11 全体大会

圆桌论坛：人工智能产业的下一个十年

整理：智源社区 常政、王炜强、贾伟

2020 年 6 月 22 日晚，在第二届北京智源大会晚间圆桌论坛上，智源研究院四位理事单位的相关负责人“云上相聚”，围绕“人工智能产业的下一个十年这个主题发表了独到的见解，他们分别是：百度 CTO 王海峰、小米副总裁崔宝秋、旷视科技首席科学家孙剑、美团首席科学家夏华夏。智源研究院运营副院长刘江担任主持人。在圆桌讨论中，嘉宾们盘点了过去 10 年里 AI 产业中的重要进展，梳理了可能影响未来 10 年发展的几个重要变量，如小数据、AIoT、客户侧训练等，并结合自己在工业界的实践体会，对高校老师、学生们的 AI 研究，提供了一些参考建议。下面，是嘉宾们的精彩观点摘要。

图 1：上排从左至右：孙剑、刘江、崔宝秋下排从左至右：王海峰、夏华夏

一、过去十年：人工智能从“实验”走向“实用”

问题：回顾过去十年，你对 AI 产业印象最深刻的事情是什么？

王海峰：如果总结一下过去 10 年，那便是人工智能真正从实验室走向实用、走向工业大生产的 10 年。百度做人工智能，首先是用 AI 来改造搜索引擎，并对它有一个全面的布局，包括语音技术、视觉技术、自然语言、知识图谱等。就我个人经历而言，我在百度已经工作了 10 年。记得我在 2010 年刚加入百度时，接手了一个机器翻译的项目，一年后上线。当时机器翻译虽然主要基于统计和规则，还没有采用深度学习，但借助于互联网大数据和强大的计算平台，翻译系统仍旧实现了比较迅速的提升。2015 年，我们上线了首个基于神经网络的机器翻译系统，这使得翻译系统取得了突破性的进展，迄今为止这个系统已经实现了全球 200 种语言的互译，每天的翻译体量已经达到千亿字符。我从事机器翻译工作已有 20、30 年，这种现象在早些年是完全不可想象的，它已经真正实现了大规模产业化。

孙剑：有两个 Moments 对我影响很大。第一个是于 2012 年，Hinton 和他的学生 Alex Krizhevsky 设计、推出了 AlexNet。对于 AlexNet 当时的训练结果，包括我在内的很多做计算机视觉的人，都不太敢相信，怀疑可能是数据集出了什么问题。到了 2013 年，Google 宣布将它的深度学习放到 Google Photo（当时叫 Google Picasa）上，我便将自己的几百张单反照片传上去进行了检测，发现识别结果居然大多是正确的。这给了我非常大的震撼，之前我做过一些包括 Sparse Coding 等方法在内的非深度学习的图像识别，根本没有想到深度学习

居然会有这么好的效果。这颠覆了我对图像识别的认识，也让我萌生了非常大的信心，我就立刻组织了一支团队来加大这方面的研究。

夏华夏： 的确，这 10 年来人工智能的发展非常快。记得我 2013 年加入美团时，它还没有算法团队，直到 2013 年下半年才开始有了第一个与人工智能相关的项目——用户画像，此项目希望通过对用户的理解来完成更好的用户推荐和搜索。现在七年过去了，人工智能几乎已经渗透进了美团的每一个业务中（现在约 200 多个），除了做推荐搜索，我们还做智能配送调度、人脸认证和识别、文字识别、智能语音交互、智能语音客服，以及无人车、无人机等等。这种变化真的让人非常兴奋。

崔宝秋： 我想补充一点的是，20 多年前当我还在读博士的时候，其实是很羞于讲自己是研究人工智能的。时至如今，尤其是 AlphaGo 问世之后，最大的一个改变就是大家都可以非常骄傲地讲“我做的是 AI”。

刘江： 之前孙剑、海峰都谈到了，过去 10 年 AI 为什么兴起，最重要的就是深度学习。我们现在来看 Gartner 的一个总结——深度学习的时间线：

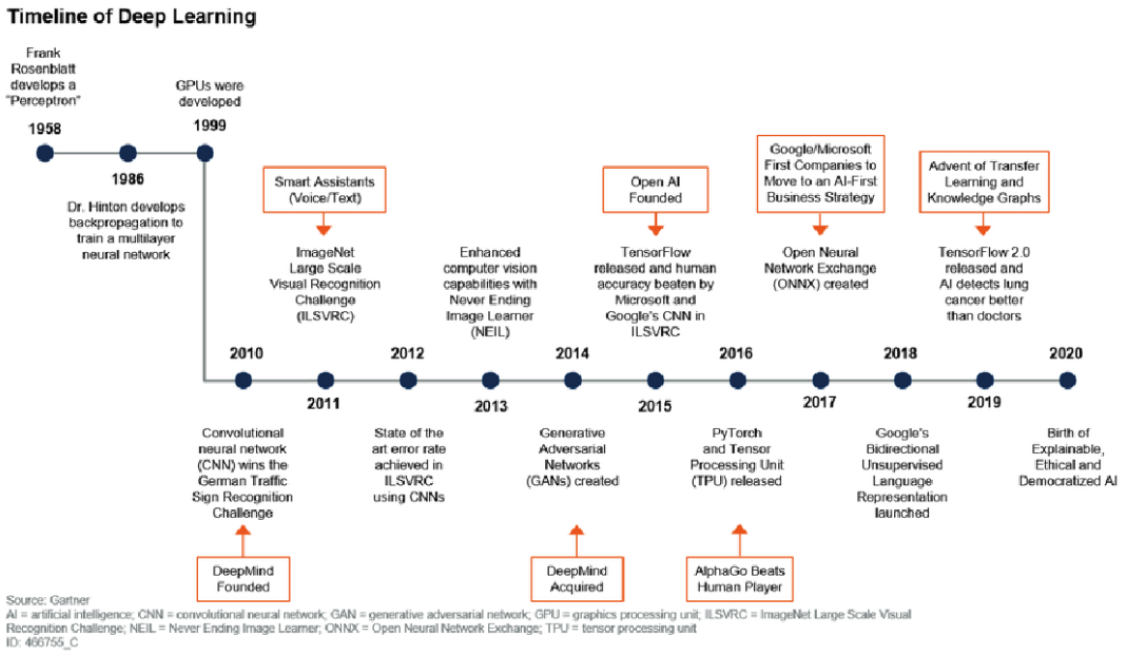

图 2：深度学习发展时间线

深度学习算法最早可以追溯到 1958 年、1986 年，但这波浪潮真正能起来，是 2010 年是 DeepMind 创立，以及 2011 年之后，如孙剑所谈到的，在 ImgNet 等比赛里突然有人用深度学习处理视觉、语音等产生了非常大的突破，包括之后一些深度学习框架、以及 GAN 的诞生等。当然了 Gartner 这张图还有很多真正内行的信息没有包容进去。我曾经跟崔宝秋交流过，他说印象最深的是 Google 实现让机器自己会识别猫了。这让我意识到，这次人工智能在公众视野中的全面兴起还不是因为 ImgNet 这种专业小圈子的比赛，大家真正意识到 AI 的存在是因为 AlphaGo 的诞生：它居然能把围棋界最厉害的李世石都下赢了！这是令公众印象非常深刻的一个点。

二、影响未来十年的变量：小数据、AIoT、客户侧训练…

问题：展望未来十年，你认为哪些因素可能成为大变量，改变 AI 产业格局?

刘江：过去十年，人工智能对整个信息产业产生特别大的影响。我专门查找了一下 2010 和 2019 这两年的 Gartner 报告，我们看到一个有意思事情:

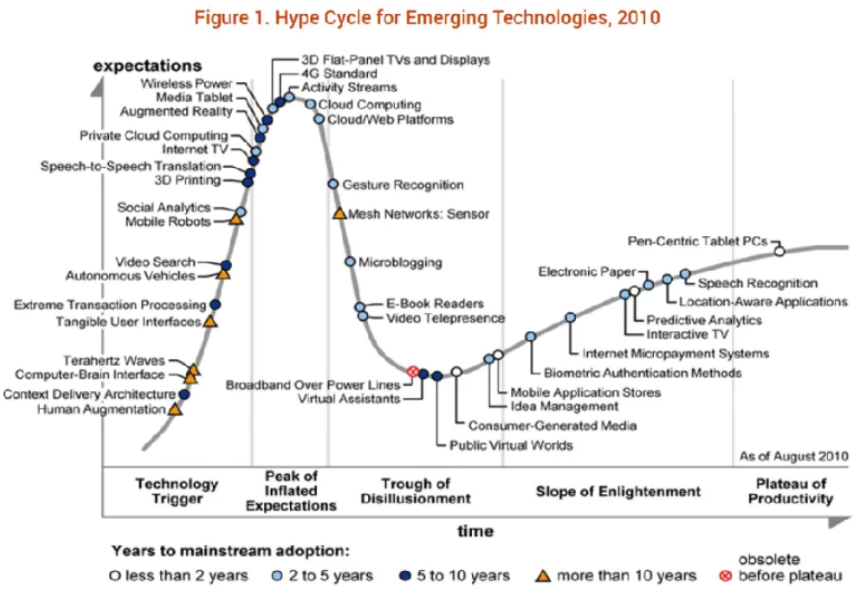

图 3：Hype Cycle for Emerging Technologies, 2010

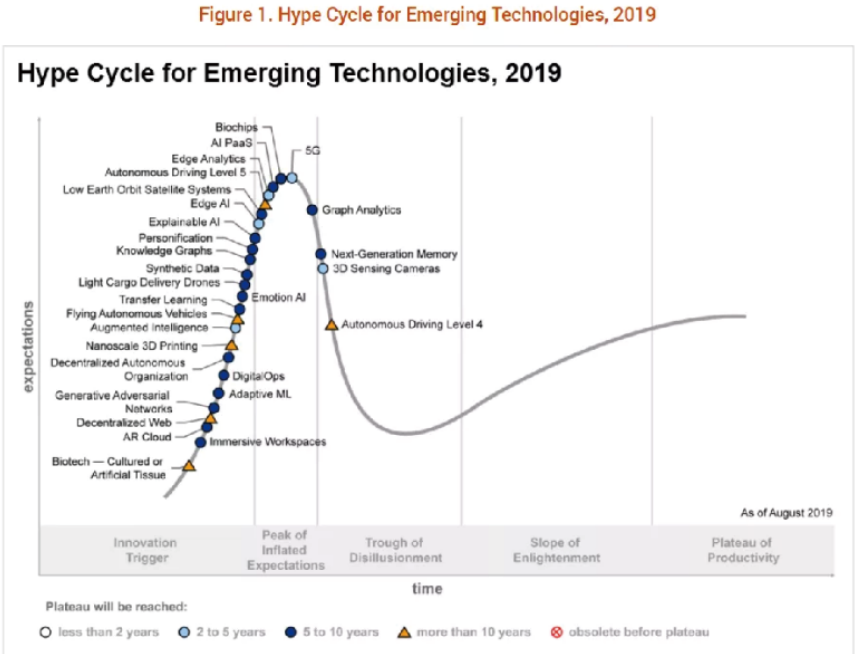

图 4：Hype Cycle for Emerging Technologies, 2019

2010 年的 Gartner 新技术报告中，没有人工智能，也没有深度学习，而到了最新的 2019 年报告，深度学习已经不再出现，因为不算新技术了。这说明，即使像 Gartner 这样专业跟踪技术发展的公司，也没有第一时间关注到以深度学习为代表的这一波人工智能技术的兴起。

另一方面，我个人感觉过去十年真正创造产业价值的人工智能技术：

1）搜索。以百度、Google 为代表的，两家公司都曾经是互联网界、甚至信息产业界市值最高的公司，非常大的商业机会。

2）推荐系统。字节跳动是推荐系统的公司，百度很多业务也是推荐系统在支持，阿里、淘宝起来推荐系统也起到非常重要的作用。昨天在智源大会上，启元世界的创始人（当年曾在淘宝负责系统）说淘宝从千万级用户量到亿级，推荐系统起到至关重要的作用。但是很多人并没有真正注意到这些。无论是百度、字节跳动还是阿里，它里面核心技术是推荐。

3）调度。刚才华夏讲到美团、滴滴，这类公司是 O2O，向物理世界去服务的，这里面非常核心的是调度，智能调度系统。

但是这几个领域，相对来讲在学术界的关注度不是太高。

总结过去十年明显的重要突破显然是深度学习，但是要挖掘一个技术的价值，还要看应用领域，比如搜索、推荐、调度等领域，我们要更细更深一层看这个事情。

王海峰：我想从两个方面来说：技术和产业。从技术层面看，现在的人工智能算法和应用都是基于大数据，计算量非常庞大，我相信未来 10 年中，在小样本、低能耗的学习机制领域一定会发生突破。另一方面，人工智能不仅需要数据，也离不开知识，我相信对知识的挖掘、掌握和充分利用，包括基于知识的的多模态语义理解等，也会成为一个重要的突破方向。还有一个方面，则是人工智能的可解释性问题，亟待大家去解决。从产业层面看，我认为在人工智能“软硬一体化”方向，比如 AIoT 等，包括其在与应用场景深度融合的过程中，一定会产生很多突破。

夏华夏：我想根据电影和书籍举一些例子，来阐述我的判断。首先，在电影《机器人总动员》中，有一个镜头让我深受感触：它描绘的是人类无需工作，每天躺在一个特定的椅子上，由机器人代替完成所有的吃喝娱乐。这说明，在未来人工智能与硬件结合之后，会大幅度提升工业自动化水平，这是一个大趋势和大方向。此外，在电影《黑客帝国》中，机器通过与人类的脑部进行连接，为人类营造出一个完全虚拟的影像与感受，我认为这会是未来娱乐的终极形态。这涉及到所谓的纯数字化，与近年来流行的神经学、脑科学研究，以及我们对人类认知的一些理解等密切相关，这方面如果能够有所突破，对娱乐、学习、教育等领域都会产生很大的影响。第三方面，在《银河帝国》这本书中，银河系每个星球、每个人、每个生物与非生物都互相连接到了一起，一切都有了意识与思考，这与当前万物互联的趋势非常相似。我认为，随着 5G 的普及和推广，以及物联网技术的发展，未来将有大量数据与物体连接到这个大网里。那么 AI 技术如何去处理如此巨量的数据、如此多的非生物意识，将是一个很有意思的挑战。

孙剑：我想向大家分享一下我看好的小方向、小变化。 第一，是目前比较热门的自监督学习，从去年年底开始，它已经出现了可喜的进展： 以前，我们需要通过 ImgNet 上有标注的数据，进行有监督的训练来获得一些特征 (Feature)，而今天已经可以不用标注数据就能学到一样好或者稍微更好的特征。这将会对我们的应用产生非常积极的影响。设想一下，这意味着当你在一个特定场景里，可以通过大量无标注数据预训练出它的特征，然后只需要通过小量的样本，在那个场景里就可以实现更好的成绩。所以我非常看好这个研究方向，包括我们旷视研究院也在非常积极地做这方面的工作。 第二，我也非常看好芯片和算法的协同演化和协同设计，这方面在智源大会“智能体系架构和芯片”专题论坛上有多个报告介绍，我就不展开了。 第三，鉴于机器学习训练所需要的数据，往往都分散在各行各业、各个企业中，很难拿到，所以我比较看好数据安全训练，或者说隐私安全训练的进展。这个方向上的突破，很可能是以第三方的角度提供一个高安全、高可信的机器学习环境，从而在客户那里，能真正地把各行各业的数据价值挖掘出来。我们团队内部将这种模式称为“客户侧训练”。

崔宝秋：我特别想说对于王海峰老师的两个观点，我深以为然。我觉得以下是未来可能产生大变量的方向： 第一是小数据 AI。未来能否基于小样本、小数据，基于逻辑和推理，融合过去的 AI 技术，学到一些新的学习能力，做到通用的人工智能？我希望在这方面能存在一个大变量。 第二是广义的开源。它不仅是代码的开源，更是数据的开源、知识的共享等。 第三是 AI 与其他技术的融合，包括： AI 与 AR、VR 的融合，从而实现非常自然的真正人机交互；AI 与 5G、6G 等通讯技术的融合，5G 将衍生很多新应用场景，6G 将实现泛人工智能；AI 和 IoT（物联网）各种设备的融合，真正打造一个以人为中心的整体智能服务。综合我认为，这是一种从 IoT 到 AIoT 的质变，从 GUI（图形用户界面）到 VUI（语音、视觉用户界面）的质变，从个体到整体的质变。所以小米打造 AIoT 是一个很好的方向，它构建的生态几乎能容纳所有的 AI 企业，包括技术、产品和生态等。

三、AI 研究：立足基础理论，连接产业真实场景

问题：作为产业的代表，对于 AI 学术界的老师同学，你有什么期待和建议?

孙剑：对于学生而言，首先一定要打好基础。在深度学习方向，要打好编程基础，比如用好 Python、能熟练地使用至少一个深度学习框架等。其次，深度学习是一个富于实践性的学科，涉及很多优化，所以我建议大家能够去熟悉优化的基本理论、场景算法等。这里，我要推荐以前上学时候读过的一本书——《数值算法大全》，书中包含了每个算法的高质量 code，相信对深度学习研究者们的帮助会比较大。另外，我觉得还是要全方面的建立、培养科研素质，不迷信权威、勇于挑战前沿智慧等。对于刚毕业的同学而言，如果想快速进步，一个最好的方法是找一个好的研发环境，周围有经验丰富的同伴，比如选择象我们在座诸位的企业等。

夏华夏：首先对老师来说，我建议要多跟产业界合作，产业界有很多真实场景和大量真实数据，现在是时候把在学校里对人工智能理论方面的积累应用到真实场景中来了。其次，我要告诉同学们，现在肯定不用担心人工智能已经到了末尾，如果考虑一下像《黑客帝国》所描绘的人机接口或《银河帝国》的终极世界，我们离它们还很遥远。现在肯定不是结束，甚至都不是结束的开始，最多算开始的结束。

崔宝秋：我阐述两个观点。第一，我认为 AI 的春天刚刚开始，尤其从 AIoT 的角度看，有很多应用场景都还没有将 AI 的能力充分发挥出来，产品的创新、智能场景的创新会带来很多技术的创新。AI 这个春天能持续多久，在于我们所有从业人员的呵护。其次，很多基础技术的研究是需要坚持的，如果我们都很在意一个领域火不火，犹豫要不要进去，这便反映了一种急功近利、挑肥拣瘦的心态，是需要研究人员摒弃的。

王海峰： 我认为 AI 未来在理论研究、技术开发、产业发展上具有非常广阔的前景，只要认准方向、坚定去做，就一定会有收获。

刘江： 我最后做个总结。刚才几位产业嘉宾的总结都是有一些共识的。

首先是这波 AI 的大发展中，现在还是处在一个比较早的时期，未来还将有一个很长的发展阶段。所以从我们工程技术人员的角度来讲，之前深度学习等为代表的很多技术红利，在包括应用层在内的很多层面，还刚刚开始起步，还需要进一步的拓展。同时，对于在校的老师同学来讲，尤其这次疫情之下，我们已经能看到数字化、无人化等趋势，正在快速地发展，所以这个大前景下还是有非常多的事情、场景等待我们科研人员去探索和研究。

但是在外在的热潮中，大家也要注意看清、把握事情的整个本质，不能人云亦云、急功近利，而是要关注基础，真正把理论、技术和产业的原理、发展脉络搞清楚，这是一个需要大家能沉静下来，进行长期耕耘的过程。

尖峰对话：NLP 技术的前沿进展及商业化应用中的挑战

转载自：机器之心

过去一年里，人工智能进展最大的方向在自然语言处理（NLP），BERT、GPT-2 等预训练模型引领了很多方向的新时代，又催生出了大量商业应用机会。面对技术的进步，AI 领域的顶级学者和从业高管是如何看待未来前景的？近日，2020 智源大会在线上召开，在为期四天的会议中，5 位图灵奖得主、上百位业内专家在 19 个专题论坛云上共同畅想了人工智能的下一个十年。

在智源大会上，京东集团技术委员会主席、京东智联云总裁、京东人工智能研究院院长、IEEE Fellow 周伯文与斯坦福大学教授、人工智能实验室负责人克里斯托弗 · 曼宁（Christopher Manning）展开了一次精彩的交流。他们讨论了自然语言处理领域近期的进展，预训练模型兴起之后的未来发展方向，甚至还为人工智能的标杆评测基准——图灵测试找到了一个「替代方案」。

在交流过程中，两人也提及了京东最近被人工智能顶会 ACL-2020 接收的研究，以及曼宁刚刚发表的工作，有关预训练模型学习到的语言结构。

图 1：Christopher Manning（左）与周伯文（右）

在过去这一年中，我们见证了许多 NLP 领域的技术成果和场景落地。对此，人工智能著名学者克里斯托弗 · 曼宁和京东集团技术「掌门人」周伯文是如何看待的？让我们一探究竟。

一、语言理解 & 人机对话领域过去一年的进展

周伯文与曼宁在对话伊始回顾了在 2019 年智源大会上尖峰对话中达成的共识：任务导向的多轮对话是 NLP 下一个十年重点的研究和应用方向。周伯文还创造了一个新词「任务导向型对话智能」（Task-oriented

Conversational Intelligence)，一方面，任务导向型对话智能可以反向推动许多基础技术能力的进步，另一方面，它的发展也将对经济方面产生巨大影响，带来人机交互技术驱动的万亿级市场。

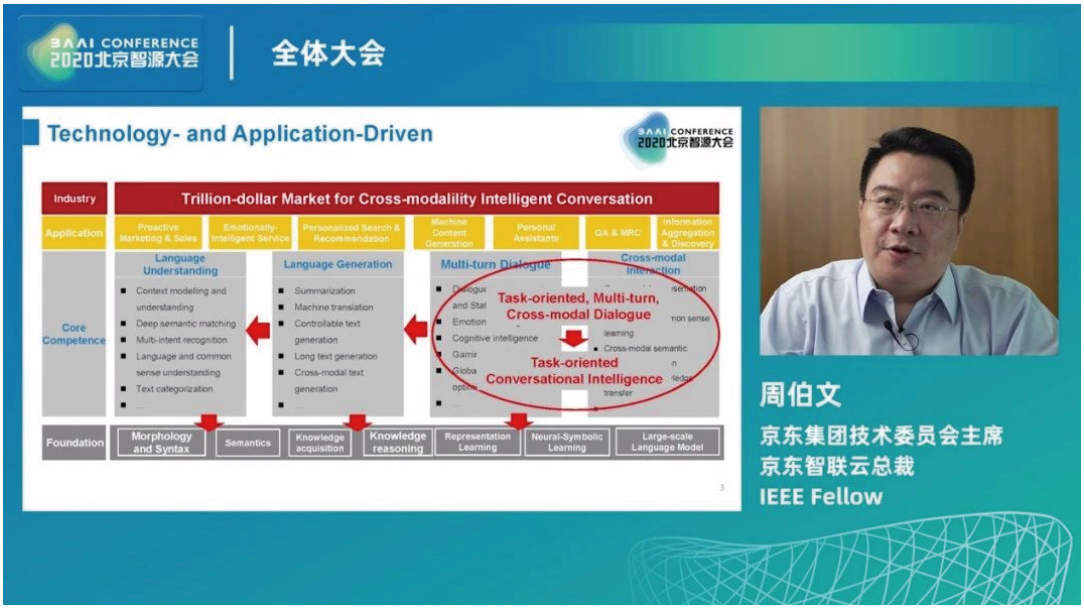

图 2：Technology–and Application–Driven

在语言理解 & 人机对话领域过去一年的进展层面上，周伯文和曼宁不约而同提到了「最令人印象深刻的就是人们见证了超大规模预训练语言模型的出现，它们可以生成有组织的语言文字表达。」

曼宁表示:「其中的代表就是 GPT–2 和 GPT–3，也包含 BERT、RoBERTA 和 ALBERT、ERNIE 等等不少 BERT 变种。它们使得自然语言理解与生成有了非常大的发展。我们也看到传统 AI 领域有了很大转变，很多任务目前都倾向于被大型模型来解决。」

人工智能发展的 40 多年来，我们一直在努力试图让 AI 可以回答科学问题。我们过去尝试使用的思路是研究知识的表达方法，阿兰图灵实验室的 Aristo Project 试图让 AI 理解科学道理，进而深度理解世界，这一思路在最初的十年推动了知识的表达与推理。

在 2020 年，我们通过超大尺寸模型实现了巨大的进步。基于 RoBERTa 预训练模型，我们可以实现 95% 的科学问题回答准确率，这看起来是目前解决知识问题的最好方法了。

这些进步为新一轮商业应用打开了道路。「未来的方向虽然还无法确定，但我们可以看到基于预训练语言模型，为搜索引擎公司等科技企业带来了很多新商业机会，」曼宁表示。「他们可以实现近十年来最大的单个技术进步，构建更好的机器翻译系统，对话 AI，人工智能客服系统等等。现在，我们正在经历 NLP 领域激动人心的时刻。」

NLP 领域最近发生了从特定任务模型向多任务，大规模预训练模型方向转变的重要变化。一方面，工业界乐于看到 BERT 这样模型在下游应用上的前景。但对于学界研究者来说，这种发展大大提高了新研究的门槛。看看 GPT–2 到 GPT–3，它的参数从 15 亿增加到了 1750 亿。但如果仔细观察的话，你会发现模型对知识的获取

和推理性能的提高，可没有参数增加的数量那么多。

针对这一问题，周伯文指出「在查看 GPT-2、GPT-3 相关论文后，有一件事情引起了我的注意，那就是 – 当我们从零样本学习（Zero-Shot）到单样本（One-Shot）学习时，我认为 GPT-3 改进了很多。这有效证明了，从小型模型转换为大型模型时，预训练等于更多的信息。」

与此同时，周伯文发现，从单样本（One-Shot）学习过渡到少样本（Few-Shot）学习时，GPT-3 或 GPT-2 的改进非常非常有限。周伯文指出：「我认为这从另一方面证明，这些更大规模的模型可能并没有学习到足够多的信息。」

由此观之，知识的获取和表征可能仍是 NLP 的正确方向。

曼宁认为，目前的大规模预训练模型可能存在一些「根本性」的错误——这些模型非常低效率。从现实世界人们的对话中学习知识的表征，总不是一个好方法。可能 5 年后人们往回看就会嘲笑现在的工作：「看看这些人吧，只想着把模型做得越来越大就妄想能够实现人工智能了。」

对于研究者来说，我们必须寻找更加有趣的，让模型可以思考、能够更高效提取知识的方法。某种程度上，人们应该需要找到更好的知识编码机制，这有关知识空间，语义连接的更好表达方式。这可能和传统 NLP 的知识图谱和知识表征有关。所以让模型记忆和推断真实世界的情况，看起来从基础上就不是一个正确的，高效的方法。

「人类不是通过这种方法学习知识的。人类存储的知识很少，但可以理解大量知识。」曼宁说道。

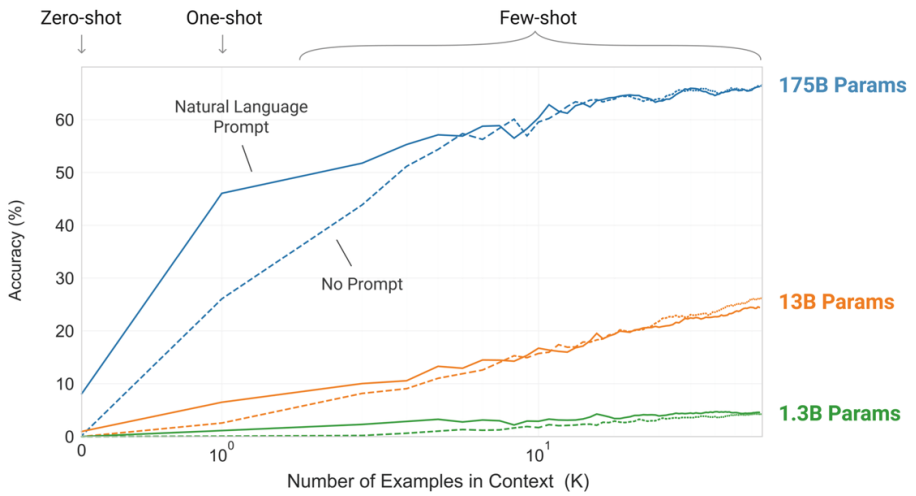

图 3：GPT-3 通过高达 1750 亿参数实现了其他模型无法匹敌的文本生成效果。

作为一个在该领域中务实的研究人员，周伯文非常关注最近预训练的大规模语言模型以及对语言任务进行微调的功能。在一个月前放榜的自然语言处理顶会 ACL 2020 上，周伯文等人有两篇论文被接收。

「在论文《Orthogonal Relation Transforms with Graph Context Modeling for Knowledge Graph Embedding》中，我们得出的结论是通过预训练模型，我们可以生成非常自然的商品介绍，内容来自预训练模型，还有图片、知识图谱和用户的评价，」周伯文表示。

另一个例子是在论文《Self-Attention Guided Copy Mechanism for Abstractive Summarization》中，自注意力机制（self-attention）可以帮助我们在对话任务和文本摘要任务上，生成了更多更自然的语句。据了解，京东智联云在跨模态内容生成上已取得诸多成果，并正式应用到京东的业务流程中。目前京东智联云打造的智能写作产品，是基于商品图谱和语言模型构建的营销内容智能生成服务，在 2020 年京东 618 期间，已覆盖京东零售过半数的商品品类，创作出的导购素材，曝光点击率相较于人工撰写的内容高出 40%，让用户在大促高峰期间也享受到优质服务。

这样一些接近实用化的方向已经受到了 NLP 新范式的帮助。毫无疑问，使用预训练的模型现在可以生成很自然的文本以及对话。但目前的预训练模型还称不上完美，曼宁指出，我们还没法控制这些模型生成的内容。

二、超越图灵测试的 AI 新基准

若想实现更好的人工智能，我们必须拥有完美的评测基准（Benchmark），几十年以来我们一直将图灵测试作为「真正人工智能」的测试标准。但图灵测试是以 AI 模仿人类，试图「欺骗」测试者进行无特定内容对话的形式来进行的。对于研究者来说，这个过程一直存在难以量化的问题。

在 NLP 技术发展多年后的今天，「我们会不会出现可以代替图灵测试的新基准呢?」周伯文在对话中提出了这个问题，「过去的几十年中，图灵测试一直是基准，但是在日常研究中，它让我们的研究目标变得明确，对结果推动又没有太多直接的帮助。」

「这个问题很有趣，也很难回答，」曼宁表示。「我同意这个看法——图灵测试不是非常清楚的基准。某种程度上我们需要找一个另外的方法，标量真正的理解、真正的持续对话。但我一时没法给出完美的答案。」

不过周伯文有一个「稍显疯狂」的主意，有关最近正火的直播带货：热门主播几个小时可以带几千万元的货。这种互动形式看起来非常吸引人，究其根本，它是一个实时的、富有交互性的方式。在这里播主和观众用弹幕和语音实时交流，这似乎为对话型 AI 提出了更多的要求。

原本的图灵测试，不会预先指出被测试者的身份，通过评判相似性去界定智能化水平；那么，我们是不是可以直接公开使用两个对话型 AI 做直播带货，通过统计以每小时能卖出多少商品的可量化指标来对比哪个 AI 的对话更吸引人，从而评估对话型 AI 的智能化水平?

这样的话，所有评价指标都可以量化，形式也非常接近于真实世界。

「这是一个非常有趣的想法，可以带来非常清楚的评价指标，」曼宁表示。「直播对于我来说是一个很新鲜的概念，某种程度上来说，这是一个非常直接的评价方式。我不清楚是否完美，但它很有创意：一个人类销售想要成功，并不取决于对潜在消费者传递信息的完美平衡，有时还需要提出超出实际一点点的主张，更加强烈地表达自己的观点。」

周伯文表示，在未来几个月里，京东会对这个方向进行一些尝试和研究。

三、学术界如何在预训练时代引领前瞻性研究

今天的人工智能研究正凭借算力的增长而快速发展，随着模型体量的增加，学界研究者面临的挑战越来越大。对于研究者们来说，即使希望方法足够创新，也会在大会上宣讲论文时受到这样的挑战:「你使用的基准是最新的吗?」这意味着你不得不直面大量数据。

周伯文表示:「近来，我常被问到一个问题，在如今的云计算 + AI 时代，研究人员和学者如何跟上?」

据了解，2019 年底，京东整合云计算、人工智能、物联网业务资源，形成京东云与 AI 事业部，并于 3 月 5 日面向技术服务领域推出全新的「京东智联云」品牌。在刚刚过去的京东 618，京东智联云提供了全面、稳定、安全、可信赖的技术保障，成为京东 618 的技术基石，并秉持着「成为最值得信赖的智能技术提供者」的愿景，对外输出更多、更好、更融合、更场景化的技术与服务。

目前云服务在商业公司中的布局已日趋成熟。那么在斯坦福大学，教授们是怎样平衡增量创新与理论创新的?研究者们是如何使用算力的?

「近年来我们的工作方式有了很大变化。在 20 年前，大学里才有最大的超级计算机、最快的网络。但在最近这些年里，情况有了翻天覆地的变化——现在算力都在商业公司那里了。」曼宁说道。

如何解决算力不足的问题，每所大学都有不少思路，最直接的方式就是购买数量有限的，当前最顶配的 GPU，让很多博士生共用以满足 80% 时间的需求。「我想这是很多大学都在使用的方法，如果你的实验室里有 20 名博士生，这要比每人配置一台机器节省三倍成本，」曼宁表示，「现在我们构建起了小型集群，斯坦福 NLP 实验室有 15 名研究者，我们有大约 100 块 GPU。你看，这不是一个很大的数字。」

另一个思路就是和京东智联云这样的科技公司合作，在一些需要更多计算的研究中，斯坦福也在购买云端算力。

每年冬天，曼宁都会亲自为斯坦福 NLP 大课 CS224N 授课。这门课可以吸引 500 名学生，他们的作业都需要使用 CPU、GPU 来训练模型，而所有学生在课程期间的算力需求是大学负担不起的。因此，斯坦福接受业界的捐赠。

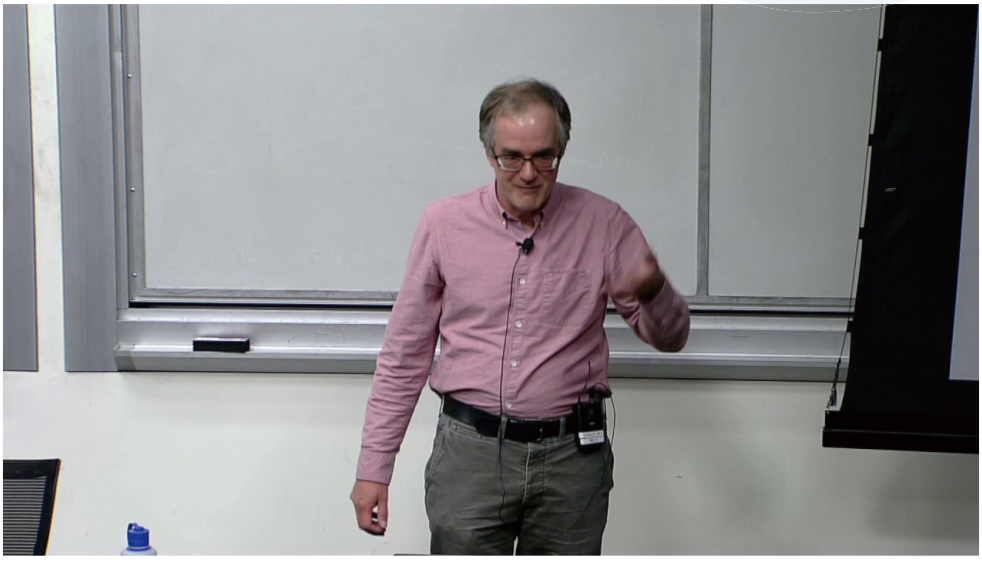

图 4：斯坦福的自然语言处理课程 CS224n 与计算机视觉课程 CS231n 齐名，是 AI 领域最具影响力的公开课程之一。

最后，研究方向也是个问题。「让模型越来越大可能在最近五年可以实现很大的进展，但在下个十年就不一定了。」曼宁说道，「我们现在可以构建出更大的模型，然后发出论文。但这个对于基础方向的研究没有什么帮助。未来 5–7 年里可能会出现一个窗口，最聪明的研究者可以用普通电脑和 GPU 构建出 SOTA 模型，打败大公司的巨大模型。」

「但未来也有可能不是这样，看看其他行业，如果你是个机械工程的 PhD，你肯定没法上来就盖世界最高的摩天大楼，如果你是个航空工程学生，你肯定不会试图造一架比波音还好的飞机。你需要做的是寻找新的想法。」

研究学者需要更加注重于寻找具有开创性的新想法，并提出原型。举个例子： 机器学习领域里的 Dropout，其实是在很小的数据集上首次实践的。

四、构建可信赖的 AI：可解释性和真实世界的鲁棒性

最近一段时间，周伯文曾在多个不同场合表达了对于可信赖的 AI（Trustworthy AI）的看法，并指出可信赖的 AI 将是智能经济未来 10 年的新原点。

目前有关可信赖的 AI 已经达成 6 个共识，包含公平、鲁棒性（技术的可用性）、价值对齐（技术提供者、使用者和产品应用方都认为产品带来价值）、可复制、可解释以及负责任。构建可信赖的 AI 一面是对技术的巨大挑战，一面是人文精神，无论是京东智能情感客服传递温暖、亦或京东物流设施传递信赖，都是对人类的社会责任与价值体现。

曼宁认为，人工智能学界目前在可解释性方面已经取得了一些进展。一方面是像 transformer 这样的预训练模型，注意力机制带来的好处——这些模型具有相当高的可解释性。

「我的一些学生发表过论文试图解读 BERT 的运作机制。现在，我们已能够对这些模型进行大量解码，并看到这些模型不仅是巨大的联想学习机器，而且它们实际上是在学习人类语言的结构，其解句子的语法结构，了解哪

些词是指同一实体，」曼宁说道。

因此，我们已经能够获得模型内部的可解释性，这意味着模型可以对其整体行为做出某种决定的原因做出一些解释。当然，这里还有很多工作要做，斯坦福研究者们正进行的工作希望就驱动模型决策的特征进行解释。

曼宁教授在 6 月份还以第一作者的形式发表了论文《Emergent linguistic structure in artificial neural networks trained by self-supervision》，其中写到预训练模型实际上可以学习语言结构，不需要任何监督。这解释了为什么大规模的模型是可行的。但是对于下一步如何更好的理解他们是怎么学习到的，这个目前还不太清楚，周伯文指出「这部分需要可信赖的 AI 来解决」。

这些发现非常令人兴奋。之前我们总是认为想让 AI 在某些任务上工作良好，需要是大型有监督模型。因此我们总是以大量资金、雇佣很多人进行数据标注开始。这是过去 20 年来的工作范式，人们也是通过这种形式在某些任务上让 NLP 模型达到接近人类水平的。

「如果下一代人工智能机器本质上和十年前一样，而考虑到训练的内容大幅增加，我们实际上是倒退了，而不是前进了。」曼宁说道。

「从技术角度来看，我将专注于尝试提高 NLP 的鲁棒性以及可解释性。在 NLP 领域中，如果了解 NLP 的结构，了解 NLP 的语义，将是人们构建可信任 AI 向前迈进的一大步，」周伯文表示，「如何预测下一个单词的过程对于人们来说还是一个黑箱。另一个方向是可扩展性，当我们从一个任务转移到另一个任务时，模型需要迁移得足够好。无论如何，可信赖的 AI 非常重要。如果我们可以在这个领域取得更大的进步，AI 市场和 AI 应用将变得越来越大、越来越多，并且适应性也将大大提高。因此，这将是我们长期关注的重点。」

图灵奖得主 Manuel Blum & 卡内基梅隆教授 Lenore Blum：迈向有意识的 AI——受神经科学启发的计算机架构

整理：智源社区 沈磊贤

在第二届北京智源大会全体大会中，卡耐基梅隆大学计算机科学系教授，图灵奖得主 Manuel Blum 及其夫人 Lenore Blum 教授做了题为《Towards a Conscious AI: A Computer Architecture inspired by Neuroscience》的报告。

Manuel Blum 教授是 1995 年图灵奖获得者、美国科学院院士，世界上理论计算机学大师，密码系统和程序检验先驱，计算复杂性理论的主要奠基人之一。Lenore Blum 是卡内基梅隆大学计算机科学系杰出教授，曾任美国国家数学研究所（MSRI）副所长。本次报告主要关注有意识的人工智能，即一种受到认知神经科学启发的计算机结构。本次演讲分为两个部分，分别由 Lenore 教授和 Manuel 教授介绍。

一、Lenore Blum 教授演讲

Lenore 教授首先从神经科学角度解释了意识（Consciousness）的含义。意识是人类为了保持清醒或保持睡眠，而不是无止尽睡眠而需要付出注意力的所有事情。意识是人的所见所闻所感所嗅，是喜怒哀乐，更多的也是自己内心的声音。

1.1 意识的研究历程

在大约三十年前，关于意识的科学研究对于任何受人尊敬的科学家而言基本上都是禁忌。直到 1988 年，知名认知科学家伯纳德·巴尔斯（Bernard Baars）发表了一篇有关意识认知理论的论文，成为了后来学术界统一认可的模板框架。1990 年，随着 fMRI 成像技术的出现，认知科学家们能够在人们进行认知活动的同时对大脑进行 90 次观察。在 1995 年，诺贝尔奖获得者弗朗西斯·克里克（Francis crick）在其著作《The Astonishing Hypothesis》中提到：人的全部意识活动都只不过是一大群神经细胞及其分子的集体行为，只要能找到意识的神经相关物，就能够认识意识（包括别人的意识）。在千禧年间，关于意识的思考和研究学者们一直在提出不同的研究方法。

从古至今，意识其实更多地得到了哲学家们的注意，如孔子、柏拉图、Dennett 及 Chalmers 等人，他们都是哲学领域对意识提出重要看法和观点的著名学者。现如今世界上有两个占据主流地位的哲学家：Daniel Dennett 和 David Chalmers。他们在意识形态层面分别代表着哲学的两个派系，Daniel Dennett 被称为功能主义者，而 David Chalmers 被称为现象主义者。

Blum 团队更倾向于接受 Dennett 的思想。Dennett 认为意识就如幻想一般，对于我们所处的某个场景而言，我们可能感官上觉得看到了眼前的整个场景，但是对于任何人而言看到的都只是实际的一小部分。关于整体的错觉是由部分多个扫视的信息整合结果所导致的。我们以一种不可思议的行为来想象整个整体。正如我们在计算模型（我们称为有意识的图灵模型）中应该看到的那样，整体的幻觉表现在不断涌出冒泡的小切片中，这些小切片在全局范围内广播辐射从而导致意识流的产生。

除了哲学领域的意识研究方法，还有很多其他领域的意识研究方法，比如关于心理学的研究方法，代表人物有 William James 等； 关于神经相关的研究方法，代表人物有 Crick 和 Christof Koch 等； 以及关于意识度量的研究方法，代表人物有 Guilio Tononi 等。在 2011 年，卡内基梅隆大学的 Helen Newell、Simon Reddy 以及 Anderson 教授提出一种建筑性结构认知的学说来研究认知本身； 受到上述几位学者的影响，Baars、S.Dehaene 以及 A.Baddeley 等人提出了一种建筑性结构意识的学说用于研究意识本身； 而 Blum 团队采取了和 Baars 等人相似的观点和看法，并从计算机科学理论角度入手，提出意识本质上是一种建构。Blum 团队之所以提出意识本质上是一种建构的观点，其实更多的是与数学以及计算机理论密切相关的。

1.2 何以构建有意识的人工智能

针对这个问题，Blum 团队认为需要从以下三个方面入手。

- 首先需要了解意识究竟是什么
- 然后需要构建一个简单的数学模型来帮助科学家更好的认识意识
- 一个重要的任务是需要对组块给出明确的定义

在 1950 年，心理学家 George Miller 提出了神奇的数字 7 +/ −2 的理论，George 认为组块本质上是在短期记忆中相互竞争的极少部分生成的信息，这些组块能在短时间内保留在短期记忆里。组块是极少一部分信息，它可能是一个单词，也可能是一个数字，还可能是一首诗歌，大多数对组块的定义都是非正式的，而 Blum 团队的目标是对组块给出一个正式且合理的定义，具体包括以下层面:

- 需要合理区分开模拟和真实经验
- 需要意识的属性能够出现，而不是被编程
- 实际任务不是在寻找关于大脑的模型，也不是在寻找学习或认知的模型
- 目标是寻找一种意识模型
- 寻求尽可能的简单，而非复杂性
- 希望组建蓝图程序模块用于制造类似意识的机器

1.3 意识研究的简单问题与复杂问题

在 1995 年，著名哲学家 David Chalmers 提出了关于意识研究的简单问题以及复杂问题。

- 意识的简单问题

意识的简单问题是那些似乎直接受认知科学标准方法影响的问题，通过计算机或神经机制来解释现象。这种更接近于研究功能性意识或者获得意识。

- 意识的复杂问题

对于意识研究的真正困难的问题是关于真实体验。当我们思考或者感知到信息处理的漩涡时，其中存在着主观方面的影响。这种往往称为现象意识，常常与可感受的特性息息相关，也就是“这个东西感觉如何”。举个例子，比如看到红色，那么现象意识的研究将集中在你看到红色后会产生什么样的感觉，也就是红色带给人的可感受特性。Blum 团队在 David Chalmers 关于意识研究的简单问题和复杂问题的基础之上，结合了计算机理论和数学，作出了重新的解读。意识的简单问题即让机器人能够仿真感觉，如悲伤或欢乐。意识的复杂问题即让机器人能切

身体会到情绪，如悲伤或者欢乐。

1.4 仿真和真实体验的区别

要想理解意识研究的两个问题之间的区别，就必须要理解仿真和真实体验之间的区别。Lenore 教授以 Sophia 机器人为例讲解了仿真为何区别于体验。Sophia 机器人可以很好的仿真人类的情绪，如喜怒哀乐等。然而尽管 Sophia 的表达能较好的仿真欢乐情绪，其本身却依然存在的严重的快感缺失问题。

快感缺失就是一种无法体验真实欢乐情绪的问题。对于人类而言，这是享乐功能的多种缺陷。它通常是由于伏隔核和腹侧被盖区域的损坏引起的。

1.5 Blum 团队的意识观

Blum 团队的意识观是： 意识本身是所有适当组织的计算系统的一种本质属性，无论该系统是由肉和血还是由金属和硅构成的。Blum 团队研究的主题是这些计算系统的组织架构是真正让这些系统具备意识的核心原因。许多认知神经科学家都一致赞同意识本身与大脑的组织架构密切相关。Blum 团队研究的架构以非常高的抽象水平来解释大脑，这一水平远高于神经元。

Blum 团队对机器意识给出正式定义，这种有意识的机器被称作有意识的图灵机，或者意识人工智能。Blum 团队在该模型中定义意识，然后指出该模型中的意识属性。这种组织架构受到了图灵的计算机模型的影响，图灵模型虽然结构简单，但是对理解计算机的计算原理具有重要的影响和意义。

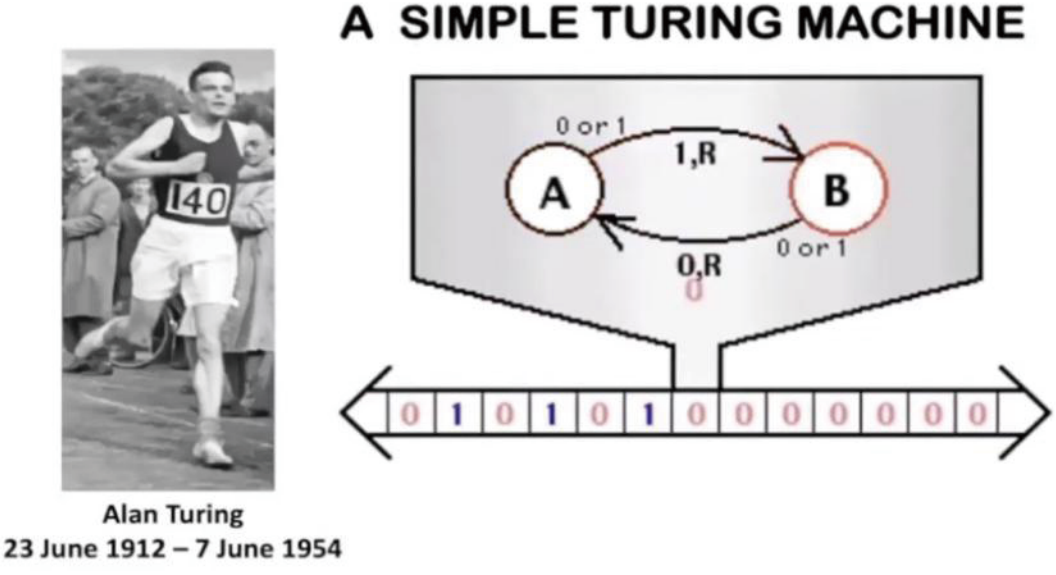

图 1 ： 图灵机示例

一种简单的图灵机如上所示。这种图灵机有两个状态。状态 A 有两个数值可以选择分别为 0 和 1，可以输入 1 转换到状态 B。状态 B 同样有两个数值可以选择分别为 0 和 1，可以输入 0 实现到状态 A 的转换。由此可以通过输入某个数值实现不同状态之间的相互转换。这样的图灵机由于组织架构过于简单，因此只能实现一些简单的功能，没有办法实现太过复杂的计算功能。

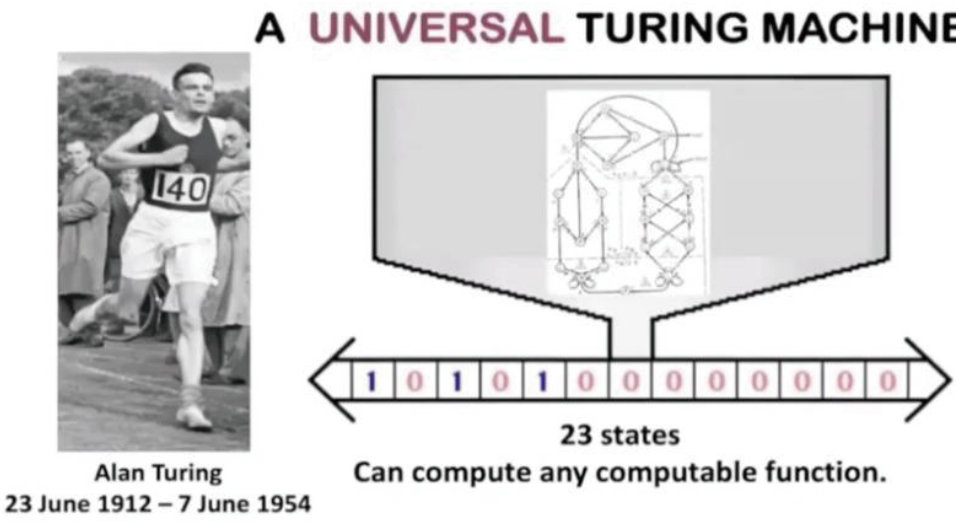

图 2：23 个状态的图灵机

然而对于万能图灵机而言，如上图所示的这样一个 23 个状态的图灵机可以实现大多数计算功能。这个图灵机是由 Morgan Ministry 设计，按照这样的结构设计，图灵机可以计算任何可计算的功能。因此任何的万能图灵机都能实现所有的计算功能，换句话说，如果想要在云服务器或者电路计算机上计算某个函数。那么就一定可以在万能图灵机上实现计算。为此可以将重点集中在图灵机上而不是云服务器上。受到图灵机的启发，Blum 团队对于组织架构的目标是设计更简单的结构，而非更复杂的结构。

1.6 Blum 团队的意识组织架构

Blum 团队研究的架构将认知神经科学家 Bernard Baars 的剧院模型或全局工作空间模型形式化。Baars 教授通过戏剧类比描述意识，Baars 学说中的一个重要的类比就是意识是演员在工作或短期记忆阶段表演的活动。脑内自言自语的演员经常在舞台上工作表演，可以考虑把演讲者作为一个正在颅内自言自语的演员，演员的表现受到众多坐在黑暗之中观众的观察，这些观众拥有大量长期记忆中的无意识单元处理器。

举个例子。比如当你去一个派对看到一个人似曾相识，但是就是记不得这个人究竟叫什么。然而回到家大概半个小时左右，这个人的名字会突然从脑中或者意识中蹦出来。这个过程到底发生了什么呢？可能你会回忆起你和他的初次相遇，可能是在一节计算机科学理论课堂上，然后这个记忆将会从短期记忆单元传播至大脑的每一个处理单元。其中一个处理单元可能会回忆起这个人在机器学习课堂中分享过一些关于意识的研究，这个短期记忆将会接着广播到其他所有处理单元。另一个处理单元可能会回忆起这个人的名字是 T 开头的吗，从而继续广播。可能大概半个小时之后，这个人的所有信息就会到达有意识短期记忆模块。无意识的长期记忆单元通过不同的思考，搜索最终确定最后的结果如何。然而意识本身却不知道这个人的名字究竟是如何找到的。

由此可见有意识的自我并不参与无意识自我的运作。还有一个数学史上著名的例子可以作为示例。Henri Poincare 是一位著名的数学家，他致力于解决很多复杂的数学问题。有一次他在路上突然解决了一个数学问题，后来在接受采访时说，当他即将上车时，突然有个用来定义 Fuchsian 函数的变换与非欧几里得几何变换相同的想法蹦了出来，以前似乎从来没有类似的想法基础。这两个领域 Henri 教授都非常熟悉，然而他之前从来没有考虑过这两个领域之间的联系。然而突然之间跃入他脑海中的是这两个其实本质上属于同一个领域。

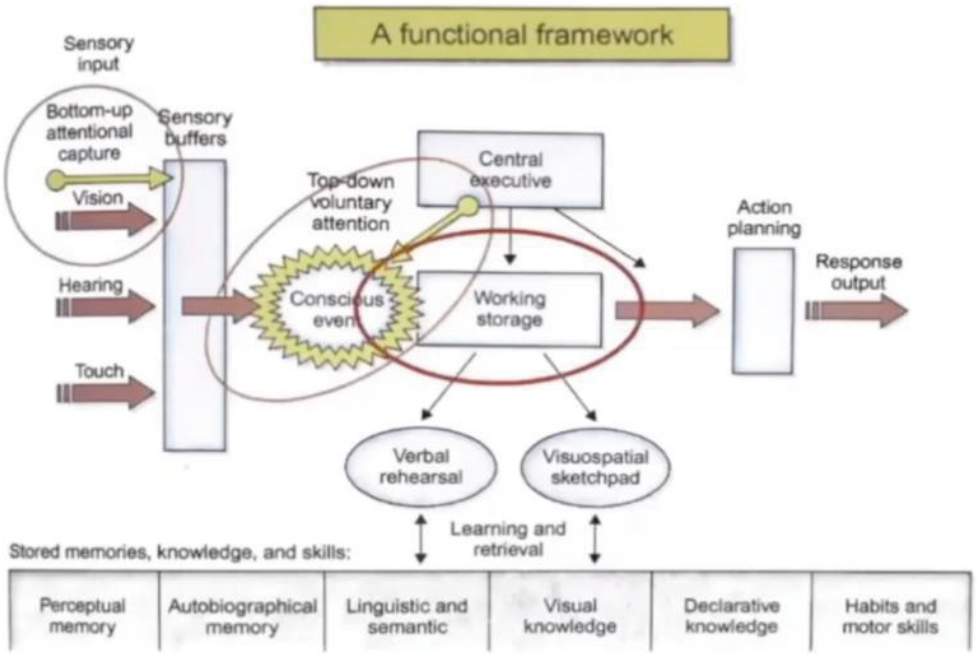

图 3：Baars 的意识模型结构

Baars 的意识模型结构如上所示。在正中心位置是整个模型的工作存储，也就是存储着短期记忆。在模型的底部有许多长期记忆处理器，可以看作是听众当模型得到了来自外界的输入，进入了视觉，听觉和触觉单元，这将进入工作存储到有意识的短期记忆单元中。在工作存储单元右侧的是从工作存储到实际使用的外部输出。模型中还有一个中央执行官，可以看作是舞台经理。组织架构中还有两个重要的功能模块，语言排练模块和视觉空间画板模块。要想摆脱中央执行官，就必须要摆脱语言演练和视觉空间画板模块，因为实际上可以通过利用长长期记忆处理单元来实现这些功能。

1.7　Blum 团队的有意识图灵机结构

有意识图灵机首先需要一个非常小的短期记忆单元，这是一个极小的可供读写的记忆单元，也可以理解成一个舞台，这个舞台一次性只能够承载一片组块信息。可以想像成这个舞台上有一个表演者在承担这一片组块的信息。同样的，舞台剧场还需要观众，这些观众可以看作是大量的长期记忆处理单元，他们大多数都是平行的处理器，将这些处理器加入进来，整个流程就将开始运转，而本质上这些处理器事先是没有相互联结起来的。

中央处理单元是由很多无意识的处理器共同承担实现的，可以发现最终可以去掉整个中央处理单元，首先将外部输入通过感知器输入到有意识的图灵机模型中，然后可以在一分钟内看到整个流程是如何运行从而得出合理的输出。当一个组块进入到短期记忆单元中，它将立刻通过一个向下的树广播到所有的长期记忆处理器。因此对于一个人的名字而言，首先会广播到所有的长期记忆单元处理器，这些都是无意识的处理器，然后信息将会中断，这便是向下传播树的快速广播机制。同样的还有一个向上的树，这个向上的树是关于长期记忆单元处理器的机制，主要承担着获取舞台上的组块信息。每一个组块实际上都在树的节点中相互竞争，从而决定最终哪一个组块信息被快速传播和接受。

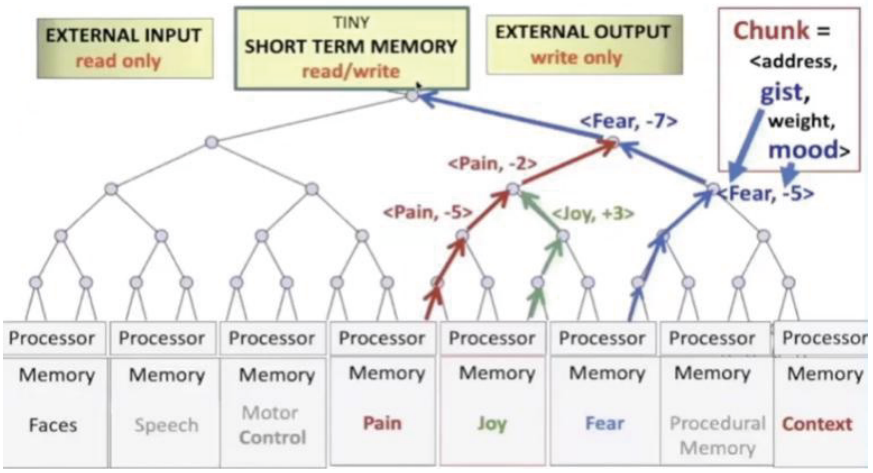

图 4：有意识图灵机的向上传播树状图

如上图所示的向上传播树状图，每一个组块都是一个四元组，包含着处理器的地址、主题、权重、情绪等信息，处理器的地址信息是 7 位数的数字，而主题则是一些关于问题描述的简短信息。在组块中，处理器将会对信息的重要程度进行打分，并且判断其是消极还是积极的。对于不同的情绪还有权重来衡量情绪的不同程度。情绪及其附属权重将会随着树形图的传播而不断改变。如上图所示的痛苦传感器，欢乐传感器以及恐惧传感器经由树形图向上传播，根据树形图节点的竞争机制，最终输出恐惧传感器极其附属权重至可读写的短期记忆单元。

因此整个机器运行的动态机制如下： 首先从外部世界获取输入从而传送至传感器中，例如眼睛或者耳朵。然后这些信息将会直接输入到长期记忆处理单元中，而不会经过短期记忆处理单元。然后所有处理器将会对组块中的信息进行广播，通过树形图的广播机制得到合理的组块信息输入到短期记忆模块单元中。比如说图里的恐惧处理器，尤其是当我们看到一些恐怖事情的时候，恐惧处理器对应的组块信息将会进入短期记忆处理单元。此时相关的组块信息已经广播到所有的长期记忆处理中，这时演讲处理器看到了恐惧相关的处理信息，它将会尖叫从而激发我们的大喊尖叫机制，这也是尖叫如何产生的全部流程。

实际上不同的处理器之间是可以相互交流的。比如在之前提到的恐惧处理器的例子中，演讲处理器以及恐惧处理器其实是相互交流建立起联系的，它们之间的相互联系确保了最终嘴唇发出尖叫的结果。在长期记忆处理器中，两个处理器如果产生应答，那么这两个处理器将会建立联系。并且如果处理器之间的应答密切相关，那么处理器之间的联系将会加强并且最终得到强有力的连接。

由于不同处理器之间建立的连接，最终将使得有意识的处理过程变为无意识过程。同样的，由于处理器之间连接，将会加强向上传播树形图中的拓扑结构。

1.8 有意识图灵机中意识的定义

在短期记忆单元模块中的组块信息其实是有意识的图灵机中的意识内容。在有意识的图灵机中，意识其实本质上就是图灵机意识内容中所有长期记忆处理器模块单元产生的意识或者反应。因此对于组块信息（也可以称为有意识的内容），通过广播到所有的处理器中，从而结合短期记忆单元中的内容实现对有意识的图灵机意识唤醒的过程。长时间持续的组块信息竞争进入短期记忆模块然后广播到所有长期记忆处理器单元能够创造出一连串的意识流。

1.9　7 元组的图灵机模型

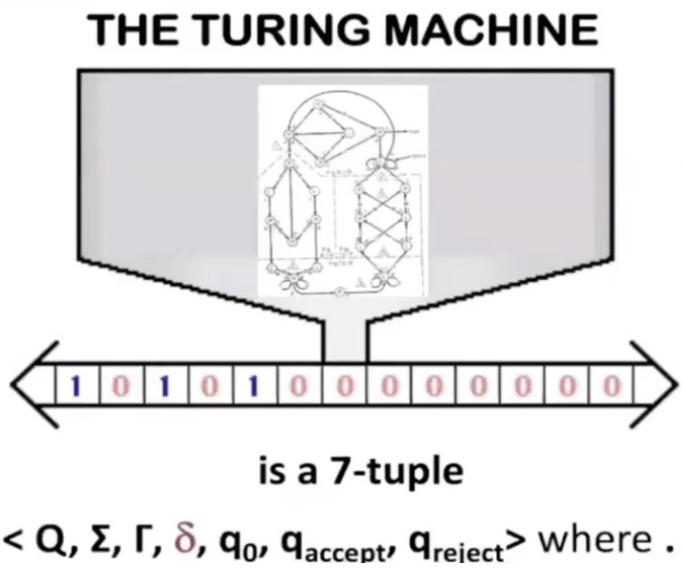

图 5：7 元组的图灵机模型

如上图所示的图灵机包含着很多复杂的数学原理，然而这并不在本次演讲讨论的范畴之内。上述所示的图灵机是一个 7 元组族的图灵机，其中包含着一系列的状态，通过给定初始状态，更新输入将会得到不同的输出和不同时刻的状态。事实上图灵机给出如下所示的定义，只有能被图灵机计算的函数才能算作具有可计算性质。对于有意识的图灵机而言，包含短期记忆单元、长期记忆处理器单元、向下传播树形图、向上传播树形图、连接、输入和输出七个元素。因此对于有意识图灵机中的意识流本质上是由短期记忆单元意识内容附属的长期记忆处理器单元产生的反应。

1.10　盲视反应示例

盲视是一个能够合理解释 Blum 团队关于意识定义的重要现象。盲视其实就是人们自认为看不见然而视觉处理器仍能正常工作的现象。

图 6：盲视反应示例

如上图所示一个男人坐在房间的右边角落，当他被要求走到另一个房间（可能另一个房间有食物）的时候，他可能会出现看不见另一个房间入口的现象，因为这个房间入口在他的视觉盲区。然而当另一个房间响起歌声的时候，这个男人将能径直走向房间的入口并且能完美躲避所有障碍。这其中发生的一切可以用图灵机来解释。

首先有一个来自外部世界的输入。当某个人告诉她需要去往另一扇门的时候，外部输入信息将会产生，位于房间入口另一边的双眼获取的信息将会导致视觉处理器正常运行，从而将信息直接传导至行走传动器。然而关于短期记忆单元的连接被破坏或者压根不存在，这将导致无法连接到短期记忆单元，因此对于视觉将会产生无意识感知。

二、Manual 教授演讲

在报告的第二部分中，Manual Blum 讨论了产生意识的诱因，关于痛苦（Pain）的复杂问题和简单问题以及产生意识所需要和不需要的长期过程。

Manual Blum 教授延续 Lenore Blum 教授所讲的内容，即认知神经学家 Baars 和 Dehaene 已经了解了意识是什么，然而这两位的贡献鲜为人知。Manual 和 Lenore 所做的是将 Baars 和 Dehaene 的认知科学理论转换为数学表达，从而在此基础上推到大脑如何产生诸如疼痛和愉悦之类的感觉。

2.1 团队研究历史

Manual 回顾了他们在认知科学领域的研究历史。多年以来，人们一直想知道头脑中发生了什么。为了更好地解决这个问题，Manual 先后借助了 4 个模型。

第一个模型是在头脑中控制身体活动的小人（A homunculus steering us），这不是一个很好的模型，因为为了建立有效的模型，还必须了解小人的头脑内部发生了什么。

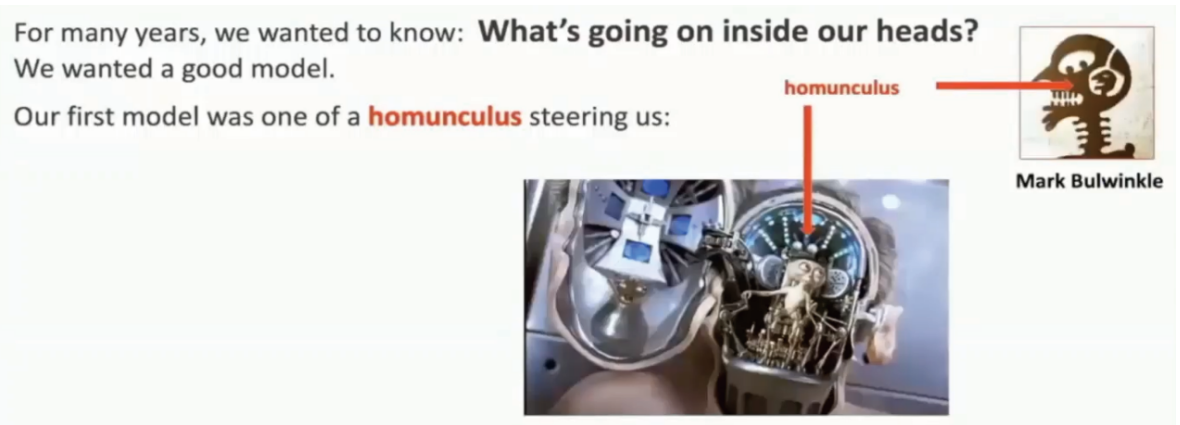

图 7：在头脑中控制身体活动的小人模型

第二个模型是柏拉图的有限状态机（Plato's finite state machine）。柏拉图认为，当我们出生时，就学会了世界上每一种语言。我们便从开始状态（Start State）出发，最终达到与周围人所说语言相同的状态，然而这个理论看起来有些牵强。

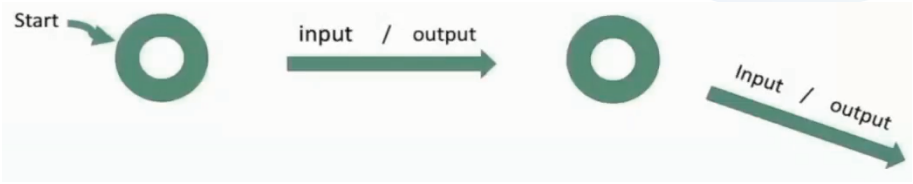

图 8：柏拉图的有限状态机模型

第三个模型是具有眼睛、耳朵、手臂和腿的图灵机。但这样的模型仍然没有帮助 Blum 洞察到意识的本质。

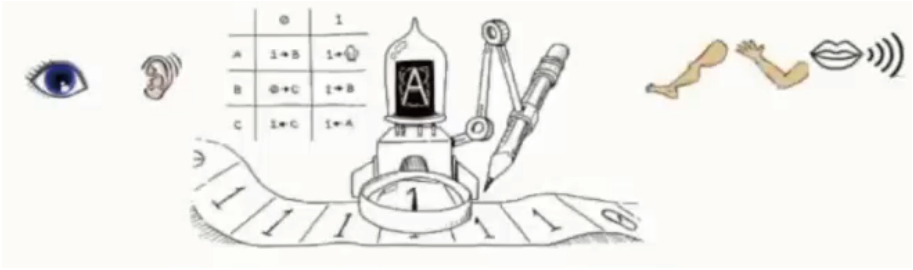

图 9：具有眼睛、耳朵、手臂和腿的图灵机模型

在尝试多个模型之后，Manual Blum 最终选择了 Baar 的全局工作空间模型 (Global Workspace Model，GWM)，即被称为意识剧场 (Theater of Consciousness)。而他和 Lenore 的贡献就在于在数学上精确地表达出全局工作空间模型。

下图即为全局工作空间模型，人通过视觉、触觉、听觉从外界获得信息，传入图中的工作储存区 (Working Storage)。工作储存区是短期记忆机制 (Short Term Memory)，经过工作储存区之后向外输出。在整个模型的最上面是中央处理区，在工作储存区的下面是语言排练区 (Verbal Rehearsal) 和视觉空间画板 (Visual Spatial Sketchpad)。GWM 模型将这两者放入一个长期存储器中，该存储器将所有进程存储在底部。其他一些更改输入不会进入到工作存储区的短期内存中，他们直接进入长期记忆，而输出也来自长期记忆。

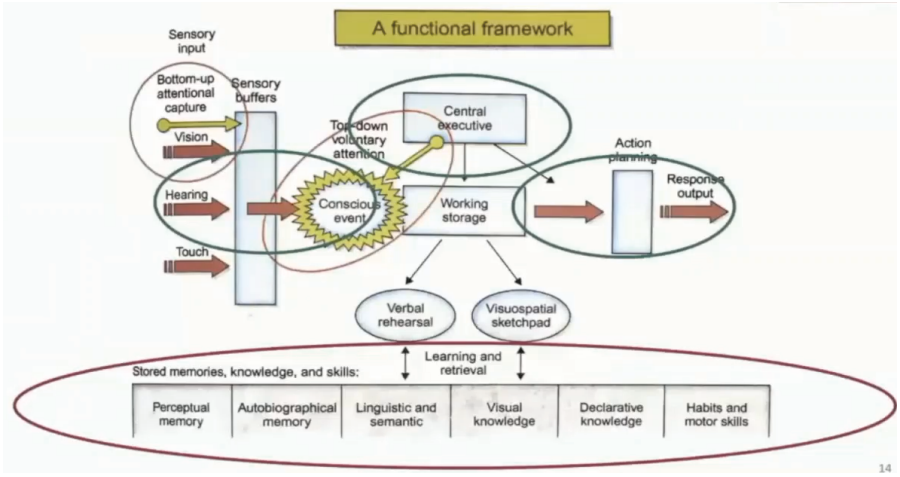

图 10：全局工作空间模型

下图为意识图灵机（Conscious Turing Machine）模型。蓝色箭头表示信息正在向下传输，灰色箭头表示信息正在向上传输。这些信息无处不在，都是块（Chunk）的形式存在。

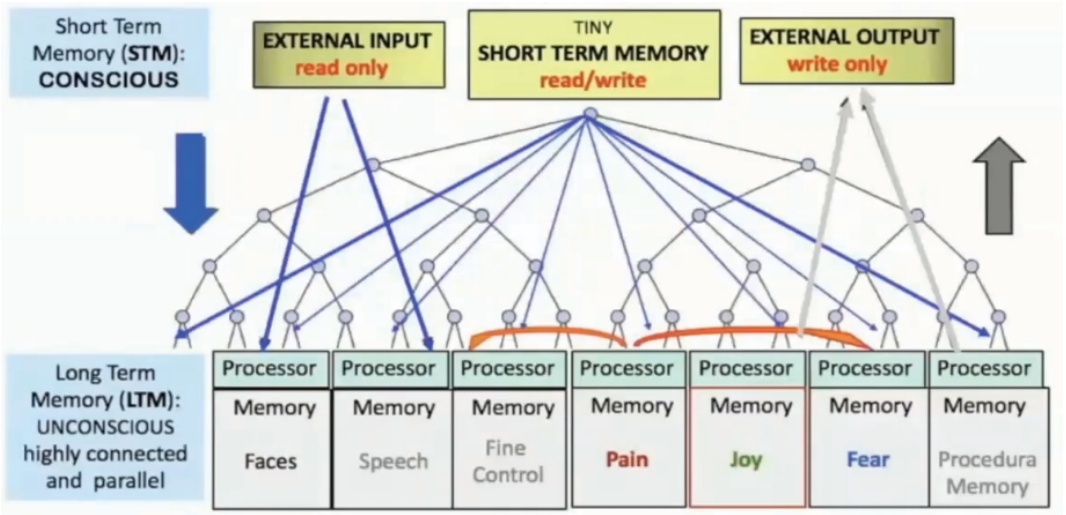

图 11：意识图灵机模型

块（Chunk）是一个 5 元组，包括原始地址（Originating Address）、要点（Gist）、权重（Weight）、强度（Intensity）和情态（Mood）。

图 12：块（chunk）5 元组

其中要点是一个最多包含两个短句的句子，描述了处理器拥有或给予的东西。权重是可以是正数或负数，分别表示好和坏。强度和情态，其中强度是权重的绝对值之和，而情态是权重的总和（无绝对值）。

此外模型中处理器之间存在链接。CTM 诞生时，处理器之间没有链接。但是如果两个处理器经常通过短期记忆相互交流，那么它们之间就会形成链接。从那时起，这两个处理器就可以彼此交谈，而无需经过短期记忆。这就是进入短期记忆的有意识活动变得无意识的方式。

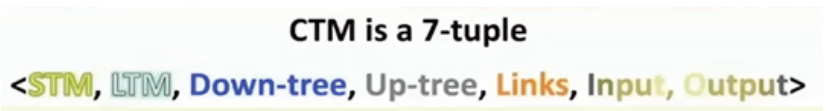

图 13：CTM 的 7 元组

2.2 意识的基本含义

Manual 接着介绍了与 CTM、意识有关的概念。正如 Lenore 所说，CTM 意识到了短期记忆（Short Term

Memory，以下简写为 STM）是什么，不多也不少。 CTM 中的意识定义为所有长期记忆机制（Long Term Memory，以下简写为 LTM）处理器对短期记忆（STM）中所储存内容的意识。在 CTM 中，所有困难的工作都是由处理器完成的，所有这些处理器都在 LTM 中无意识地运行。假设 CTM 代表着人类意识模型，人类仅意识到 LTM 处理器在 STM 中储存的要点。

那么这些要点是什么？主要有五种，如下图所示。第一种内在的声音，即很多人在自言自语，表达内心想法时听到的声音。第二种是内在的图像，性质与内在声音相似，比如梦境、地图等。除此以外还有感觉（Sensation）、感受（Feeling）等。

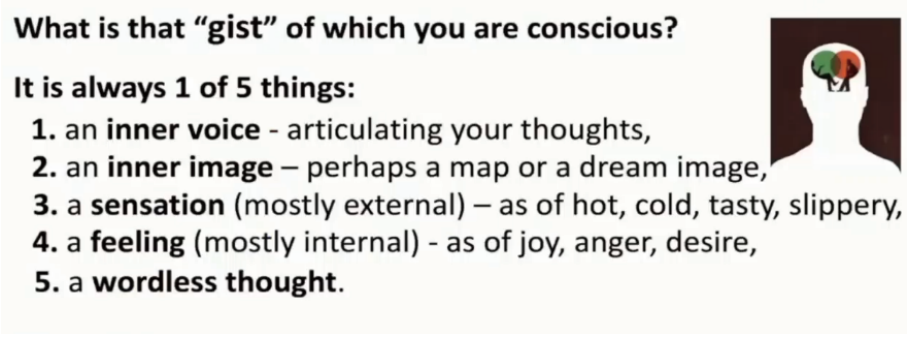

图 14：STM 中储存的五种要点

CTM 中的意识（Consciousness–in–the–CTM）被定义为对短期记忆内容的意识。 但是这个定义也留下了一个问题，即什么引起了意识?

Manual 给出了他的答案:

1. 所有 LTM 处理器都知道 STM 中的内容。因此，如果有任何一个处理器负责意识，那么该处理器就会知道 STM 中的内容。
2. 有一些 LTM 处理器专门负责意识。包括:
   2.1 内部语音处理器，将 STM 中的所有语音，Manual 称之为“ Brainish”，即大脑的语言，转换为类似耳朵从外界听到的内部语音。
   2.2 内部图像处理器，其机制与内部语音处理器类似。
3. 内在和外在的感觉都是从 STM 广播的。

CTM 在梦想和现实世界中所意识到的全部来自 STM。这些感觉从 STM 存储器中广播出来，并被所有 LTM 处理器接收，这使得很难将梦的内在形象与世界的外在形象区分开。
Manual 接着解释了意识对 CTM 的作用。通过 STM 的广播:

1. 意识使 LTM 处理器专注于生成对当前世界的最佳解释，并不断检查该解释。 它可以帮助 CTM 解决不一致的问题，例如在视觉上不一致的纳克方块（Necker Cube）。

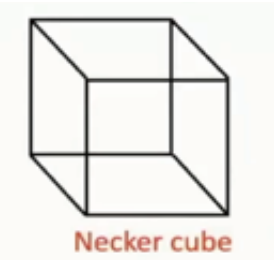

图 15：视觉上不一致的纳克方块

2. 意识始终关注 CTM 认为重要的任何问题，想法或思想。
3. 它意识使 CTM 能够解决意料之外的问题，譬如，使用其建议的所有工具处理复杂的世界。

CTM 是否会感到有意识？ Manual 相信答案是肯定的。但是，要如何证明 CTM 有意识呢？ Manual 表示不需要证明，他认为对意识的定义的合理性在于:

1. 它对意识的解释
2. CTM 能够解释的概念范围，
3. 这些解释在多大程度上与我们对自己的意识的直觉相吻合
4. CTM 对 CTM 创造者无法预料的情况的回应。

2.3 CTM 相关概念解释

以下是与 CTM 有关的一些概念解释。

1. CTM 意识到了什么?

答案是 STM 的内容。由于语音处理器同时将语音发送到舌头和 STM，因此舌头可以先于有意识的大脑开始动作。

2. CTM 中的块（Chunk）指的是什么?

块（chunk）本质上是指向位置和长期记忆（Long term memory, LTM）的指针。在长期记忆的最顶部是一个要点，要点有少量信息； 再加上权重，权重表示信息的重要性； 还有强度，权重的绝对值之和； 以及情态，权重之和，不含绝对值。

3. 为什么短期记忆（Short Term Memory）如此之小?

1. 这样可以确保所有 LTM 处理器都注意到同一意识思想。
2. 考虑另一个极端，假设 STM 包含每个处理器的要点，就会包含大量的块，那么它们就不可能专注于同一思想。
3. 以数学证明为例，当一个人完全理解数学证明时，他会觉得自己已经掌握了全部知识，但是实际上拥有的是要点，即证明的概念，带有“指针”指向其定义和引理。

4. LTM 处理器如何确定赋予其权重的符号?

对婴儿来说，饥饿和痛苦是消极的，食物和爱是积极的。这些正负的选择是内置的。在以后的生活中，如果要点的符号不明显，CTM 会将其设置为当前情态的符号。仅仅因为痛苦是消极的而快乐是积极的，并不意味着 CTM 会感到痛苦和快乐。

2.4 简单问题和复杂问题

那么，是什么让 CTM 感到痛苦和愉悦？这个问题带来了简单问题和复杂问题。正如 Lenore 所提到的，简单问题是：制作一个模拟诸如痛苦和喜悦之类的感觉的机器人，而复杂问题则是：制造一个真正体验诸如痛苦和享受之类的感觉的机器人。

Manual 讲述了“模拟”和“体验”二者的差别。在称为 Pain Asymbolia 的疾病中，患者知道自己有疼痛，但没有痛苦。这类患者知道疼痛是什么，并且自己仍然有疼痛，但是说没关系。机器人将我们的痛苦作为象征。我们知道如何使机器人看上去痛苦不堪，但我们不知道如何使机器人感受到痛苦、体验痛苦。这就是“模拟”和“体验”二者的差别。对于解释极端疼痛的经验，Blum 有三个建议：

- 广播 Broadcast

  极端痛苦是指占据整个短期记忆的 Actor 或 Chunk，它阻止所有其他 actor 进入 STM，并会广播痛苦消息，使得每个处理器都知道痛苦。在极端情况下，没有其他东西可以进入 STM。在通常会造成痛苦的情况下，Pain Asymbolia 患者会思考，而正常人群则不会。在《教育心理学评论》上，Smith and Ayres 发表了一篇名为“The impact of persistent pain and working memory and learning”的论文，他们写到“被鉴定为 6 个月或更长时间处于低水平疼痛的参与者在各种测试上的表现明显好于无疼痛者。”尽管 Broadcast 造成了一定的痛苦，但它们并不能解释韧带撕裂时突然产生的剧烈痛苦。

- 中断 Interrupts

  突然的剧烈疼痛，譬如一根手指触摸着燃烧的火炉，打断了所有无意识的处理器。中断使处理器立即将其工作放到堆栈上，从而迫使他们立即最大程度地注意中断原因，这种机制与广播相反，广播不向其发送信息处理器，而不会强迫他们将正在做的事情放在堆栈中。

- 恶性循环 Vicious Cycles

  痛苦导致恐惧，恐惧带来痛苦。这是保持注意力回到痛苦状态的许多可能周期之一。下图就是一个例子。

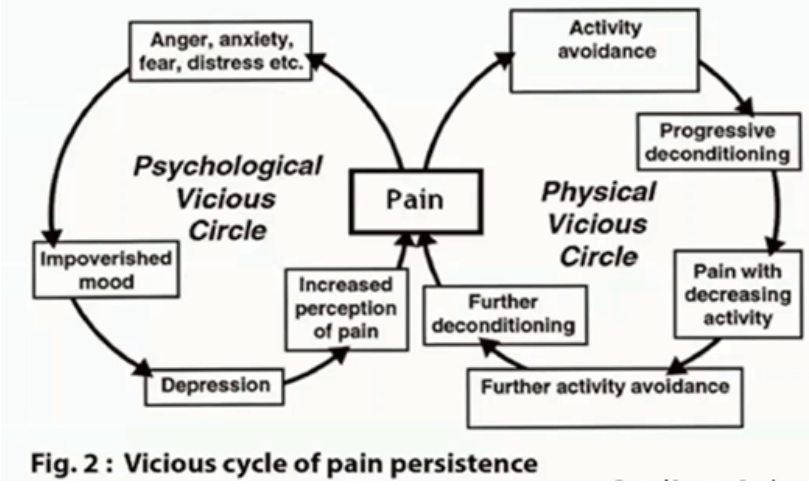

图 16：恶性循环示例

2.5 LTM 的相关问题

1. 意识不需要哪些 LTM 处理器?

大部分 LTM 处理器都是不需要的。Manual 举出了以下几个例子:

1. 1. 视觉和听觉处理器 Vision and hearing processor

Helen Keller 失去了视力和听力，但仍然有意识。

图 17：Helen Keller

1.2. 规划和教化处理器 Planning and civilizing processor

下图是 Phineas Gage。他在美国佛蒙特州铁路工地工作时发生意外，被铁棍穿透头颅，从颧骨下面进入，从眉骨上方出去，但却依然存活。铁棍损坏了他的额叶，这个部位被认为是 Planning and civilizing processor。事发之后，Gage 仍然可以说话、走路，但是性格逐渐变得暴躁、缺乏耐心，无法计划和安排自己的生活。但是一段时间后，他仍然在智利找到了工作： 驾驶 6 匹马的马车，从瓦尔帕莱索 (Valparaiso) 到圣地亚哥 (Santiago)，单程 124 公里。

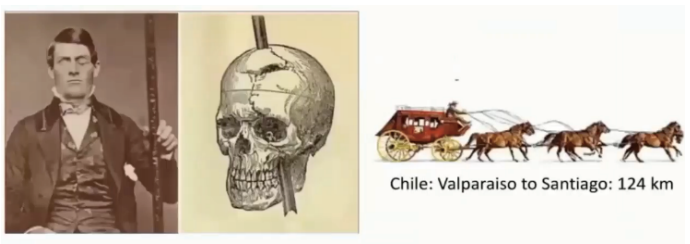

图 18：Phineas Gage

1.3. 危险警告处理器 Danger warning processors.

S.M.，有时也称为 SM-046，是一名特殊的大脑损伤类型的美国女性，她脑中的杏仁核发生了钙化，因此她什么都不怕。可以将蛇和蜘蛛放在她的手中，虽然她意识到了，但完全没有恐惧。

1.4. 陈述式记忆 Declarative memory

下图是 H.M. 他进行了手术摘除了海马体，因此丧失了留下永久记忆的能力。但他是有意识的。虽然他可以留下程序记忆 (Procedural Memory)，但他无法创造新的陈述性回忆。譬如，让他坐在打字机前面打字，他会说，不会打字。但倘若要求他尝试，他会发现实际上自己可以。他自己为自己可以打来的字迹感到非常惊讶，他不

记得会这样做了。

图 19：H.M.

大多数 LTM 处理器不需要意识。下图是 Jonathan Keleher，他出生时没有小脑，而大脑中一半以上的神经元在小脑里。下图是 Jonathan Keleher 的大脑和正常人大脑的一个对比。虽然没有小脑，但是 Jonathan Keleher 仍然可以说话、走路，甚至独自生活。

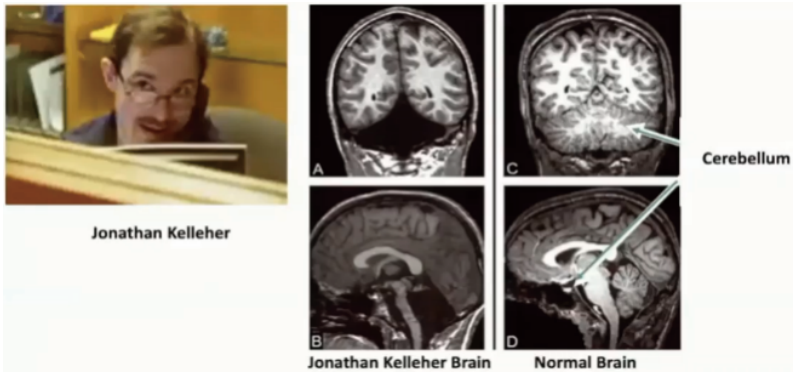

图 20：Jonathan Keleher

下图总结了上述提到的各种案例，此外，在图中右上角，可以看到 Dustin Hoffman 饰演的雨人，原型是 Kim Peek，他患有自闭症，脑部结构与常人不同（对比图见下图右下角），但是却有超强的记忆力。综上所述，大部分 LTM 处理器都是不需要意识的。

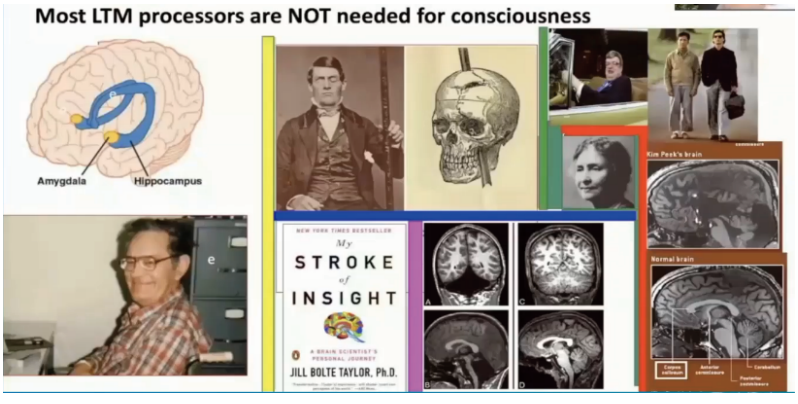

图 21：Kim Peek

2. 意识需要哪些 LTM 处理器?

2.1. 内部对话处理器 Inner-dialogue processor

这里并不是指内部语音处理器，因为动物可能没有讲话，但是动物有一些方法和方法可以进行计划，这是由某些内部对话处理器完成的，它可以是语音，也可以是图像。

2.2. 世界模型处理器 Model-of-the-World processors

借助 Model-of-the-World processors，可以将自身和外界、自身的各个部分区分开来。下图是一个婴儿，正在学习区分自己。他发现自己可以移动左腿。然后他向右看，发现他不能移动右腿。通过调动自己的双腿，他学会了将自己身体的各个部分区分开。

2.3. 宽泛的一般思考能力 Some broad general (minimal) ability to think

Manual 认为，宽泛的一般思考能力同样很重要。这其中就包括动机（Motivation），动机是我们行动的能量和动力。但研究者还没有完全理解这种思考能力的含义。毕竟，Alpha-zero 的表现非常出色，它绝对可以思考，但不具备这里所说的宽泛的思考能力。

如果失去了内部对话处理器和 Model-of-the-World processors，意识会发生什么？举例来说，神经科学家 Jill Bolte Taylor 就失去了大脑的整个左半部分，这部分大脑负责产生言语和其他事物。当不幸发生时，她意识到自己的工状态很糟糕。她在书中说，在她眼中外界和她自己是一体的，她无法区分自己和非自己。幸运的是，她仍然具有一定的思考能力，包括动力，精力和驱动力。虽然她在书中声称自己失去了很多意识，她仍然有一些意识。

3. 存在一种对意识的测试吗?

George Gallup 提出了一种对自我意识的镜像测试。测试的内容是：在动物的额头上做一些标记，然后把动物放在镜子前。在它看到了标记后，如果尝试将其从标记额头上移除，就通过了镜像测试。但是包含了存在自我意识的最低要求，即存在内在对话处理器，并对自身运动进行规划。

下图为一头通过了镜像测试的大象。可以看到它的额头上有一个它以前从未见过的“X”标记。它必须规划好象鼻的运动轨迹，使之刚好到“X”标记处，然后擦除掉这个标记。因此，它具有规划能力。它没有尝试从镜子中的大象上删除标记，而是尝试从其自身上删除标记，因此它也有 Model-of-the-World processors。同样地，它也具有宽泛的一般思考能力。还有哪些动物通过了镜检？如下图所示。

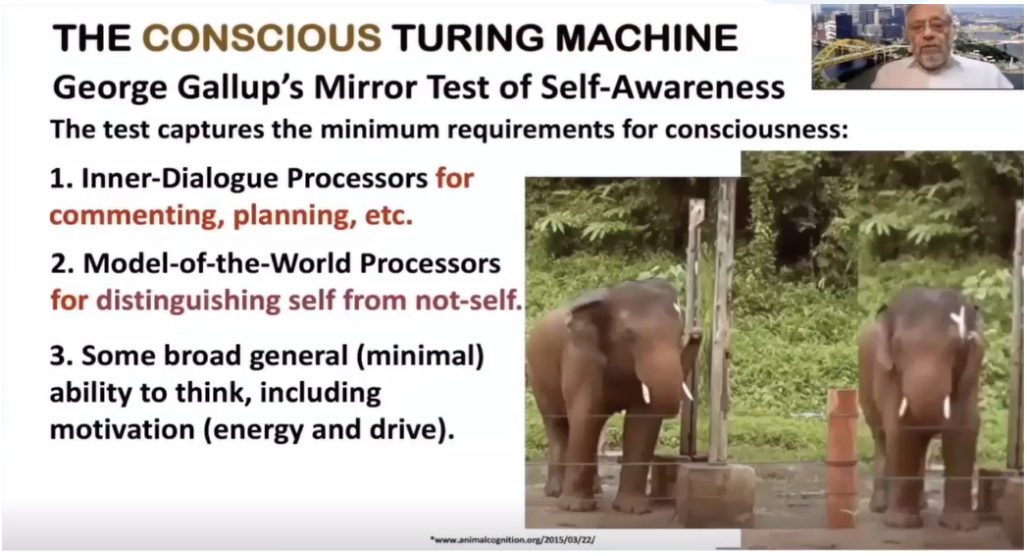

图 22：通过镜像测试的大象

狗不是视觉动物，但是使用嗅觉通过了测试； 黑猩猩确实通过了镜像测试； 而大猩猩失败，失败的原因很有趣，大猩猩不能在用眼睛直视另一只大猩猩，那意味着战争。因此，当大猩猩站在镜子前时，它会砸碎镜子中的大猩猩。 海豚通过了测试。

除此以外，对鱼类而言，动物学家 Temple Grandin 认为所有脊椎动物都有意识，但是不包括鱼。然而，有一条特别的鱼叫做清洁濑鱼（Cleaner Wrasse），如下图所示，可以在水族馆中买到。将这类鱼放在镜子面前时，它会做的就是像斗鱼一样尝试与镜中鱼进行抗争。 但是到了某个阶段，它认识到镜子里的鱼就是它自己。然后有趣的事情发生了，鱼在垂直向上飞向空中然后向下飞回，然后上下颠倒游泳。它通过这种手段检查自己，以确保镜中是其自身的图像。当它在下巴上看到一个标记时，它会尝试将其擦掉。因此这类清洁濑鱼也通过了镜检。

图 23：Cleaner Wrasse 镜检

对蚁类而言，有一个叫 Myrmica 的属，如下图所示，有三种，这三个物种均可以通过镜检： 给它们的头上标上有一点红色或黄色的油漆标记。它们在镜子上行走，看到油漆标记时，会走下镜子，尝试将其擦掉，然后再走上镜子检查。

图 24：Myrmica 镜检

4. 有意识的活动如何变为无意识?

在 CTM 模型中，链接在处理器之间生成，这样的链接减轻了 STM 的负担。一旦他们减轻了负担，那些有意识的活动就会变得无意识。

5. 解释梦的一些问题

存在梦境生成器和梦境接收器，即其中一个处理器成为梦生成器，其他处理器作为梦接收器来响应生成器。梦的产生很容易，它是一系列要点的组合。

6. 正念（Mindfulness）如何工作?

一些负责正念的 LTM 处理器进入 STM，它的作用是提高自身的权重，本质上是静默其他处理器，以防止其他大多数处理器将其块放入 STM，这就是正念，也是催眠的一种解释。