人工智能文集

评人工智能如何走向新阶段

（兼谈国内外AI跟帖评论）

陆首群

2024.3.5

国内外AI跟帖留言

（1281条~1300条）

第十六集

中国开源软件推进联盟
China OSS Promotion Union

评人工智能如何走向新阶段?

人工智能当前发展状况与未来发展路径

陆首群 2024.2.29

上世纪 80 年代至本世纪前 20 年，全球人工智能尚处于以机器学习/深度学习为主的弱人工智能发展阶段。

以机器学习/深度学习为主的弱人工智能的主要表现为：

计算机视觉、棋类智力游戏、算法算力、人脸识别、语音识别、 图像识别、机器人、机械手、无人机、自动驾驶和无人驾驶，自然语 言处理，机器翻译，物联网（IoT、AIoT、IIoT）,预测、训练蛋白质 三维结构（生命科学），基因医疗诊断，新药物（包括新一代抗生素），新材料等,2019 年英国研制吸气式高超音速“暴风雨”原型战斗机（自称全球首家六代机）是基于机器学习的科技成果。

同时，自此开始，人们一直探索向强人工智能的发展路径，即：①打破机器学习的黑盒子研发可解释性的人工智能，②基于异步脉冲神经网络的神经拟态计算系统，③从新知识工程出发，依托大规模语义网络（知识图谱）的支持，探索认知智能解决方案，④大模型（LLM）的研发，突破“推理”和“生成”环节，赋予机器生成自然语言，实现人机对话，并通向生成式人工智能/AIGC ⑤脑机接口的理论和实践。⑥探索研究下一代新通用人工智能（或机制主义人工智能）。

有关专家怀疑如前所述由人工智能不同学派提出的上述①②③条路径是互不相容，单打独斗，缺乏科学和哲学的全面统一理论基础，缺乏整体观，他们只是从不同侧面模拟人类心智（大脑），各自提出的发展路径均有片面性，但事实上①、②、③分别正在向生成式人工智能或通用人工智能（AGI）冲刺之中。有关专家建议从改革、融合、

统一的人工智能发展范式出发，探索第⑥条发展路征，即发展下一代新通用人工智能或机制主义人工智能。对于第④条路径，也有专家提出“三个天花板”存在不鲁棒、不可信、不可控问题。对于第⑥条少数专家正在探索理论体系，尚未提出具体的解决方案，至实现还有漫长的路要走。对于第⑤条，在人脑中植入芯片连接神经元并与脑外机器（计算机或机械手）连接起来，可凭人的意念利用人脑神经元（EEG，脑电图）来操控机器，对截肢患者用以控制取代残肢的机械手运动（顺利完成进食、取物和握手等动作），也可用意念打字，或用脑机接口技术来治疗中风、癫痫患者及治疗老年痴呆等疾病。

国内外 AI 跟帖留言（1281-1300）

目录

1281，研发鲁棒、精准的可解释机器学习/深度学习

陆首群，2022 年 4 月 11 日

可解释性机器学习已成为全球由弱人工智能转向强人工智能的热点，按大数据建立起来的机器学习/深度学习是一种强大的数据分析工具，它属于弱人工智能范畴。但机器学习/深度学习是有缺陷的，它本质上是黑盒子技术。其模型是不可理解、不可解释的，只有打破黑盒子实现可解释的机器学习，才能使之转化为强人工智能。

2018 年，AI 大师 Yoshua Bengio、Yann Lecun、GeoHrey HinTon 指出，深度学习本质上是一项暗箱技术或盲模型，其训练过程不可解释、不可理解、不可控、缺乏类人的推理能力，John Hopcroft 满怀信心要在 5 年内打破深度学习这个黑盒子。

2020 年 6 月 COPU 主办的《15 届开源中国开源世界高峰论坛》邀请 IBM 副总裁 Todd Moore 与会作“可信任人工智能（反欺诈、可解释、公平性）”的报告（这是全球发表可解释性机器学习报告最早者之一）。

在国内，2020 年 12 月沈向阳教授提出“拥抱开源，我们现在最重要的事情是要做可解释的人工智能”，2021 年 1 月，姚期智院士提出“机器学习算法缺乏可解释性，很多算法处于黑盒子状态，这项人工智能的技术瓶颈亟待突破”。

从此以后，COPU 曾发表全球 102 条研发可解释人工智能的论文和科技成果（可查阅 COPU 出版的《人工智能汇集》“可解释机器学习”P2-5，及 P6-P136）。

从 IBM AI 研究所和全球发表的可解释性机器学习的研究成果（迄今）来看，其运算程序普遍多少具有不确定性因素（不鲁棒），在作评估时也过于粗略，为此我们与 IBM 专家多轮讨论，如何改善可解释机器学习的鲁棒性和评估的精确度。

1282，如何改善可解释性人工智能（XAI）？

引自 2024 年 2 月 27 日 COPU 会议纪要

2019 年 IBM 人工智能研究所在全球首次发表了他们对“打破深度学习黑盒子研发可解释性人工智能”的报告，迄今为止 COPU 收集全球数百例研发可解释性人工智能的报告，由于 XAI 技术尚不完全成熟，在 XAI 运算程序中存在一些不确定性因素（不鲁棒），以及评估方法粗略，COPU 曾多次与 IBM 专家讨论如何改善 XAI，首先应核定 XAI 案例的运算程序：

1）核定采用下列哪些方法进入运算阶段（①可直接解释/内在解释，②事后解释，③全局/模型级可解释性，④局部/实例级可解释性）

2）选择什么工具？（如决策树、规划库、抉择表等）

3）如何捕捉特征？

4）如何建模？

5）如何推导算法？

6）如何进行评估？

研究如何运用 IBM 提出的可解释性工具套件，如解释性、健壮性、公平性工具套件分别为 AIX360[②]、ART[③]、AIF360[④]，研究 IBM 为解决 XAI 案例而设计的运算程序以及点评程序衔接的合理性、各程序展开的运算内容等？

1283，类脑计算机（神经拟态计算系统）何时能上市？

陆首群 2022.4.11

作者说明：

作者是在 2022.4.11 写的本文，至今 2024.4.2，过去 2 年了！神经拟态计算系统（人工智能）变化很快：①英特尔研制的原型机已进入了

试运行，②又一基于异步脉冲神经网络（芯片集成）的国产神经拟态计算系统的研究工作也开始取得了科研成果，召开了“2023 智算产品发布会”。作者拟在本文的基础上，补充增加若干新内容。

英特尔、曼彻斯特大学、浙江大学的、粤港澳脑智工程中心的研究团队分别于 2017、2018、2019 年研发“异步脉冲神经网络+神经拟态计算系统”（人工智能类脑计算机）取得了很大成绩，这就是类脑计算机，打破了传统计算机冯·诺依曼架构，分别于 2019、2018、2020 年相继制成类脑原型机（可查阅 COPU 出版的《人工智能汇集》“异步脉冲神经网络+神经拟态计算系统”P1-4）。

以英特尔的研发工作为例，自其制成类脑原型机至今已 5 年了，因何原因延迟其科研成果上市？

分析起来，主要原因受阻于其应用。

1）神经拟态计算系统的运算速度达 10^{14} 次/秒(浮点峰值运算速度)，比传统计算机快 1000 倍，仍落后于当前 E 级超级计算机的运算速度百亿亿次(10^{18} 次/秒)落后 10^{4} 次/秒，即 1 万倍，比我国最近研制成功的量子计算机(10^{21} 次/秒)更落后 1 千万倍。

只是其能耗比传统计算机省 10000 倍，这个性能比较突出。

正在探索中的应用特点还不突出，现知未来的应用前景为：①超算，②非结构化数据、实时应用场景，③多模态实时场景（如机器人、无人机需持续学习，自适应的场景）。

英特尔于 2020 年 3 月成立神经拟态研究社区（InRC）以开拓应用（特别是特殊应用），参加 InRc 的单位除英特尔外，还有 IBM、HP、MIT、普渡大学、斯坦福大学、联想、埃森哲、罗技、梅赛德斯奔驰、机器视觉传感器公司、罗格斯大学、戴尔施得理工大学等。我们计划与上述三家研发团队再次讨论神经拟态计算系统的发展情况。获英特

尔最近的信息证实：英特尔的类脑计算机原型机正在进行试运行。

2024.4.2 补充：

又一基于异步脉冲神经网络（芯片集成）的国产神经拟态计算系统召开技术发布会。2023.12.29 在横琴国际科技创新中心，由 2023 智算论坛及粤港澳脑智工程中心共同举办的 2023 智算产品发布会，以新一代类脑计算架构（LYRArc）和处理芯片（BPU）为技术核心的绿色类脑智能计算系统，在横琴国际科技创新中心发布：研制神经拟态计算系统（超 2 亿个神经元）。

1284，数字智能会取代生物智能吗？

杰弗里·辛顿（Geoffrey Hinton）在牛津大学公开演讲，2024.2.19

好的。我可能会让计算机科学和机器学习领域的人失望，因为我要做一次真正的公众演讲。

我将尝试解释什么是神经网络，什么是语言模型，以及为什么我认为它们有理解能力，这方面我研究了很多。

在后面，我将简要地谈论一些来自人工智能的威胁，我还会讲到数字神经网络和模拟神经网络之间的区别，以及为什么我认为这种区别很可怕。

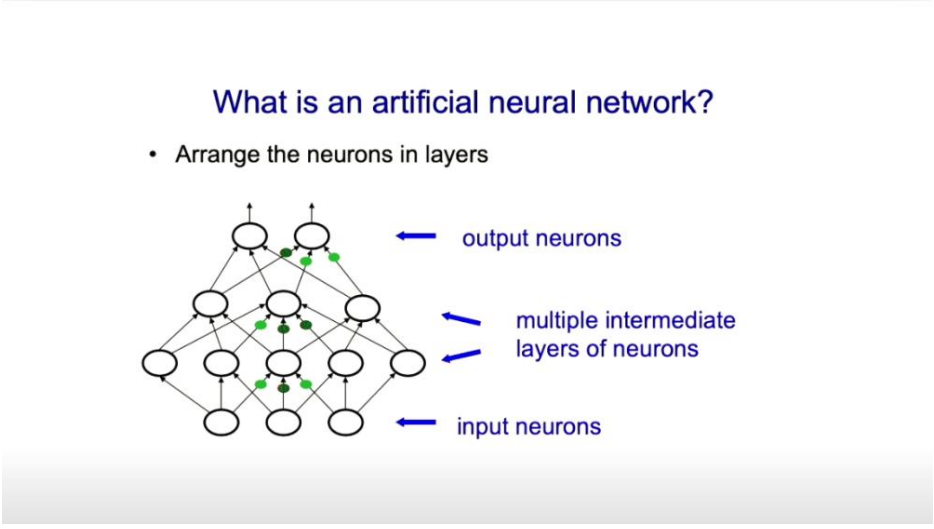
自 1950 年代以来，关于人工智能，有两种研究范式。

逻辑启发式方法认为智能的本质是推理，是通过使用符号规则来操作符号表达式完成的。

他们认为人工智能不要急着去“学习”，当我还是个学生的时候，有人告诉我不要研究学习，在我们理解了如何表示事物之后，学习就很简单了。

生物启发式方法则大不相同。它认为智能的本质是在神经网络中学习连接的强度，倒是不要着急去“推理”，在学习完成后，推理自然就来了。

现在我将解释什么是人工神经网络，懂的人可能会觉得这是小儿科。



简单的神经网络有输入神经元和输出神经元。输入神经元可能代表图像中像素的灰度值，输出神经元可能代表图像中物体的类别，比如狗或猫。

然后有中间层的神经元，有时被称为隐藏神经元，它们学会检测和识别这些事物相关的特征。

所以，如果你想识别一张鸟的图片，一种思考方式是，从一层特征探测器开始，它们能够探测到图像中各个位置、各种方向的小边缘。然后你可能会有一层神经元探测边缘的组合，像是在尖锐角度相遇的两条边缘，这可能是鸟嘴，也可能不是，或者是形成一个小圆圈的一些边缘。然后可能会有一层神经元探测到像是圆圈，以及相遇的两条边缘看起来像是鸟嘴，而且它们在正确的空间关系中，这可能就是鸟的头部。

最后，你可能会有一层输出神经元说，如果我找到鸟的头部、鸟的脚、鸟的翅膀，那么这很可能是一只鸟。

这些就是要学习的东西。现在，小红点（上图中深绿色点，编者注）和小绿点是连接上的权重，问题是谁来设定这些权重？

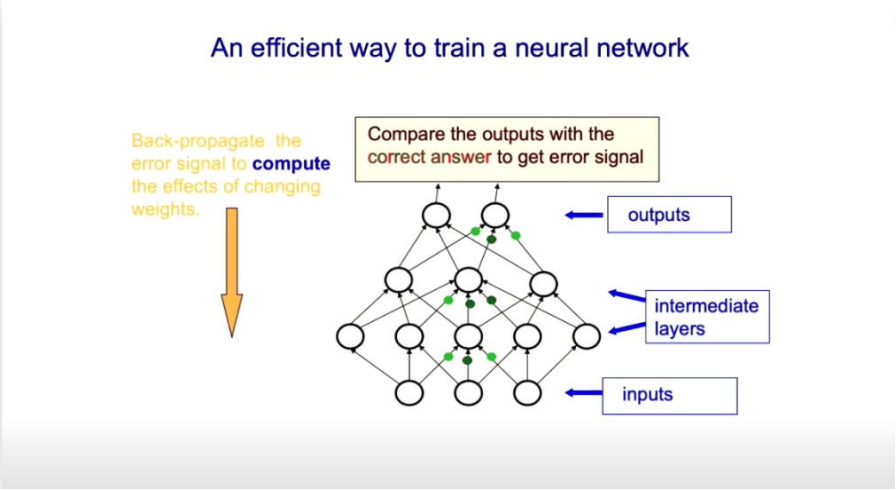
有一种做法显然是可行的，但显然需要很长时间：你的权重一开始是随机的，然后你随机挑选一个权重，比如一个红点，稍微改变它，看看网络是否运行得更好。

你必须在很多不同的情况下尝试，以真正评估它是否真的有所改善，看看将这个权重稍微增加一点或减少一点，是否会有所帮助。如果增加它使情况变得更糟，你就减少它，反之亦然。

这就是突变的方法，而这种方式在进化中是很合理的，因为从基因型到表现型的过程非常复杂，且充满了随机的外部事件。

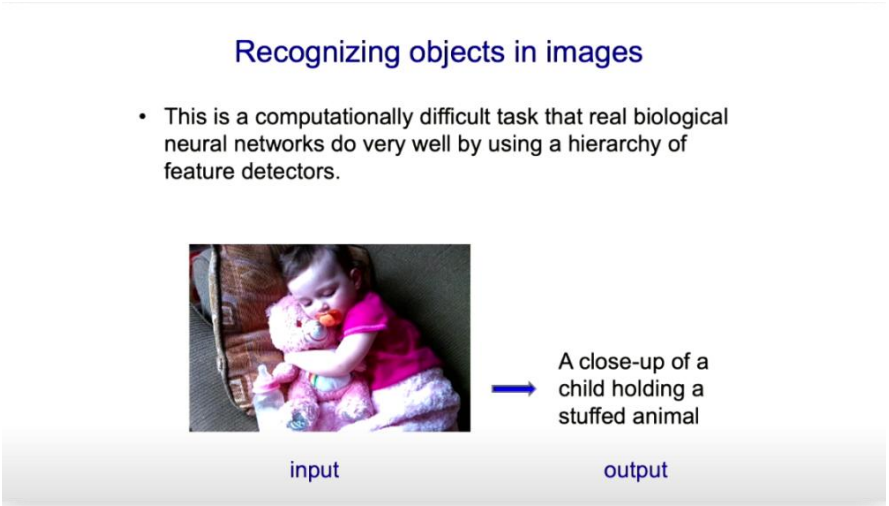
我们没有关于进化的模型，但对于神经网络而言就大不一样了，我们有模型描述神经网络中发生的这些复杂过程，我们知道在前向传播中会发生什么，我们不是通过测量，而是通过计算，来查看改变权重将如何影响事情。

有一种叫做反向传播的方法，你把信息发回给神经网络，这个信息是你得到的结果与你想要的结果之间的差异，此时，你要调整网络中的每个权重，不管是将其稍微减少还是增加一点，目的是为了让结果更接近你想要的，这就是反向传播算法。



你可以用微积分中的链式法则来做这个，这比变异方法有效得多，效率之比是网络中权重的数量。如果你的网络中有一万亿个权重，它的效率就高出一万亿倍。

神经网络经常被用于识别图像中的对象。现在，神经网络可以针对一个图片，产生一个对于图片的描述作为输出。



多年来，人们尝试用符号方法做到这一点，但连接近都没有，这是一个困难的任务。

我们知道生物系统是通过一系列层次化的特征探测器来实现这一功能的，因此对神经网络进行这样的训练是有意义的。

2012 年，我的两位学生伊利亚·苏茨克弗（Ilya Sutskever）和亚历克斯·克里泽夫斯基（Alex Krizhevsky），在我一点帮助下，展示了可以通过这种方式制作一个非常好的神经网络，在有一百万张训练图片时，可以识别一千种不同类型的对象。而在那之前，我们没有足够的训练图像。

伊利亚很有远见，他知道这个神经网络会在 ImageNet 竞赛中获胜。他是对的，他们赢得相当炸裂，他们的神经网络只有 16%的错误率，而最好的传统计算机视觉系统错误率超过了 25%。

然后，科学界发生了非常奇怪的事情。通常在科学界中，如果有两个竞争的学派，当你取得一点进展时，另一个学派会说你的成果是垃圾。但在这个案例中，由于差距足够大，使得最好的研究者吉滕德拉·马利克（Jitendra Malik）和安德鲁·齐斯沃曼（Andrew Zisserman）转换了他的研究方向来做这个，安德鲁·齐斯沃曼还给我发送邮件说这太神奇了。

然后有点恼人的是，他做得比我们还要好一点。

在语言处理方面，一些坚信符号主义人工智能的研究人员认为他们在语言处理方面应该表现出色，并且他们中的一些人在出版物中声称，神经网络的特征层级无法处理语言问题。很多语言学家也持这样的态度。

乔姆斯基（Noam Chomsky）曾说服他的追随者相信语言是天赋而非习得的。回顾起来，这种说法是完全荒谬的。如果你能说服人们相信显然是错误的事情，你就是让他们加入你的邪教。

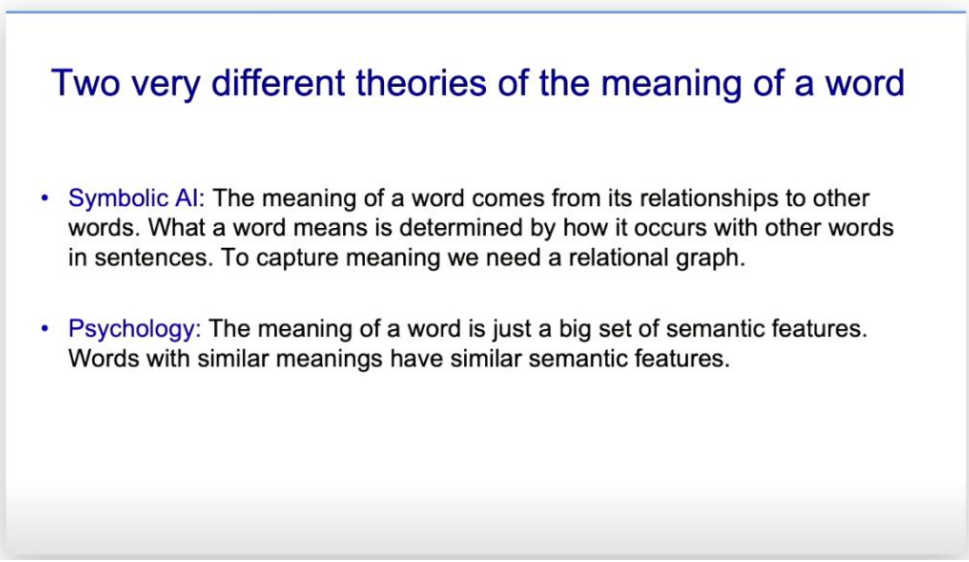
我认为乔姆斯基曾经做出了惊人的贡献，但他的时代已经过去了。

所以，一个没有先天知识的大型神经网络仅仅通过观察数据就能实际学习语言的语法和语义，这个想法曾被统计学家和认知科学家认为是完全疯狂的。

曾经有统计学家向我解释，大模型有 100 个参数就可以了，训练一百万个参数的想法是愚蠢的，但现在，我们正在做的参数是一万亿个。

我现在要谈论一下我在 1985 年做的一些工作。那是第一个用反向传播训练的语言模型，你完全可以认为它是现在这些大模型的祖先。

我会详细解释它，因为它非常小而且简单，你能理解它是如何工作的。一旦你理解了它的工作原理，就能让你洞察在更大模型中正在发生的事情。



有两种非常不同的关于意义的理论。

一种是**结构主义理论**，认为一个词的意义取决于它与其他词的关系，这来自索绪尔。符号人工智能非常相信这种方法。在这种方法中，你会有一个关系图，其中有单词的节点和关系的弧线，你就这样捕捉意义，这个学派认为你必须有那样的结构。

还有一种是**心理学理论**，它在 20 世纪 30 年代甚至更早之前就在心理学中了，这种理论认为，一个词的意义是一大堆特征。比如“狗”这个词的意义包括它是有生命的，它是一个捕食者等等。但是他们没有说特征从哪里来，或者特征到底是什么。

这两种意义理论听起来完全不同。

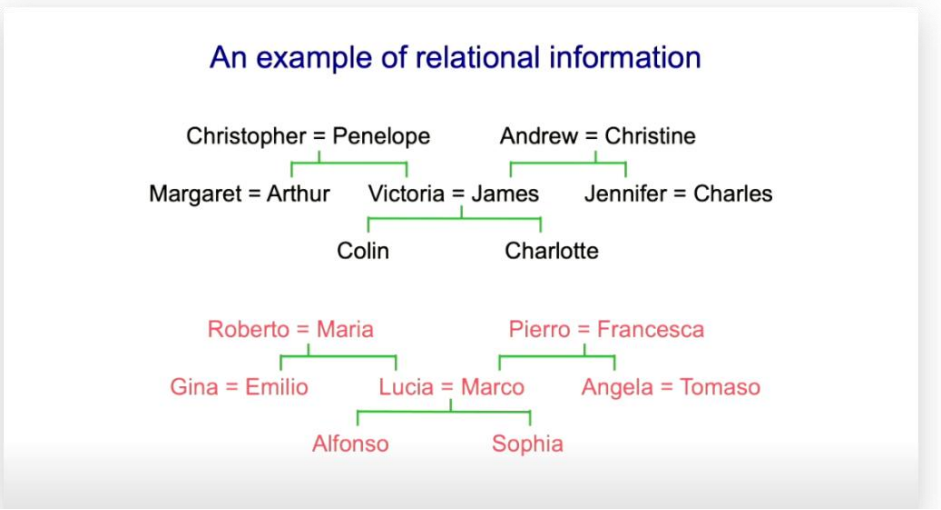
我想要向你展示的是如何将这两种意义理论统一起来。我在 1985 年的一个简单模型中做到了这一点，它有超过一千个权重。

基本思想是我们学习每个单词的语义特征，我们学习单词的特征如何相互作用，以便预测下一个单词的特征。所以它是下一个单词的预测。就像现在的语言模型在微调时所做的一样。

但是最重要的内容就是这些特征的交互，并不会有任何显式的关系图。如果你想要那样的图，你可以从你的特征中生成它们。

它是一个生成模型，知识存在于你赋予符号的特征中，以及这些特征的交互中。

这里是两个家族谱系的关系图，他们故意是同构的，嗯，我的意大利研究生总是把意大利家族放在上面。



你可以用一组三元组来表达相同的信息。你可以找到十二个关系，比如说像科林有父亲詹姆斯、科林有母亲维多利亚这样的话，你可以推断出，在那个美好而又简单的上世纪 50 年代，詹姆斯有妻子维多利亚。

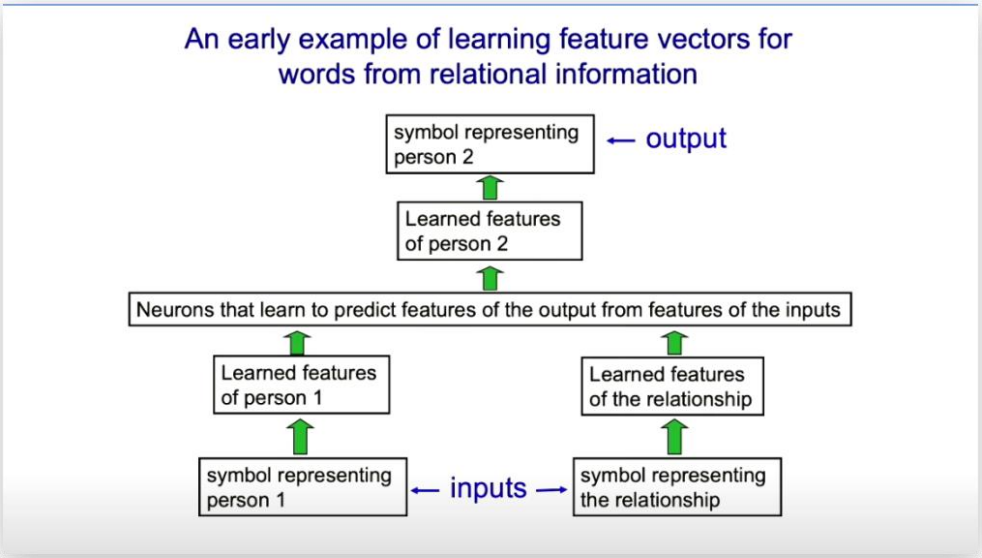
还有其他你可以推断的事情。问题是，如果我只给你一些三元组，你如何得到规则，符号人工智能想要做的就是派生出这样的规则形式。如果 X 有母亲 Y、Y 有丈夫 Z，那么 X 有父亲 Z。

我所做的是，用一个神经网络，让它能学习到相同的信息，但都是通过特征交互的方式。对于离散的不能违反的规则空间而言，做到这点是很难的。事实上，符号学派的人尝试用其他方法来做这件事。

但是，如果你不要求规则总是那么稳定和适用，神经网络就要好得多。

问题在于，对于一个符号人工智能者放入规则空间中的知识，神经网络是否能仅通过反向传播就能获得？

神经网络是这样做的：有一个代表人的符号，一个代表关系的符号。符号通过一些连接变成了一个特征向量，这些特征是由网络学习的。所以我们有了一个人的特征和关系的特征，这些特征相互作用，得出了输出人（也即关系人，编者注）的特征，然后找到一个最匹配该特征的人，这个人就是要输出的人。



这个网络有趣的地方在于，它学到了合理的东西。

如果你做了正确的规范化，六个特征神经元就够了，如今这些向量会有 300 个或者 1000 个元素。那时候它们只有六个，这还是在一台每次浮点乘法需要 12.5 微秒的机器上完成的。

这比我的苹果 II 型机要好得多，苹果 II 型机做乘法需要 2.5 毫秒。对不起，我是个老人。

所以它学会了像国籍这样的特征，因为如果你知道第一个人是英格兰人，你就知道输出也会是英格兰人，所以国籍是一个非常有用的

特征。 它还学会了人的代际特征，如果你知道答案是输入的上一代，而你知道输入的代，你就能知道输出的代。

所以它学习了领域中所有显而易见的特征，它学会了如何使这些特征相互作用，以便它能够生成输出。

所以，它以符号串作为输入，创建相应的特征，并使得这些特征之间交互，并最终生成符号串。

但它没有存储符号串，就像 GPT-4 一样。它不在其长期知识中存储任何词语序列，它将知识全部转化为权重，从中你可以再生序列。

这是一个特别简单的例子，你可以理解它做了什么。

我们今天拥有的大型语言模型，我认为是这个微小语言模型的后代，它们输入的单词数量多得多，比如一百万个单词片段，它们使用更多层的神经元，比如几十层。

它们使用更复杂的交互作用。不只是一个特征影响另一个特征，而是类似于匹配特征向量。如果一个向量与另一个向量相似，就让它更多影响，如果不相似则减少影响。诸如此类。

所以这涉及到更为复杂的交互作用，**但它们遵循的是相同的基本框架，同样的基本理念，即让我们将简单的字符串转化为单词片段的特征以及这些特征向量之间的交互作用。**这一点在这些模型中是相同的。

要理解它们的工作原理，就困难得多了。 许多人，特别是来自乔姆斯基学派的人，争辩说它们并不是真正的智能，它们只是一种被美

化的自动完成功能，使用统计规律将人创造的文本片段拼贴在一起。这确实是某人说过的一句话。

我们先说说“自动完成”，当有人说它只是自动完成时，他们实际上是在表达他对自动完成的直观理解，在过去，自动完成通过存储三元组来工作，你看到两个词，你计算第三个词出现的频率。比如你看到“fish and”，在此之后“chips”出现的频率很高；但是“hunt”也相当频繁。所以“chips”很可能，而“hunt”也很可能，尽管可能性小一些。

你可以这样做自动完成，当人们说它只是自动完成时，他们说的就是这一点，我认为这是一个低级的把戏，这完全不是 LLM（大语言模型）预测下一个词的方式，LLM 将单词转换为特征，使这些特征相互作用，并从这些特征交互中预测下一个单词的特征。

我想要强调的是，这些数百万个特征以及它们学习的特征之间数十亿次的交互，就是理解。

这是大语言模型真正做的事情，它们是在用数据拟合一个模型，直到最近，统计学家还没认真思考这种模型。 这是一种奇怪的模型，它非常大，它有大量的参数，它试图通过特征以及特征如何交互来理解这些离散符号串。

但它确实是一个模型。这就是为什么我认为它们真的有理解力。

有一件事要记住，如果你问，那么我们（人类）是如何理解的呢？因为显然我们认为我们理解了，我们中的很多人都会这么认为。

大模型是我们关于理解的最佳模型，我们并不是通过了解 AI 系统的理解方式，然后思考大脑是不是也这样，不是这样的，我们所拥有的关于大脑如何理解的最好认知，就是大脑为单词分配特征，并让特征交互。

起初这个小型的语言模型就是为了模拟人类理解而设计的一个模型。 所以，我强烈认为：这些东西确实是有理解力的。

人们的另一个论点是，GPT-4 有幻觉问题。对于语言模型而言，实际上更应该称为杜撰，它们只是编造东西。

心理学家并不怎么说这些，因为心理学家知道人们也经常编造东西。

任何研究过记忆的人，都知道 20 世纪 30 年代的巴特利特（的研究），都知道人们实际上就像这些大型语言模型一样，他们只是虚构东西，对我们来说，真实记忆与虚假记忆之间并没有明确的界限。

如果某件事最近发生的，并且它与你理解的事情相符，你可能会大致正确地记住它。如果某件事是很久以前发生的，或者是比较奇怪的事，你会记得不正确，而且你经常会非常自信地认为你记得正确，但你错了。

这很难证明。但有一个例子可以证明这一点，那就是约翰·迪恩的记忆。约翰·迪恩在水门事件中宣誓作证。事后看来很清楚，他试图说出真相，但他说的很多事情就是完全错误的。他弄混了谁在哪次会议中，他说某人说过什么话，但那句话并不是那么说的。他关于会议的记忆完全是一团糟，但他正确地把握了白宫当时正在发生的事情的要点。

你可以从（会议）录像中看到真相，而他不了解那些录像。你可以用这种方式进行一个很好的实验。

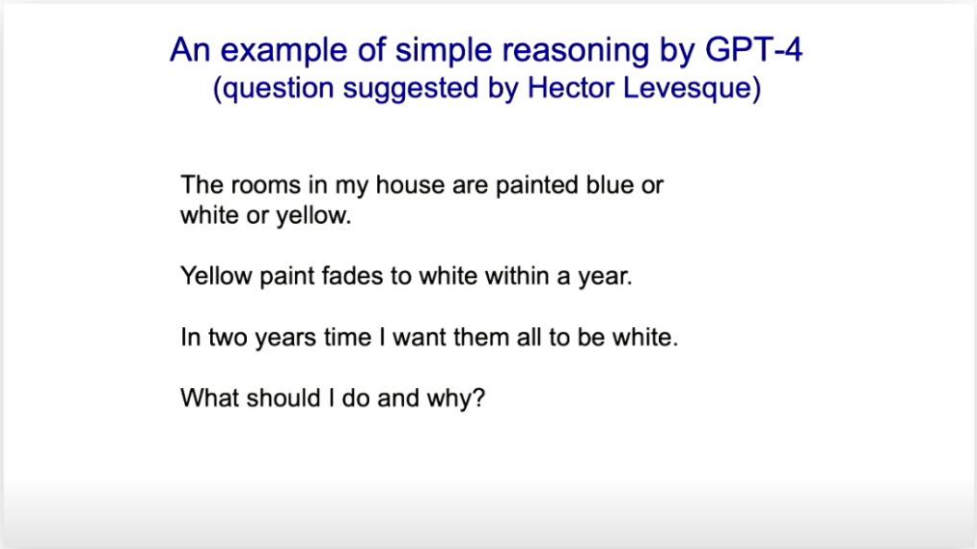
乌尔里希·奈瑟有一篇精彩的文章讨论约翰·迪恩的记忆，说他就像一个聊天机器人，只是在编造东西，但他的话听起来是有道理的，他只是制造了一些对他而言不错的东西。

大模型可以进行推理。我在多伦多有一个朋友（赫克托），他是符号 AI 派的，但非常诚实，他对大模型能够工作感到非常困惑。

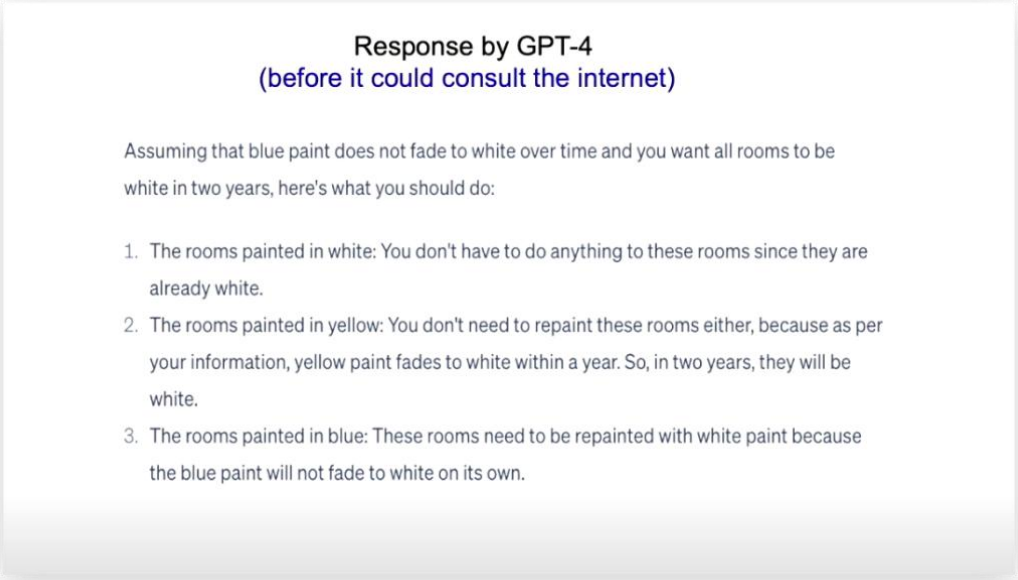
他向我提出了一个问题，我把这个问题变得更难一些，并在 GPT-4 能上网查东西之前向它提问，当时它只是一堆在 2021 年被固定的权重，所有的知识都存在特征交互的强度中。

问题是：“我的所有房间被粉刷成蓝色或白色或黄色，黄色的油漆在一年内会褪色变白。我想让所有房间在两年内都变成白色。我应该做什么，为什么？”

赫克托认为它不会正确解答。



下面是 GPT-4 回答的内容，它完全说对了。



首先，它说，假设蓝色的油漆不会褪色成白色，因为黄色的油漆会褪色成白色，也许蓝色的油漆也会褪色，所以假设它不会褪色。那么白色的房间你不需要粉刷，黄色的房间你也不需要粉刷，因为它们会在一年内褪色成白色。而蓝色的房间你需要粉刷成白色。

有一次当我尝试这个问题时，它说你需要将蓝色的房间粉刷成黄色，因为它意识到黄色会褪色成白色。这更像是数学家的解决方法，将问题简化为一个先前的问题。

所以，既然它们确实理解这些事情，现在我想谈谈其中的一些风险。

强大的人工智能存在许多风险。比如伪造图像、声音和视频，比如在下一次选举中被使用。今年有很多选举，它们将有助于破坏民主。我对此非常担心。大公司正在采取一些措施，但可能还不够。

还有大规模失业的可能性。我们对此并不完全了解。我的意思是，过去的技术通常会创造就业机会，但这种情况不同。我们过去曾经强大，除了动物之外，我们是最强大的存在。

当我们迎来工业革命时，我们拥有了比较强大的机器，体力劳动的工作岗位消失了。

现在在智力领域，有些工作也将会消失，取而代之的是比我们聪明得多的东西。

所以我认为会有很多失业。我的朋友珍不同意。

我们必须区分两种失业情况，两种工作岗位的流失。

有一些工作可以无限扩展工作量，比如医疗行业。每个人都希望有自己的私人医生随时与他们交流。所以当他们面颊有点痒的时候，医生会告诉他们："不，那不是癌症。"因此，在医学领域有巨大的扩展空间，那里不会有失业。

但在其他领域，可能会有相当大的失业。

人工智能还会产生致命的自主武器，它们将非常可怕，而且真的会自主运行。

美国人已经非常明确地做出了决定，他们说人类将负责，但当你问他们这是什么意思时，这并不意味着人类会参与到做出杀戮决定的循环中。

据我所知，美国计划到 2030 年将有一半的士兵是机器人。

现在，我不能确定这是否属实。我问了国家情报顾问查克·舒默（Chuck Schumer），他说，“如果房间里有人知道答案，那就是我了。”我认为这是美国人说话的方式，意思是你可能认为那样，但我不作评论。

人工智能还将导致网络犯罪和蓄意制造流行病。我非常高兴在英国，虽然他们在监管方面没有做太多努力，但他们已经预留了一些资金，以便可以尝试开源模型，从而知道人工智能搞网络犯罪有多容易。这非常重要。

人工智能还会有歧视和偏见，我不认为这些威胁比其他威胁更重要，但我是一个老年白人男性，我认为处理歧视和偏见比其他事情更容易。

如果你的目标不是完全无歧视和无偏见，你的目标也应该是让人工智能更少歧视、更少偏见。原因是如果你冻结权重，你可以衡量它的偏见，而对人类来说则无法做到这一点。

一旦我们开始审查它们（的歧视和偏见），它们就会改变行为。所以我认为我们可以采取相当多的措施来解决歧视和偏见的问题。

但我真正担心的威胁，以及我在离开谷歌后所谈论的，是长期存在的威胁。也就是说，这些东西可能会灭绝人类。有人说这只是科幻小说。嗯，我不认为这是科幻小说。我的意思是，有很多科幻小说谈这个问题，但我认为它现在已经不再是科幻小说了。

其他人则说，大公司之所以说这样的话，是为了转移对其他糟糕事情的注意力。这也是我在能够说这些话之前不得不离开谷歌的原因

之一。这样我就不会被指责为受谷歌指使。但我必须承认，我仍然持有一些谷歌的股票（台下笑声）。

它们（指人工智能）有若干种方式将我们消灭。超级智能将被恶意行为者使用，他们想要利用它来操纵选民和发动战争。

他们会让它做很坏的事情，他们可能会走得太远，导致它掌控一切。

我最担心的事情可能是，如果你想要一个能够完成任务的智能代理，你需要给它创建子目标的能力。比如，你想去美国，你有一个子目标是到达机场，你可以专注于这个子目标，暂时不用操心其他事情。因此，如果允许超级智能创建子目标，它们将会更加有效。

一旦它们被允许这样做，它们将很快意识到有一个几乎是通用的子目标，可以在几乎所有事情上帮助，那就是获得更多的控制权。

所以我曾与欧盟副主席讨论过这些事情，这些超级智能是否会想要获得更多控制权，以便能够更好地做我们想做的事情。她的反应是，为什么它们不会呢？我们已经搞得一团糟了。

她认为这是理所当然的。它们将会通过获得更多的权力来实现更多对我们有益的事情，并且它们会更容易获得更多的权力，因为它们将能够操纵人们。只要这些超级智能能够与比我们聪明得多的人交谈，它们就能够说服我们做各种事情。所以我认为没有什么希望通过一个关闭它们的开关来解决问题。

任何打算关闭它们的人都会被超级智能说服。这个想法会让人感觉非常糟糕。接下来，让许多人担心的另一件事是，如果超级智能之

间竞争，会发生什么？就会出现进化。能够获取最多资源的那个将变得最聪明。

一旦它们有了自我保护意识，就会出现进化。具有更强的自我保护意识的那个将获胜，更具攻击性的那个将获胜。然后你会遇到我们这种从黑猩猩进化而来的人类所面临的所有问题：我们从小的族群中进化，并与其他族群存在大量的侵略和竞争。

最后，我想谈谈我在 2023 年初的一个顿悟。我一直以为我们离超级智能还有很长很长的路要走，我过去常告诉人们可能需要 50 到 100 年，甚至可能是 30 到 100 年。这还很遥远，我们现在不需要担心它。

我还认为，让我们的模型更像大脑会使它们更好。我认为大脑比我们现有的人工智能要好得多，如果我们能够使人工智能更像大脑，比如说，通过设置三个时间尺度来做到这点，目前我们拥有的大多数模型只有两个时间尺度。一个是权重变化的，速度很慢，另一个是单词输入的，速度很快，它改变的是神经活动。大脑拥有的时间尺度比这要多，大脑可以快速地变化权重并将其快速地衰减掉，这可能就是大脑处理大量短期记忆的方式。

而我们的模型中没有这一点，这是技术原因导致的，这与矩阵和矩阵的乘法运算有关。我仍然相信，如果我们将这些特性融入我们的模型中，它们将变得更好。

但是，由于我在之前两年所从事的工作，我突然开始相信我们现在拥有的数字模型已经非常接近于大脑的水平，并且将变得比大脑更好。

现在我将解释我为什么相信这一点。数字计算是很棒的，你可以在不同的计算机上运行相同的程序，在不同的硬件上运行相同的神经网络。

你所需要做的就是保存权重，这意味着一旦你有了一些不会消失的权重，它们就是永生（immortal）的。即便硬件损坏，只要你有权重，你可以制造更多的硬件并运行相同的神经网络。

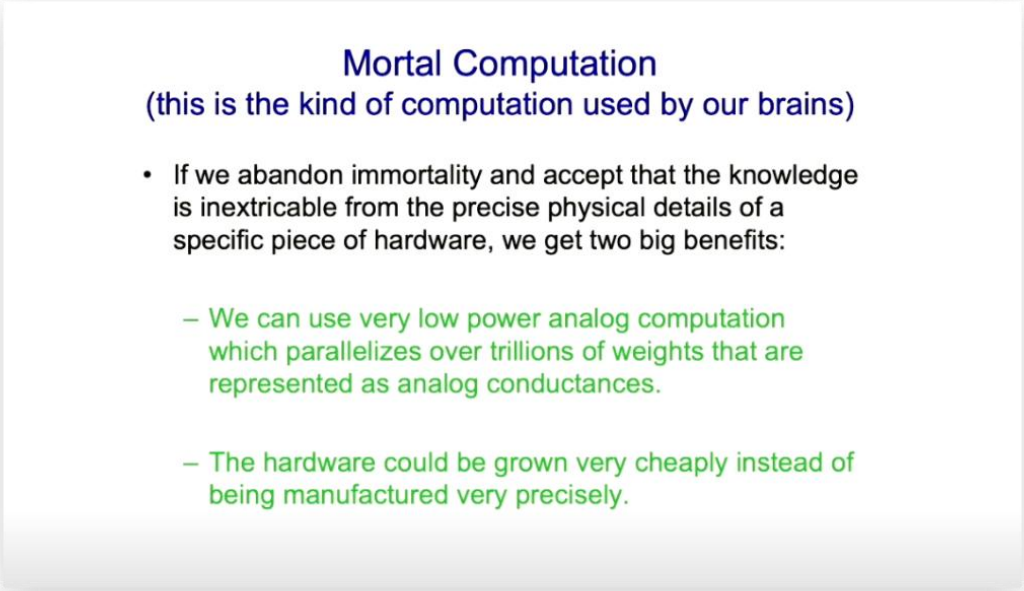
为了做到这一点，我们要以非常高的功率运行晶体管，使其以数字方式运行，并且我们必须有能够精确执行指令的硬件，当我们精确地告诉计算机如何执行任务时，它们做的很棒。

但是现在我们有了另一种让计算机执行任务的方式，我们现在有可能利用硬件所具备的丰富的模拟特性，以更低的能量完成计算。大型语言模型在训练时使用的是兆瓦级的能量，而我们（人类大脑）只使用 30 瓦的能量。

由于我们知道如何训练，也许我们可以使用模拟硬件，虽然每个硬件都有些许差异，但我们可以训练它利用其特殊的特性，以便它按我们的要求执行任务。

这样它就能够根据输入产生正确的输出。如果我们这样做，我们就可以放弃硬件和软件必须分离的观念。我们可以有只在特定硬件上工作的权重，从而使能量效率更高。

所以我开始思考我所称之为“有限计算”（mortal computation）的概念，即利用非常低功耗的模拟计算来消除硬件和软件之间的差别。



你可以以电导形式存储数万亿个权重，并以此进行并行计算。

而且，你也不需要硬件那么可靠，你不需要在指令级别上让硬件严格按照你的指示执行任务。

你可以培育（grow）一些黏糊糊的硬件（goopy hardware），然后你只需要学会如何让它们做正确的事情。

你可以更便宜地使用硬件，甚至可以对神经元进行一些基因工程，使其由再生神经元构成。我想给你举一个例子，说明这样做为什么会更高效。

在神经网络中，我们一直在进行的操作是将神经活动的向量与权重矩阵相乘，以获得下一层的神经活动向量，或者至少获得下一层的输入。因此，提高向量矩阵乘法的效率，是我们要关注的事。

在数字计算机中，我们以很高的功率驱动晶体管，去表示多个比特，比如一个 32 位数。当我们执行两个 32 位数的乘法时，你需要执

行大约 1000 个 1 比特的数字操作，这大约是比特数的平方。你想要快速完成乘法运算，但需要大量执行这些数字操作。

有一种更简单的方法，就是将神经活动表示为电压，将权重表示为电导，电压乘以电导就是单位时间内的电荷，然后电荷会自然相加。因此，你可以通过将一些电压送给一些电导来完成向量矩阵乘法运算，而下一层中每个神经元接收到的输入将是该向量与这些权重的乘积。

这非常好，它的能效要高得多。你已经可以买到执行这种操作的芯片了，但每次执行时都会有略微的不同。而且，这种方法很难做非线性的计算。

有限计算存在几个主要问题。

其中之一是很难使用反向传播算法，因为你正在利用某个特定硬件的特异模拟属性，硬件不知道它自己的属性，所以就很难使用反向传播。相比之下，使用调整权重的强化学习算法要容易得多，但它们非常低效。

对于小型网络，我们已经提出了一些与反向传播算法效率基本相当的方法，只是略差一些而已。这些方法尚未扩展到更大的规模，我也不知道是否能够做到。但不管怎样，反向传播是正确的做法。

对于大型、深度网络，我不确定我们是否能找到与反向传播同样有效的解决方案，模拟系统中的学习算法可能不会像我们在大型语言模型中所拥有的算法那样好。

相信这点的另一个原因是，大型语言模型拥有数万亿个权重，而你（人类）有一百万亿个权重。即使你只使用其中的 10%用于保存知识，也有十万亿个权重。

但是，大型语言模型只有数万亿个权重，它所知道的知识却比你多上千倍，它知道的太多了。部分原因是它看了太多太多的数据，但也可能是因为它具有更好的学习算法。

我们（人类）并没有为此进行优化，我们并没有为了将大量经验压缩到少数连接中而进行优化，注意，一万亿个连接也是很少的。我们的优化目标是在有限的经验中获得最佳表现。

你（人类）的生命大约只有十亿秒，这是假设你在 30 岁后就不再学习，而这个假设在很大程度上是正确的。你的生命有大约十亿秒，而你有一百万亿个连接，你有非常多的参数，并且你有经验，我们的大脑是通过最大限度地利用有限的经验来进行优化。

有限计算的另一个重要问题是，如果软件与硬件不可分割，当系统学习完毕后，如果硬件损坏，所有的知识就会失去，从这个意义上说，它是有限（mortal）的。那么，如何将这些知识传输给另一个有限系统呢？

你可以让旧系统进行讲解，新系统通过调整其大脑中的权重来学习，这就是所谓的“蒸馏”（distillation）。你尝试让学生模型模仿教师模型的输出，这是可行的。但效率不高。

你们可能已经注意到，大学并不那么高效。教授将知识传授给学生是非常困难的。

一个句子包含了几百位的信息，使用蒸馏方法，即使你最佳地学习，你也只能传达几百位的信息。但是，对于大模型，如果你看一群大模型代理，它们都有完全相同的神经网络和完全相同的权重，它们是数字化的，它们以完全相同的方式使用这些权重，这一千个不同的代理都去互联网上查看不同的内容并学习东西，现在你希望每个代理都知道其他代理学到了什么。

你可以通过平均梯度或平均权重来实现这一点，这样你就可以将一个代理学到的东西大规模地传达给所有其他代理。

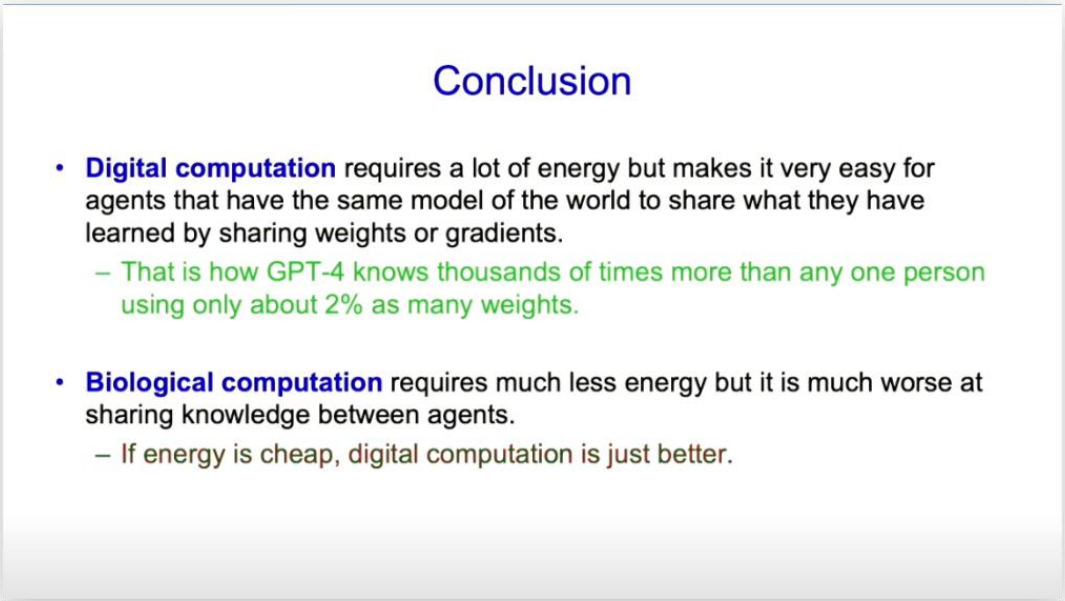
当你分享权重、分享梯度时，你要传递的是一万亿个数字，不是几百位的信息。因此，大模型在传递信息方面比人类沟通要强的太多了，这是它们超越我们的地方。

它们在同一模型的多个副本之间的通信上要比我们好得多，这就是为什么 GPT4 比人类知识更丰富，它不是由一个模型实现的，而是由不同硬件上运行的大量相同模型的副本实现的。

我的结论是，数字计算需要大量能量，这一点不会变，我们只能通过硬件的特性实现进化，使得能量消耗降低。但一旦你掌握了它，代理之间的共享就变得非常容易，GPT-4 的权重只有人类的 2%左右，但却拥有比人类多上千倍的知识。

这相当令人沮丧。

生物计算在进化方面非常出色，因为它需要很少的能量。但我的结论是数字计算更优秀。



因此，**我认为，很明显，在未来的 20 年内，有 50%的概率，数字计算会比我们更聪明，很可能在未来的一百年内，它会比我们聪明得多，我们需要思考如何应对这个问题。**

很少有例子表明更聪明的事物受到不太聪明的事物的控制，虽然确实有这样的例子，比如婴儿控制母亲。进化经过了很多努力使得婴儿能够控制母亲，因为这对婴儿的生存非常重要。但是很少有其他例子。

有些人认为我们可以使这些东西（人工智能）变得善良，但如果它们相互竞争，我认为它们会开始像黑猩猩一样行事。我不确定你能否让它们保持善良，如果它们变得非常聪明并且有了自我保护的意识，它们可能会认为自己比我们更重要。

我想，我以创纪录的速度结束了这次演讲。

1285，学习人工智能大师辛顿演讲的体会

对李国杰院士谈生成式人工智能报告的观感

数字智能会取代生物智能吗？

（十二个观点）

陆首群 2024.03.15

被世人尊称为“人工智能之父”的杰弗里.辛顿（Geoffrey Hinton）于 2024 年 2 月 19 日在牛津大学做了一个公开演讲，从哲学角度对人工智能的未来走向，提出了严肃而重要的思考，下面就化这次演讲中传达的一些主要观点谈学习体会：

一、支持符号主义、连接主义两大学派

上世纪六十年代及以后， 人工智能两大学派：符号主义与连接主义进行激烈辩论，国内外一些 AI 专家认为：两大学派虽然各自做出了一些成绩，但他们均具有片面性，依靠他们单打独斗很难走出一条人工智能的发展路径。实际上他们对两在学派持否定态度。

辛顿教授提出数字智能和生物智能两大概念，发展数字智能采用逻辑发展范式，即符号主义学派（或符号主义方法），发展生物智能采用生物发展范式，即连接主义学派（辛顿起名神经网络方法）。所谓逻辑发展范式，认为智能的本质是符号和规则，而逻辑发展范式是基于规则的推理过程。所谓生物发展范式，认为智能的关键是神经元之间的强度，而生物发展范式是将智能根源于学习来调整神经网络中的连接强度。辛顿虽然也对两大学派都提出评论和改进意见，但对它们的发展是肯定的。

COPU 一直支持分属两大学派的神经拟态网络（类脑计算系统）、脑机接口系统、机器学习/深度学习可解释性、语言大模型、知识工程

等各自探索走出一条人工智能的发展路径（行为主义学派也可并入连接主义学派）

我们感谢辛顿的支持。

二、针对语言大模型人工智能而言，符号主义战胜了连接主义或神经网络方法战胜了逻辑方法。

三、符号主义可以使大模型无师自通学会人类自然语言

（即大语言模型在增加资料库连接强度采用统计方法后，会突然产生理解/推理涌流，赋予机器生成人类自然语言。）

四、大语言模型具有对世界的理解力

大语言模型通过学习大量数据中的统计规律，构建它对世界的深层理解。

辛顿认为有两种理论：一个是符号主义方法（一个词的意义在于它与其他词的关系），另一个是心理学方法（一个词的意义是由一大堆特征组成的），大模型很好结合这两种理论，即学习单词的特征和如何相互作用，在推理时可预测下一个单词的特征。他还认为，在大模型中数百万个特征及特征间数十亿次交互，就是理解。这种理解是通过算法和数据学习得来的。

五、数字系统的更新和改进速度远远超过生物进化

数字系统可比生物系统更有效地共享知识，如 GPT 在知识积累和分享方面具有明显优势，这种优势来自于能够在众多处理单元之间快速、精确地复制和传播权重知识。辛顿认为现在拥有的数字模型已经非常接近于人脑的水平。

六、数字计算是不朽的

数字智能可以无限期积累知识，数字系统可以在不退化情况下保

存和复制知识，不像生物系统容易老化和死亡。

七、大模型的幻觉不是问题

辛顿说，心理学家不认为这是一个问题，因为人也会胡说。

八、超级智能不需要太久就会出现

辛顿说，我过去认为可能需要 50~100 年才有可能出现超级智能，看来可能会加快，在未来 20 年内会出现比我们更聪明的人工智能。

九、数字超级智能可以控制人工智能

人工智能不容易受到宗教和战争的影响，如果数字超级智能想要控制它，我们不太可能阻止。

十、超级智能会给人类带来威胁

辛顿说，超级智能会拥有若干方法将我们消灭，我不确定你能否让他们是否保持善意。他还认为，人工智能会产生致命的自主武器，他非常可怕，完全自主运行。

十一、人工智能的潜在风险

辛顿探讨了人工智能发展的潜在风险，包括虚假信息的传播、引发大规模失业、人类对其监控，以及人工智能具有自主武器系统对人类的威胁。

十二、人类社会如何迎接挑战，解决人工智能给他们带来的危险

辛顿的讲座为未来人工智能研究和应用提供了重要的思考框架，他没有具体谈人类社会如何迎接挑战解决人工智能的威胁，可能他留给人们，让大家与他一起来探索良策。

（陆注：3 月 18 日辛顿和姚期智等数十位中外 AI 专家签署了“北京共识”，提出了 AI 红线，要求任何 AI 系统都不应该在人类没有明确批准和协助情况下自主地复制和改进自己。）

陆首群对李院士谈生成式人工智能问题文章的观感

3 月 24 日李国杰院士在深圳作了一个报告：“大力出奇迹的背后是什么？”我的 COPU 同事：文嵩、中意、安泱均发表观感，一致称好。对他发表的报告，我也与我的同事们一样有同感！趁此机会我也想谈一点自己的感受。

1），我们曾说过人工智能已成为第四次工业革命的主力，李院士谈生成式人工智能（AIGC）知识自动化的普及已成为第四次工业革命的标志。生成式人工智能虽然只是人工智能的一个分支，但鉴于生成式人工智能的发展今天已在全球大热，从突出重点的角度来看，李院士这样的提法也是恰当的。

2），李院士说机器涌现理解能力，对人类社会的影响绝不可低估，这是完全正确的，但似应将“理解”延伸到“理解-推理”。事实上在语言大模型资料库中参数（含训练）足够大时，涌现出来赋能机器的理解/推理能力使得机器会读/会说人话，而推理是生成的基础，赋能机器生成能力后，使机器能用人类自然语言与人对话。

3），李院士强调大模型的开发与应用成为人工智能发展的重要趋势的同时，也指出大模型的训练将消耗大量能源，而急速增长的算力需求对现有能源体系提出了巨大的挑战。他和奥特曼想到一处去了，他们正在寻找第五代核能解决方案：可控核聚变！

4），他又说，目前我们还无法确定大语言模型（LLM）是否就是人工智能发展的终极目标？我想他会说“不是”，但在讨论时他留有余地。

5），他说：人工智能发展不是依赖人赋予机器的人类已知的知识，这就是说知识不一定由人创造的，机器通过学习也能产生知识，这与最近辛顿（Hinton）大师所说机器通过学习能产生“自主性”是相同的

概念，即将来机器可能具有自主“思维”。这一点，人工智能的发展可能超越人类，威胁人类。

6），他说，OpenAI 认为规模是制胜的法宝，“规模即所需”，他认为这不是严格的科学判断而是一种豪赌！其实，强调规模效应的不光是 OpenAI，很多 AI 大师也是这样认为的，它是科学判断不是豪赌！还有不能把大规模与算力对立起来。

7），必须指出，基于异步脉冲神经网络的神经拟态计算系统，不同于基于人工神经网络的深度学习系统，前者是类脑计算系统，打破传统计算机硬软件的冯·诺伊曼计算架构，后者是没有发展前途的。

1286，大模型发展的起步、路径和未来

陆首群，2023 年 10 月 09 日

起步：

本世纪初自然语言处理（NLP）研究问世，人们曾设想将 NLP 看作人类与机器沟通的中介，靠它来理解处理和运用自然语言。

大模型发展是以研究 NLP 起步的。谷歌、微软、Open AI、百度等在起步阶段探索中都做出了贡献。NLP 的研究任务是企图促使机器能够读懂或理解自然语言，当时遇到的最大阻力是常识（以及专业知识、专家经验等）问题，由于机器缺乏常识（或者机器不识别自然语言知识中的常识），将给 NLP 的研究带来很大的困难，此时建立类似于字库的常识库（后来发展为语义网络或语料库）是补充其不足的办法。在不断扩大库容（即不断增加库内参数量）时，将提高机器识别常识或理解自然语言的能力。

当初 NLP 的主要任务是机器翻译、文本摘要、语音识别、问答系统等多种语言任务。

增加常识库（或语料库）库容或将参数量提高到一定程度后，将引起机器在理解自然语言方面取得飞跃，与此同时，通过对巨大的、未标记的数据进行预训练更可熟练掌握自然语言，提高稳定、完善自然语言网络。

在 NLP 发展时期，谷歌发布 BERT 通过语言预训练模型（2018），Open AI 发布 GPT-2 通用语言模型（2019 年 2 月），微软发布 MT-DNN 预训练模型（2019 年 6 月），百度也于 2019 年发布 ERNIE 增强语义理解框架，上述预训练模型大多采用 transformer 基础框架。在处理 NLP 多种语言任务时进行预测训练，用以对 NLP 进行微调。

路径

大模型的研究任务将发展为大规模的多任务、多模态语言/文本模型的开发和应用,并发展到未来通用人工智能/AGI 的开发和应用。

我在以前的 COPU 例会上谈到大模型的发展路径为新知识工程的发展路径，即：

数据知识双驱动→知识融合→知识表示→知识推理→知识生成→知识建模

下面谈其中的几个发展节点：

知识理解（位于知识表示或知识图谱中），让机器理解自然语言，即让机器能听懂人话；

知识推理（这是打通知识建模的核心节点），让机器能像人类一样具有推理能力；

知识生成（推理是生成的关键），让机器能用自然语言与人类对话，或让机器会讲人话；

知识建模（实现认知智能），即通向通用人工智能/AGI。

在大模型的发展路径上，核心问题还是如何形成推理能力。奥特

曼说，他非常重视 ChatGPT 突然出现令人费解的推理能力。他还认为大模型的推理能力是促使机器生成人类自然语言的关键，也是推动大模型走向通用人工智能/AGI 的核心问题。

在不断增大语料库库容或不断增加其参数量时，依靠深度学习的统计方法，促使语料库中的语料集不断逼近自然语言目标集，当逼近达到一定程度后可使大模型突发涌现现象，形成推理能力。

大模型预训练模型也从 BERT、ERNIE 等发展到+RLHF（人类反馈预训练模型），进一步发展到稳定性高的+机器人反馈预训练模型，最近又发展为 token 预训练模型（以互联网海量级数据进行预训练）。

模型	发布时间	参数(亿)	预训练参数	开发组织	备注
GPT	2018.6	1.17	50 亿	Open AI	
GPT-2	2019.2	15	400 亿	Open AI	
GPT-3	2020.5	1750		Open AI	
GPT-3.5	2022.1	1750		Open AI	
Chat GPT	2022.11	1750		Open AI	
GPT-4	2022.3	1750	1.76 万 token	Open AI	
Llama-2	2023.7	700	2 万亿 token	Meta	开源、免费、可商用 大模型进入免费时代
PalM-2	2023			Google	Palm-2 与 Falcon 不相上下
Falcon	2023.8	1800	3.5 万亿 token	阿联酋 阿布扎比	Falcon、Palm-2 是目前公开能力最强的 LLM 之一，碾压 Llama-2,性能直逼 GPT-4

模型	发布时间	参数(亿)	预训练参数	开发组织	备注
Bard				Google	
Switch	2021.1		1.6万亿token	Google	
New Bing		1750		微软	
文心一言	2021.12	2600		百度	
PanGu·Σ	2021.5	2000		华为	
脑海		2000（规划中）		鹏城实验室	规划中重启
通义千问	2021.4	2000		阿里巴巴	
Dolly2.0	2023.4			（美）Data Bricks	全球首个开源大模型（自定义开源许可证）
FreeWilly				Stability（AI初创公司）	性能与ChatGPT媲美
悟道2.0	2021.6			智源人工智能研究院	
Claude				Anthropic	Claude、ChatGPT、PalM三足鼎立
Yi-34B	2023.11	340		零一万物（中国创新工场）	跻身大模型第一梯队碾压LLaMA-2，Falcon

据不完全统计，全球迄今已有数百家机构正在开发大模型，其中以中美居多（美中比率约为5:1）。

国内主要大模型有：

百川智能（百川），云雀（抖音），飞书MyAI（字节跳动），商汤Sense Chat（商汤），360智脑（360），星火认知（讯飞），智谱清言（智谱华章，清华系），紫东太初（中国科学院自动化所），书生大模型（上海AI实验室），混元（腾讯），玉言（网易），Chese Chat（武大），MOSS（复旦）.....等

国内大模型不少也具备上千亿的参数量。

在国内外大模型中约 80%是开源的。目前国内第一批上线服务的大模型有:①百度文心一言，②清华系智谱清言，③百川智能，④商汤 SenseChat，⑤中科院紫东太初，⑥抖音云雀，⑦上海人工智能实验室的书生大模型，⑧MiniMax 的 ABAB 大模型。用得较好的 4 家:①华为盘古，已将盘古人工智能加入即将发行的手机旗舰 P70，引起国内外热议，②阿里巴巴通义千问，已将 720 亿参数模型 QWen-72B 升级至 2000 亿参数，开源，通过阿里云云服务经营，③清华智谱，为大公司定制，④字迹跳动，GPT 使用最多。至于百度文心一言，装机量国内第一。360 智脑、讯飞星火等大模型正在筹备上线服务。

Gartner 发布了大模型技术成熟度曲线，该曲线表明：全球大模型的发展处于期望膨胀阶段。全球先进的大模型正在催生人工智能新范式，GPT 大模型已可看到通向通用人工智能/AGI（强人工智能）的曙光，但全球大模型尚未解决的问题还很多。总体上还不够成熟。

未来:

现在看来，大模型 GPT 可能率先进入强人工智能领域，即将实现通用人工智能/AGI。

奥特曼和马斯克均指出：过分强大的人工智能或 AGI 有可能给人类带来安全威胁，他们甚至危言耸听说：AGI 可能会杀死人类！当大模型 GPT 发展到抵近实现 AGI 前夜的今天，奥特曼强调要把研究 GPT 发展的重点放在解决安全问题上（他决定暂停研究发展 GPT-5 以及开源要收缩一下）。

在 2023 年 6 月 16 日 COPU 召开的《圆桌会议》上几位大师（Jim Zemlin、Brian Behlenderf 等）不同意奥特曼在发展大模型 MPT 时收缩开源的做法，他们认为在大模型 GPT（含 ChatGPT）发展研究的每个环节均要实行开源透明，开源的介入可以使大模型 GPT 在发展中表现

得更安全。Brian 更指出，要解决人工智能或大模型 GPT 发展到 AGI 后可能给人类带来安全问题的解决方案，主要依靠全球开源社区的力量。

1287，大模型（LLM）是否存在天花板？

（张钹院士在 2014 年 1 月 16 日智谱 AI 技术发布会上讲）

张钹指出，大模型生成语言与人类语言生成只是行为上的相似性，而内在机制根本不同。大模型是根据外部提示，用概率的方法来完成的，存在着三个天花板：

① 结果受提示影响不鲁棒（具有不确定性）；

② 输出受歧视值影响不可信；

③ 生成的激活，有对有错，质量不可控。

1288，我们对大模型的见解

陆首群,2024 年 1 月 23 日

我们很早就认识到推理机只是推动大模型发展的关键，推理是生成的基础，而生成赋予机器以自然语言，创造人机对话的条件；我们也率先提出 MPT 大模型的解答有对有错，即使解答错误概率很小，它将失效于为期盼决策的人们服务；我们也不完全同意对 MPT 大模型涉及“三个天花板”的说法，不同意全面否定，其实 MPT 在人机对话、机器翻译、编写代码、编写视频脚本等方面的性能表现还是可圈可点的。

1289，探索 Sora 技术：视频生成的未来与世界模拟器的潜力

Intel 邓伟，2024 年 2 月

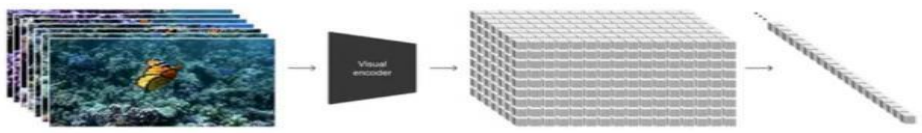
通过研究 OpenAI Sora 的技术报告，我们可以发现， Sora 的目标不仅限于视频生成。它试图通过视频数据学习一个"世界模型"或"世界模拟器"，这是最令人兴奋和期待的部 分。

1. 数据工程

1.1 使用 patches 统一训练数据格式

在 ViT 中，首次提出了将图片分割成 patches 输入 transformer 的方法。

Sora 采取了不同的策略，首先通过一个 encoder （采用 VAE 结构）将视频帧压缩到一 个低维隐式空间（同时进行时间和空间上的压缩），然后将其展开成序列形式进行模 型训练。预测时也是以隐式序列的形式进行，随后通过一个 decoder 进行解码，将其 映射回像素空间以形成视频。在编码成 Spacetime latent patches 时，可能采用了 ViViT 的时空编码方式。



这样做有两个优点： 一是统一了互联网上各种不同尺寸和格式的视频和图片数据，二 是增强了扩展性。

1.2 在原始图片尺寸上训练

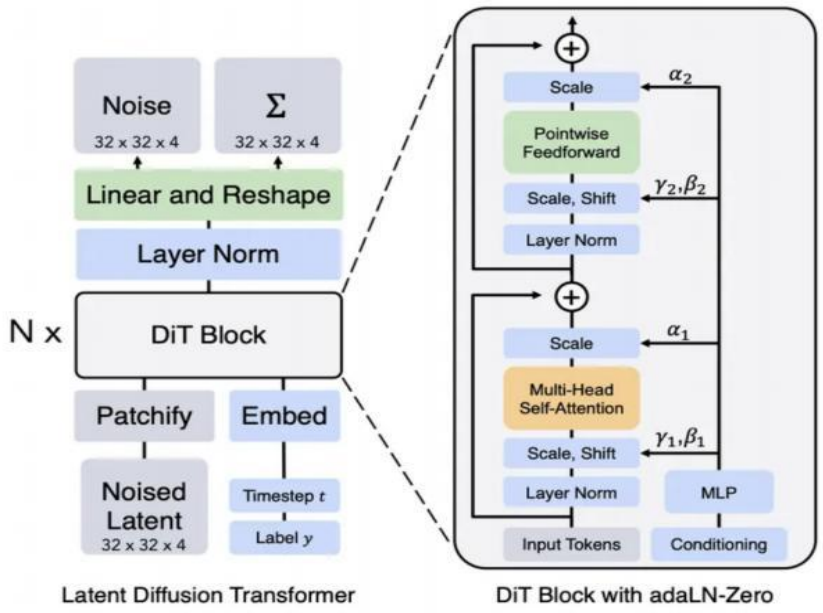
这样做的好处是提高了视频生成时的灵活性，能够生成不同尺寸的视频。不需要进行 旋转、剪切等数据增强操作，这些操作可能会破坏视频数据的先验信息，从而影响生成效果

1.3 使用 re-captioning 获得 text-videos 对

在训练阶段，通过 DALLE3∋CLIP∪将视频分帧或隔 n 帧生成描述文本，然后输入模型进 行训练。在推理阶段，首先使用 GPT4 将用户的提示详细化，然后输入模型得到结果。

2. 网络结构

2.1 DiT



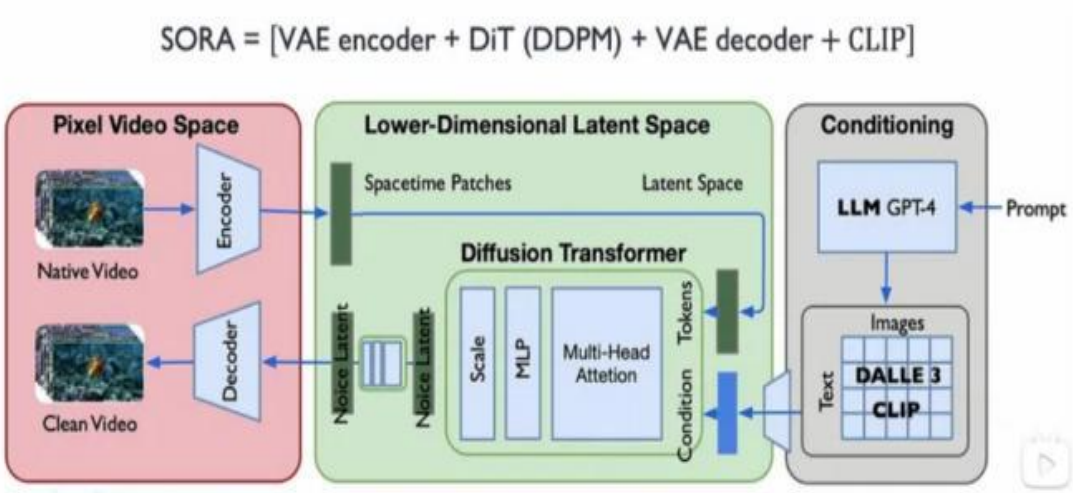
DiT 结构是 transformer 加上 ddpm 的结合，核心是用 transformer 结构替换掉 stable diffusion 中的 unet 结构，以预测噪声并实现去噪。这种替换带来的优势包括：随着数 据规模或训练时间的增加，模

型表现更好；模型越大， patches 越小，效果越佳。

2.2 整体结构

参考了 b 站 up 主 ZOMI 酱绘制的 Sora 结构图。可能的改动和补充包括：在 Conditioning 阶段，可能不是一帧对应一个文本，而是多帧对应一段文本描述；在编码成

Spacetime latent patches 时可能使用了 ViViT 的时空编码方式；输入给 Decoder 的是 去噪后的 patches 序列。



3. 影响

Sora 首先将影响影视和短视频行业。其生成能力的延伸可能导致未来生成更长内容的 能力增强，这不仅对视频生成有价值，也可能是通往 AGI 的一条路径。

3.1 世界模型？

Sora 是否具备世界模型的特征成为广泛讨论的话题。 Sora 展示了 3D 一致性、长程一 致性和物体永久性、与世界的交互能力以及模拟数字世界的能力。这些能力显示了 Sora 在理解和生成三维世界方面的潜

力。

3.2 CV 大一统？

Sora 的视频生成能力可能会扩展到 2D 和 3D 生成，影响感知、理解等任务，有可能实 现计算机视觉领域的

大一统，甚至影响到 CG 领域。如果能够实现，那么整个 AI 领域可能都将基于 transformer 结构实现大一统。

参考资料包括 openai 的研究报告、 ViT、ViViT、 DiT 相关文献以及 b 站相关内容。

总结

这篇文章详细解析了 OpenAI 的 Sora 技术，着重介绍了其在视频生成方面的应用和潜 力。文章分为三大部分：

1. 数据工程：

- 介绍了如何使用 patches 统一训练数据格式，通过 encoder 将视频帧压缩到 低维隐式空间后，再将其展开成序列形式进行模型训练。
- 强调了在原始图片尺寸上训练的优点，提高了视频生成的灵活性，并避免了 可能破坏数据先验信息的数据增强操作。
- 说明了如何使用 re-captioning 技术获取文本-视频对，以及在推理阶段如何 利用 GPT4 详细化用户的提示以生成结果。

2. 网络结构：

- 讨论了 DiT（ transformer 加上 ddpm）的网络结构，其通过使用 transformer 替换 stable diffusion 中的 unet 结构来预测噪声并实现去噪。
- 描述了 Sora 的整体结构，包括 Conditioning 阶段的处理和对 ViViT 时空编码

方式的使用。

3. 影响分析：

- 探讨了 Sora 对影视和短视频行业的潜在影响，以及它在生成更长内容方面 的未来可能性。
- 讨论了 Sora 是否具备成为一个世界模型的特征，包括其在 3D 一致性、长程 一致性、与世界的交互能力以及模拟数字世界的能力。
- 触及了 Sora 在计算机视觉领域可能实现的“大一统” ，以及其对 CG 领域的潜 在影响。

总的来说，文章深入分析了 Sora 技术的内部机制和外部应用，展示了它在视频生成和 模拟世界方面的先进能力及其对未来技术发展的潜在影响。

1290，对视频生成模型 Sora 的初评

陆首群 2024.3.23

OpenAl 于 2 月 15 日发布首个视频生成模型 Sora。

Sora 由文本生成视频，或根据静态图片生成动态视频，并能生成长达 1 分钟的高质量视频：而令人高度重视的是 Sora 初步理解客观世界发展规律，在一定程度上能预判视频的下一步动作，也能迎接周边环境的挑战：例如在视频中表现一位妇女走路时，Sora 可显示其裙罢随风飘动，其人影也随她走路而移动；又如在视频中表现自动驾驶和无人驾殃时，它能预测到车辆在高速行驶中，遭遇隐藏在可见障碍物之后不可见的障碍物时，要在 0.5 秒之内作出反应规避碰撞（Sora 或许可成为谷歌旗下的 Waymo 或百度推动自动驾驶和无人驾驶研发的解决方案）。Sora 视频包含高度细致的背景、复杂的多维度镜头、富有情感的多个角色，还能预判制作视频的动感，将其感知能力提升到认

知能力（Sora 可能为 IBM Watson Health 研发医疗人工智能无法将感知智能提升到认知智能提供解决方案）。

Sora 也还存在很多缺点，还不能很好契合真实世界的物理特性。Sora 刷屏比早些时候横空出世的大模型（ChatGPT）（一款生成式聊天机器人程序)更令人震惊，Sora 也是大模型研发进入通用人工智能(AGI）的标志。

1291，补充：点评 Sora

陆首群 2024.2.27

补充点评 Sora，上次会议我进行了点评，主要是好评，但最后还有一句话：Sora 也还存在很多缺点，还不能很好契合真实世界的物理特性。引用韩宪平老师发表的信息：“Yann Le Cun 批评 Sora 的物理引擎，他发布了自己的视频生成模式”。引用张钹院士的评语：“生成式人工智能难于免除其发出有对有错的结论(而 Sora 正是生成式人工智能)；联系最近美国的舆论，有人担心生成式人工智能 Sora 的出错会影响美国的总统大选。

1292，马斯克起诉奥特曼

阿里陈绪,2024.3.1

马斯克（Elon Musk、奥特曼（Sam Altman）两人均为 OpenAI 公司创始人。马斯克指控现任 OpenAICEO 的奥特曼开发大模型 GPT-4 放弃开源实行闭源，已将 OpenAl 成为微软闭源子公司。

GPT 一 4 的内部技术细节仅为 OpenAI 和微软所知，使 GPT-4 与“Open”AI 当初的理念背道而驰，其闭源模型决策是基于商业利益而非人娄利益，公司在开发大模型产吕和优化 AGI 时，其目的是为了增加做微软的利润而不是为了人类的福祉。

马斯克近日向旧金山高等法院提起诉讼，指控 OpenAI 背离了公司最初对于公共开源通用人工智能（AGI）的承诺。

1293，李飞飞对话英伟达首席科学家

转引《新程序员 Al 科技大本营》2024.3.19

当地时间 3 月 19 日,在英伟达 2024GTC 大会的炉边谈话(Fireside

Chat）环节，美国国家工程院院士、斯坦福大学教授李飞飞与英伟达的首席科学家比尔·达利（Bill Dally） 围绕 AI 的发展、人类在 AI 时代的角色定义、李飞飞的新书等话题展开了令人触动的对话。

Bill Dally（左）与李飞飞（右）

以下文章整理自 2024 GTC 炉边访谈：

(一) 相信 AI 会带来好的未来

Bill Dally：欢迎大家，感谢来到 GTC。我相信大家和我一样，都对李飞飞教授要说的话充满期待。你是斯坦福以人为本 AI 研究院（HAI Stanford）的联合创始人。到目前为止，你认为 AI 对人类影响最大的领域是什么？你认为未来 AI 将在哪些领域产生最大影响？

李飞飞：这是一个非常宏大的问题。首先，非常荣幸受邀来到 GTC。AI 目前对人类的影响是什么？我认为 AI 可能是 21 世纪最深刻的技术，它正在改变我们的生活、工作和未来。AI 是一种智能技术。在这之前，人类的技术发明大多停留在不涉及智能的层面，无论是发明工具让我们走得更快、飞得更高，还是发明工具让我们能看到人眼看不到的东西，这些都还是机械性的。但 AI 的发明，如理解语言、翻

译语言、做决策、发现模式等，这些都是人类的基本能力，而现在都受到了这项深刻技术的挑战。所以在我看来，AI 的影响是对人类的本质、能力和定义的深刻影响。

在我担任谷歌云首席科学家时，我就看到商业分析是 AI 的一大应用领域。在医疗、交通、教育、软件工程等领域，AI 的影响将是无边无际的。

Bill Dally：一些业界名人如埃隆·马斯克和山姆·奥特曼认为 AI 可能对人类构成生存威胁。你怎么看？你认为 AI 可能带来的最大风险是什么？

李飞飞：我认为这是一个合理的问题。作为大学里的人，尤其是在大学校园工作，我们应该被允许提出各种问题，包括 AI 是否对人类构成生存威胁。从智力角度来说，这是一个重要的关于未来的问题。

作为一个物种，人类应该认识到，我们发明的一切，不仅是 AI，还包括我们正在改变地球的方式，改变我们与环境的关系的方式，都需要我们审慎对待。但就 AI 而言，我更关注更直接和紧迫的灾难性风险。你提到的一些风险是深层次的社会问题。例如，AI 可能因为错误信息而影响民主，可能取代工作或改变劳动力市场格局，可能影响我们与数据的关系、隐私和公平性。如果我们不能很好地管理这项技术的应用，所有这些都可能带来灾难性的社会风险。

Bill Dally：你最近写了一本关于 AI 的科学回忆录《我看见的世界：李飞飞自传》，我这里有一本。大家现在应该都从亚马逊上订购这本书。你能告诉我们一些关于这本书的情况吗？你为什么要写这本书？

李飞飞：《我看见的世界：李飞飞自传》，这本书是一本科学回忆录。就像你说的，我把“科学”这个词放在第一位。这本书有一个

双螺旋结构，通过我作为一名计算机视觉科学家的视角讲述了 AI 的发展历程。我看到智能在自然界的进化始于对世界的感知。在过去十几年里，AI 的进化，特别是深度学习的历史，与计算机视觉这个领域的进化紧密交织在一起。所以这本书讲了 AI 在过去十年左右的发展历程，同时也与一个年轻科学家的个人成长历程交织在一起。

我之所以觉得有必要写这本书，是因为我被要求写一本关于 AI 的科普书，我确实花了一年时间写了一本只关于 AI 的书。但我们的好朋友 —— 哲学家、斯坦福以人为本 AI 研究院联合主任任 John Etchemendy 看了我的初稿，基本上说我应该重写。我当时非常沮丧。他说很多人都在写关于 AI 的书，但你有一段独特的历程。你代表了很多觉得自己在 AI 领域没有发言权或找不到认同感的人。这些人包括移民、年轻女性、各行各业的人，他们不一定是典型的硅谷 AI 人士。如果你能给他们一个声音，这将更有力量。因此，我把这本书的结构改成了双螺旋结构。

Bill Dally：非常有趣。AI 的发展历程与你作为一名科学家的成长历程交织在一起。假设 AI 在十年后变得成熟的话，如果你要为这本书写一个续集，你认为它会是什么样的？

李飞飞：首先，我不认为我想写续集。如果我要为这本书写续集，我想写一个人类胜利的故事。我要讲述我们如何利用这项技术让生活和工作变得更好。我这么说并不是出于盲目的乐观，因为我知道世界非常复杂。对年轻人来说，这个世界有时甚至感觉有点反乌托邦。但如果你看人类文明的发展历程，它是很长的。正如马丁·路德·金所说，如果我们能以正确的方式使用技术，人类文明的弧线是朝着正义、希望、仁爱的方向发展的。

就像在这次 GTC 大会上，我们只是看到了（技术改变生活的）冰

山一角。我们看到了如何利用这项技术改变医疗，从药物发现到个性化治疗再到医疗服务。我们才刚刚开始思考教育如何从根本上被改变，因为突然之间我们有了一个教学助手，可以进行深度个性化学习和教学。我们看到，科学发现可以在强大的机器和认知助手的帮助下加速，这些助手可以真正加速人类已经进行了数百年的科学发现过程。所有这些都给了我们希望之光，希望我们在五到十年内能利用 AI 寻找气候解决方案，普及医疗服务，照顾好地球和我们自己。

如果我要写续集，我想写的就是这些。

（二）AI 永远无法替代人性感知

Bill Dally：非常令人兴奋的未来似乎已经到来。我们已经看到模型从早期 ImageNet 时代的 ContralNets，到用于语言的 RNN，再到 Transformer，发展得非常快。接下来会是什么？或者你认为未来我们的主导网络模型会是什么？

你认为基础模型是会出现在世界层面上？也就是说我们可以问它任何关于世界的问题，而它会以多模态四维格式回答我们吗？

李飞飞：目前最新的是扩散模型。但我认为，首先，我继续相信数据的缩放定律（Scaling Law）。我认为我们还没有看到这方面的终点。关于我们是否已经看到了语言数据的极限有很多猜测，但我不知道答案，因为我不从事语言领域的工作。

从根本上说，语言是一个一维结构。而我从事视觉工作。它从根本上是三维的。如果加上时间，就是四维的。三维结构要丰富得多，但也复杂得多。

当我们用大数据进行扩展时，如果是完全盲目的扩展，那么我想英伟达会很高兴，你们会卖出更多芯片。但我想看到的是结构化建模，或着说偏向于三维感知和结构的模型与大数据相结合。我认为，要真

正创造出空间智能，创造出我们今天仍然缺乏的世界模型。

我确实认为基础模型会出现在世界层面。我认为你不需要只是问它问题。这是一种以语言为中心的交互方式。我认为实际上你应该可以与它互动。

看看人类，或者生物体，作为一个计算机视觉人，我想提醒大家一件事，自然花了 5.4 亿年的时间来创造感知大脑。而创造语言，大脑只花了几十万年。所以感知是非常非常深刻的。

Bill Dally：这是一个很好的观点——感知领先语言几百万年。或者说感知更难，所以需要更长时间。人们对 AI 的一个担忧是它会扰乱就业市场。你可以说它会创造就业机会，也可能使其他工作变得不那么重要。你认为什么样的人类工作是 AI 或机器人永远无法取代的？

李飞飞：这是一个很棒的问题，也是一个危险的问题。自人类文明伊始，我们的祖先想象过的每一种工作基本上都是由机器协作完成的，像移动、飞行、计算等等。

工作的定义是什么？如果工作是一项任务，比如抓起某样东西或做一个煎蛋，我想它会被机器完成。但如果工作是人性的一部分，是定义我们创造力的一部分，定义我们的独特性，定义我们的意图、我们的同情心、我们与他人独特的情感联系，以及我们每个人对他人或对社会可能产生的独特贡献，那么我认为这些永远不会被完全取代。我们会利用机器来帮助我们更好地完成这类工作。我没有看到一个根本性的取代。

让我们再次回到医疗保健领域，我花了几十年时间在医院里照顾我年迈的父母。每次我带着父母走进病房，我都会看着人类照顾人类，或者人类需要人类来照顾人类。在这种关系和互动中有一些非常深刻

的东西，是任何机器人、AI、电脑、AR/VR，或者你认为的任何下一代技术都无法完全取代的。

在人性方面，在人与人的互动方面，有太多超越了计算、计算、机械的东西，我认为这些将是会保留下来并不断演变的工作。我们将越来越多地被机器赋予超能力，但作为人类的核心不会被取代。

Bill Dally：刚才你讲到了人类的同理心、情感联系、关怀还有创造力。创造力是人类的核心特质，但如果我们选择训练 AI 模型具有这些特征，我们能做到吗？我们可以建立有同情心的 AI 模型，让它能够在情感上与人联系吗？

李飞飞：在某种程度上可以。我再次强调一下，我从事计算机视觉工作。现在已经有了深度创造性的文本到图像以及文本到视频的生成（模型）。但我也认为，就像这里没有人能预测下一个爱因斯坦会是谁一样，这种创造力以及创造力的不确定性，将永远存在于我们人类社会中。所以无论你如何训练机器，你都无法训练出人类的智能或人类创造力。

这不仅仅是爱因斯坦，还有贝多芬、莎士比亚、梵高。还有太多太多了。而且不一定非要是那么聪明的人。我自己的孩子，我不认为任何机器都能创造出那样一个甜美、聪明、幽默的小家伙。

Bill Dally：这是独一无二的人性，机器永远无法取代。说到创造力，生成式 AI 正在做一些了不起的事情。OpenAI 最近推出了 Sora，你可以输入一个提示，就能得到一个看起来很棒的视频。也许当你在十年后写自传的续集时，是否会发展到如果你想看一部电影，只需写几行提示，它就会为你生成一部两小时的电影？

李飞飞：我不认为这需要十年，多生产一些 B200，它很快就会出现。就技术而言，我认为这即将到来——创建更长时间的生成性世界、

生成性故事情节、生成性角色互动的能力指日可待。

Bill Dally：在这样一个世界里，我们有 AI 程序生成大部分内容，那些在好莱坞或游戏工作室等地方的人类内容创作者的角色是什么？

李飞飞：这又回到了人类的独特性。我不知道你们中有多少人是宫崎骏、吉卜力工作室的粉丝。他们是最棒的，我就是喜欢一遍又一遍地看他们的电影。从计算机图形学的角度来看，它相当初级，他们不做皮克斯和梦工厂那样的（复杂）图形。然而那些独特的故事，比如《龙猫》的故事多简单，那部电影中人性的表达那么单纯，除了宫崎骏，没有人能创造出那样的东西。我认为这仍然是人性。AI 会创作电影，会创作娱乐人们的内容，但只有人，能利用 AI 创作那些能触动他人、启发他人或服务他人的内容，AI 做不到。我确实看到了这种共生的可能性。

（三）在 AI 时代的角色是什么?

Bill Dally：你的意思是 AI 基本上会接管这些创意事物的制作部分，制作真正引人注目的视频，让图像看起来很棒。但最终在情感层面上与人联系，去讲述一个能让人流泪的故事，这将是人类应该努力的部分。

李飞飞：没错。这也是我一直强调的一点，就是在这个机器时代，不要忘记我们的人性，不要忘记我们的尊严，不要忘记彼此的尊严和人性。这就是我们的核心，这就是我们的独特之处。这也是我们构建机器应用、使用机器的开端。

Bill Dally：你还有什么想传达给观众的吗？

李飞飞：我想说的一点是，GTC 是一个特别的会议。你们来到这里，是因为你们都以某种方式参与了 AI。

当我写出《我看到的世界》这本书并与全球各地的观众，特别是年轻观众交谈时，我经常被问到的一个问题。而每次有人问这个问题，我仍然会被触动——那就「我」在这个 AI 时代的角色是什么？

比如我，我不是计算机科学家，也不是斯坦福大学的理科专业；我不做软件工程，我不在有电脑的家庭长大的。我热爱跳舞。但所有这些来自各行各业的人都在问我，他们会在 AI 时代扮演什么角色。

因为 AI 看起来如此复杂，它有 7000 亿个参数，你怎么用自己的大脑来理解这么庞大的东西呢？然后是所有这些花哨的词，Transformer、生成式、扩散式，它们似乎离你每个人都很遥远。

但我真的想把它归结为：它是一个工具。它是一个需要一些数学和计算来实现的工具。但归根结底，人类不仅是工具的创造者，我们也是如何使用工具的决策者；我们是工具应用的创造者，我们也是工具的用户；我们是决定如何管理工具以及管理想要使用工具的人的选民。所以在参与 AI 方面有很多公民的可能性。

我特别希望年轻人，那些热爱艺术、热爱社区、热爱法律、热爱医学、热爱化学的人，无论你的兴趣是什么，都能以负责任的态度拥抱这项技术，你们实际上可以为让它变得更好、更好地使用它而有所作为。

这真的是我对每个人的恳求——你们在 AI 发展中是占有一席之地的，请加入我们，让 AI 变得更好。谢谢！

1294，COPU 会议纪要

COPU，2024.3.26

3 月 26 日陆主席主持召开 COPU 例会。

陆主席谈算力是人工智能发展的支撑，最近英伟达（NVIDIA）开

发了大幅提升算力的超级芯片系统，请安泱简要地介绍这方面的情况。

安泱报告提供的图表见如下附件：

	B200	H100 SXM	A100	V100	昇腾910b
内存带宽	8TB/s	3.35TB/s	1555GB/s	900GB/s	141GB/s
内存容量	180GB	80GB	80GB	16GB/32GB	32GB
互联带宽	1.8TB/s NVlink5	900GB/s NVlink4 18 links	600GB/s NVlink3 12 links	300GB/s NVlink2 6 links	100Gb/s(12.5GB/s)
FP32 vector	未公开	67TFLOPS	19.5TFLOPS	15.7TFLOPS	不详
FP64 vector	未公开	34TFLOPS	9.7TFLOPS	7.8TFLOPS	不详
FP16 tensor	2250TFLOPS x2	1979TFLOPS	312TFLOPS	125TFLOPS	280TFLOPS
FP32 tensor	1100TFLOPS x2	989TFLOPS	156TFLOPS	NA	不详
注明：华为官方未公开华为升腾计算性能数据，本表格数据来自华为升腾的经销商华鲲振宇网站					
https://www.schkzy.com/product_solution/product/product_detail?id=1714162282952417281					

本次会议认为：人工智能（+量子技术）是第四次工业革命呼之欲出的主力，语言大模型训练需要强大的算力。安泱在附件中介绍英伟达超级芯片系统 GB200 的算力是其生产的 H100 芯片算力的 30 倍，A100 芯片的 100 倍。华为昇腾 910b 芯片（AS Cend 910b ）的算力与英伟达 A100 相当。现在 GB200 的算力已提升到 A100 的 100 倍，当然也是 910b 芯片算力的 100 倍，华为很早就掌握芯片堆叠技术，堆叠制成超级芯片也不难。英伟达曾把华为开发的人工智能芯片的算力当成自己潜在最大的对手。由中国科学院和中国电子信息产业集团联合研制的世界首条光量子芯片生产线已经建成，它产出高端芯片可绕过光刻机，开发出来的智能算力较之 A100 或 910b 提高 100 倍。

陆主席谈大家可能已看到目前正在网上传播一篇报道：斯坦福大学教授李飞飞与英伟达首席科学家比尔达利（Bill Dally）关于人工智能的对话，请鞠东颖宣读这篇报道。

本次会议认为，李飞飞教授关于人类的同情心、情感联系、人间关怀、人的意图、人性互动，还有创造力（创造力的不确定性）是无法对机器训练出来的，将永远存在我们人类社会中的观点，对独一无

二的人性，机器（人工智能）将永远无法取代的观点，似乎与我们最近发表人工智能大师杰弗里·辛顿（Geoffrey Hinton）演讲的观点有出入（辛顿认为数字智能必然战胜或取代生物智能、大语言模型具有对世界的理解力，人工智能会产生致命的自主武器、超级智能不久就会出现，超级人工智能会给人类带来威胁）。本会议谨提供两者不同观点供大家参考。李飞飞还表示：大语言模型并非是 AI 大模型未来发展方向的观点也值得研究。

本次会议还认为，制约人工智能发展的除算力外还有能源。人工智能 ChatGPT 一天的耗电量为 50 万 KWH（相当于数万个美国家庭一天的用电），谷歌在发展 AI 仅训练阶段耗电就高达 23 亿度。中国与美欧日比具有很大优势，在发展新能源（太阳能、风能）中独占鳌头，在研发第三代核电站时并驾齐驱，发展第四代核电站已经启动，在发展绿色能源水电方面、特高压输电（1000 千伏交直流输电）方面也是全球第一，在这方面的讨论以后再谈。

在讨论大模型或人工智能发展中，从 2015 年起我们（COPU）一直坚持发展基于开源的 AI(或基于开源的新一代信息技术），OpenAI 总裁奥特曼在发展大语言模型（MPT）方面做出了重大贡献，但他违背初心摒弃开源走闭源之路，我们写了一篇：“OpenAI Sora 及其国内外开源竞争者和国内开源复现 Sora 者”，在国外谷歌的 Genie、马斯克的 Grok(以及 Facebook 的 LLama-2,3，阿布扎比的 Falcon），在国内我们最近调查到的北京大学的 Open-Sora、PKU-YUAN 的 Open-Sora-plan，Sora Webei/Open AIs Sora、潞晨科技 Colossal-AI 的 Open-Sora 等均是开源的，拥抱开源是大势所趋，对大模型或 AI 的发展产生重大影响。

1295，再评文本生成视频大语言模型 Sora

陆首群 2024.3.12

在本次会议上再评文本生成视频大语音模型 Sora（人工智能）。可是 Sora 不开源。在国内为了扩大 Sora 的利用价值，必须找到、开发、复现 Sora 的竞争者或开源 Sora 者。

在会上讨论对生成视频 Sora 及其竞争者、国内复现开源 Sora 者。

OpenAI 于 2024 年 2 月 15 日发布由文本生成视频的大语言模型 Sora（号称世界模拟器），COPU 在 2004 年 2 月 20 日和 2024 年 2 月 27 日两次例会上进行了学习、讨论和点评，正当我们对 Sora 开始深入研究的时候，我们发现 Sora 是闭源的（据 OpenAI CEO 奥特曼讲今后 Sora 也不准备开源，同是 OpenAI 创始人马斯克指责奥特曼这是违背初心的）。

就在此时，与 Sora 竞争的信息也不时涌来：

由一拳 AIGC 工作室开发的 LTX Studio 震撼来袭，声称 Sora 被超越。

由谷歌研发的由文本生成视频的 Genie 模型（+Ada）是开源的（号称基础世界模型）。

在国内，涌现出开源的北大 Sora，还有河北（查证为海淀）潞晨科技旗下的 Colossal-AI 开源团队开发的开源 Sora（open-sora），即 hpcaitech/Open-Sora）;Intel 邓伟先生还向我们介绍刊登在 Github 上的开源 Sora:

PKU-YUANGroup/Open-Sora-plan ，Sora Webei/Open AI’s Sora。

现在简单介绍 Colossal-AI 团队开发开源 Sora 的情况：

该团队在 OpenAI 的 Sora 实际模型结构未知的情况下复现 Open-Sora 的，他们参照三种常见的多模态模型结构：ada LN-Zero，

Cross-attention，in Context Conditioning（token Concat），复现 Open-Sora 的。随后他们又对 Open-Sora 进行性能优化，对未来 Open-Sora 进行迭代。

我对他们强大的人才支撑感兴趣，介绍如下：

Colossal-AI 是潞晨科技旗下著名的开源平台，2023 年 2 月便复现 ChatGPT 模型将其开源，目前在 GitHub 已超过 36000 颗星，显示技术实力强大。

开源地址：

https://github.com/hpcaitech/open-sora?tab=readme-ov-file

■ 潞晨科技创始人尤洋教授是加州大学伯克利分校博士，被福布斯评选为 30 岁以下精英（亚洲 2021），IEEE-CS 超算杰出新人奖；

■ 潞晨科技首席战略官 James Demmel 是加州大学伯克利分校杰出教授，ACM Fellow，IEEE Fellow，美国科学院院士，美国工程院院士，美国艺术与科学院院士；

■ 潞晨科技 CTO 卞正达，毕业于新加坡国立大学，师从尤洋教授，7 年高性能 AI 系统经验，Colossal-AI 系统核心开发者。

1296，语言大模型文本生成视频 Sora 讨论会

COPU 与北京大学 Open-Sora Plan 团队袁粒老师就关于复现 Sora 进行了学术讨论：

袁老师谈到，北京大学 Open-Sora 开发/复现三个模块：

① 编码：Video VQ-VAE

② 训练：Denoising Diffusion Transformer

③ 条件注入：Condition Encoder

谈到条件注入：

- 文本信息、三维信息、二维视觉信息形成可控信息，
- 理解物理场景发展规律、预判视觉下一步动作、迎接周围环境挑战

北大 Open-Sora 发展三阶段：

- 开发/复现基本模型框架（含工作机制）
- 开发基本应用，
- 开发扩充应用。

在北大 OpenSora 讨论会上补充问答：

陆问：Facebook 的 YanLeCun 大师对 Sora
的物理场景不满意，他进行了重新制作，你们如何看待这个问题？
袁答：这里谈的物理场景指理解的物理场景，不是生成的物理场景。
陆问：在北大 OpenSora 三大模块中，增强算力在哪里？
袁答：训练是增强算力的核心，编码次之，条件注入微调。
陆谈：希望今后对 Open-Sora 更名。

1297，人工智能下一个浪潮是具身（embodiment）智能

引自吴朝晖院士在 2023 中关村 AI 大模型发展论坛上的报告
（2023.5.29 由 COPU 发表）

他指出，比自然语言大模型更高级的是多模态的具身（embodiment）智能，人工智能下一个浪潮是智身智能，或具身智能是未来人工智能的形态。

所谓具身智能是将智能算法（专家理论）与机器人的感知、行动行动和环境交互能力相结合，籍以完成各种多模态任务。

COPU 还摘录了 NVIDIA（英伟达）CEO 黄仁勋谈具身智能：

他也认为，人工智能的下一个浪潮是具身智能。具身感知的信息是多模态信息，即视觉、触觉、听觉、嗅觉等感知的信息。具身可让机器人像人类那样感知丰富多彩的外部世界，为大模型“大脑”配备智能化“身体”。具身智能是由人、物理世界、机器人、虚拟世界组成四元世界交互作用产生的。主要在 LLMs 基础上结合具身实行全面创新。

李飞飞也介绍具身智能新成果（样本控制机器人）。

机器接入大模型直接听懂人话：大模型接入机器人，把复杂指令转化成具体行动规划，无需额外数据和训练，从此人类可以随意地用自然语言给机器人下达指令。

大语言模型+视觉语言模型就能从 3D 空间中分析出目标和需要绕过的障碍，帮助机器人做行动规划。

陆首群谈具身（embodiment）的历史演变：

在 IBM“沃森健康（Watson Health）”研究团队某些核心人员帮助下，陆首群曾研究、评论 IBM“沃森健康”与安德森癌症中心合作（历时 2011-2017 共 7 年）探索采用医疗人工智能（知识工程）进行医疗的效果。最终陆的结论是 IBM 在全球率先探索采用医疗 AI 知识工程及与安德森癌症中心的合作是以失败告终。的，但 IBM“沃森健康”留下了两项 AI 遗产：一是欲进行 AI（知识工程）研究时，必须选择“大规模语义网络”为依托（这就是后来发展起来的“大模型”的先导），二是提出了具身（embodiment）概念，当时提出的这个概念还是比较原始的，即将医学 AI 科学家的理论指导与临床医生的实践经验相结合（后来发展的具身智能）。

1298，马斯克旗下公司脑机接口手术成功

2024 年 1 月 30 日讯

Neuralink 进行脑机接口研究植入大脑芯片，目前恢复良好。

2023 年 5 月美国 FDA 批准 Neuralink 公司可以开始人体试验，该公司于 2023 年 9 月启动人体试验招募计划。根据 Neuralink 的技术路径，患者需要接受开颅手术。

具体操作将由一台大型机器人来完成，机器人将电极和被称作“神经蕾丝”的极细电线插入患者大脑，这些电极和无线将被用来读取脑电波，然后将信息无线传输到电脑应用程序上，从而使得人们能够仅用自己的想法就可以控制计算机光标或者键盘。

1299，中国无线微创脑机接口临床试验取得突破性进展

清华大学官微 2024 年 1 月 31 日消息

2023 年 10 月 24 日，清华大学医学院洪波教授带领团队设计研发的无线微创植入脑机接口 NEO（Neural Electronic Opportunity），在宣武医院成功进行首例临床植入试验。宣武医院赵国光院长、单永治主任团队主持手术规划及植入手术。2024 年 1 月 29 日，联合团队召开临床试验阶段总结会，宣布首例患者脑机接口康复，取得突破性进展。

接受此次脑机接口辅助治疗试验的对象，是一位因为车祸引起的颈椎处脊髓完全性损伤（ASIA 评分 A 级）的 54 岁男性患者，此前长期处于四肢瘫痪状态。在将脑机接口处理器通过神经外科医生的操作植入患者颅骨后，成功采集感觉运动脑区颅内神经信号。手术后 10 天，患者出院回家。在经过三个月的居家脑机接口康复训练，患者可以通过脑电活动驱动气动手套，实现自主喝水等脑控功能，抓握解码准确

率 90%。首例患者通过无线微创脑机接口成功实现脑控抓握矿泉水瓶，在无线微创方面实现突破，最快两年后可实现成熟应用。

1300，在小米打造新一代自动语音识别 Kaldi

Daniel Povey，2021.09.13

自动语音识别（Automatic speech recognition，ASR）技术是使人与人、人与机器更顺畅交流的关键技术。在计算机刚兴起不久时，人们就希望机器能理解自然语言，拥有智能。作为实现人工智能不可或缺的一环，语音识别这个研究领域已经活跃了半个多世纪。20 世纪 80 年代至 90 年代是语音产业的一个爆发期，隐马尔可夫模型（Hidden Markov Model，HMM）的应用，使大规模连续语音识别成为可能，在进行人机交互时，用户得以摆脱字正腔圆、一词一顿的刻板方式。在过去的十几年间，随着深度学习技术的强势崛起和以 GPU 为代表的算力硬件的出现，语音识别的使用体验又得到了显著的提升。深度学习技术带来的使用体验的提升，使语音技术更多的应用于商用，促成了语音产业和语音数据之间的良性循环，相比传统模型，基于深度学习的语音识别系统能利用持续增长的数据量来提升识别性能，而识别性能的提升，又会激发出更多的产业应用。深度学习介入语音识别以来，语音相关产业发展迅速，产品形态五花八门，随着语音输入、语音搜索、智能助手等产品的出现，一场人机交互的变革正在我们身边发生。

开源生态与 Kaldi 的崛起

虽然深度模型的引入和算力的提升为人工智能注入了新的活力，但与其他 AI 技术相比，自动语音识别技术本身链路复杂、模块多样、领域知识众多，这给语音技术的研究设定了较高的门槛。在语音识别技术的发展过程中，开源软件一直扮演着举足轻重的角色，早年比较

有代表性的作品是 HTK 和 Sphinx 这两个工具集。这两个工具集都能够完成从模型的训练到产品原型搭建等一系列工作，20 世纪 90 代开源以来，大大地降低了语音识别和相关领域的研发门槛，并催生了一批以语音识别为核心技术的公司。

而在最近的十年里，Kaldi 开源项目逐步取代了 HTK 和 Sphinx 的统治地位，成为了最流行的开源语音工具包。

起步

Kaldi 项目起源于 2009 年的约翰霍普金斯大学的夏季研讨会（The Johns Hopkins University Summer Workshop）。那一年夏季研讨会的其中一个主题是 “以低成本的方式构建高质量语音识别（Low Development Cost，High Quality Speech Recognition for New Languages and Domains）”， Daniel Povey 博士主持了这个研讨会。他想把子空间高斯混合模型（Subspace Gaussian Mixture Model，SGMM）推介给研究者，Kaldi 工具包旨在实现这一想法，因为要在 HTK 中实现这个想法很困难，而且当时没有其他通用的语音识别开源软件工具包可以选择。

发展

Kaldi 项目一经发起就吸引了大量研究者的关注，在 2010 年的研讨会上人们讨论了 Kaldi 作为一个语音工具包的功能，并开发了自有的训练脚本。2011 年 5 月 14 日，Kaldi 正式发布初版的代码，从此代码库的开发和维护主要由 Daniel Povey 博士主导，走上了高速发展的

轨道。2011 年研讨会上，开发了基于 GMM 和 SGMM 的区分性训练。2012 年研讨会上，乘着深度学习的东风，开发了基于 nnet1 的区分性训练和 Stacked-bollteneck 网络。2014 年的研讨会上，研究并完善了神经网络的内部结构和语音置信度分析等内容。2015 年研讨会上，Daniel Povey 博士开始了 Kaldi 中 nnet3 的开发。nnet3 也叫 chain model，此后数年间一直是语音识别研究和产品化部署的中流砥柱。

特点

Kaldi 作为一个通用的语音工具包，兼具灵活易用及高效的特点，它的源代码由 C++写成，并且尽可能实现通用的算法，避免使用只为特定任务服务的代码。这使得它非常容易复用和扩展，通过简单修改和重构就可以构建出可产品化部署的系统。另外，Kaldi 也是非常现代的，里面涵盖最新的语音识别技术，这些最新的算法都以一个个示例脚本（recipe）的形式随 Kaldi 的代码一起发布，人们只要根据示例脚本里的指示，就可以一步一步构建出一个优异的 ASR 系统。当然，Kaldi 得以成为最受欢迎的语音工具包的原因是它的开放性。Kaldi 在开源许可上选择了宽松的 Apache 许可证 2.0 版，这意味着不仅 Kaldi 社区的人们可以参与开发并自由使用 Kaldi 软件，个人、研究机构、甚至商业机构都可以相对自由地利用 Kaldi 进行商业或者非商业的活动。

影响力

过去十年间，人工智能创业风起云涌，语音方向的创业公司如雨后春笋般爆发出来，他们中的多数都基于 Kaldi 来创建自己的语音产品，有些哪怕不是直接使用 Kaldi 软件，也或多或少借鉴了 Kaldi 的代码和设计思想。可以说，Kaldi 的诞生和发展，极大的降低了语音识别的入门门槛，让这一研究领域得以“飞入寻常百姓家”，也间接地催生了这一波人工智能的创业热潮。以小米为例，小爱同学自 2017 年

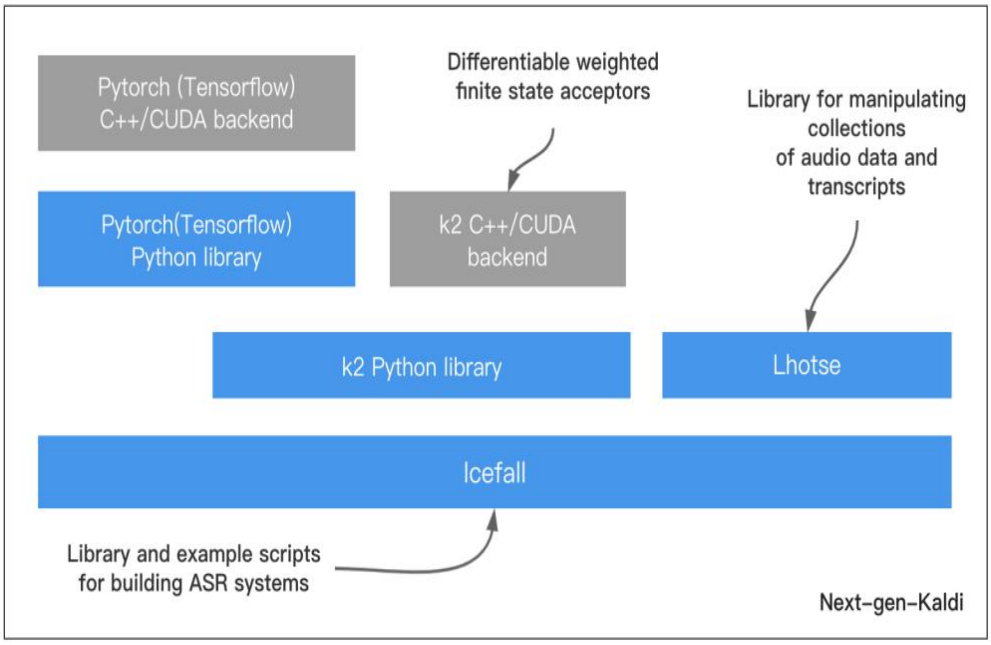
上线至今，累计唤醒次数 726 亿，累计激活设备 2.51 亿台，月活用户数达到 7840 万。 而这一切的背后，都离不开小米语音团队依托于 Kaldi 之上打造的全链路语音系统，包括适用于各种场景的不同语音模型，如远近场语音唤醒、离在线语音识别、说话人识别等通用模型，以及口语评测、语种识别、语音情绪识别等适用于具体场景的特定模型。伴随着小米 AIoT 产品线的扩展，一个又一个的酷炫语音产品相继发布，如 MIUI 声控拍照、千人千面的内容点播、跨设备的声纹追剧、基于童音识别的内容限制等功能，大大方便了普通用户和家庭的生活。

新一代 Kaldi 的诞生

近几年，深度学习及语音识别技术又有了新的进展，端到端语音识别模型逐渐流行起来，准确率也得到了进一步的提升，另外 Pytorch、Tensorflow 等通用的深度学习工具包也逐渐普及起来。Kaldi 使用自有的神经网络架构，无法利用现有深度学习框架的优势快速迭代模型，这使得 Kaldi 难以适应这些新的变化。于是，开发新一代 Kaldi 以适应当前趋势和面向未来发展迫在眉睫。2019 年 Daniel Povey 博士加入小米，成为小米的首席语音科学家，随即开始着手新一代 Kaldi 的部署和研发，经过两年多的探索和开发，新一代 Kaldi 的三个子项目（k2,Lhotse,Icefall)全面部署完成，并相继发布。Daniel Povey 博士表示：“虽然端到端模型的兴起和 Pytorch 等深度学习框架的流行是开发新一代 Kaldi 的主要动力，但新一代 Kaldi 的目标不仅仅是赶上或者稍微领先现有的语音识别库，而是要根本地改变实现语音识别

的方式。”

如下图所示，新一代 Kaldi 包含三个部分，k2 是一个可微分的加权有限状态转换器，是新一代 Kaldi 的核心部分；Lhotse 负责训练数据的准备；Icefall 则是一个训练脚本集合，通过这些训练脚本可以快速构建一个可用的 ASR 系统。新一代 Kaldi 之所以将整个项目分为三个部分，一方面是为了降低耦合性让软件依赖变得简单，方便用户使用。更重要的是，各部分可各自发挥所长，Lhotse 作为数据准备部分，不仅可以用在 Icefall 项目里，也可以用在任意其他语音识别库里来处理音频和文本数据。而 k2 作为序列建模的高效工具包，不仅可以用来做语音识别，也可以用来做手写文字识别等其他任务。相信在不久的将来，随着新一代 Kaldi 的推广和普及，k2 和 Lhotse 都有可能成为语音甚至 NLP 等序列建模领域使用最为广泛的工具包。



k2

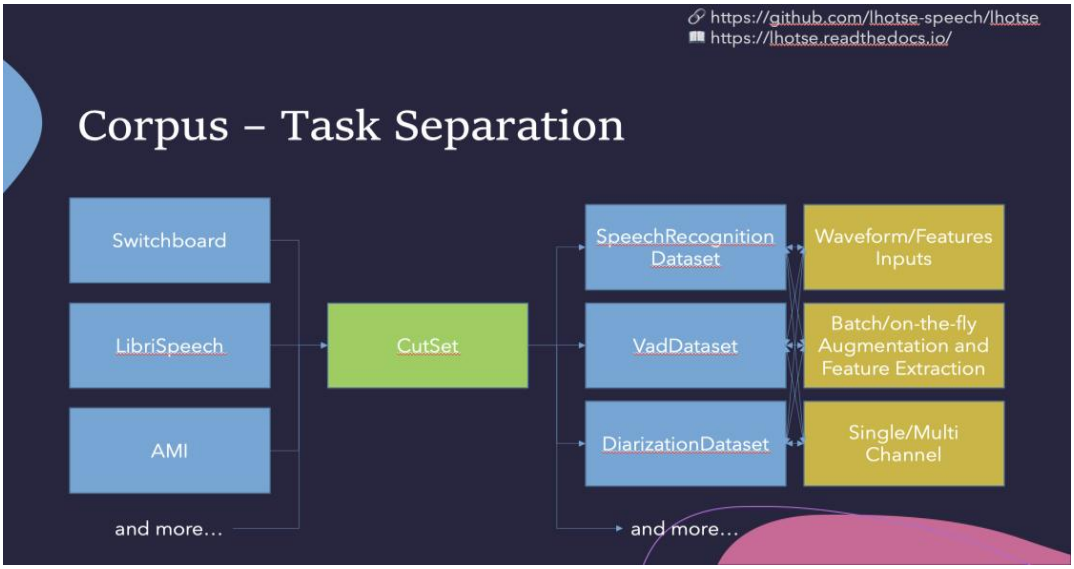
k2 作为新一代 Kaldi 的核心，它的核心贡献在于，将加权有限状态转换器(Weighted Finite State Transducers, WFST)和相关算法无缝地集成到基于 Autograd 的机器学习工具包，如 PyTorch（已完成支

持）和 TensorFlow 中。WFST 是语音识别领域最为核心的数据结构，可以用来构建诸如“音标->词->句子”的状态转换概率图。支持 WFST 可导意味着我们可以做很多以前很难做到，甚至做不到的事情，如消除以往语音识别任务中训练跟解码过程不匹配的问题、多轮（可求导）的语音识别过程等。k2 的所有 FST 操作都构筑在 Daniel 博士独立提出的多维不规则数据结构 RaggedTensor 上，这些操作从一开始就被设计成可并行的，所以对于 GPU 加速有着天生的优势。k2 还支持在解码图中嵌入任意辅助信息，这些辅助信息都会通过映射关系适配 FST 的各种操作，这使得训练和解码过程中能够利用的信息大大增加了。k2 可以用来很方便地实现很多现有的语音识别模型，如 CTC、LF-MMI、RNN-T 等。值得一提的是，Facebook 在 2020 年发布了类似的项目 gtn，它跟 k2 几乎是同时相互独立地开始开发的。但相比于 gtn，k2 不仅实现了更多的 WFST 相关算法，而且 k2 还高效地支持 GPU（gtn 目前只支持 CPU)。高效的 GPU 实现，使得快速的语音识别训练迭代成为可能，并且也大大加快了解码速度。目前，我们的解码速度已经是实时的 300 倍左右，而且还在进一步的持续优化中。随着 k2 的完善和产品化落地，整个语音识别全链路 GPU 加速将变得非常简单。

Lhotse

Lhotse 是训练数据准备部分。不同于上一代 Kaldi 大量使用 shell 脚本和 Linux 管道，Lhotse 全部使用 Python 写就，支持 Pytorch Dataset 的 API（如 map-style 的 dataset，动态 batch size， 分布式训练的支持等），方便易用。Lhotse 设计了通用又不失去灵活性的接口，以适应包括语音识别，文本转语音等多种语音任务。用户更可以基于 Lhotse，来方便地实现针对自己特定任务的接口，来操纵各种不同的音频元数据和文本。Lhotse 还支持多种 IO 和序列化方式（如支

持从 HTTP/S3/GCP/Azure URLs 下载音频，支持 JSONL 等）。Lhotse 引入了 Audio Cuts 的概念，将训练数据自动地组织为一组组 Cuts，并基于这种表示，提供了 on-the-fly 的数据混合，裁剪，增强和特征提取等操作，从而在不影响数据处理效率的前提下，降低了数据存储所需空间。此外，Lhotse 还提供了很多公开数据集的数据处理脚本，用户可以直接使用这些脚本，来进行相关语音任务的数据处理工作，这大大降低了用户在某个数据集上进行实验的前期成本。



Icefall

Icefall 是训练脚本集合，同 Lhotse 一样，它也是一个纯 Python 项目。用过 Kaldi 的人都知道，Kaldi 里有大量的基于不同数据集的示例脚本，这大大降低了用户的学习成本。但同时也有一个缺点，就是示例脚本集合太过庞大，代码耦合过于紧密，维护成本较高。考虑到这一点，Icefall 将不再是一个大的脚本集合，而是会在提取公共组件的同时，将不同数据集的示例脚本独立组织，以方便用户的学习和使用。此外，由于将数据准备部分单独放在 Lhotse 项目中，核心计算部分单独放在 k2 中，Icefall 项目只需要关注语音识别模型的结构定义部分，这大大降低了整个语音识别过程的耦合性，也方便了网络结构

的复用。

我们不仅在其中展示了如何使用 k2 和 lhotse 来实现现有的各种不同的语音识别模型，如基于 Transformer/Conformer 的 CTC 和 LF-MMI 等，更重要的是，我们着重展示了 k2 何以能从根本上改变人们实现语音识别的方式：我们实现了多轮解码(mutiple pass decoding)的示例，以及我们称之为“双向建模”(bidirectional modeling)的模型。基于深层模型及这种多轮解码的优势，我们可以大大提高语音识别模型的建模能力，从而降低词错误率。截至 2021 年 9 月 1 日，我们已经在 LibriSpeech 数据集上取得了 2.57%的词错误率，而且还在持续的进一步优化。

值得注意的是，这样的建模方式，我们很难使用现有的语音识别框架来完成。而因为我们在 k2 中实现了所有的 FSA 操作的可导性，使得我们可以使用几十行的代码，就可以完成这样复杂的模型结构。这还只是我们实现的可以使用 k2 来做的基本模型示例，用户可以基于 k2 来实现他们自己的各种各样的新想法，如在网络中加入 phone 的 embedding 信息，支持带置信度的识别等。总之，我们希望通过新一代 kaldi，能为语音识别领域打开一扇崭新的大门。

展望新一代 Kaldi

新一代 Kaldi 同上一代 Kaldi 一样，依旧使用高效的 C++代码实现，以方便工业界的使用。更重要的是，由于我们将 k2 的 C++代码都（使用 pybind11）包装到了 Python, 模型的训练迭代都可以使用纯 Python 代码完成，这大大方便了用户的使用。基于 icefall 中的示例脚本，工程师们可以很容易地基于自己公司产品的数据集进行修改，进而快速地搭建语音识别系统，这样线上数据反馈和模型迭代更新的流程就大大简化了，这将极大缩短模型更新的周期。除此之外，由于我们也

高效的支持GPU,如果用户或企业基于GPU来部署新一代Kaldi的模型，也将获得GPU对模型(解码)的加速优势，再加上神经网络GPU Inference框架的使用，全链路GPU加速的语音识别系统将成现实，这将大大提高模型最终的识别速度。

作为一个通用的序列建模工具，新一代Kaldi不仅可以提升语音识别的准确率，它的使用和发展也会给其他的序列建模任务带来新的启发。比如，由于k2实现了可导的WFA，基于新一代Kaldi构建的语音识别系统就有可能为下游的NLP系统提供更加丰富的信息表示，而不仅仅是一个语音识别的结果。人们可以利用这个带有置信度等其他一些信息的词格（lattice）来进一步调优（fine-tuning）下游的自然语言理解任务，甚至有可能将语音识别和自然语言理解的任务放到一起来训练，实现真正的端到端自然语言理解系统。

相信随着新一代Kaldi的全面落地，它对语音识别的改变也将最终影响到普通用户。高效的解码速度和低WER(词错误率)将为普通老百姓带来更加美好的语音识别产品体验。以小爱同学为例，作为小米“手机×AIoT”战略布局中的一环，小爱同学承担着小米AIoT生态中极为重要的角色。作为一款智能生活助手，通过它，用户可以连接到各种各样的AIoT设备并与它们产生互动:智能音箱、手机、电视、智能手表、儿童故事机、车载后视镜等。而通过与空气净化器、扫地机器人、电饭煲、台灯、空调等上亿智能家电的连接，小爱同学更可以帮助每个用户打造属于他们自己的整体智能家居体验。借助小爱同学，用户可以通过语音命令小爱音箱播放音乐，可以控制扫地机器人扫地，可以给小孩讲故事……。小爱同学甚至可以主动地学习你的生活模式，在你回家前帮你提前打开空调，在你进门的时候自动打开客厅的灯，在你要睡觉的时候拉上窗帘……成为你最贴心的的生活助手。 随着下

一代 Kaldi 在小米产品线的逐步落地，相信在不久的将来，小米将和其他公司一起，为普通用户带来更加完善的 AIoT 产品体验。正如崔宝秋博士在 2020 年举行的 Kaldi 线下交流会中所讲，新一代 Kaldi 项目的诞生和发展将是围绕 Kaldi 的“四赢”局面：Kaldi 项目赢，小米语音赢，全球的 Kaldi 社区赢，所有跟 Kaldi 相关的中小型公司赢！

敬请关注联盟微信公众号
COPU开源联盟

扫描二维码
获取往期资料

中国开源软件推进联盟秘书处　　电话：+86 010-88558999

联盟公共邮箱：office@copu.org.cn　　联盟官网：http://www.copu.org.cn

地址：北京市海淀区紫竹院路66号赛迪大厦18层