05 全体大会

圆桌论坛：AI 精度与隐私的博弈

转载自：AI 科技评论

AI 时代，大众是不是真的就没有隐私了呢?

以联邦学习为代表的新兴 AI 技术，能否实现 AI 协作，提升模型精度的同时实现数据隐私的保护。那么中国如何抢占人工智能安全发展的制高点？下一个 10 年中人工智能又将何去何从?

就上述话题，在第二届北京智源大会上，AI Time 联合北京智源研究院，邀请了**张钹院士、高文院士、杨强教授、唐杰教授、刘知远副教授进行了第 15 期论道，共同探讨“ AI 精度与隐私的博弈”。**在具体讨论过程中，采用了唐杰、刘知远发问，张钹、高文、杨强回答的形式进行。

在论坛中，几位嘉宾提到，国外以“欧盟的 GDPR”为代表的相关法律法规以及国内的相关法律规定都取得了长足进展，隐私计算技术也出现了三大主流门派：譬如说安全多方计算——少林派；安全可信计算环境方式——华山派；联邦计算——武当派。

另外，几位嘉宾还就提升模型精度是否一定意味着牺牲隐私保护？如何让更多人参与到技术创新中来，是开源还是激励机制？下一代 AI 技术应该具备哪些特点？我们如何抢占制高点？人与 AI 如何更好地协作，创造更大的价值？等等这些问题进行了讨论。

以下是演讲全文，AI 科技评论做了不改变愿意的整理。

一、隐私保护是否阻碍了人工智能的发展

唐杰：提升模型精度是否一定要牺牲隐私保护？重视隐私保护是否阻碍了人工智能的发展，对 AI 的应用和用户隐私的数据安全的担忧，是杞人忧天还是未雨绸缪，如何处理好 AI 的技术，如何提高模型的精度的同时，能够实现智能与精度以及隐私的同时保护?

张钹：今天讨论的问题是人工智能和隐私保护的关系，实际上涉及到技术和隐私保护的关系。隐私保护的原定义是：有关个人或者团体的隐私，在本人或者团体没有允许情况下，不能随便的收集、传播和利用。

但是随着技术的变化，隐私被破坏的可能性越来越大。例如，有了照相技术，就有肖像权的问题。所以，按照上述隐私保护定义，照片不能随便获取，也不能随便传播，更不能随便使用。

有了网络以后，此问题就变得非常突出，照片在网络上到处都有，隐私保护也遇到了新挑战。所以，技术确实带来了新的隐私保护的问题。

如何解决？主要有两个方面思路：一方面是隐私如何不被误用和滥用，这属于人工智能的治理问题。第二个方面是如何利用技术的手段来保护个人隐私或者团体的隐私，包括数据的安全等等。

高文：隐私保护和技术本身的发展关联性很强。如果隐私保护不出问题，也许不需要太多的技术关注，如果隐私保护不好，可能就需要技术上多想一些办法提供保护。

隐私保护是一个社会学范畴的问题。例如，家庭成员之间的 90% 的隐私是“互通”的；扩大一圈，亲属之间，可能是 80% 的隐私“互通”，再扩大一圈，例如同学关系，长期生活工作在一起的同事，可能掌握你 50% 左右的隐私。随着圈子的扩大，隐私共享的可能性越来越低。所以，在社会中“隐私”是一个相对的概念。

扩展到人工智能相关的隐私，例如人脸识别，大家之所以这么敏感，是因为在法律方面，相关“规则”不配套。所以，如何做好隐私保护，一方面尽可能从技术方面做一些贡献，同时完善相关法律法规。

杨强：隐私的问题一直是人工智能的短板，尤其深度学习模型的训练离不开大数据的支撑，大数据的获得的方法通常有两个方式，一是互联网购买，二是聚合不同的数据源，这两种方式或多或少会侵犯用户隐私。

随着深度学习的精度、数据量的增加，隐私的受威胁程度也在增加，相应的一些法律法规也在出台，例如欧洲的 GDPR。那么隐私保护法规是否会阻碍人工智能的发展？在 2018 年时，我问了瑞典的工业部长这个问题。他的回答到：“人类的技术进步是螺旋性的，我们今天在欧洲提出了非常严格的隐私保护法，也促使人工智能技术公司都遵照法律创造下一代的保护隐私的技术，这也是我们击败美国对应公司的一个办法，毕竟，美国这方面没有欧洲严格。”

瑞典工业部长这番话给我们的启示是，隐私保护法确实为人工智能和大数据规定了很多限制，但同时也激励我们发展下一代的既能保护隐私又能够提高技术的方法。除了联邦学习，还有多方安全计算、差分隐私等等技术正在探索的路上。

二、联邦学习等技术如何实现突围

刘知远：针对数据隐私愈演愈烈的趋势，我们接下来的技术突围之道是什么？以联邦学习为代表的 AI 新技术，它能否解决大数据 AI 协作与数据隐私保护之间的矛盾？这些技术的优势和局限性有哪些？如何让更多的人来参与到这些技术的创新中来，是通过开源还是有什么其他的激励机制？

高文：我赞成百花齐放：一方面，让做隐私保护技术的研究员，尽可能去寻找最好的技术，让隐私得到最好的保护；另一方面，关乎社会的诚信，要想整个技术和社会和谐快速的发展，诚信非常重要。前些年美国发展的比较快，诚信起着非常关键的作用。

而现在，中国已经走过了快速发展的阶段，在最开始原始积累时，有很多原罪性的东西不能较真，我们现也慢慢度过了那个阶段，正在进入非常规范的社会发展期。所以现在诚信非常重要。

前几天我看过一段视频，视频里提到，之所以中国的人工智能发展比美国快，是因为中国民众有限度让步了个人隐私。这种让渡使得技术在没有多少壁垒的情况下，可以快速的发展。

现在隐私保护现状是：美国比中国严格，欧洲又比美国严格，也即欧洲是要求最严的。刚才杨强教授表示，欧洲是用一种严格的办法倒逼技术发展，如果技术出现了，也许就会形成壁垒，甚至到时中国不按照欧洲的规范

操作，很可能中国的人工智能产品无法打入欧洲市场。

反过来，如果“中国规矩”过严，可能对技术积累和发展造成障碍。所以我赞成百花齐放。既不完全遵循欧洲人的做法，也不能置之不理，换句话说，要鼓励中国模式，这样中国的人工智能才能百花齐放。

张钹：我们为什么要去保护隐私？隐私保护的目的清晰非常重要。从西方角度看，他们把隐私的问题看做价值观，是一种绝对化的体现，也即个人数据只要不经本人许可，任何情况下，其他人都无权收集，无权利用和传播。

这个观点看起来很有道理，但是技术发展到今天，我认为此看法不全面。例如人脸识别，如果一条街道不装视频监控，那么这条街道可能经常会发生盗窃，引发社会不安全问题。按照西方的观点，如果街道上有一个用户不赞成安装，那么监控器就无法安装。

按照东方综合考虑个人利益和大众利益的观点。这个监控器应该装，因为能够保证大家安全。将“装”或“不装”问题简化，关键在于个人的隐私不要被滥用，只要保证这一条，我觉得就行了。此问题在中国显得比较简单，在西方掺进去了意识形态，就变得非常复杂。

综上，我的观点是要把为什么要保护隐私这个问题想清楚，很多问题迎刃而解。

刘知远：杨强老师，作为联邦学习这一代表方向的提出者，您可以分享一下如何去利用这些技术，以及如何协同大数据 AI 和隐私保护之间的矛盾?

杨强：联邦学习技术的出现是多种不同技术的聚合，一种是人工智能，一种是多方安全计算，一种是加密技术，一种是分布式这种大规模分布式计算。

我举两个例子，一个例子是谷歌，它有安卓系统，有几千万人在用它的安卓系统，那么每一台设备，都需要 AI 的模型的不断更新，过去的方式是把自己的私有数据上传到云端，云端把这些数据整合以后，再训练出一个更新的模型下发给大家，此过程就暴露了每一个用户的隐私。

谷歌在两年前提出了“set 聚合”的办法，可以不上传数据本身而只上传一些模型参数。这种加密后参数的集群加以整合，然后形成一个更新的模型，再下发给各个用户，整个的过程中服务器并不知道用户的数据，也无法推断出用户的数据，整个过程形成了闭环，就能实现联邦学习。

联邦学习在众多研究者的推动下，可以推广到很多的算法，例如，深度学习算法，逻辑回归算法，XGBOOST 算法等等。所以此领域非常活跃，近几年发展得非常快，在 to C 和 to B 的两大领域都有所发展。

另外，技术开源对联邦学习也非常重要。因为联邦学习需要多方协作，在协作的过程中，每一个参与方都要保证拿到的软件是没有“后门”的。确保无“后门”，最好的方法就是公开，让每位研究者都能检查开源的模型，用大众的力量保证开源软件是安全的，所以开源和联邦学习是分不开的。

唐杰：隐私保护非常重要，在社会共同体里，保证个人的隐私属于个人优化，但也要优化整个社会。在当前中国的情况下，也许把两者同时优化是一个比较好的方式。杨老师说的非常对，技术和法律、法规都是螺旋式发展。以美国当下的技术，其深度学习的发展确实比较领先，但是欧洲开始在法律、法规上发力，也许在未来，新的技术有可能在欧洲或中国诞生，所以这就变成了螺旋式的发展。这一点对我启发非常大，其实我也想把一个问题再抛回给杨老师。

在联邦学习中，大家都把数据和模型加密了并传递这个模型。现在假设有一方是恶意的，他传递了恶意的数据，该怎么办？怎么保证在传递数据或传递模型的过程中不被恶意攻击呢?

杨强：唐老师提的问题是现在联邦学习的一个部分：对抗机制。某个参与方是恶意的，其恶意行为表现在：此参与方可以通过加密参数推测其他参与方的数据，也有可能在数据里面加入一些恶意的机制，使得“联合模型”总体朝着对他有益的方向发展。

如何解决？有各种办法，例如可以不用差分隐私，用比较严格的同态加密的办法。也可以在使用差分隐私的情况下，添加噪音让对方无法推测其他人的数据。总的来说，这个过程是动态的，表现为道高一尺，魔高一丈，永远没有结束的那一天的。

刘知远：张钹院士团队开展了非常多针对深度学习对抗攻击方面的研究。那么请张院士来分享一下，关于隐私保护、联邦学习以及对抗攻击可能存在的研究课题?

张钹：隐私保护实有两类性质的问题，刚才我们的讨论属于企业和个人自律，即如何正确合理公平地使用隐私材料，我赞成必须靠个人和企业的自律来实行。因为个人信息也好，私人拥有的数据也好，不能按照西方对隐私的定义（没有得到本人的允许就不能用）。

举个例子，现在很多服务企业，做用户模型时用了大量的用户的材料，目的是给用户提供更好的、更个性化的服务。那么这些材料没有经过个人的许可就不能用?

回到刚才为什么要保护隐私的问题？其中涉及到保护和使用的关系，此关系非常复杂，可以规定很严，也可以规定很松。在发展的过程中，我赞成逐步改进，不一定在最开始就非常严格，要求非常完整的规则，因为这并不一定会有利于人工智能发展。

现在讨论另外一个问题，即如何防止人工智能的技术被滥用？这个问题目前是最严重的。在这里必须做两方面的工作，一方面是要建立规则，订立严格的法律限制行为；另一方面要发展，也即发展安全、可靠、可信和可扩展的人工智能技术。只有往这个方面努力，技术发展才能真正保证人工智能的安全。

三、下一代人工智能：以人为中心

唐杰：下一代人工智能的技术应该具备哪些特点。另外，在当下 AI 大热的时候，我们如何抢占制高点？尤其在中国，现在强调去除“伪论文、伪数量”，所以，我们应该如何做颠覆式的创新?

杨强：下一代人工智能一定以人为中心。当前 AlphaGo、无人机、无人车此类人工智能都没有真正考虑人的因

素。近几年越来越多的发现机器和人要协作，那么在这个过程中，人的利益有哪些我们需要关注的呢?

可解释性是下一代人工智能的关注点，这里涉及人机协作。其他方面，例如如何能够把人的智慧赋予机器，让机器站在人类的肩膀上进行学习也是下一代人工智能的一个特点。

高文：我把可解释性排在下一代人工智能的第一位，排在第二位的是高效能。因为现在的人工智能不管是追求精度，还是希望能够超过人的性能，其实付出的代价非常大。例如，最近鹏城准备投资 40 多个亿，做鹏城云脑 II，希望获得 1000 个 P 的 AI 算力来支持大模型的训练。然而“人”其实不需要这么大的算力。所以，我希望下一代的人工智能，效率高一些，不要为了训练一个模型花费这么大的算力。

张钹：我在考虑一个问题：人工智能最终目标是什么？我们现在做了哪些事情？我们花了近 60 年的时间，实际上做两件事情。

一件事是根据符号主义的思路，建立一个以知识为基础的推理的模型，并用它模拟人类的理性行为，目前有了一些进展。

第二件事就是深度学习，深度学习其实是联结主义的思路，是从神经网络的层面上试图模拟人类的智能行为。那么这两件事现在做的怎么样呢？很多人评价：这只是人工智能的序幕，大戏还没开始。为什么说它是序幕?

因为这两个方式都不可能达到真正的智能。第一个方式没有解决符号的基础问题，例如基本概念：如何告诉计算机什么叫“吃饭”？什么叫“下围棋”？你只能用符号告诉它说这叫“吃饭”，这叫“下围棋”。其实这种“告诉”没有“根基”，而这种“根基”产生于人类跟环境交互中建立起来的概念，也就是说它只有吃过饭、睡过觉，才能知道什么叫吃饭，什么叫睡觉。因此现在推理也好，符号主义的方法也好，并不是真正的智能。

深度学习更是这样。深度学习和人的感知相差太多，它只做了分类、感知等事情，并没有做到认知，也即它可以区分物体，但是它并不认识这个物体。综上，上述两种方式都走不远。

也有相当多人已经认识到，只有把这两种方式结合起来，才有可能产生真正的智能，因为这两个是互补的。具体而言，符号主义是解决高层的理性分析、理性智能的问题；联结主义解决感知的问题。所以，“结合”问题得到解决的话，目前人工智能存在的那些缺点：不安全、不可靠、不可解释、不容易扩展都能够解决。

这个思路经过这段时间大家的摸索，已经看到了一条可行的路，当然这条路还很长，因为这是从 0 到 1 的创新，我们只有加强基础研究，加强人工智能的基础研究，才能够解决这个问题。

第三代人工智能的方向也很明确，就是要把第一代人工智能的知识为基础的知识驱动的方法跟第二代人工智能的数据驱动方法结合起来。

唐杰：我们把这两个结合起来实现认知，那么过程中有没有阶段性，也即能不能分解：第一步要做什么，第二步做什么，或者是说分解成几个方面?

张钹：在认知这条路上的，大家已经做了好多工作，比如说对抗神经网络，还有在深层次的神经网络，就是通过深层的网络来学习先验的知识。例如，人要认识狗，必须要有狗的先验，计算机没有狗的先验，所以它不认识狗。

狗的先验从哪来呢？深层次的网络是通过无监督或者弱监督学习来实现。所以大家已经在往前走。那么我们也做了一个工作，把深层次的网络等几个网络结合在一起，就可以把物体的先验知识通过弱监督或者无监督进行学习，那么利用这个知识来指导分类，使得它既可以分辨物体，又可以认识物体。

唐杰：进行下一个问题，我们如何抢占制高点？从学生的角度来看，如何做科研，也就是说如何在研究上抢占制高点，如何在工程上和系统上抢占制高点。如何看待当前“伪论文和伪数量”的观点？

高文：抢占制高点和发表论文一点都不矛盾，其中研究的动机是关键，因为以前有很多学生的动机是发论文获得毕业资格，研究本身是否是新的，学生并不在乎。所以，老师的责任就很大，老师能否“指挥”学生做一个完全陌生的问题，如果学生能够做出来，那么肯定学生将来很厉害。但大多数老师没有这个经历。所以，抢占制高点最关键是瞄准一些没有人做过的东西，需要老师“监督”学生做创新，改变传统的“发论文毕业”的研究动机。

杨强：唐老师说的非常对，现在论文都成灾了，一些顶会动辄就是上万个投稿。但这个现象并不奇怪，因为大家认为一个方向有前途，必然第一件事是写文章，从而出人头地。那么有胸怀、有志向的研究者应该是在创新上面努力。创新的一部分来源于选题，而现实是，大家可能更关注文章的数量，而没有在选题方面进行更多的关注。

我曾在《学术研究》这本书里面提到什么样的研究该做：首先这个研究很新，没有发生过，例如在计算机出现的时候，语言识别就是好方向；第二是这个研究很难，乍一听感觉无从下手；第三，这个问题很容易解释，此类问题往往一句话就能描述，但对计算机来说并不容易。

有了这些条件还不够，还需要有方法把“研究课题”分解成不同的阶段，每一个阶段就是一个小目标，毕竟小目标好解决一些。另外，数据最好是很容易获得，然后这个研究才可以落地。

大家如果按照上述方式去寻找一些新的问题，然后会发现，总是有一些新的问题没有解决。在大家眼里，这种包含“新问题”的文章并不嫌多。

举个例子，例如在联邦学习领域，一有文章出现，我就立刻去读。为什么呢？因为这是一个新的领域，例如迁移学习，通用性的迁移学习领域很新，做的人却很少。所以这样的领域并不是说在今天不存在了，还是有新的领域值得大家去发现的。

张钹：我基本上同意刚才两位老师讲的内容，但是基础研究最后落脚还是落在发表论文上，我们现在发表论文数量比较多，这是一个非常大的进步。特别是反映了我们从过去发表不了文章，很少能够发表文章，到现在能够大量的发表文章，这说明我们在基础研究上的巨大进步。

现在大家对“论文数量”产生批评的态度，主要原因是：我们现在基础研究的平均水平，实际上是接近世界水平的。但是缺点在哪呢？最高水平跟世界发达国家差的太多，也就是说我们还没有一个从 0~1 的发现，这种发现在发达国家里面能够达成，我们现在还做不到。

不要对现在的现象做过多的批评，过去没有条件做事，当年我们都不知道前沿在哪，这导致是最前沿的文章都发表不出来，怎么可能去做 0~1 的发现。所以我们现在平均水平已经接近世界一流的情况下，是有条件做 0~1 的发现，也就是说可以做一些具有巨大影响力的基础研究。

现在中国在在体制上还有很大的困难。我们有很多优秀的学生，但一般情况下不敢让这个学生做非常困难、非常新的问题。因为我们的学生经不起失败，而研究探索性的问题，应该要经历相当多的失败。

在这方面，外国体制的表现是：博士生毕不了业，其实产生不了多大的影响。比如一个国外博士生在校园里做了 8 年，到期了做不出来，结果去找工作，好多公司抢着要他。为什么他如此抢手，因为他有 8 年的“工作经验”。

如果在国内，有一个学生非常优秀，但在清华大学读博士期间没有做出成果，我们如何处理这个问题？现实情况是：他根本没法毕业，虽然没有很难找到工作，但找到的工作往往不太理想。

我们现在有条件做事情了，如果不能从体制和机制上去解决这个问题，这个事情还是做不了的。从学生角度来看，也不愿意去做风险太大的工作。作为老师，因为对学生的前途负责，也不敢把任务交给他。

所以我们的在基础研究上的体制机制还是需要改进的，不然的话，不容易做成功事情。

刘知远：下一代 AI 侧重于精度和隐私保护，那么在联邦学习等等方面我们有什么样值得研究的话题?

杨强：联邦学习实际上来自几个不同领域的交叉。值得研究的话题有以下几个方面:

首先当数据分布在不同的数据拥有方手里的时候，如何能够让模型平衡增长。

第二，如果我们用一个加密机制进行参数沟通的时候，如何能够保证参数的保密的前提下，又能够把模型的速度、效率提升。

第三，如何能够做出更加有效的并且高效的加密算法，这种算法的特点不仅能够保护数据本身，同时又允许在加密的状态下进行各种运算，包括非线性运算，也即支持像深度学习此类的一种网络计算。

另外一个维度是如果我们网络有多个参与方，如何能够建立一种激励机制，例如用经济学和博弈论的观点来设计一个好的机制，能够让大家不断的有动力参与到数据联盟。

还有一个维度是人的因素：在进行协作的情况下，如何让各方遵循同一个标准，也即如何建立一些行业和国际的标准，来让大家有共同的语言交流。

唐杰：隐私保护要做精度和稳定性的提升，而下一代人工智能一个很重要的方向是可解释性。那么联邦学习平台会不会可解释性变成一个黑盒子，甚至变成一个更复杂的黑盒子，反而使得我们下一代人工智能很难实现?

杨强：听起来可解释性和隐私保护是矛盾的，但实际上并不是。例如可以保护原始的数据和模型的参数，但是模型的推理机制却是可以透明的。再例如，一个人去看病，医生往往会给你解释这个病为什么开这个药，但是医生不会给你透露是从哪些案例里面得到这些信息。所以隐私保护和可解释性是两个不同的维度，可以分开讨论。

四、总结与展望：人工智能治理问题很重要

国家新一代人工智能治理专业委员会刚发布了《新一代人工智能治理原则》，发展负责任的人工智能，那么人和AI 如何更好的协作，从而创造更大的价值。希望三位老师做一个总结和展望。

张钹：今天讨论的比较多的问题是关于人工智能的治理问题，这是个非常重要的问题。对搞技术的人员来说，只有把人工智能的治理问题想清楚了，才能够知道技术应该往什么样方向发展，应该做哪些研究工作。

高文：发展人工智能是为了造福人类、造福社会。一个技术的发展，其实是成本和社会影响的产物。如果社会影响很大，那么可以在成本上多付出一些，如果社会影响没有那么大，技术上也不要花那么高的成本。所以，在发展技术的同时，一定要看到它对社会带来的影响到底有多大。

杨强：我们大家都关注下一代的技术发展，尤其是下一代人工智能，我们现在也看到社会对我们使用人工智能有越来越多的限制和要求，那么这些限制和要求乍看上去是限制了我们的发展，但实际上是为我们提供一个机会，一个创新的机会。所以我在此鼓励各位学者和学生多多关注，尤其是在隐私以及人类与人工智能的协作方面引发的一些新的题目。

圆桌论坛：AI——The Past and Coming Decade

转载自：AI 科技评论

经典人工智能方法在未来会得到关注吗？经典方法和深度学习的关系将如何发展？新的突破点又在哪里？

针对上述话题，在 2020 北京智源大会 6 月 21 日晚间的全体大会上，北京智源人工智能研究院理事长张宏江和图灵奖获得者、智源研究院学术顾问委员会委员 John Hopcroft、AAAI 候任主席 Bart Selman 共论 AI，探讨“AI 的过去和未来 10 年”。在具体讨论过程中，采用了张宏江发问，John Hopcroft、Bart Selman 回答的形式进行。在论坛最后，两位教授还向中国的学者们就如何开展 AI 研究提出了中肯的建议。

一、艰难的 50 年和腾飞的 10 年：经典 AI 和深度学习的不同境遇

张宏江：今天我们非常幸运请到了图灵奖的获得者 John Hopcroft 教授，美国康奈尔大学的计算机科学教授 Bart Selman。今天我希望跟两位教授一起回顾人工智能过去几十年的发展，并且介绍他们对于人工智能未来十年发展的方向和途径的看法。今天的主题叫“人工智能新的 10 年”，在谈新的十年之前，请两位先谈一下对过去 60 年，尤其是过去 10 年的回顾。

Bart Selman：这个领域的最初工作主要是由对理解人脑的思维和认知感兴趣的研究人员完成的，在人工智能的早期有很多乐观主义者，但在实际研究过程中遇到了很多意想不到的困难。这个领域在近几十年来纯粹是一门学术学科，因为我们无法在任何可以与人类相比的地方获得很好的成果，第一个转折点是 IBM 在 1997 年开发出国际象棋 AI 深蓝的时候。深蓝在国际象棋中打败人类，这是当时的一个突破。

大约在 2012 年，多层神经网络即深度学习几乎都改变了整个领域，使得我们可以实现视觉识别和语音识别等任务。人们发现，深度学习算法几乎超越了所有类型的机器学习模型，算力的发展是这段时期改变人工智能领域的原因。2012 年，我们让 AI 实现了感知。我不会说感知问题已经解决，但我们离解决更近了一步。此外，这些技术还能和经典人工智能的技术结合，例如决策、规划、推理等。

张宏江：John Hopcroft 教授，您能基于 Bart Selman 教授的观点分享一下您的看法吗?

John Hopcroft：上世纪 60 年代初，人工智能刚起步。当时约翰·麦卡锡创造了斯坦福大学的人工智能实验室，有大量高素质的研究人员在研究符号逻辑。1964 年，人工智能的研究者只能训练单一的权重，原因是根本没有算力，缺乏大型数据集，连手写字符数据集都没有。事实上我拿到博士学位时，才做出了一个包含 1000 个 10x10 像素手写字符的数据集，这和今天的数据集相比是很小的，但是在那时已经是很大的了。

2012 年 AlexNet 出现了，直到那时图像识别错误率才开始显著下降。AlexNet 将图像识别错误率从 25% 降到了 15%，这是一个巨大的进步。AlexNet 有大约 8 个层级，在 2015 年图像识别大赛的冠军 ResNet 则有 1000 个层级，并且它只有 3.6% 的错误率，而人类识别这些图像时有 5% 的错误率。这些技术已经被应用到了非常多的领域，如医学、金融等。

张宏江：在上世纪 90 年代，我们发现经典人工智能方法不管用，现在我们知道是因为没有足够的算力，在过去这几年，由于算力的发展，我们获得了极大进展。那么现在，我们在推理等经典人工智能方法的探索上处于什么阶段呢?

Bart Selman：我觉得现在人们经常把 AI 等同于深度学习，或者把深度学习等同于 AI。过去十年来我们在推理等算法的研究上实际上已经有很大的进步，并且已经有了实践应用，例如英特尔、微软等企业，只是并不广为人知。在社区中，这些技术也被广泛应用。

AlphaGo 是深度学习成功的范例，但实际上 AlphaGo 使用了树搜索算法，这也是一种符号推理算法。这也是 AI 领域出现的巨大进步，我们可以组合不同的算法，例如推理、规划和深度学习方法等。

自动驾驶汽车领域在组合不同的人工智能方法的探索上更加开放。当然你也可以尝试用端到端的方法来训练自动驾驶系统，但这太困难了。深度学习一般作为自驾系统的视觉系统，但是控制系统和路径规划系统等则需要更多的经典人工智能方法。

相比较而言，深度学习确实是非常数据驱动的方法，这跟经典人工智能中我们称之为知识的方法不同。知识就像牛顿定律或万有引力定律，人类的认知需要很多真实的知识。而深度学习要获取知识并不容易，深度学习目前如此有效，是因为我们有大量的数据。我认为深度学习下一步需要学会获取知识，这是个巨大的挑战。

张宏江：人工智能是一个非常宽泛的领域，深度学习仅仅是一部分。Hopcroft 教授，您在过去的 50 年里，在计算理论上做了很多工作，您愿意和我们分享一下，从理论和算法的角度如何看待人工智能的进展吗?

John Hopcroft：我首先再谈论一下深度学习，深度学习真正的意义是在高维空间中更好地识别。比如你在看自行车的图像，深度学习不会告诉你自行车的函数是什么。如果你展示的东西看起来像自行车，但不能让你骑着它去地铁站，它仍然会把它归类为自行车。为了解决这个问题，我们需要将逻辑加入深度学习。

深度学习可以将自行车进行分解，它会告诉你自行车有轮子、链子、座椅、踏板等等。对这些部件，你可以添加逻辑，说踏板带动了链子然后带动轮子，从而自行车可以移动，车头可以让你把控方向和转弯，座椅可以让自行车有运输功能。通过添加逻辑，或许就可以构建关于自行车的函数，这个函数的意义是将一个人从一个地点运输到另一个地点。

关于人工智能理论，单阈值逻辑单元可以用一个非常简单的算法来训练。如果图像集合是线性可分的，那么阈值逻辑可以实现分类。如果图像集合不是线性可分的，应该将集合映射到更高维的空间中，使得集合是线性可分的。关于训练阈值逻辑单元的方式，你也可以不将数据映射到高维空间中，而在原始空间中运行算法，这就是构建支持向量机的技术。直到深度学习发展之前，这是人工智能领域的主要技术。

还有关于过拟合的问题，假设你有一个大型数据集，该数据集告诉我中国所有人的年龄和薪资。我想要问在某个确定的年龄和薪资上有多少人，但是不想保存整个数据集，因为我想把数据集放到手机上。因此我取数据集的一小部分，并且在适当地扩展之后，相信答案会非常接近真值。答案的真假取决于我们要问的问题的范围，并且需要多大规模的样本的数学取决于问题的复杂性。

如果大家对这些数学感兴趣的话，可以看看我所写的书。在这本书的第五章，包含了所需要的数学知识。在我的照片下面写着剑桥（Cambridge）的地方，可以得到 PDF 的链接。

John E. Hopcroft

Computer Science Department
Cornell University
426 Gates Hall
Ithaca, NY 14853
jeh at cs dot cornell dot edu
(607) 255-1179

vita

Book_with_Avrim Blum and Ravi Kannan

Cambridge

John Hopcroft 主页：https://www.cs.cornell.edu/jeh/

二、未来 10 年：数据和知识的相遇

张宏江：Selman 教授，我们知道您之前发表过美国人工智能研究未来 20 年的白皮书，可以分享您对 AI 未来发展趋势以及重点的看法?

白皮书地址：https://arxiv.org/abs/1908.02624

Bart Selman：我认为人工智能研究的未来趋势是社区化，在业界有很多大型的研究团队在快速组建。美国正在构建国家 AI 基础设施，为不同的学术研究团队提供试验台。大多数研究项目无论是人力还是财力对于个人都是无法承受的，需要大量资金用于基础设施软件和机器开发，所以必须共享资源，允许协作。除了仍然要强调知识和深度学习的结合以外，我认为自我意识学习是未来人工智能发展的重点。

张宏江：您所说的自我意识学习到底是什么意思？为什么您认为这是未来 20 年人工智能发展的首要任务?

Bart Selman：这是一个我们还没有解决的问题。人类能以不同的方式学习，例如我们去学校学习新技能，这相比于数据驱动学习而言是非常不同的类型的学习。人类不需要学习上百个例子就可以掌握新技能。自我意识

学习就是指，AI 会反省自己的学习方式，然后调整它的学习行为。

例如，AI 会反思： 我不太明白学习的微积分的知识，我想问一些问题，我想做一些微积分的练习来提高。这个过程中使用的学习例子很少，不需要数以百万计的例子。这些学习风格是人类独有的，而当前的人工智能并没有掌握。

张宏江：当我们回顾过去 50 多年的历史，人工智能有一个非常缓慢的发展过程，然后 10 年前深度学习的出现使领域飞速发展。现在要发展自我意识学习，您认为我们还要再花费 50 年时间吗？您认为自我意识学习中，算力仍然是最重要的因素吗?

Bart Selman：我认为可能要再花费 10 年到 20 年的时间，这很难预测，毕竟一些研究需要基础层面的突破。现在和过去 50 年非常不同的是，我们已经有了数据，所以我对未来还是非常乐观的。

我觉得算力还是必要的组成。我认为自我意识学习，也就是说将深度学习和知识、推理结合起来，还需要很多新的 idea。在机器翻译领域，深度学习做的非常好，可以达到 90% 以上的准确率。问题是最后的 10% 的提升，可能需要相当不同的方法。在自动驾驶系统中，机器做出的决策必须是非常准确的，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 10% 的提升，我们还需要做额外的工作。

张宏江：数据和知识之间是有差别的，这我很赞同。Hopcroft 教授，您可以就人工智能的未来谈谈你的看法吗?

John Hopcroft：农业发展需要很长时间，而制造业发展只需要几百年，人工智能只用了 50 年的时间就发展到了今天的地步。所以我认为接下来的发展会非常快，就像你建立工具来支持更有效的工作。

在未来，我们需要解决可解释性问题、偏见问题、常识问题、责任问题、持续学习问题，此外还包括 AI 取代人类工作后出现的各种社会问题。

张宏江：Selman 教授，您对于我们何时能实现通用人工智能有什么见解呢?

Bart Selman：未来 10 年，我们将更加注重专业能力。但到未来 20 年，我们会集中发展常识、知识、真正的语言理解等领域。真正的语言理解，即机器会像人一样阅读。一旦这个问题解决了，人类就能实现下一次人工智能发展的飞跃。机器一旦掌握如何理解语言，对我而言，就相当于实现了通用人工智能。

张宏江：Hopcroft 教授，您有什么要补充的吗?

John Hopcroft：在接下来的 10 年里，人们可能会聚焦于如何利用工程技术的最新成果解决特定问题。但我们也需要基础研究，甚至是人工智能领域外的基础研究。在过去的 25 年里，基于人类大脑发育的研究也是很重要的。现在人们知道在儿童生命的头两年，大脑会学习如何学习。在未来，真正的进步可能来源于生物学，但真实情况仍然是不确定的。

张宏江：Selman 教授，您能否谈谈，美国近期在人工智能的政策上做出了哪些正确的决定?

Bart Selman：我认为美国做对的决定是，对人工智能领域的研究坚持投资 60 年，即使什么成果也没有，这是一件非常困难的事情，我们不一定知道下一步的发展方向。

如果我们继续投资于各种各样的研究项目，会得到下一个突破。而如果只是投资到某些喜欢的领域，特别是目前最喜欢的领域，可能不会实现突破。研究突破是革命性的，不是进化性的。我相信这就是人工智能的未来，更大的规模，以及更加协作的人工智能研究项目。

张宏江：您能否也说说中国在人工智能支持政策上有哪些做对的地方?

Bart Selman：我认为中国在人工智能支持方面做得很好，如果要提建议的话就是，不要局限于一个领域。看看人工智能的历史，有很多非常困难的问题是从人工智能的历史中就提出的，实际上还没有解决。所以我鼓励新一代研究者熟悉历史，走向比数据驱动更广泛的领域，例如类脑学习、one-shot learning 等等。除了深度学习，还有一些非常重要的领域，即使你专注于深度学习，也不要太狭隘。

三、给中国研究者的建议

张宏江：您建议允许人们探索不同的领域，即使可能会失败，也可以从这些失败中学习。最后一个问题，你们会给中国的人工智能研究人员提什么建议呢?

John Hopcroft：我的建议就是远离指标，中国的研究人员非常感兴趣于发表的论文数量和得到的研究资金数量。远离这些指标，并关注其它的一些更有价值的层面。

张宏江：非常好的建议，所以不要只追求数字。那 Selman 教授，您有什么建议呢?

Bart Selman：我会建议中国的研究人员多关注有创造意义的研究，并愿意考虑其他人没有考虑到的问题，我也同意 Hopcroft 所提的建议。看看那些没有解决的问题，看看那些最难解决的问题，下一个突破比下一个增量式论文更有价值。

张宏江：非常感谢 John Hopcroft 教授和 Bart Selman 教授，我们回顾了过去几十年在人工智能上取得的突破，以及我们对下一个 10 年的展望。在未来，我们或许会看到，人工智能社区走向全球化，经典人工智能未来可期，并与深度学习协同发展，人工智能的各个领域会迎来百花齐放的盛况，新的突破点潜藏其中，或者在领域之外会带来惊喜。