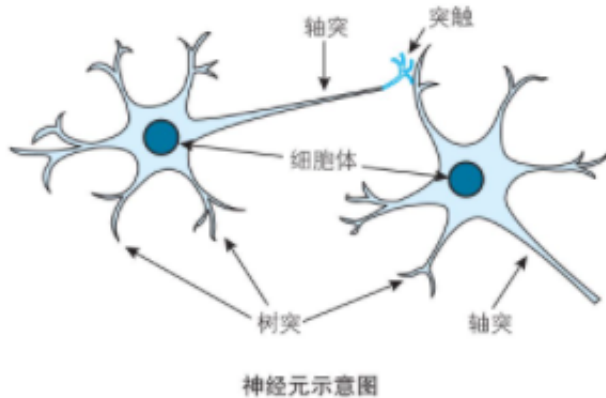


ChatGPT的出现，让人们从科技和人文的角度都开始重新思考人类与机械、人类与科技、人类与人工智能之间的关系。作为一个大型人工智能语言模型，ChatGPT是怎样理解人类的逻辑的？我们要怎么看它，它又如何观照这个世界？一起来看看这篇文章，希望能为你带来启发。



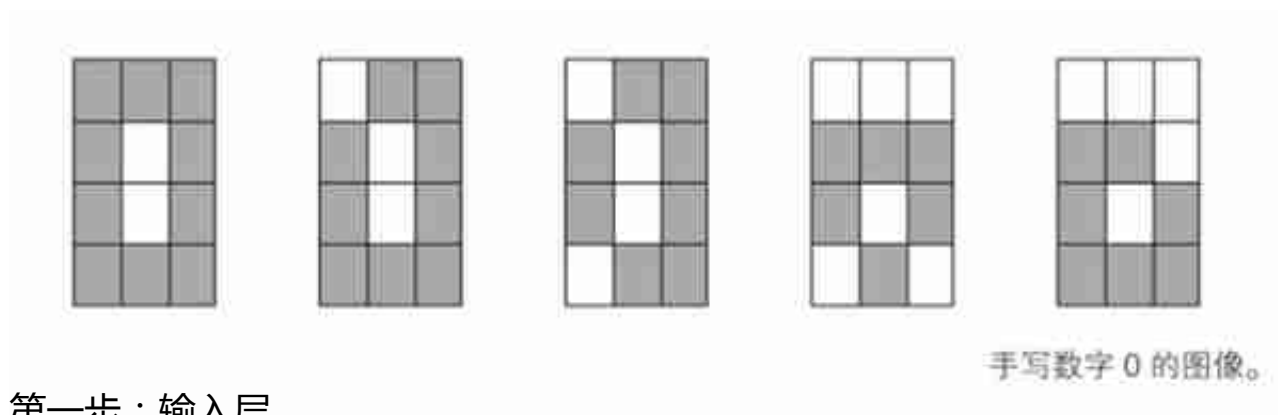
神经元主要由细胞体、轴突、树突等构成。树突是从其他神经元接收信号的突起。轴突是向其他神经元发送信号的突起。由树突接收的电信号在细胞体中进行处理之后，通过作为输出装置的轴突，被输送到其他神经元。另外，神经元是借助突触结合而形成网络的。

图片来自《深度学习的数学》一书。

以上原理，用计算模拟和解释，就是：神经元在信号之和超过阈值时点火，不超过阈值时不点火。

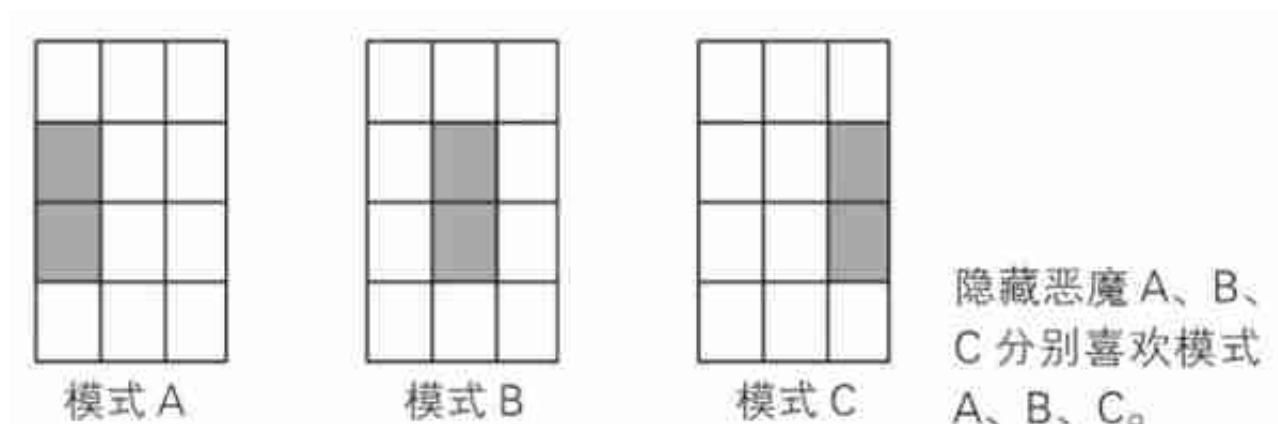
20世纪五六十年代，[Oliver Selfridge](#)创造了名为“鬼域”的概念。这是一个图案识别设备，其中进行特征检测的“恶魔”通过互相竞争，来争取代表图像中对象的权利。

“鬼域”是生动的关于深度学习的隐喻，如下图：[Oliver Selfridge's Ghostly Demons](#)

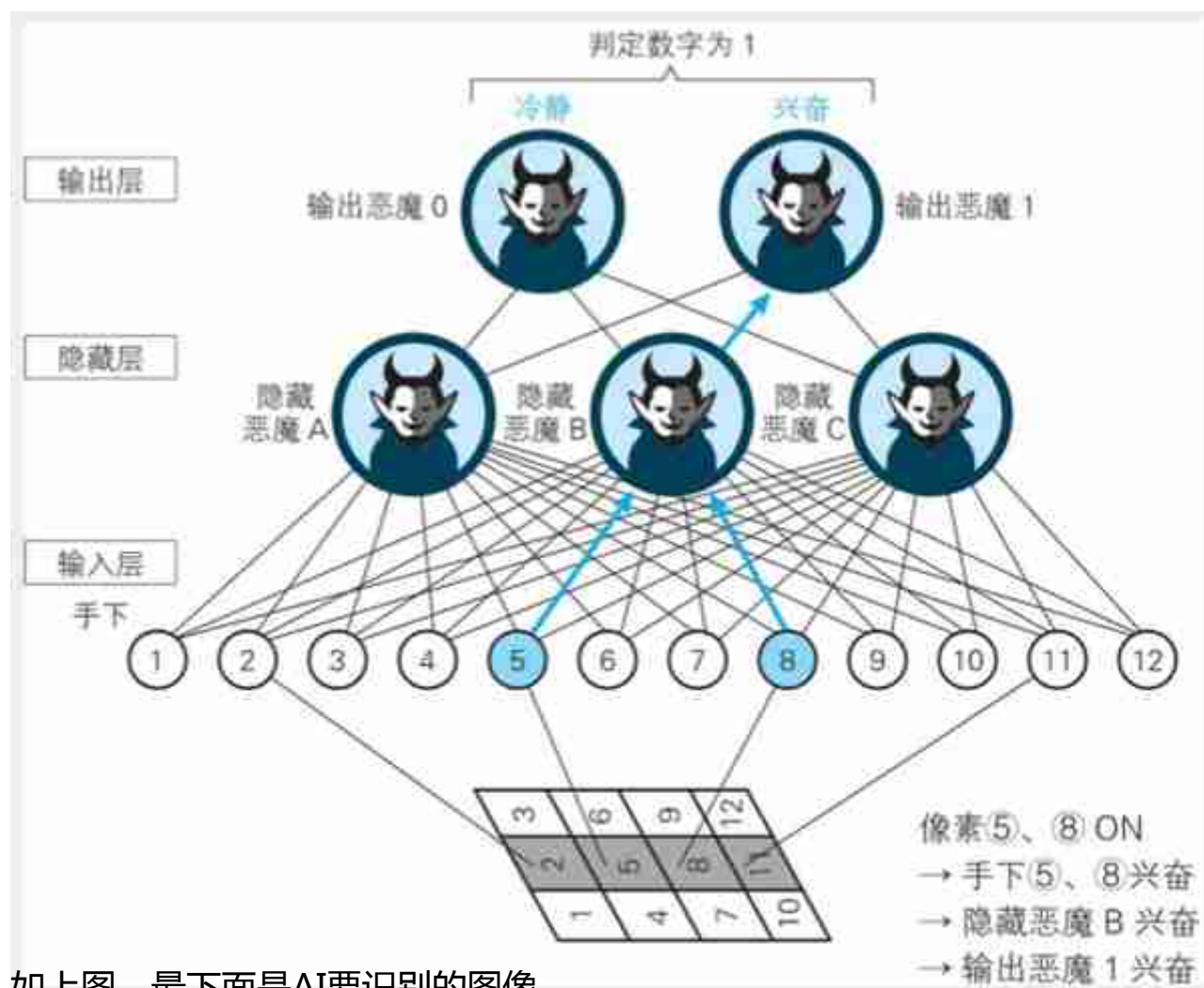


第一步：输入层

12个格子，相当于每个格子住一个人，分别编号为1-12。如下图。



不同的模式对应着相应的数字格子的组合。如下图。模式A对应的是数字4和7，B对应5和8，C对应6和9。



如上图，最下面是AI要识别的图像。

首先，输入层的2、5、8、11点火；

然后，隐藏层5和8所对应的特征被提取，“模式B”点火；□□

最后，输出层的1被对应的“模式B”点火。□

所以，“大恶魔”识别出图像为数字1。□

在上面的例子里，AI可以精确地识别出0和1，但它并不懂0和1，它的眼里只有像素。

可这么说，似乎过于拟人化了。人类又如何懂0和1呢？

人类不也是通过双眼输入，通过迄今仍是宇宙间最大谜团的大脑神经网络（更加复杂、强大且节能的隐藏层）提取特征，然后通过大脑的某个部位再进行自我解释的吗？

辛顿曾在采访中提及，认知科学领域两个学派关于“大脑处理视觉图像”的不同理念：□□□□□

一派认为，当大脑处理视觉图像时，你拥有的是一组正在移动的像素。如同上面的演示；

另一学派偏向于老派的人工智能，认为是分层、结构性的描述，脑内处理的是符号结构。

辛顿自己则认为以上两派都不对，

“实际上大脑内部是多个神经活动的大向量。”而符号只是存在于外部世界。

不管怎样，神经网络的模型有用，并且非常有用。□□□□□

不愿意和外行分享专业话题的辛顿，用如下这段话生动介绍了“神经网络”：

首先是相对简单的处理元素，也就是松散的神经元模型。然后神经元会连接起来，每一个连接都有其权值，这种权值通过学习可以改变。神经元要做的事就是将连接的活动量与权值相乘，然后累加，再决定是否发送结果。如果得到的数字足够大，就会发送一个结果。如果数字是负的，就不会发送任何信息。

你要做的事就是将无数的活动与无数的权重联系起来，然后搞清如何改变权重，那样就行了。问题的关键就是如何改变权重。

## 八

神经网络和深度学习经历了并不算短暂的黑暗期。

从上世纪80年代开始的整整30年间，只有很少一部分相关研究者义无反顾地投身其间，他们饱受怀疑，也几乎拿不到科研经费。

也许是由于这个原因，深度学习三巨头辛顿（Hinton）、本吉奥（Bengio）、杨立昆（LeCun）似乎都和加拿大有些关系，他们退守在那里研究、教学、读书。这倒是很符合那个“傻国家”的气质。

一个让人“心酸”的细节是，2012年辛顿带着学生在ImageNet 图像识别比赛上拿

了冠军，商业公司蜂拥而至。辛顿教授开出的商业报价，只是区区一百万美元。（后来谷歌以4400万美元“中标”。）

“老派” AI，使用明确的一步步指令指引计算机，而深度学习则使用学习算法从数据中提取输入数据与期望输出的关联模式，正如上一节的演示。

众所周知，漫漫长夜之后，随着人类计算机算力和数据的指数级增长，深度学习一飞冲天，从阿尔法狗一战封神，再到ChatGPT征服全球。□□□□

为什么是Open AI，而不是DeepMind？我对此略有好奇。

OpenAI的联合创始人兼首席科学家伊利亚·萨特斯基弗，是辛顿在多伦多大学带的学生。

他似乎延续了辛顿对深度学习的信仰，并且勇于全力下注。□□□□

辛顿认为 “深度学习足以复制人类所有的智力”，将无所不能，只要有更多概念上的突破。例如“transformers”利用向量来表示词义的概念性突破。

此外，还要大幅度增加规模，包括神经网络规模和数据规模。例如，人脑大约有100万亿个参数，是真正的巨大模型。而GPT-3有1750亿个参数，约比大脑小一千倍。

神经网络模仿了人类的优点：处理有大量参数的少量数据。但人类在这方面做得更好，而且节能许多倍。

先行一步的DeepMind，其发展方向和速度，除了陷入与谷歌的“商业VS科研”的两难纠缠，还不可避免地受到哈萨比斯的AI哲学观的影响。

哈萨比斯认为不管是ChatGPT，还是自家的Gopher，尽管可以帮你写作，为你绘画，“有一些令人印象深刻的模仿”，但AI“仍然不能真正理解它在说什么”。

所以，他说：“（这些）不是真正的意义上的（智能）。”

哈萨比斯的老师，MIT的Poggio教授更尖锐地指出：深度学习有点像这个时代的“炼金术”，但是需要从“炼金术”转化为真正的化学。

杨立昆反对炼金术的提法，但他也认为要探究智能与学习的本质。人工神经元受到脑神经元的直接启发，不能仅仅复制大自然。

他的观点大概是，工程学实现了的东西，也只有通过科学打开黑盒子，才能走得更远。

“我认为，我们必须探究智能和学习的基础原理，不管这些原理是以生物学的形式还是以电子的形式存在。正如空气动力学解释了飞机、鸟类、蝙蝠和昆虫的飞行原理，热力学解释了热机和生化过程中的能量转换一样，智能理论也必须考虑到各种形式的智能。”

几年前，巅峰时刻的哈萨比斯就表达过，仅靠神经网络和强化学习，无法令人工智能走得更远。

类似的反思，也发生于贝叶斯网络之父 Judea Pearl。

他说，机器学习不过是在拟合数据和概率分布曲线。变量的内在因果关系不仅没有被重视，反而被刻意忽略和简化。

简单来说，就是：重视相关，忽视因果。□□□□□

在Pearl看来，如果要真正解决科学问题，甚至开发具有真正意义智能的机器，因果关系是必然要迈过的一道坎。□□□□□□□□

不少科学家有类似的观点，认为应该给人工智能加上常识，加上因果推理的能力，加上了解世界事实的能力。所以，解决方案也许是“混合模式”——用神经网络结合老式的手工编码逻辑。□□□□□□□□

辛顿对此颇为不屑，一方面他坚信神经网络完全可以有推理能力，毕竟大脑就是类似的神经网络。另一方面，他认为加入手工编码的逻辑很蠢：

它会遇到所有专家系统的问题，那就是你永远无法预测你想要给机器的所有常识。

AI真的需要那些人类概念吗？阿尔法狗早已证明，所谓棋理和定式只是多余的夹层

解释而已。□□□□□□□□

关于AI是否真正“理解”，真正“懂得”，真正有“判断力”，辛顿以“昆虫识别花朵”为例：

“昆虫可以看到紫外线，而人类不能，所以在人类看来一模一样的两朵花，在昆虫眼中却可能截然不同。那么能不能说昆虫判断错误了呢？昆虫通过不同的紫外线信号识别出这是两朵不同的花，显然昆虫没有错，只是人类看不到紫外线，所以不知道有区别而已。”

我们说AI“不懂”什么，会不会是过于以人类为中心了？

假如我们认为AI没有可解释性，算不上智能，可会不会是即使AI解释了，我们也不懂？就像“人类只有借助机器检测，看到两朵花的颜色信号在电磁波谱上分属不同区域，才能确信两朵花确有不同。”

从十几岁开始，就相信“模仿大脑神经网络”的辛顿，仿佛有某种宗教式的坚定。

于是，在某个路口，哈萨比斯略有迟疑，而伊利亚·萨特斯基弗则和辛顿一路向前，豪赌到底。

辛顿的人生哲学是“基于信仰的差异化”，他的确也是如此实践的。

如今，尽管哈萨比斯认为ChatGPT仅仅是更多的计算能力和数据的蛮力，但他也不得不承认，这是目前获得最佳结果的有效方式。

## 九

对AI路线的分歧，不过是一百多年来某类科学暗涌的延续。

相当长的岁月里，在大雪纷飞的多伦多，辛顿几乎是深度学习唯一的守夜人。

他本科在剑桥大学读生理学和物理学，其间转向哲学，拿的是心理学学士学位，后来再读了人工智能博士学位。

辛顿等人在统计力学中得到灵感，于1986年提出的神经网络结构玻尔兹曼机，向有隐藏单元的网络引入了玻尔兹曼机器学习算法。

如下图，所有节点之间的连线都是双向的。所以玻尔兹曼机具有负反馈机制，节点向相邻节点输出的值会再次反馈到节点本身。

玻尔兹曼机在神经元状态变化中引入了统计概率，网络的平衡状态服从玻尔兹曼分布，网络运行机制基于模拟退火算法。