第169封信 | 人工智能出了错，会怎样？

吴军

小师弟，你好！

几周前，有这样一则报道。

IBM著名的沃森（Watson）医疗机器人捅了个大篓子，给患者开错了药。许多医生发现，沃森给出了多个“不安全，不正确的治疗意见”，甚至给有出血症的癌症患者开出了易出血的药。今天，全世界有200多家医院使用沃森，其中大约1/3在中国。

得到产品团队看到这则新闻，希望我聊一聊人工智能出差错之后，会有什么样的严重后果。无独有偶，那几天我也看到了同一类的新闻，讲的是亚马逊公司的人脸识别软件出了错，把20多位国会议员识别成了罪犯。虽然新闻里没有给出具体的原因，但是根据目前全世界的图像识别的水平，这样大面积地出错，不应该是识别率不够高的问题，而是出现了bug，也就是说要么亚马逊的人脸识别程序有明显的漏洞，要么工程师在实现算法的过程中，不小心产生了一个容易被激活的错误，以至于能够将500多个国会议员识别错20多个。

上述这样的例子其实并不罕见，虽然相比人工智能成功的时候，比例不算很高。计算机出错和人出错有一个非常大的区别，后者出差错常常体现在个案上，比如让一个人来识别国会议员，他几乎难以做到全对，但是认错人也就是认错几个，通常不会成批出错。

一个开错药的医生，影响到的只是一个医院的个别病人。计算机则不同，它们通常不出错，出错就是成批出错，造成大错。比如历次股灾，都会出现计算机狂抛股票最后让股价一落千丈的情况。人在设计计算机操作股票时，为了确保在暴跌时它们的委托人能够立即止损，会在瞬间抛售掉手中的股票，这种操作有时会触发另一台计算机的止损条件，导致更多的抛售，于是就出现雪崩式股灾。

在1987年11月的“黑色星期一”，纽约证交所就出现了上述情况，最后交易所不得不停掉了所有的计算机操作，改用手工操作。虽然各大证券公司更新了它们的智能操作程序，但是到了2001年“9·11”后的股灾，2008年年底的股灾，甚至在2010年金融危机已经过去后第一次莫名其妙的股市暴跌，在很大程度上都是因为计算机操作导致股价雪崩式下跌。

为什么人工智能容易产生大面积失败的情况呢？它潜在的巨大失败其实和它经常获得的巨大成功来自同一个原因，就是人工智能本身是一个网，它的“智慧”来自于网络效应。我在很多公开的讲座中，以及在《智能时代》一书中讲过，今天的人工智能是基于大数据的，而数据的收集和共享，本身具有网络效应。

此外，今天很多智能应用，它们的决策虽然是在不同的计算机上进行的，但是，这些计算机使用的是同一个程序，比如在北京西单地铁站的智能摄像头，识别出一个罪犯，这个罪犯即使跑到王府井、东单、圆明园，还会被识别出来，未来的智能摄像头甚至会迅速互相通信，通知各地这个罪犯来了。但是，如果这个程序识别错了，西单、王府井、东单等地都会出现误识别。这个人如果去和计算机解释，会百口莫辩。

更可怕的是，今天由于信息交流非常通畅，在人工智能领域的研究是完全无国界，学者们的通信是实时的，只要一个好的机器学习方法被发现，发明人会在第一时间将论文摘要放在互联网上，一周后，关注它的实验室就开始验证结果，如果管用，马上就用于了产品开发，因此坦率地讲，全世界使用的人脸识别、下围棋和自动驾驶，背后的机器学习算法都大同小异，这一方面让全世界的人工智能研究进步很快，但是它们的缺陷也会迅速地被复制到全世界。可以毫不夸张地讲，如果哪天无人驾驶汽车在道路上占了大多数，一个bug可能会导致周围上百辆车彼此相撞，因为这个bug可能会在很多无人驾驶汽车里同时被激活。就如同在股灾时，恐慌性抛售会相互激活一样。

那么是否可以通过提高技术水平尽可能地避免上述情况发生呢？当然可以，实际上每一次股灾后，证券公司都要修复它们系统中的bug，这杜绝了很多灾难的发生，但是依然会有人们想不到的没有把漏洞堵上的情况。我们生活的世界其实远比我们想象的要复杂得多，以至于很多情况在第一次出现之前，没有人能够想到。

上个世纪70年代美苏冷战时，美国开始建造区域性导弹防御系统。军工产品的设计、开发和测试要比互联网产品不知高出多少个数量级。但是，第一次测试时，它就闹了笑话，对方的导弹还没有发射，这边的反导弹就射出去了，后来发现原因是月亮升起来了。事后我们说起这件事，觉得是一个很容易想到的情景，但是事先那么多人谁都没有想到。当人们没有经历过一件事之前，即使这件事出现的可能性很大，也不会提前想得到。黑天鹅并非罕见的物种，但是在它被人们发现以前，大家真想不到天鹅可以是黑色的。

为什么人通常可以避免这样雪崩式的灾难呢？最主要的原因有两个。

首先，人的思维总的来讲是独立的，我们每一个人的判断虽然受到他人的影响，但是不会完全相同。这样的缺点是难以形成合力，特别是需要凝聚力量的时候，显得像是一盘散沙。但是它也带来好处，就是虽然小错不断，但是发生大错的可能性不大。历史上灾难性的大错误，通常是人们经过洗脑后，变成了机器。

其次，人在遇到未知的麻烦时，不会像机器那样陷入死循环，而是会根据其它价值来审视当前的情况。当Google的无人驾驶汽车在路上遇到不认识的小沙袋时，会按照预定的方案避开它，结果和侧后方的大巴士相撞了。而人遇到这种不认识的东西，会在很短的时间里判断压上它是否有翻车的危险，如果它的高度足够低，可能就直接压过去，而不是避开。这就是人其它的价值判断用于未知情况的好处。

当股票发生雪崩时，一部分人并非像其他人那样赶快逃命，而是当下跌到一定程度时，会激活他对市场价值的判断，开始购进股票。当所有人都开始涌入加州淘金的时候，会有人想到卖水或者卖牛仔裤。2008年股灾时，一位精明的投资人制定了一个简单易行的操作方式。他在股市跌到一定程度后，每次比上回买进的时候跌几个百分点，他就用手上的现金买入几个百分点的股指基金。他讲，如果跌100%，跌到零，他相当于用零成本把整个美国股市买了下来。注意，他买的是股指基金，而不是单只股票。在历史上，美国的股指总会涨回来，因此它不怕短期下跌。大家千万不要用这种方法去操作单只股票，因为单只股票经常会跌到零。当然，这样操作的基本条件是手上有现金。事实上，总会有些人在大家欢呼股市创新高时会遵守纪律兑现一部分收益，在大家觉得世界末日正在到来时逐渐买入。每一个人凭自己的经验和知识，在麻烦时作出有利于自己的决断，避免了很多大面积的灾难。

我们讲，人是万物之灵，是有一定道理的，人类用智慧避免了很多灭顶之灾，虽然也在不断犯错误。在未来的智能时代，人要做的是保持多样性，而不是像机器那样为了高效率遵从一种思想，一种方法。人还要在不伤害自己的前提下不断地试错，以发现未知里面的奥秘。

结合前三天讲的内容。下一代全盘接受上一代的经验有什么好处和坏处呢？好处是上一代做得很好的事情，下一代可以走捷径，轻而易举地掌握相应的技能。但坏处是，上一代没有解决的问题，下一代也难以解决，甚至上一代已经解决的问题，在下一代变了种，后者也无能为力。

也就是讲，这样教育出来的人就如同机器一样，有一个系统的bug，就陷入了一代代走不出的死循环。我们为什么说在教育上，死记硬背不是好方法，经过思考后理解了，形成自己的思想非常重要，因为人不是机器，注定要用自己的能动性解决一些前人没有解决的问题，要比前人走得远一些。

这周我们谈了人如何走出死循环，最重要的是，不能一遍又一遍，一代人又一代人地走老路。如果是那样，我们就和有bug的机器没有什么两样了。

祝近安