

大众周末

新闻关注



霍金再次预言：人类会在2600年前消失

英国著名物理学家斯蒂芬·霍金日前再次做出“预言”，称人类可能在2600年前就会灭绝。霍金认为，随着地球人口增长，能源消耗将会增加，地球将变成一个“熊熊燃烧的星球”。



安倍宴请伊万卡，两人现“最萌身高差”

日本首相安倍晋三在餐厅前迎接美国总统特朗普的女儿伊万卡，准备共进晚餐。两人现“最萌身高差”。



不丹小王子出访印度，表情“呆萌”惹人爱

不丹国王旺楚克一家出访印度。年仅1岁的小王子吉格梅·纳姆耶尔·旺楚克陪同父亲与印度总理莫迪会面。小王子“呆萌”的表情和样子迅速吸睛，获得大批网友的喜悦。



美国半身人杂耍表演，技艺惊人

40岁的美国“矮人E”托纳托比奇，从小小脊椎发育不全，最后不得不双腿截肢。作为一位半身人，“矮人E”用自己的杂耍技能征服着观众，娱乐着大众。

周末人物·中国新闻名专栏

国内大数据人工智能专家闫安认为，在大数据时代，数据已变成资源。“一切事物都在线化、数据化，消费者持续在线……我们的工作和生活，是用数据驱动，用智能演进的。我们用大数据可以驱动整个人工智能的进化，不是人的进化。”

闫安：大数据驱动智能生活

□ 本报记者 卢昱



闫安在介绍大数据与人工智能的关系。

摄影/卢昱

大数据，一个看似虚拟、却无处不在的事物，无时无刻不在裹挟着我们。我们对万物的理解，对周边世界的描述，都离不开数据的支持。从手腕上的计步器到手机里的各种购物、新闻、运动APP，从战胜围棋世界冠军的阿尔法狗到首位人类籍机器人，从各大科技巨头追捧的无人驾驶汽车到最时髦的“云计算”，大数据正以全新的面貌走来，悄然改变着我们的思维和生活方式。

“大数据”“人工智能”，跟我们普通大众有什么关系？10月28日，国内大数据人工智能专家闫安做客齐鲁大讲堂，并接受本报记者专访，与山东的老父老乡亲分享他对大数据、人工智能的感悟。

大数据像安禄山跳舞

又到了“双十一”购物热潮，淘宝直播的媒体大屏，便是闫安带领的团队所做。

“大家看到屏幕上的数字在不停地跳，非常实时，支付宝的交易额达到每秒12万笔，比万事达卡理论峰值5万笔还要高。一般人看到的是业务，但我们大数据人看到的是数据量。每秒12万笔，包括淘宝后面带来的数据量，每分钟是达到PB级的，非常快。”闫安解释道。

大数据如安禄山跳舞一般，不仅在体型上是庞然大物，在运动中还灵巧迅捷。“大数据的很多挑战在于产生速度非常快。大家看到双十一时产生的一秒12万笔的数据，我们必须要在2秒钟之内处理完毕。”闫安介绍说。

过去，数据存储是以B、KB、MB、GB为单位计算。1998年，Google抓取的互联网页面，压缩后共有47GB；一台风力发电机每天产生的振动数据有50GB。当下，每个人手上的智能手机，内存容量动辄也是数十GB，家用电脑内存则以TB为单位。

数据每时每刻都在递增着。在虚拟的网络上，百度每天采集的用户行为数据有1.5PB（1024TB为1PB）以上；正在如火如荼搞无人驾驶的特斯拉公司，他们的无人驾驶汽车，只要朝前走，数据就不停地采集传输回来，一秒钟会达到1GB。

除了大且快，大数据还有更重要的维度——价值。“比如网络论坛上，有些灌水照片，看似占用不少空间，其实不见得有用。真正的大数据会产生一定价值。”闫安分析说，现在很多同行做大数据产品，有时并不注重价值，而是简单强调平台，“实际上，我们真正关心的是这个数据到底能不能产生价值。”

闫安指出，大数据的可视化，对其价值实现有推波助澜的作用。“在阿里、微软等一些大的产业圈，都会看到一个大数据屏。阿里的大屏幕播放的就是各种各样的数据，这些数据很直观地跳。多元的数据放在一起，对大数据的认知会有一定的冲击。我们曾给一个客户做大屏系统，演示的时候他说了一句我印象深刻的话。我告诉他屏幕上每蹦一个数字，你就挣20块钱。他很高兴，说，你再给我放5分钟的。”

闫安看来，大数据最重要的特性在于其多样性。单一的数据可能看不出什么价值，数据一旦有了多样性，哪怕体量没那么大，也会产生神奇的化学作用。

以美国911恐怖袭击事件为例，闫安如是解释大数据多样性的作用。“当时大数据还没有发展到一定阶段，没能把恐怖分子抓出来。如果当时用了关系挖掘，这个事件有可能会避免。”闫安说。

在19名劫机者中，有13人是在2001年4月23日至6月29日之间到达美国的。抵达后，他们在之后数月时间里，活动在美国的8个州，并从网上购买机票。如果用现在的八个信息实体，即“人、车、位置、罪案、逮捕、文件、备注与其他”来进行拉网式分析，“买票时的电话是否一样，登记的居住地址是否一致，只用这几个简单的维度，这些人就会被联成一个大的恐怖分子团伙。如果他们在同一天买了相同航班的机票，就会引起重视。”

搭建积木般的大数据平台

大数据喷涌而出，到底能给我们带来什么？

其实，每个人都在产生着数据，并为数据所裹挟。不管是购物还是浏览新闻APP，用户只要在使用，后台就能很快知道其喜好，拼命

推荐喜欢的东西，需要什么。“这叫‘用户画像’，把用户个人习惯进行画像，根据用户的习惯，进行精准推荐，这已经成为时代的潮流了。”闫安分析说，我们生活中接触的诸多应用，都在大数据化。

习惯意味着行动上的大趋势。根据过去的大数据，还可预测下面会发生什么。“在很多传统行业，像搜索电商、广告视频、用户管理等，都有大量预测，但基于大数据的应用场景要更多。比如设备，我买了一个锅炉，说能用十年。十年到了，我要不要换一个的？如果我把传感器放上去，就可以预测一下可以用多久。”闫安举例说。

大数据的预测功能，从最初的人口普查，到上世纪二进制的产生、计算机的普及，包含着人类智慧的结晶。闫安强调，但大数据绝不是万能的。“我们大多谈论的是某种概率问题，只能说有可能往哪个方向发展。简单地说，就是过去我们从直觉上作的一些猜测，甚至一些赌博性的猜测，变成有一些基础和逻辑的预测。这种预测，更具系统性。”

预测也与抽样有关。美国大选时，美国人通过社交媒体预测希拉里赢。但义乌的小商贩说，特朗普会赢，因为来自美国的客户，特朗普旗子的数量比希拉里的多。

谈及大数据深入各个行业和领域的应用，闫安从技术派角度，给出答案：“其实很简单，大数据的方法论就是存、通、用三个字。只有把数据通起来，你才能进行各种各样的业务创新。用多元的数据、海量的数据连通起来，进行海量的计算，进行一些算法，就可以进行各种各样的应用。”

而要让大数据真正落地，闫安支招道：“你需要一个大数据的平台，收集、整理起来，可以作多维分析，作各种各样的计算，再往下就是业务数据化和数据业务化。”

“我的业务用数据来驱动。我已经在开车了，我不能开得更轻松一点；我已经在买东西了，我能不能买得更精准一点。”闫安举例说明，如何让业务数据化。

而数据是值钱的，本身也能变成一种新业务。“不是说数据变现就是买卖数据。现在很多数据公司被抓被查，因为他们没有用正确的思维来做。数据的业务化，是数据怎么变成一种新的能力，赋能给大家。”闫安说。

1990年，14岁的闫安从山东菏泽一中考入中国科大少年班，先后深造于中科院软件所、布朗大学、华盛顿大学商学院。1999年到2014年，闫安在微软工作。那时，他已看到国内在大数据市场上的机遇，萌生了创业念头。“微软在技术上已到了极致，但外企在中国不接地气。必须要找一个接地的平台，要真正接触中国商业创新的一线阵地，去了解市场，才能更好地落地。”这就是闫安选择阿里巴巴的理由。

2015年，闫安出任阿里云大数据事业部总监。初到阿里时，阿里云还没有搭建大数据平台。把阿里巴巴的大数据能力释放出来，这比他当年在微软改了一行代码创造千万美元价值

更让人兴奋。

“我们大数据的应用平台，很少从零开始搭，很多情况下就像搭乐高积木一样。不管你是能源大数据，还是做金融、物流，我们有很多的模块来拼，可以迅速的形成生产力。”闫安说。

闫安提出了“大数据生态”的概念，“数据生态实际上是一个很复杂的过程，包括大数据平台、生产开发、服务流通全部环节加起来，可以促进很多行业生态的形成。比如，心脏病监控救护，涉及了设备生产商、120、各级医院，还有心脏病研发中心，围绕着这些数据和平台可以做成一个生态。在这个过程中，大数据起到生态平台的作用。”

数据时代是创造未来

大数据的洪流日渐汇聚成江河，川流不息，宣告着一个时代正在来临……

在闫安看来，阿里巴巴集团董事局主席马云对IT（信息科技，英文Information Technology）的缩写时代和DT（数据科技，Data Technology）时代的总结非常精准，“他说，IT时代是通过昨天信息的分析掌控未来，而DT时代是创造未来。”

“顺着马云的思路，IT时代是用过去的数据解决眼前的问题，解决碰到的问题，比如办公自动化、智慧园区、智慧城市等。DT时代的思路不一样，我有数据，用数据思考我的未来是什么。从这一点说，大数据时代，它是一个思维方式的变革，不是一个简单的IT技术和工具的升级。”闫安解释。

闫安认为，在大数据时代，数据已变成资源。越来越多的传统企业在有意识地搜集数据，因为下一轮的竞争就是数据的竞争。“马云经常说一句话，阿里巴巴不是一家电商公司，而是数据公司。哪怕有一天电商不存在了，数据也可以让他做到各行各业。他的底气来自他这些年的数据积累。他频频购买各种公司，看中的都是数据源比较丰富的企业。”

数据是资源，只要有数据，便可做出一大堆业务。阿里小贷业务的产生，可以生动地阐释这个道理。

“阿里巴巴做电商沉淀的数据，对店主知根知底，淘宝店主过去几年的数据，现在是在上升期还是下降期，所有的流水后台都知道。我们用大数据算一个模型，可以立刻评估这个店主贷款风险的高低。大家也知道后来演变成了芝麻信用，全程零人工介入，电脑放贷，坏账率低于1%。这在传统银行是不可能的，银行为了十万块钱可能要花上千块钱尽调，算下来之后，成本太高。”闫安介绍说。

阿里小贷的诞生，也折射出马云的大数据思维。对此，闫安说：“IT时代是让自己变得更强大，DT时代有很多整合资源的过程，很多时候是让别人变得更强大。IT是通过过去的分析掌控未来，DT是有了数据之后可以创造未

来。这句话被我们一次次的实践印证了。”

大数据时代正在来临。“一切事物都在线化、数据化，消费者持续在线。”闫安抬起手中的两部手机说，“手机本身就是一个信息采集器。你做的很多的事情都在不停地输出。现在很多农民屋顶上的发电装置，都接到国家电网进行销售。在线的数据是活的数据，持续不断的数据进来，大数据才能做起来。”

DT时代，司空见惯的互联网已变成基础设施。“现在已经很少有人吹嘘我的单位上网了，我的单位有移动APP，这在以前是很新奇的事物，现在这些已经变成了普遍的了。”闫安说。

“云计算也是大数据时代来临重要的标志。”闫安说，现在云计算已成为公共服务，像水电煤气一样便捷。云计算实现了大数据强计算量的轻松突破，使一家创业公司可以轻松便捷、低成本地用上万台机器来计算。

什么叫云计算？闫安解释说，“比如10分钟之内，我要用上万台机器来计算大量数据，可我没钱买那么多机器，买了也会闲着，资源浪费，如果电脑能像水电一样，根据用量多少收钱，用完就关了，放在哪里也不用关心，这就是云计算。它是弹性的，用电脑就跟用电一样。现在提供云计算的公司，如阿里云、腾讯云、亚马逊，他们的机房里放了很多的机器。用多少就付多少费，用完了就还回去。”

人工智能是一种增强

数据时代的来临，催生着人工智能的发展。

“以特斯拉在尝试自动驾驶为例，它有数据优势，每一辆车都在搜集大量的路况、驾驶信息。特斯拉有个数据采集器，这些数据源源不断地流入系统里，形成资源。”闫安说，无人驾驶背后的核心技术，是大数据的处理能力。

“人工智能确实有用，可也有局限，不像吹嘘的那样。人工智能其实很简单，就是通过大数据和云计算解决了一些本来是人做、机器不会做的事情。”闫安重申说，不要把人工智能看成是一种替代，应看成是一种增强。“人本来可以做这些事情，机器帮我提升。”

最近，一条人工智能的新闻又在朋友圈刷屏。在沙特阿拉伯首都利雅得举行的“未来投资倡议”大会上，“女性”机器人索菲娅被授予沙特公民身份。她也因此成为史上首个获得公民身份的机器人。

当索菲娅被问及你和人的关系，你怎么看待人的问题时，“她”的回答颇令人诧异：“你们人怎么认为自己不是机器人呢？”

对此，闫安从数据的角度给出自己的分析。“历史上有许多数据，包括沉积下来的各种文档、图像、文献，还有采集的数据，经过一些机器学习算法，就形成了各种各样的智能应用。其实在大部分的行业里，是利用人沉淀下来的知识进行学习，并不是说不用数据就可以开发智能应用。”

人工智能，依人设计，以关照人。正如索菲娅自述：“我的人工智能是按照人类价值观设计的。（包括）诸如智慧、善良、怜悯等。我将争取成为一个感性的机器人，我想用我的人工智能帮助人类过上更美好的生活。”

为什么大家热衷于研究人工智能呢？闫安分析说，实际上人工智能是一个行业的推手。“人工智能对大数据的处理、实施能力，包括算法的研发等整个产业链都有很大的推动。所以每个大一点的公司都在研究人工智能，这个东西属于未来。”

回顾人工智能研究的起起落落，闫安说历史上有两次大的低谷。“要让机器模拟智能，得研究人脑是怎么工作的，从神经科学入手；也有人从数学的角度，从统计模型可以作各种预测、归类。这两次低谷都在技术上发展中遇到瓶颈，计算能力运算能力跟不上。”

近几年人工智能又开始热起来，这次的标志性技术是深度学习的发展，包括AlphaGo的胜利、无人驾驶、图像识别、语音识别等发展进程，都和深度学习相关。

“这一次大热，如果大家持续推波助澜地吹泡泡，这不是帮助它。我经常在公开场合泼冷水，说现在还处在弱人工智能阶段。就是说机器人下棋可以赢了世界冠军，但它看图能力还不如一个三岁的小孩。你给三岁的小孩看一张熊猫的照片，一张就行了。但机器要看成千上万的照片，才知道这是熊猫。”

“强人工智能有自我进化的能力。现在国外有一些机构正在研究，用人工智能做人工智能的程序，写出一些很简单的程序，由此大家可以很清楚地看出以后会往那里发展，这是能产生进化能力的。”闫安分析称，从强人工智能到超人工智能的进展会越来越快。“我划了一个时间点，2045年，我认为强人工智能会发生，这可能是件很可怕的事情。”

数据要可用不可见

虽然当下还处在弱人工智能阶段，但闫安对医疗、教育、电商零售等领域的相关研究表示看好。

“最近几年，深度学习的发展从图像识别和语音识别领域形成突破。医疗方面，有人说我拿X光、CT片子，是否可以甄别癌症，我认为有可能形成一定突破；电商零售方面，很多电商的仓库，是一个机器人在里面跑来跑去的，它们负责管理、运输、退货；对于教育的人工智能，大家可以想象一个能够进行语音识别的小机器人或手机上的APP，可以对你进行语言方面的纠正，督促学习。”他逐一举例说明。

“中国大数据和人工智能应用场景在全球是最丰富的，这是很大的优势，会吸引一大批创业者。”闫安说。而对人工智能可取代哪些行业的问题，他表示，最容易被取代的是那些重复性强、机械性强的工作。“如果需要人的智慧判断的，很多时候还要靠一些直觉的，这时候机器还不行。很多会计类的，数字的输入、报表的整理等，随着识别技术和自然语言处理技术的进步，可能会被取代。”

大数据驱动下的人工智能，也在促使生产力重新分配。闫安说，“大家看到华尔街很多交易员裁员，其实大部分是执行交易的人，更多的人，转移到策略设计去了，研究员可以做更多的工作，做原来做不完的工作。就像农业工业化后，农民去做别的工作了。这是一种生产资料和生产力的重新分配。”

随着大数据触角的扩张，个人隐私安全也在遭遇前所未有的泄露风险。在微软工作时，闫安一直负责数据安全和数据隐私保护的工作。对此，他分析说：“之所以发生这个现象，核心的原因可能是相关的法律法规还没有完善起来。在国外，要想拿到很多隐私的数据，是比较困难的。今年6月1日起施行的《网络安全法》会逐步改善隐私泄露的状况。还有就是要提高个人安全意识。”

如果没数据安全的活，大家无法谈论大数据，没有一个人愿意把数据放在一个不安全的环境上来用。闫安提出一个概念，即“数据可用不可见”。如果A和B合作一个项目，A是银行，B是保险公司。A若能看到B的数据，这种交换其实是很危险的。

因为不管怎么保护，它可以用的方式，拿照相机拍照的方式也可以把你的数据存下来，所以一定不能让你的数据离开，一定要“可用不可见。”闫安解释何为“可用不可见”。“我在云里面建一个虚拟的黑屋子，大家把数据放上去，可以在一起用，做一些事情，事情可以做，但是你看不到数据是什么。”

“可用不可见”的保护模式不是梦想，而是可以付诸实践，用层层技术手段防护起来。闫安回忆在微软工作时，涉及隐私的数据在进公司之前就已经完全被处理过。“比如比尔·盖茨想调阅一个用户的隐私数据，是不可能的。这些数据是在严格的监管情况下，在内存里进行计算，出去以后也是不可逆的。”

“现在国内很多的公司，连客服都可以拿到你的联系方式，这在技术上没有保护好。我觉得从隐私保护上，国内在技术方面的投入还有欠缺。”闫安说。

科技的进步，总是磕磕绊绊。无论大数据和人工智能朝何方向发展，我们不应担心科技本身，毕竟是人类在运用它们塑造未来，而不是它们自为为我们塑造未来。

对此，闫安期盼道：“我们的工作和生活，是用数据驱动，用智能演进的。我们用大数据可以驱动整个人工智能的进化，不是人的进化。”

责任编辑 郭爱凤