展方向的思考，分别展现了他们严谨理性和浪漫宏大的科学观和世界观。

下面，我们对 Alan Kay 本次的主旨演讲以及圆桌讨论进行了整理，希望可以启迪并激励人工智能相关行业的从业者在下一个十年中勇立潮头，把握时机，在这个最具有活力的科学研究领域中做出真正具有革命性的伟大成果。同时为了帮助大家深入理解演讲的内容，推荐大家参阅 Alan Kay 于 2019 年在英国的麦克阿瑟基金会峰会发布的题为「How ?」的白皮书（https://internetat50.com/references/Kay_How.pdf），本次演讲的部分内容、思想正是节选自这篇白皮书。

一、演讲全文

1.1 问题引入

首先，Alan Kay 介绍了他讨论问题的纲要，主要包含如下 5 个方面:

- 我们可以从人类解决各种「重大问题」的历史中学到什么?
- 当前科研和工程领域的生态存在什么问题?
- 对于政府来说，应该如何使用科研经费，如何构建新一代的研发机构?
- 如何建立学者的评价机制?
- 下一代「研究社区」。

1.2 理解世界的思维框架：7 个语境 +12 个重大挑战

步入讨论正题后，Alan Kay 介绍了他问题视野的起点——文化视野，认为它能将所有的人类问题囊括其中。目前全世界已经存在着数以千计的不同文化，而我们中的每个人都生活在自己的文化系统中。Alan Kay 认为我们都应该对社会承担责任，比如除了日常做的事情之外，还需分配出时间来思考一些关于学校教育、关于如何塑造下一代孩子价值观的问题，毕竟人类很多分歧即使是善意的分歧，都是由不同的世界观造成的。Alan Kay 于是提出了「丰富性」概念，它指的是扩展我们与事物产生情感上的联系的能力。当然了，这里也存在一个最基本的问题： 我们如何谋生，以满足我们日复一日的基本需求。综上所诉，Alan Kay 介绍了他思考问题的七个语境:「人文」、「社会」、「下一代」、「世界观」、「学校教育」、「丰富性」、「谋生」 (如图 1 所示)，以及彼此之间的逻辑。

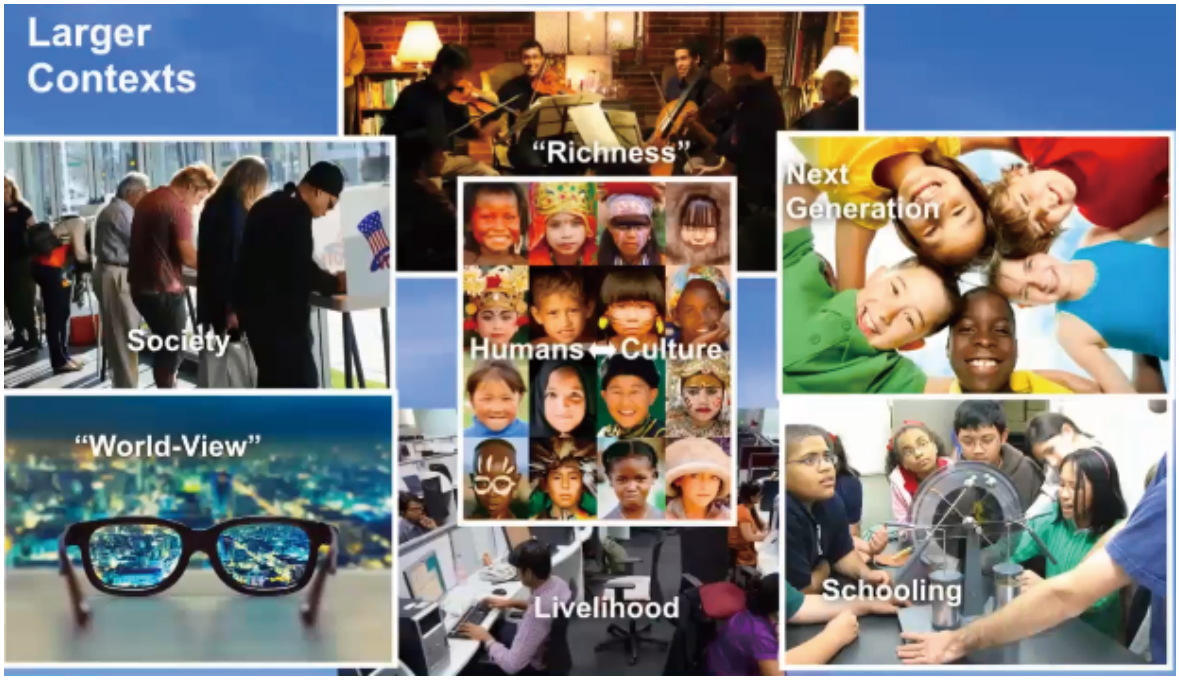


图 1: 思考世界的更广阔的语境。图中的条目分别为「人文」、「社会」、「下一代」、「世界观」、「学校教育」、「丰富性」、「谋生」

接下来，Alan Kay 进一步归纳了人类社会亟待解决的 12 个重大问题： 健康、食物、气候、水资源、居住环境、能源、教育、生态、污染、合作、人口、权利，并将之前梳理的 7 个语境概念嵌入每个问题之中，便形成了如下图 2 所示的思想框架。

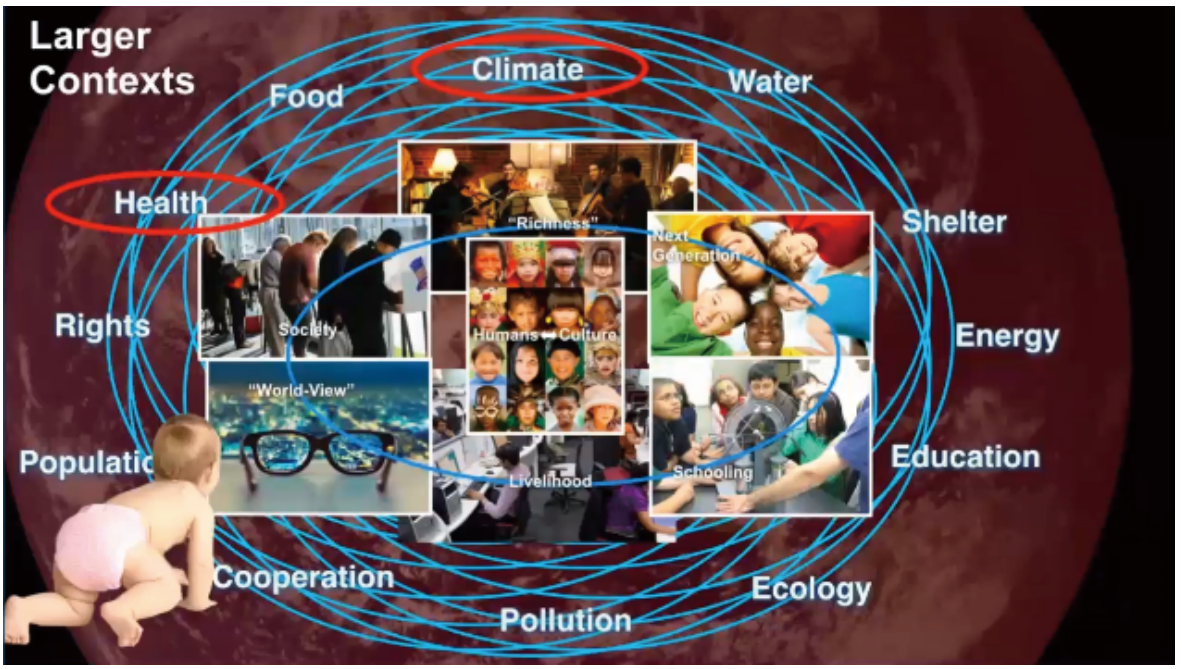


图 2：当今世界面临的重大挑战。图中列举出了 12 种重大挑战——健康、食物、气候、水资源、居住环境、能源、教育、生态、污染、合作、人口、权利

Alan Kay 指出，我们需要认真应对这些问题，但实际上很难只专心处理其中的某一个问题（而不同时考虑其它的问题）。例如，「健康」问题是十分重要的，但它也会被地球上的其它问题所干扰，比如我们居住的星球自己也正在以各种方式「走向死亡」，比如健康问题背后还隐藏着对人类有着更大影响的问题——气候问题。

1.3 用“新思维”解决“旧思维”制造的问题

那么，我们该如何面对和解决这些重大问题的挑战呢？ Alan Kay 认为关键在于思维模式的变革，他在演讲中引用了爱因斯坦的一句名言：我们不能用制造问题时所用的相同思维来解决问题。

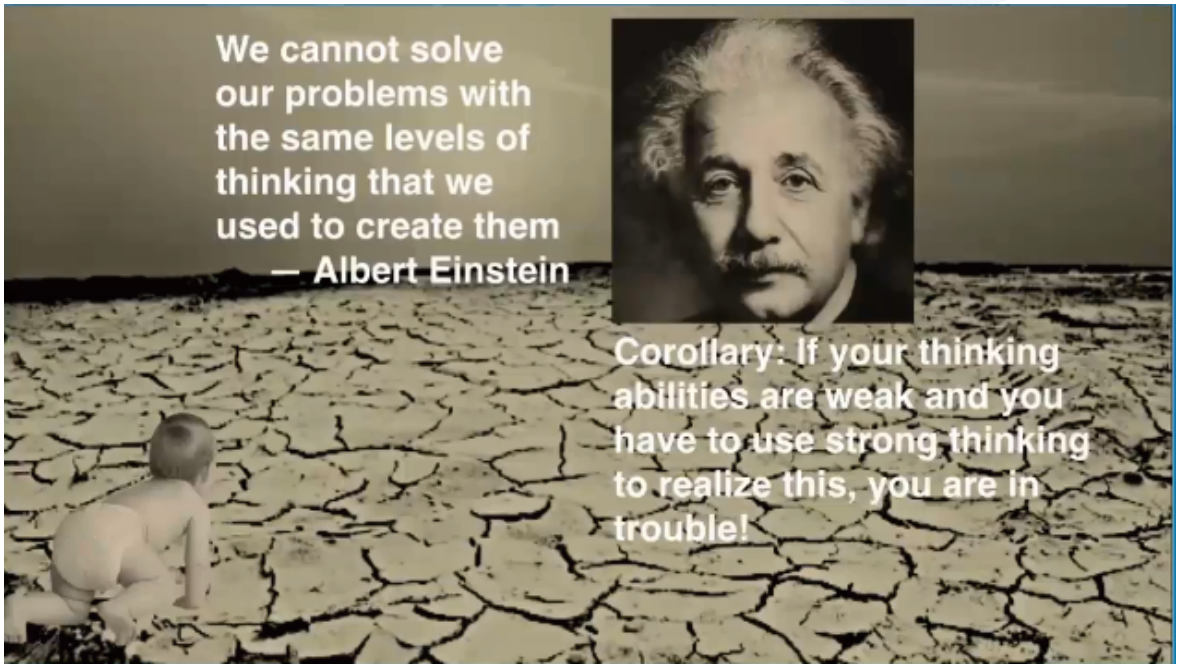


图 3：我们无法用创造问题时所用的相同思维来解决问题

面对人类社会如今的局面，我们的思维往往也是常规的。既然「常规思维」解决这些问题行不通的，我们便需要学着如何以一种更好的方式学会「疯狂地思考」。Alay Kay 以这个问题“假设一些孩子在今年出生，到本世纪末年至八旬的成长过程中，地球会变得更好还是更差呢?”为例，解释了该如何打破常规思维。

图 4：记忆中童年的绿色世界

首先，在我们很多人童年的记忆中，我们的绿色星球是这样的（如图 4 的视角 1）。

图 5：透过眼镜看到过去的绿色世界

但现在，我们认为地球已经没有以前那样的「绿色」了（如图 5 的视角 2），但如果我们戴上眼镜，我们假设仍然可以看到过去的绿色景象。

图 6：三种看待世界的视角——过去的绿色世界，当今的灰色世界，未来可能被破坏的世界

而当我们想象未来的情况时，眼镜中的景象又有可能成为右边眼镜中的情景（如图 6 的视角 3，由于疫情或者我们之前提到过的气候等问题所造成）。Alan Kay 认为，我们在思考“地球是否变好时”，需要发挥想象力，能够同时处理这三种视角。

图 7：由于威胁而产生的「非常规」思维

Alan Kay 指出，尽管科学是我们想象力的放大器，但是大多数人并非科学家（尤其是欧美国家选举出的政客们），并不喜欢跳出他们的「常规思维」。从另一个角度来看，多年以来经常会发生的一种现象是：每当有某种「威胁」，比如战争、流行疾病等被公众（包括政客在内）所承认，往往会招致很强烈的反应，会催化出一些「非常规」思维。

到这里，我们不妨补充参考一下 Alan Kay 在白皮书「How ?」关于新思维模式的描述，比较有代表性的是“全系统思维”（Whole Systems Thinking），“以最大的尺度和最复杂的方式看待大多数事物”。

接下来，Alan Kay 列举了历史上一系列基于“威胁”而诞生的重大项目，例如因为“大萧条”而产生了「帝国大厦」。

他认为建造帝国大厦是历史上最杰出的设计和规划工作。帝国大厦，从拆除之前的场所到建造起整个新的建筑只用了不到一年的时间，而完成这项工程的建造者也不超过 3,000 名。这项工程之所以能以这样的方式被完成，是由于「大萧条」严重打击了美国，而此时正是帝国大厦需要资金的时候。所以建造帝国大厦的动机，一方面是（由于资金紧张）需要尽快完成这项工程；另一方面，需要向人们展示建造这样一座高耸入云的摩天大楼的真实过程是怎样的——通过工程的力量提振人们的信心。

接下来，Alan Kay 介绍了「第二次世界大战」期间的项目，包括「原子能计划」、「布莱切利公园密码破译」、

美国和英国共同参与的「雷达」项目等。而到了和「冷战」期间，美国有大量的资金继续提供给了此类项目：1957 年，其中一部分资金用于资助美国国防部高级研究计划局（ARPA），而在 1962 年 ARPA 成立了信息处理技术办公室（IPTO），这是一个深入研究信息系统的、非结构化（Non-Structured）、与众不同的研究部门，采用了自下而上的研究组织方式。

结合白皮书「How ?」，Alan Kay 关注这些项目的主要原因在于：**这些项目规模庞大，动用了大量人力物力财力。它们都解决了以前认为不切实际或异想天开的问题，完成速度之快令人瞠目结舌。这些项目涉及各行各业的顶尖人才，他们可以自由选择发现和解决问题的方式和方法，并且随时获得相应的资金支持。**

概而言之，它们都是基于“非常规”思维的工作模式下取得了杰出成果。接下来，Alan Kay 以 1971 年成立的施乐帕克（Xerox PARC）为例，从多个角度阐释了构建优秀科研团队、科研社区以及相关评价机制的全新思维方法，这些思想之光、经验法则对于我们探索当前的 AI 科学研究，非常值得借鉴。

1.4 施乐帕克的科研奇迹

1.4.1 研究结果的好坏与资助者密切相关

施乐帕克是施乐公司资助的一个研究小组，如下图 10 中右侧的「冷战」部分所示，它脱胎自之前 ARPA 的雷达研究小组，该小组从之前的项目中学到了如何大规模地将科学与工程相结合，从而将科学研究与未知事物相结合。

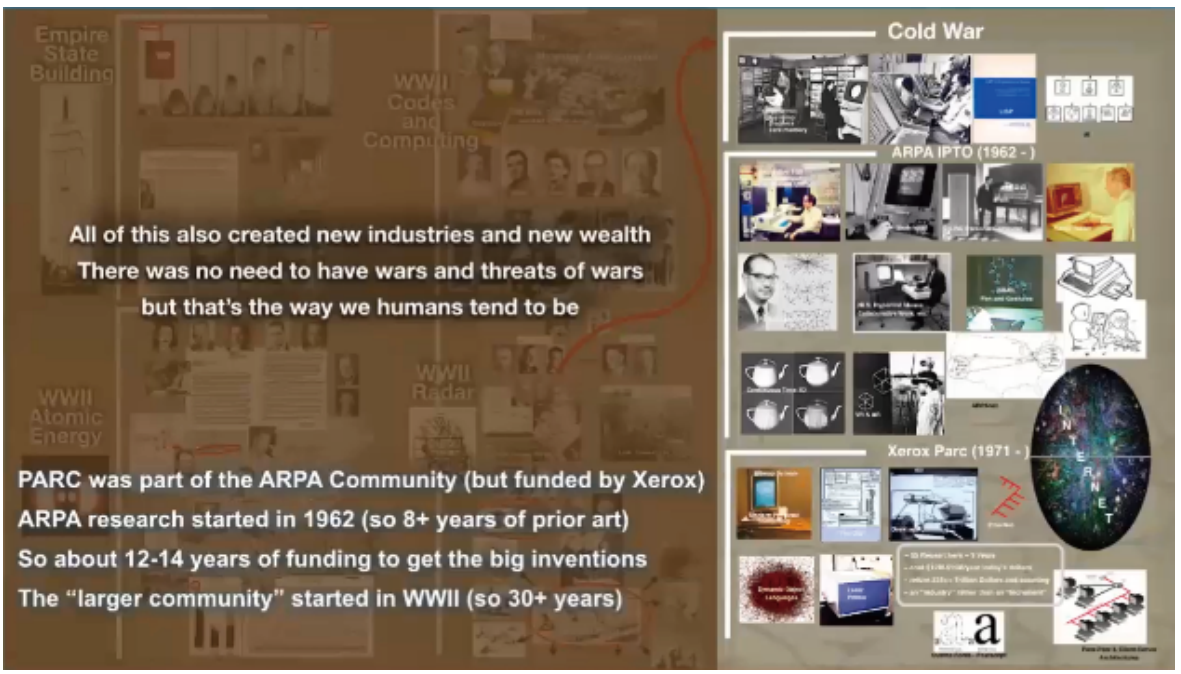


图 8：向 ARPA 社区投资的巨大回报

如图 8 右下方所示，Alan Kay 列举出了 PARC 的「8.5」个重要发明：Modern Personal Computer（个人电脑）、The GUI（图形用户界面）、WYSIWYG & DTP（所见即所得文本编器）、Real OOP（面向对象）、Laser

Printer（激光打印机）、Outline Fonts & Postscript（大纲字体和页面描述语言）、Ethernet（以太网）、Peer-Peer & Client-Server（点对点 & 客户机服务器），还有 Internet（因特网）因为是合作项目，便算 0.5 个。

令人惊讶的是，所有这些成果只不过是 25 个左右的研究人员在大约 5 年的时间内完成的。同时，完成这些研究所花费的经费并不高昂，按照现在的标准来计算，它们不过花费了 1,200 万 –1500 万美元。但是这些研究成果带来的巨大收益已经超过了 40 兆美元。这些成果创造了一系列崭新的「工业种类」，而不是仅仅对现有行业做了「加法」。由于这些发明，诞生了许多股市估值非常高的美国公司。

Alan Kay 认为 PARC 的成功，充分说明了「结果的好坏与资助者的好坏密切相关」，这里的「好」（Goodness）并不是指钱，而是指资助者真正明白本次演讲中所讨论的「如何达成我们从未见过的目标」有何意义。

1.4.2 「愿景」而非「目标」

Alan Kay 认为 施乐帕克的研究理念深受 Licklider 的影响。

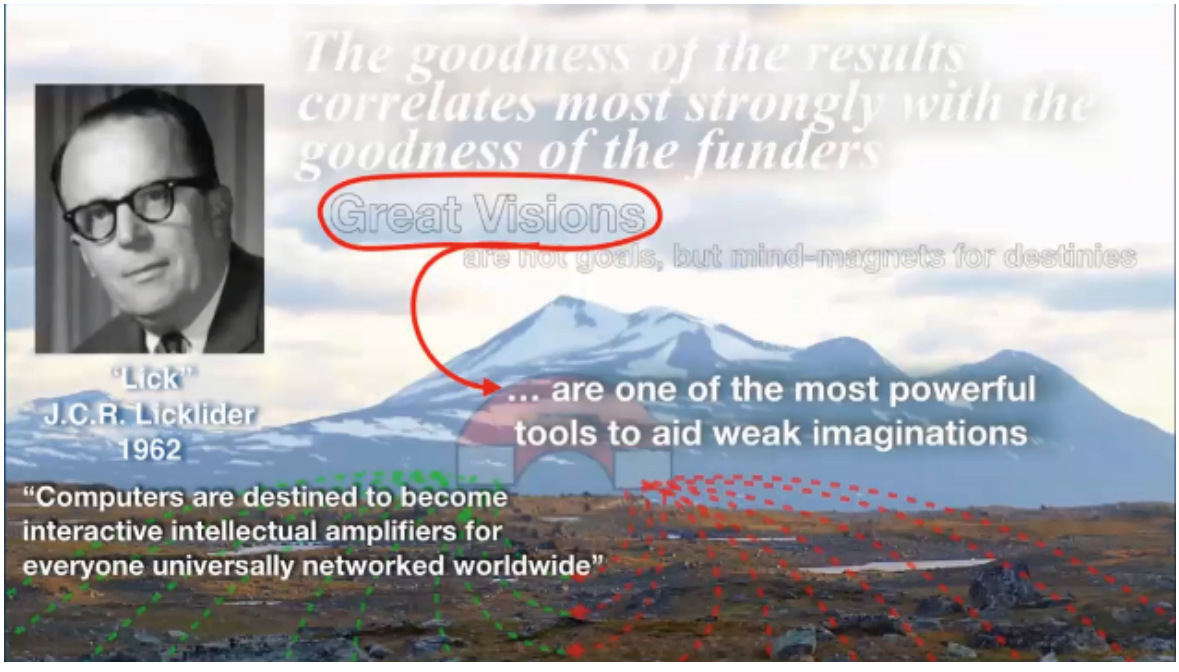


图 9：Licklider 的愿景为 PARC 带来自由开放的学术风气

Licklider 于 1962 年创立了 ARPA IPTO，他十分有远见，但对于心中的愿景，不会设定具体的目标。

Alan Kay 指出正是 Licklider 的愿景才造就了我们今天的生活。Licklider 认为「计算机注定要成为遍及全球的所有人类的交互式智能放大器」。但这就是他所说的全部「愿景」，如果你问他将如何做到这一点，他会说「我也不知道，但是我只需要将钱投给那些可以帮助我们实现这一愿景的人」。如果这其中 30% 或 40% 的资助结果是成功的，那么我们就将彻底改变世界。而事实正是如此！

所以，这些「愿景」就好比高山背后所隐藏的磁场，研究人员可以感知到这些磁场，所有的研究人员都会朝着

磁场驱使的方向运动。而你实现这种「愿景」的方式可能有很多种。所以当你陷入困境时，你需要明白大多数对问题的描述都是在当下的环境下产生的，但是大多数情况下，你需要创造新的环境才能找到解决方案。所以，你需要从具体的目标和问题退回到宏达的愿景，愿景会还以事情本来的面貌，并且让你能够有更广阔的思维。

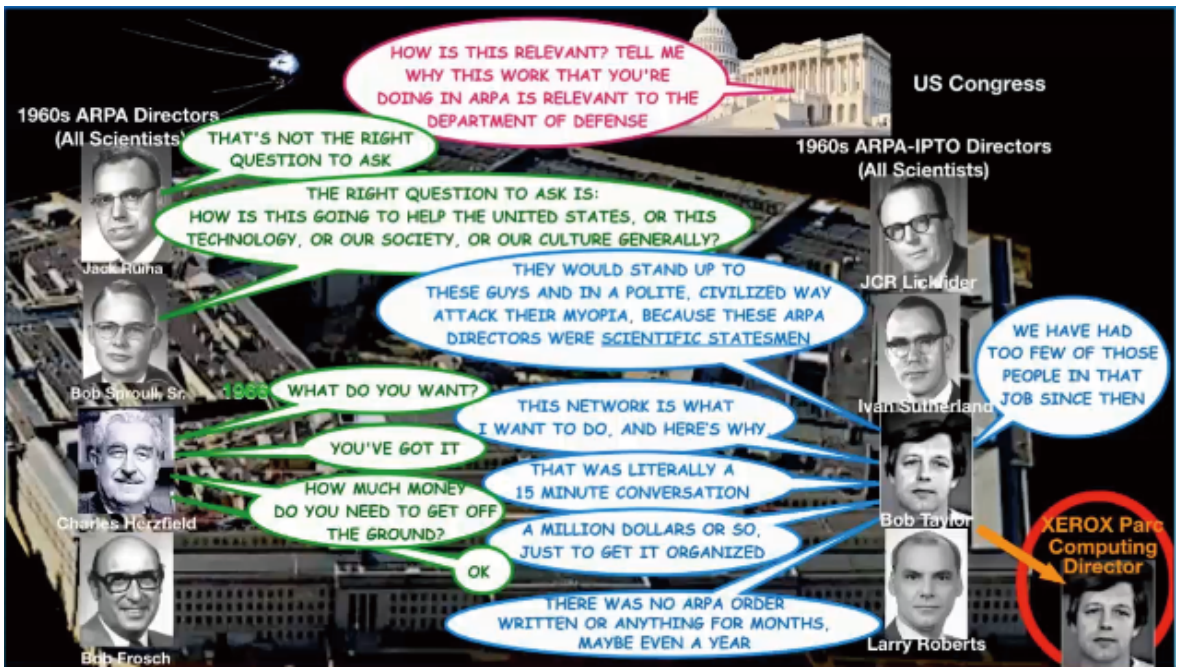


图 10：ARPA 的科学家管理文化，以及对「愿景」的坚守

那么，当时这些愿景发挥了怎样的作用呢？如上图所示，这是美国国防部的五角大楼和国会（每个国家都有相类似的机构）。1957 年，苏联发射了人造卫星，美国感到国家安全受到了威胁，这种恐惧促使美国在 同一年设立了 ARPA。当时所有的 ARPA 研究主管都是科学家（尤其是物理学家），而不是官员。随后，ARPA 设立了 IPTO，它仍然由科学家所领导，领导人员每 2 年更换一次。他们认为，如果领导者在位太长时间，那这份事业就会变成他的「工作」，而他真正的工作应该是做一名科学家。在这里，人们要做的只是为国家所服务，在为国家服务两年之后，你就可以离开了。否则你就可能会开始买房并且因为按揭而分心，你会开始担心会不会丢掉自己的「工作」（不能纯粹地为愿景而努力）。

这里还面临的一个关键问题是：ARPA 认为实现自己的「愿景」需要花费 10、15 年，甚至 20 年的时间，这势必需要花费大量经费来培养下一代研究人员，让他们继承这一愿景，这难免会面临来自国会的压力，但 ARPA 最终贯彻了这一路线。Alan Kay 回忆，施乐 PARC 实验室中几乎所有的计算机研究人员都来自于 ARPA 项目，当时大家都很年轻，Alan Kay 作为组里年纪最大的也仅仅 30 岁。

1.4.3 决策不应只是自上而下

PARC 采用的是自下而上项目管理模式，这同样追溯到 ARPA。Alen Kay 介绍，美国国会曾经想果对 ARPA 扮演一个监督机构的角色，**自上而下对项目进行管理**。他们会让你说明清楚为什么你在 ARPA 所从事的这些工作与国防部有关。但与现在不同，当年 ARPA 的主管们会直接告诉国会「不，这并不是你们该问的问题」，因为他们并不在乎被解雇。他们认为，真正该问的问题是「这些工作是否对国家、对技术本身、对我们的社会和文化

有所帮助」。比如 IPTO 的研究主管之一 Bob Taylor 会勇敢地站出来，以礼貌、文明的方式批驳他们的短视。

下面是一段 1966 年前后，Bob Taylor 与当时 ARPA 负责人 Charles herzfield 于的对话。Charles 问 Bob:「你想要的是什么?」Bob 答道:「我想做的就是这样的网络」。Charles 并没有继续问这个网络是什么。他只是说，好吧！ Bob 后来回顾这段往事，他说「毫不夸张地说，这是一段时长仅仅为 15 分钟的对话」。Charles 当时直接问 Bob 需要多少启动资金。Bob 说大概需要 100 万美元（相当于今天的 600–700 万美元）。在这次会谈之后，Bob 立刻开始专心研究，好几个月都没有收到 ARPA 任何的官方指示。而这里 Bob 研发的「网络」，正是我们如今熟知的互联网。有趣的是，「互联网」这种如今在地球的各个角落都在使用的技术，在立项的时候竟然没有一份申报书。「互联网」的诞生是由于两个科学家之间互相信任，以及 Llcklider 那样的「愿景」。值得注意的是，在这之后，Bob Taylor 成为了施乐 PARC 计算机科学实验室的创始人和主管，将这一套工作模式运用到了工业界的实验室，并邀请了包括 Alan Kay 在内的美国 20 多位顶尖计算机科学家加盟了 PARC。

1.4.4 将「责任」与「控制」分开

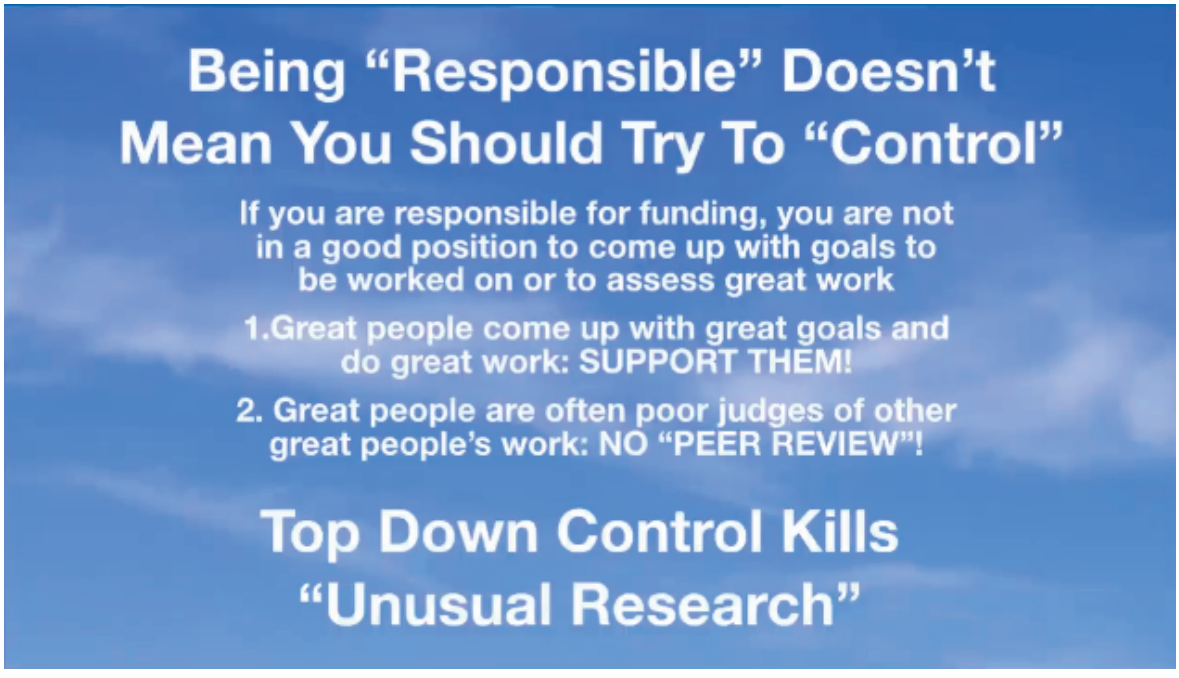


图 11：自顶向下的控制会扼杀「与众不同」的研究

下面，Alan Kay 归纳了一个自认为很关键的论点:**「负责」并不意味着试图「控制」**，他认为这是传统官僚机构中的主管们往往都会犯的一个错误。如果你买了房子，你就会担心自己被解雇，所以你需要对你的老板绝对负责。但是，你基本上无法做到「控制」。因为在「与众不同」的研究中，如果你是资助者，往往会对项目了解不足，因而想不出什么好问题。你会在错误的位置上做错误的工作。

因此，你无法自顶向下地管理这样的项目。对于那些提出了绝妙的研究目标并作出了卓越工作的人，你所需要做的就是支持他们。作为负责人，你需要放手去资助这些用于思考的人，而不要想结果应该是如何。你要知道，大约 60% 的工作都不会成功。

1.4.5 不要进行「同行评审」

对于科研项目的评价机制，Alan Kay 认为，卓越的研究者往往会对其它卓越的研究者的工作给出较差的评价，这可能是由于他们之间存在竞争，或者他们的研究背景不同。对于这种顶级的研究而言，同行评审机制是行不通的。同行评审只对普通的研究工作有意义，而找到真正意义上的「同行」也是十分困难的。

1.4.6 只资助最优秀的人和团队

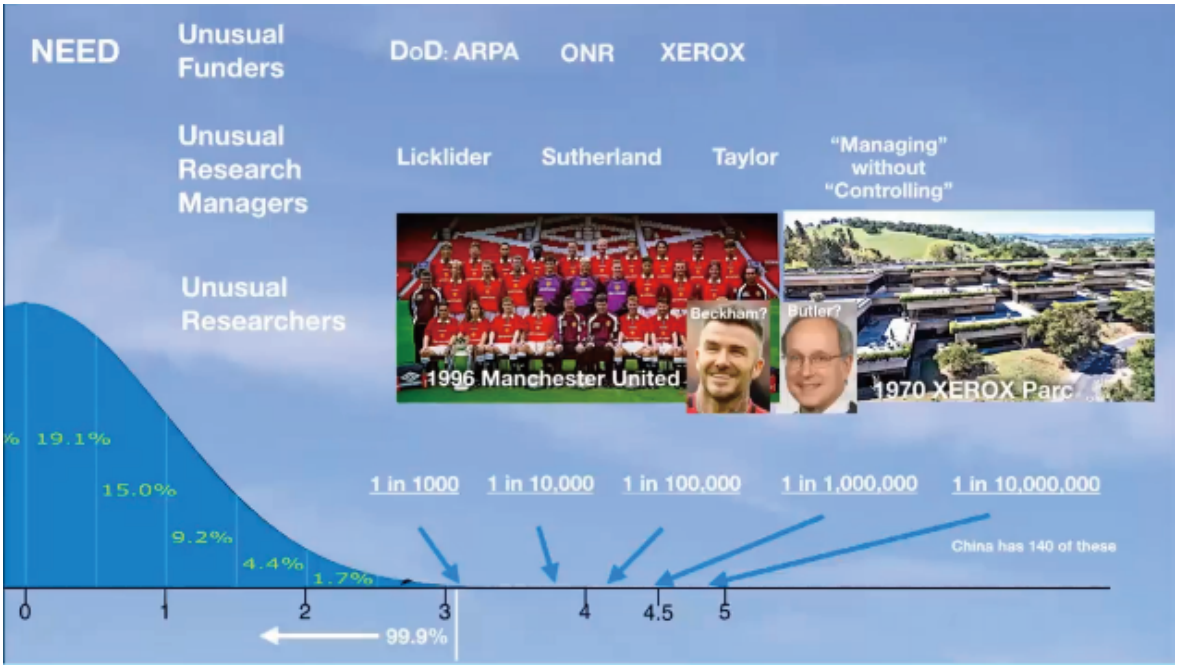


图 12：开展与众不同的研究的三要素——与众不同的资助者、研究管理者、研究人员

那么，如何做出「与众不同」的研究呢？ Alen Kay 指出这需要与众不同的资助者、研究管理者，以及非同寻常的研究人员。上世纪 60、70 年代，在 ARPA、ONR 和 PARC，有着 Licklider、Sutherland、Bob 这样与众不同的研究管理人员。有趣的是，与众不同的研究人员要相对容易找到一些。但如果没有前两者作为保障，这些研究人员也无法成功。如果你想组件一只冠军足球队，你可以雇佣世界上最好的球员（比如，贝克汉姆），只要你不是随意雇佣球员就行。对于施乐 Parc 实验室来说，他们找到了 Butler（编者注：即 Butler Lampson，PARC 的软件开发负责人，因在分布式个人计算环境及其实现技术领域的贡献获得 1992 年图灵奖）。如果你仔观察上图中左下角的钟形曲线，当你为你的研究实验室寻找出色的研究人员时，这些人有的是千里挑一、有的是万里挑一，……，有的则是千万里挑一。而在中国，这样千万里挑一的研究人员有大约 140 人，这个比例大约处于正态分布 5 σ 的位置。 如果你想要组建好的实验室，你需要找到这些人并且好好培养他们。

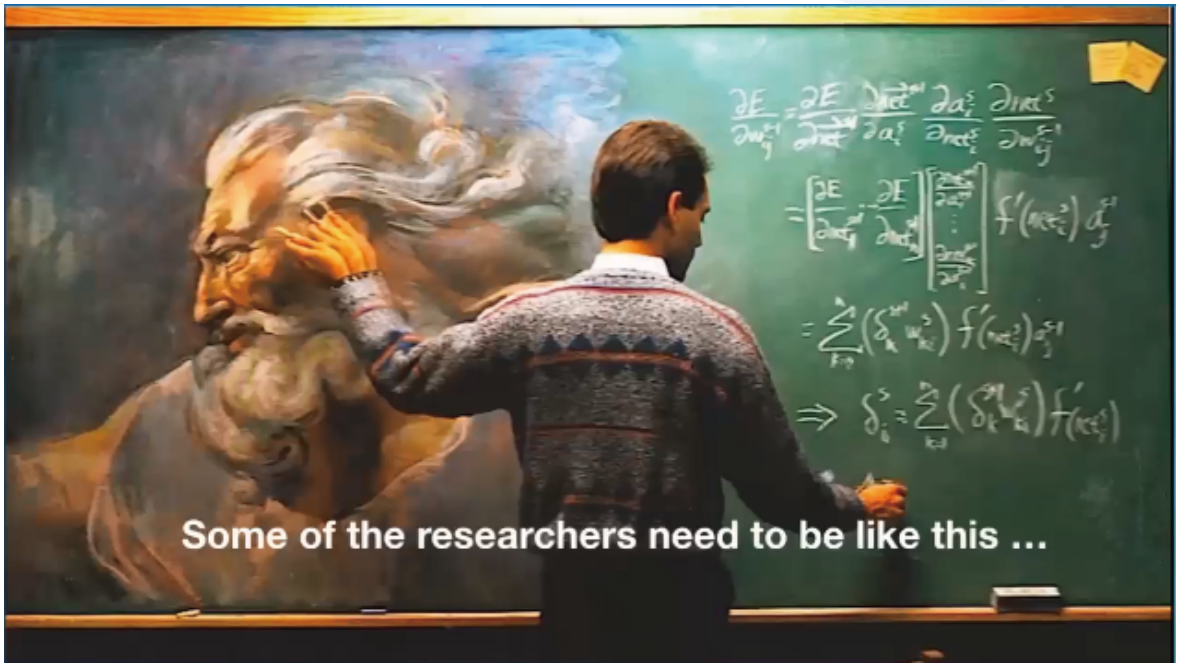


图 13：左手艺术，右手科学

有一些研究者是顶级的科学家，有的是顶级的工程师。如上图所示，这是 Alan Kay 再 40 年前建议苹果公司做的一则广告，它展示了我们的个人电脑是怎样的。而 Alan Kay 则希望它成为下面这样:

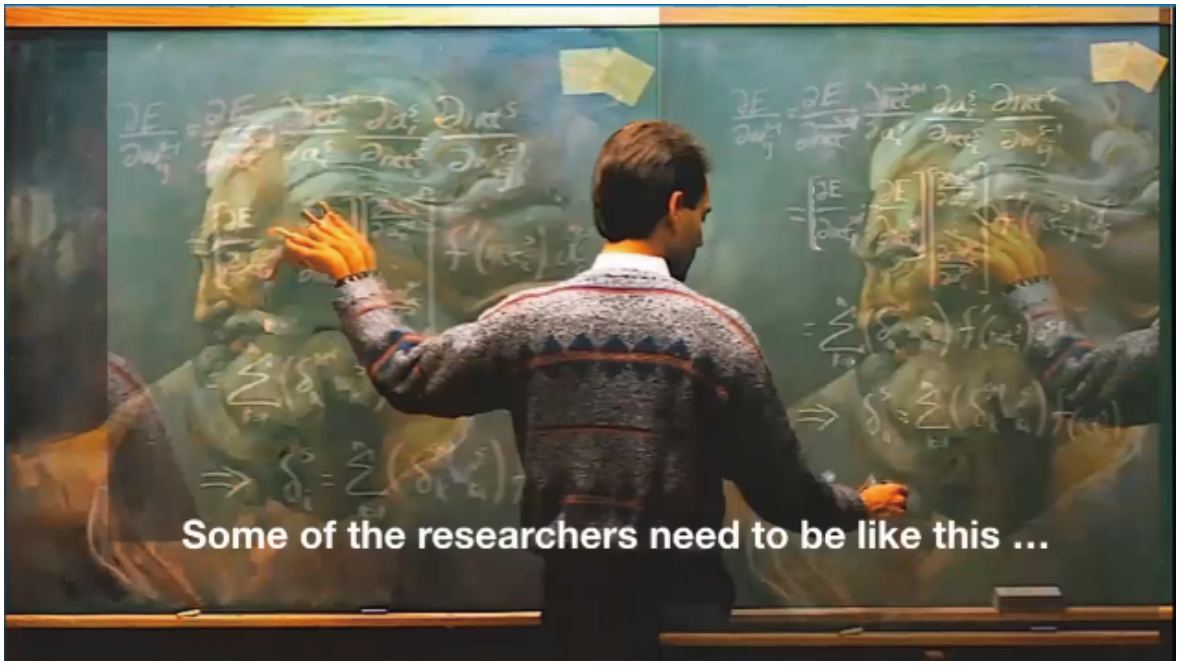


图 14：艺术与科学的交融，多元化思维的碰撞

在这里，左脑和右脑之间并没有很明显的界限； 在这里，真正的思维过程应该是能够流畅地将不同种类的思维

模式融合在一起，将美学思想融入到科学思想之中。

1.4.7 疯狂的科研投资——不计回报

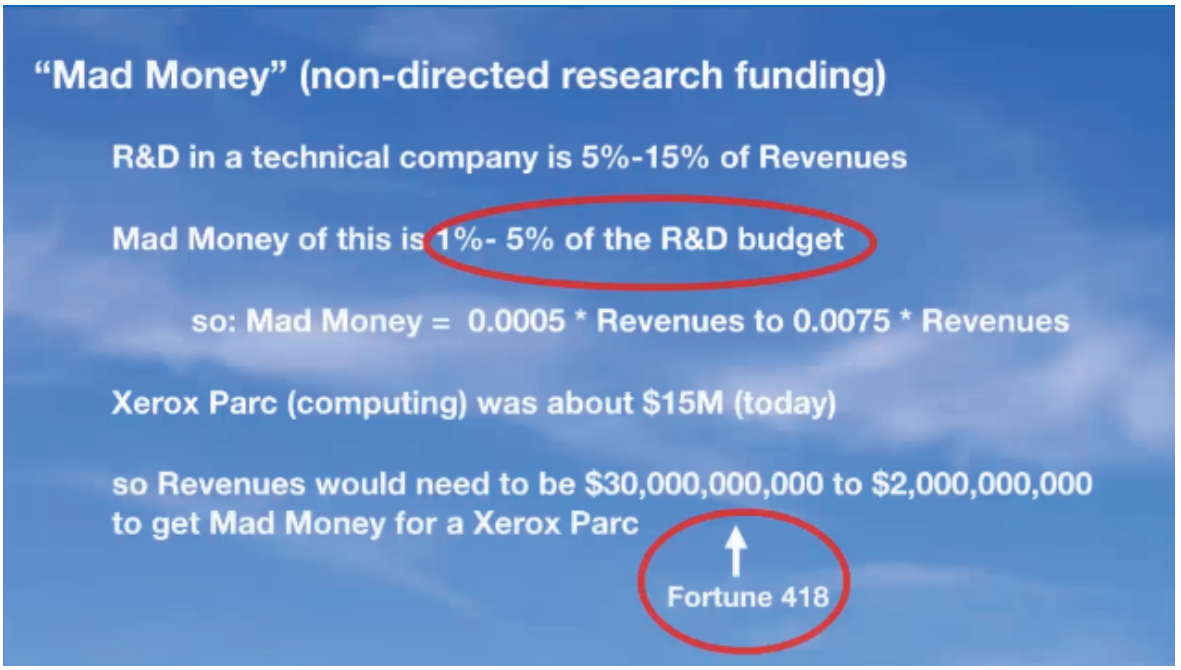


图 15：科技公司的疯狂学术投资——以施乐 Parc 为例

需要指出的是，上述关于施乐帕克的种种研究理念，均精选自 Alan Kay 的白皮书「How ?」(https://internetat50.com/references/Kay_How.pdf)，Alan Kay 总共归纳了十九条经验法则，感兴趣的朋友们不妨查阅全文。在本次演讲中，Alan Kay 最后还着重了介绍了其白皮书并未提及的一个新概念：Mad Money「疯狂的投资」，它指的是如果投资的项目并不成功，你也不在乎，这种投资并不针对任何目的。所以，我们首先要问的一个问题是：对于任何的组织或者国家来说，Mad Money 的体量应该有多大？对于一些组织程度极高的国家来说，他们可能会说「我们不存在 Mad Money，我们对每一分钱都有计划」。这对于创新来说是不利的！对于大多数公司来说，它们想向股东证明它们明智地使用了这笔钱，并带来了投资回报。它们不喜欢展示 5 年内还没有投资回报的项目。

Alan Kay 指出，对于科技公司而言，研发的费用往往占收益的 5%–15%，这些钱大多数都用在了产品上。而其中 Mad Money 则占研发预算的 1%–5%。所以，Mad Money 只占公司全部收益中极小的一部分。今天，施乐 Parc 实验室的这项预算为 1,500 万美元。即使以最低的标准（研发预算占公司总收益的 5%，Mad Money 占研发预算的 1%）进行计算，《财富》排行榜上名列 418 也可以负担起施乐 Parc 的 Mad Money 费用。而事实上，并不只有一家这样的机构。如果更高的标准（研发预算占公司总收益的 15%，Mad Money 占研发预算的 5%）来计算，想要负担起施乐 Parc 的 Mad Money 费用，《财富》排行榜名列 1,000 左右的公司也可以做到。

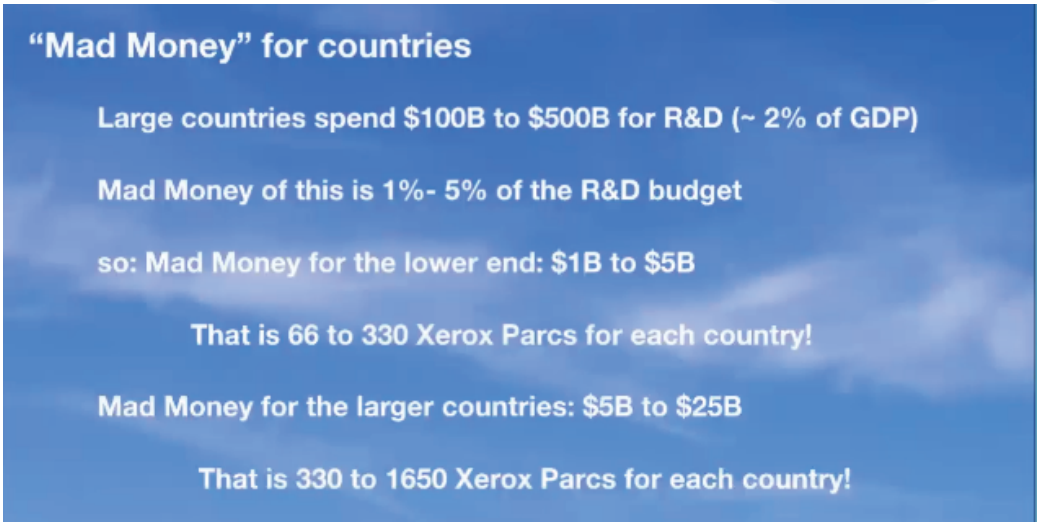


图 16：科技大国的疯狂学术投资

我们再来看看各个国家对「Mad Money」的投入。对于日本和德国这样的科技大国来说，它们每年会投入 1,000 亿到 5,000 亿美元用于国家的科技研发，中国每年大概投入了 2,400 到 2,500 美元，美国大概会投入 5,000 亿美元左右。无论具体数额如何，这一比例往往占大国 GDP 的 2%。如果我们将其中相同的比例（1%–5%）分配给 Mad Money。 那么 Mad Money 的下限就达到了 10 亿到 50 亿美元。此时，你不应该担心这些投资的结果会如何。那么对于每个 Mad Money 预算达到 10 亿美元的国家来说，就可以建立 66 个以上的施乐 Parc 实验室。而对于中国和美国这样体量更大的国家来说，他们一年可以建立 330 到 1650 个施乐 Parc 实验室。这看起来很值得去做，不是吗?

1.5 用非传统方式迎接全球的巨大挑战

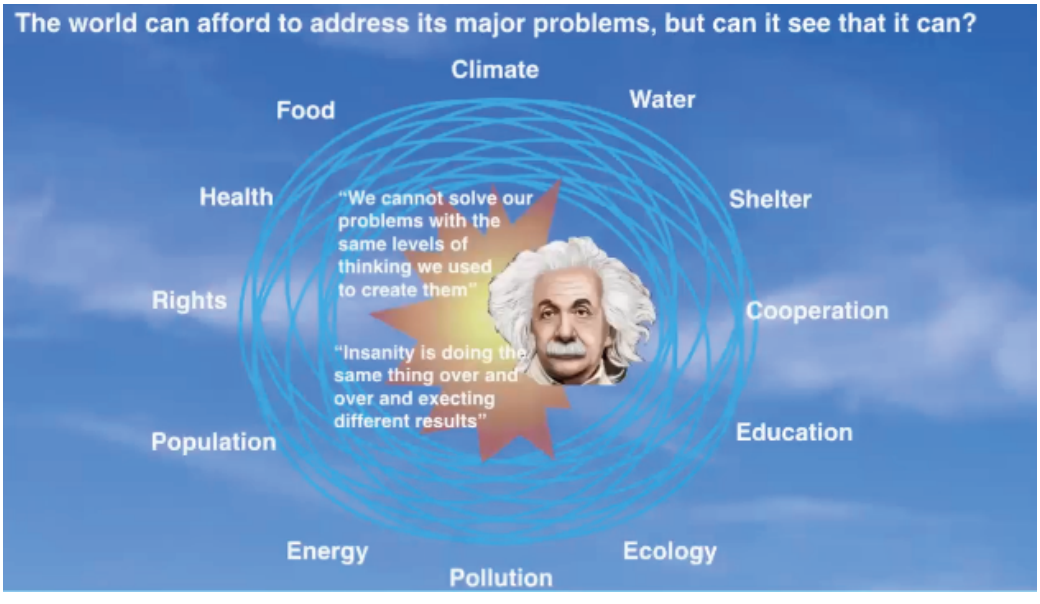


图 17：全人类联合起来就足以应对各种重大挑战，但是人们意识到了吗?

顺着 Mad Money 的思路，Alan Kay 进一步发出了这样的倡议：全球各国家（尤其是各个大国）共同行动起来，共享大量的研究基金、潜在的科研人才，从而以一种非传统的方式应对这些我们所面临的巨大挑战。但问题是，世界上的大多数人并没有意识到我们面临着这些问题，或者他们看不到这些问题带来的威胁、恐惧或其它的代价。

Alan Kay 由此用爱因斯坦的一句名言作为结尾，他认为这句话不仅概括了过去 40、50 年中计算机科学研究领域的弊端，而且对于我们面临的大多数问题，尤其是上文提及的全球 12 各个重大问题来说，值得用来警示的。这句话便是:

愚蠢之处在于，我们一遍又一遍做同样的事，却期待不同的结果发生。

二、张宏江、黄铁军与 Alan Kay 对话精彩实录

Alan Kay 演讲结束后，北京智源人工智能研究院理事长张宏江、北京智源人工智能研究院院长黄铁军，与 Alan Kay 进行了深入的交谈。下面是他们部分对话内容的精彩实录:

张宏江：在您的文章「How ?」(https://internetat50.com/references/Kay_How.pdf) 中，已经对刚才演讲中的诸多观点进行了很好的总结。在这篇文章中，您列举出了 19 条关于研发和如何应对挑战的规则，这些在您刚才的演讲中也有所提及。但是也有一些您在演讲中没有提到的部分，即对于那些想遵循您提出的 19 条规则的人来说，仍然存在着很大的障碍（例如所谓的「鸵鸟悖论」），您能谈谈对这些障碍的看法吗?

Alan Kay：相关的参考资料请参阅维基百科有关「认知障碍」(Cognitive Bias) 的文章。认知障碍与大约 200 种人脑中的错误思维有关，这是人类学、心理学在过去几百年中的研究成果，它解释了一些有趣的问题，例如「我们的思考能力为什么如此之弱」、「为什么我们需要经过大量的训练才能很好地进行思考」、「为什么科学只在过去的 200 多年才得以迸发，而在数百年前或数千年前却没有得到迅速的发展」。「确认性偏差」(Confirmation Bias) 是科学家们所熟知的一种认知偏差，而科学家们也会犯这样的错误。「确认性误差」指的是，如果你坚信某种理论，你就会有意识地寻找能够巩固这种信念的证据，并且赋予这种证据的权重要远远大于你赋予那些反对该信念的证据的权重。在大家对疫情的反应中，我们也看到了「鸵鸟综合征」(Ostrich Syndrome) 的存在。这些问题都是根植于所有哺乳动物中的。大多数动物在遇到危险时，可能基本上会选择躲藏起来或者进行反击，这些同样也是人类本能的一部分。然而，人类应该学习克服这些本能，比如令人恐惧的战争。

张宏江：正如您刚才在演讲中提到的，当人们真正感到恐惧或危险时，会采取行动，比如美国在苏联的威胁（发射卫星）下建立了 ARPA。那么我的问题是，当前已经持续了数月的新冠疫情已经造成了许多经济层面、民生层面上的问题，你认为这是一个足够大的威胁吗？它是否能唤醒人们？人们是否已经开始采取正确的应对措施?

Alan Kay：是的，它似乎正是一种这样的威胁。而对新冠疫情的处理与政府的组织形式有关。新西兰行动得就很快很好，而美国至今为止仍然在一条错误的道路上渐行渐远。但总体上来看，我们希望主流的人们能做出正确的应对。人们往往不能想象（新冠的后果），直到他们的亲朋好友患上这种疾病、孩子死于这种疾病，或者自己患病。(以战争为例）战争在人类历史上屡见不鲜，过去人们常常受到它的威胁，因此，英国人（二战前）开始研制雷达，用一些雷达阵列来感知国境之外的战役，当时其它一些国家并不相信将会与德国发生战争——政治家们都试图避免战争的发生。但是雷达的创造者们还是完成了这项发明工作，并成为该领域的领导者。当年，

我的同事们也对此做出了反应，在基金的资助下开展了雷达的研制工作。但直到二战真正到来之前，雷达在英国、美国都不是政府优先考虑的项目，但雷达还是在这两个国家建立了起来。在美国，由于政府的资助，一些创建雷达的人已经成为了亿万富翁。这些项目在美国实际参战之前由一些物理学家完成，包括 MIT 也参与了其中。但当时这个项目并没有能启发美国国会应对当下的问题。在那段日子里，美国奉行的是孤立主义。

张宏江：您在演讲中提及了现代计算机的发展历史，您怎么看过去 60 年的人工智能的发展?

Alan Kay：这个问题很好回答。如果你是一名生物学家，你就会对人脑的工作方式有所了解。人的大脑中有大量非逻辑的部分，就像机器学习本质上是一些运算的很快的相关因素一样。而现在我们要想完成一些工作，需要真正的通用人工智能。所以这是现在我们做错的了的地方。对于那些想要继续在这个领域进行研究的人来说，他们应该关注更加困难的部分，即「认知部分」。而在过去的 40 年中，在美国，认知科学研究人员受资助的情况就要差得多。机器学习这些年来取得的成绩令人钦佩。但是令人疑惑的是，对于整个人工智能领域来说，必要的进展还很小。

张宏江：本届大会的主题是「人工智能的下一个十年」，请问您对人工智能下一个十年的预测、期望、愿景是什么呢?

Alan Kay：首先，我们应该严肃地看待「智能」一词。就像在「计算机科学」一词中，我们需要认真对待「科学」。在上世纪 60 年代我们将其称之为「科学」，但 70 年代之后大部分美国 CS 专业毕业生学到的是「工程」。当 CS 在施乐帕克实验室取得成功之后，许多人都想加入到这一领域中来，从成功中分一杯羹。有趣的是，现在一些研究人员会很快地写出一篇研究论文，机器学习领域也是如此，因为论文中一些表示矩阵数学的部分，其实对应的是相同的感知机。如果你仔细看这些论文的数学细节，会发现它们并没有那么有趣。Judea Pearl 将现在的机器学习称作「曲线拟合器」。但如果你从事的是人工智能工作，就需要做更多的事情。我们不能将人工智能的冰山一角等同于「AI」(即使你用到了大量的计算机、芯片)。我想对年轻的研究人员说，请保持本心！不要在意论文的发表，不要在意你的博士学位，那些真正做出巨大突破的人并不在意这些事情。我并不反对资助者资助其它的人，但是如果资助者忽视了那些能做出突破性贡献的人，那将是一场灾难。我认为，在过去的 20–25 年中，计算机领域经历了许多这样的灾难。

张宏江：当您在 PARC 从事图形化用户界面 (GUI) 的相关工作时，有些人认为这其中并没有太多数学上的工作，比如它并没有用到太多微分方程的知识。请问您是如何说服自己的领导，告诉他们这是一项重大突破，并会彻底改变计算机的交互模式，从而导致今天我们在互联网领域的移动设备的革新?

Alan Kay：我从没有说服我的老板关于任何事。因为 Bob Taylor 只是出于直觉就雇佣了我。Bob Taylor 从来没有告诉 PARC 中的研究人员应该如何去做某个项目。

张宏江：请问您如何引导您的想法被研究社区所广泛接受?

Alan Kay：大多数我的想法并没有被社区所接受。

张宏江：那么面向对象编程 (OOP) 和图形化用户界面 (GUI) 呢?

Alan Kay：你们口中所说的 OOP 大部分都与我最初描述的 OOP 大相径庭，我失去了它。在 Parc 诞生的 GUI，则是另一个很有趣的故事。Steve Jobs 使用的是施乐帕克实验室所放弃的技术。

人们可以看到 Jobs 与众不同的思维。如果我们也可以做到这一点，我们会被我们能在 PPT 和 Keynote 中所做的事情所震惊。此时所有的东西都活了起来，它们并不是一成不变的演示结果，它们所有的部件都是可编程、可验证的。如果你纵观这些从 PARC 中诞生的技术，你会发现它们（即使是在想法上）都没有竞争者。激光打印机就是其中一个例子，所以公司立刻开发了对它的研发工作。以太网也是如此，当时也没有任何与它相似的东西，所有人都在试图构建一个局域网，而以太网是唯一行得通的方案。人们都认为现在我们有了与当年 PARC 所发明技术不同的新（OOP）编程技术和语言，其实那是基于我们的技术，通过使用接口改造而成的一个子集。而令人遗憾的一件事情是，目前的网页和网页浏览器，与我们 80 年代所放弃的技术相比，都显得太弱了。

张宏江：Steve Jobs 窃取了你的 GUI 技术，并将其用在了 Mac 中，你欣赏这种做法吗？

Alan Kay：我们开发的是公众可见的技术，在 Steve Jobs 发布带有 GUI 的 Mac 的前两年，我在「科学美国人」上的相关论文就已经发表了。我对 Steve Jobs 和 Bill Gates 都提到过该技术，这并不是什么秘密。我希他们能把完整的 GUI 想法都用上，而不要只使用其中的一部分，从而把事情弄糟。我们在 Parc 工作过的人，没有一个是想要变得富有的。不要使用「x,y,z」来衡量你自己，我根据自己的品质和向某件工作中投入的努力来评价我自己。我只能控制这一点，而无法控制别人对我的看法。此外，在 PARC 工作过的人（或者将范围扩大到整个 ARPA 社区），彼此都有着很稳定的联系，我们都很热爱 Licklider 提出的「愿景」。记得在招聘时，Taylor Bob 会倾向于寻找那些拥有很深厚科学背景的艺术家。

张宏江：2015 年，你曾经访问过企业孵化器 Y Combinator，你在他们的标语上「涂鸦」了几笔。他们的标语是「Make something people want」，而你在「want」（想要）上画了个叉，将其改成了「need」（需要），请问你有何见解？

Alan Kay：我们会想要（want）糖、脂肪、盐，也想要陪伴，这是我们的生理需求。生财之道在于，是为人类任何想要的东西制造一个技术放大器。而这种技术放大器在技术革命的过程中，实际上也制造了违禁的药物。我们不得不拥有这些东西。而我们所需要（need）的东西则是类似于教育之类的事物，所有人都需要深入的教育，而并非所有的孩子都想要接受这种教育。营销人员并不喜欢「需要」这种概念，而喜欢销售人们「想要」的东西。教育者们试图弄清楚人们「需要」什么，并试图找到帮助人们得到「需要」之物的方法。涉及人们「需要」之物的工作要远远比涉及人们「想要」之物的工作繁重得多。如今，我们的社会陷入了一种困境，即失去了在人们「想要」之物和「需要」之物之间的平衡。

张宏江：最后，我想问一个问题，您认为过去 20 年间最伟大的产品是什么？请用一句话概括。

Alan Kay：我们可以在网上找到大多数有趣的东西，在公司中，这些东西可能就会成为产品，现在你可以免费地使用这些产品。在过去的 20 年间，在计算机领域，我认为最有趣的东西是「Croquet」，目前的版本是过去 20 年中五项深度研究工作的产物。它是一种检验麻省理工学院在上世纪 70 年代撰写的论文理论是否有用的系统。这个系统涉及到像互联网一样庞大的内容，以及大规模的拟时间计算的部署工作。这项工作的重要性与 TCP/IP 协议是同一级别的。（编者注：关于 Croquet 项目详细资料，可参阅：https://en.wikipedia.org/wiki/Croquet_Project）

黄铁军： 在中国，即使我们每年有 5% 的预算用于「疯狂的投资」(Mad Money)，但是中国仍然没有诞生 PARC 实验室这样的机构，这是为什么呢?

Alan Kay： 我认为问题在于，研究者需要思考他们真正的目标是什么。

黄铁军： 我认为，造成这种现状的原因是，与你所提到的 Mad money 相比，其余 95% 的投资是「常规投资」(Normal Money)。但是聪明的人是有限的，大多数聪明的人接受的是 Normal Money，他们无法将自己的时间用于自由探索，这是中国没有诞生 PARC 和真正创新工作的原因。

Alan Kay： 微软的创始人之一 Paul Allen 曾经决定捐赠它的数十亿资产，参照施乐 PARC 的成功经验，雇佣了一些施乐的前员工，成立了 Interval 研究院。之后，他们邀请我参观了 Interval，我发现研究院中的大部分事情与 PARC 是完全不一样的。我环顾四周，一些之前在 PARC 工作效率很高的人到了这里却十分低产。Paul 不仅仅满足于资助公司，他想成为公司团队中的一员。我跟 Paul 说，你已经成功的成为了亿万富翁，但是难道你不承认你在一生中从未做过任何真正的计算机科学研究吗？ 你现在应该已经发现这个问题了。你不应该在公司中多管闲事，否则的话你的 Interval 永远无法达成 PARC 曾经的成就。

对于美国研究生院毕业的年轻科学家来说，房地产市场的形势并不乐观，通货膨胀使得房价十分高。这意味着，斯坦福毕业的研究生在找到工作时也无法负担起在帕洛阿尔托买房的费用。他们必须去谷歌上班拿高薪，即使是这样，他们也很难买得起房子。几乎没有人愿意待在家中做研究，因为这样的话，谁能资助你呢？计算机领域的资助几乎都需要工程项目申请书之类的东西，在申请书中，你需要写出你将如何解决问题。而在 ARPA 中，我们可以在信封上随手写下我们想要研究的问题，而如果他们觉得你是个不错的研究者，他们就会资助你，这就是他们对这个问题的回应。在 ARPA，没有人让我去证明任何一件事情，无论是当我在读研究生或是在 PARC 工作时都是如此。当我博士毕业一年后，他们就给了我相当于一年 200 万美元的预算。他们认为我有潜力。当你的研究没有成功时，这些支出不是失败，而是开销。而官僚机构存在的一大问题就是，他们雇佣了大量处于钟形曲线中间区域的普通员工，这些员工需要被管理。突然之间，人与人之间有了等级差异，你也就负有了一定的义务。美国有一个麦克阿瑟基金，他们设立了天才奖学金，用于资助 35 名艺术家 5 年内的开支，而并不在乎这些艺术家做了什么。事实上，「Artist can not do their art」(艺术家不能完成他们自己的艺术)，这正是艺术家的定义。我们应该吸引最好的艺术家们，然后为我们获得的 30% 的产出而感到高兴!

三、结语

从「李约瑟问题」到「钱学森之问」，我们不停地在探索培养新一代中国杰出科学家的道路。如今，Alan Kay 在本届智源大会上提出的诸多观点也从另一个角度给出了破局之法。从人文社会的宏大思考到科学研究的针砭时弊，Alan Kay 发表了自己对于人类未来科学发展走向的野望，留下了一代大师对「后浪」们的期许。

2020 北京智源大会已然落下帷幕，同时也开启了中国人工智能研究下一个十年的浩大画卷，我们期待人类科学迎来百花齐放的春天。人工智能，这颗人类现代科技桂冠上的明珠，正静待着它新的主人!