# 三个AI：从人工智能到增强智能

聚合的AI，经过训练的多达152层的深度学习神经网络，在图像识别上的错误率降到3.57%，比很聪明的斯坦福博士更低；自适应的AI，通过四种人工智能技术顺利实现不同语言间的实时翻译；隐形的AI，Hololens将把人类带入全息计算的未来。

在新智元与中信证券合办的人工智能产业研讨会上，微软亚洲研究院常务副院长芮勇发表《从人工智能到增强智能》的演讲。芮勇博士讲了“三个AI”：聚合的AI，经过训练的多达152层的深度学习神经网络，在图像识别上的错误率降到3.57%，比很聪明的斯坦福博士更低；自适应的AI，通过四种人工智能技术顺利实现不同语言间的实时翻译；隐形的AI，Hololens将把人类带入全息计算的未来。最后，芮勇说，今后不是人VS机器，而是人与机器双方优势互补，通往“增强智能”。

芮勇现任微软亚洲研究院常务副院长。他还是国际电气电子工程学会会士（IEEE Fellow），国际模式识别学会会士（IAPR Fellow）、国际光学工程学会会士（SPIE Fellow）和国际计算机协会杰出科学家（ACM Distinguished Scientist）。


【芮勇】非常高兴有这个机会参加中信证券和新智元联合举办的这么一个活动，跟大家交流一下人工智能这个话题。刚才彭总提到分析师今后的就业趋势，我个人是不担心的。人工智能，在座的各位不要对它有任何的担心，或者是恐惧，我觉得人工智能就是一个工具，它能让今后诸位在做分析的时候，做得更加精准。

三个AI 的天下

AI Magazine Volume 27 Number 4 (2006) (© AAAI)

# A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence

## August 31, 1955

*John McCarthy, Marvin L. Minsky,*
*Nathaniel Rochester,*
*and Claude E. Shannon*

在人工智能50年大会上，5位1956年Dartmouth人工智能夏季研究会的与会者再相聚

照片从左至右：Trenchard More, John McCarthy, Marvin Minsky, Oliver Selfridge, 以及 Ray Solomonoff (Photo by Joseph Mehling '59)



今天想跟大家聊的就是从人工智能到增强智能。2016年确实是人工智能很有意思的一年，因为1956年以前还没有一个词叫人工智能。1956年夏天在美丽的达特茅斯学院，一群人造出了人工智能这个词，今年是这个词被造出来第60年，我们希望60年后有一个升华。


大家谈到AI就想到了人工智能，过去60年又衍生出来几个也是用A和I打头的英文单词，也是人工智能今后发展的一个趋势和属性。

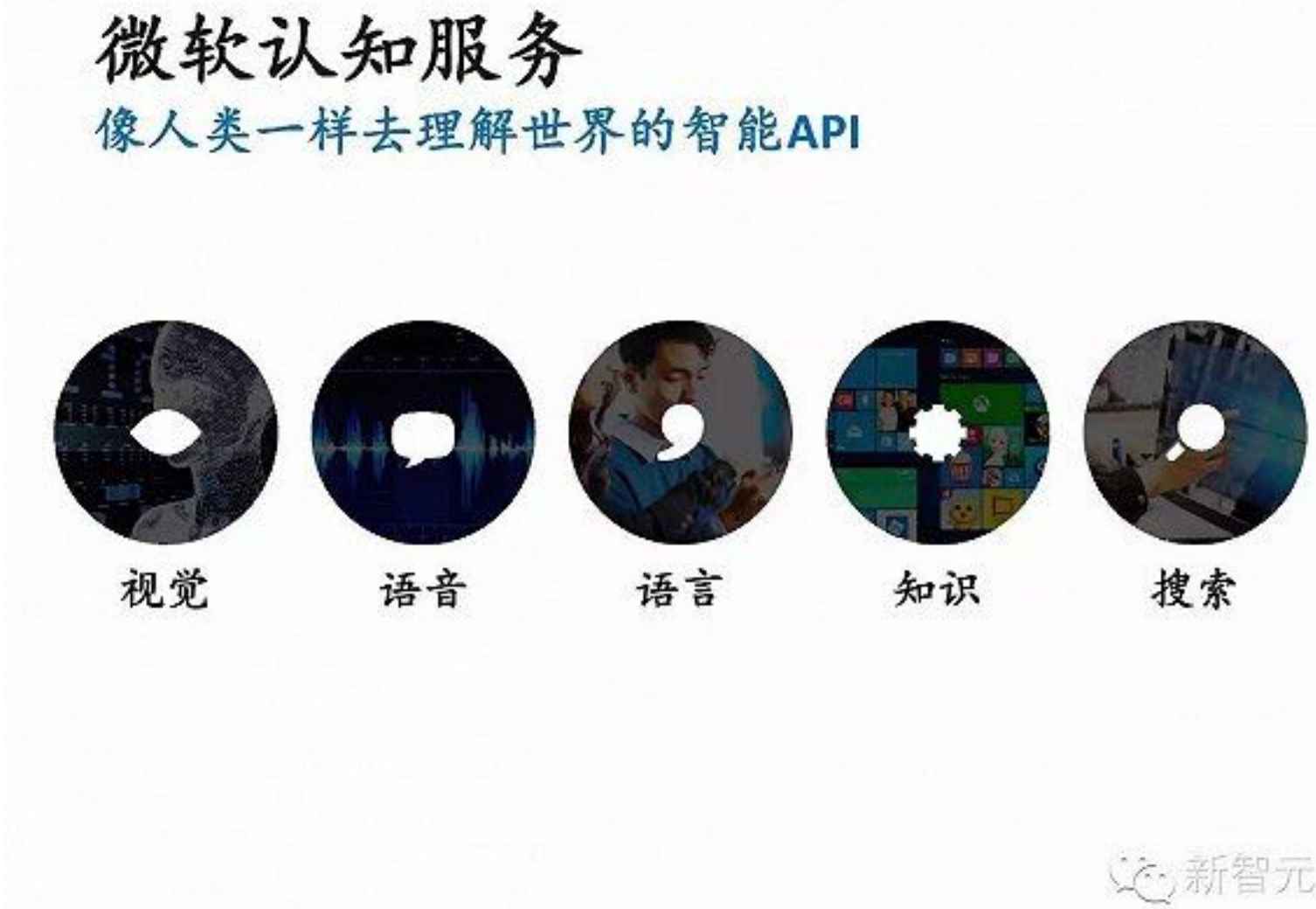
第一个也是A和I，Agglomerative Intelligence叫做聚合的智慧，什么意思？其实就是把很多人类的智慧加以提炼，用大数据去挖掘，让机器去学习。

第二个词是Adaptive Intelligence，也是一个A和I，叫自适应智能。我希望这种机器智能不要让人总是考虑这个环境是不是懂我在做什么，那个环境下可能就不懂了，我们希望这个人工智能也会自适应，它自己要理解人希望在什么时候做什么事情。

第三种叫Ambient Intelligence，这个更加高深一点，是一种隐形的智能。这里指的是，是不是有一些智能摄像头之类的东西，让所有的智能自然而然地就发生了，这就是更高层阶段的智能。

深度学习究竟可以有多深


聚合的智能，我们人类有很多的知识，这些知识以大数据的形式存在，那么通过机器学习的方式，把大数据中间的知识挖掘出来，充实自己。这能让机器做什么?




我们人类是可以通过眼睛去看，耳朵去听，嘴巴去说的。我们是不是也可以让计算机能够像人类一样，有听觉、视觉、触觉。微软做的一个项目叫认知服务，把它放在云上，如果一个云只能存数据，可能还不够智能。如果这个云通过公开API，可以让第三方开发出来很多像人类一样可以去听，可以去说的APP。微软云认知平台就是为了达到这样一个目的，把微软过去20年在人工智能上的一些成果转化为一些API，供第三方调用，包括视觉、语音、语言、知识和搜索。


其中，计算机视觉包括人脸识别，视频检测，性别识别。年龄估算是去年4月份红遍大江南北的APP。在座的各位都上传过自己的相片，或者上传过朋友的相片，通过人脸认知API，写20行代码去调用，这个APP就写出来了。

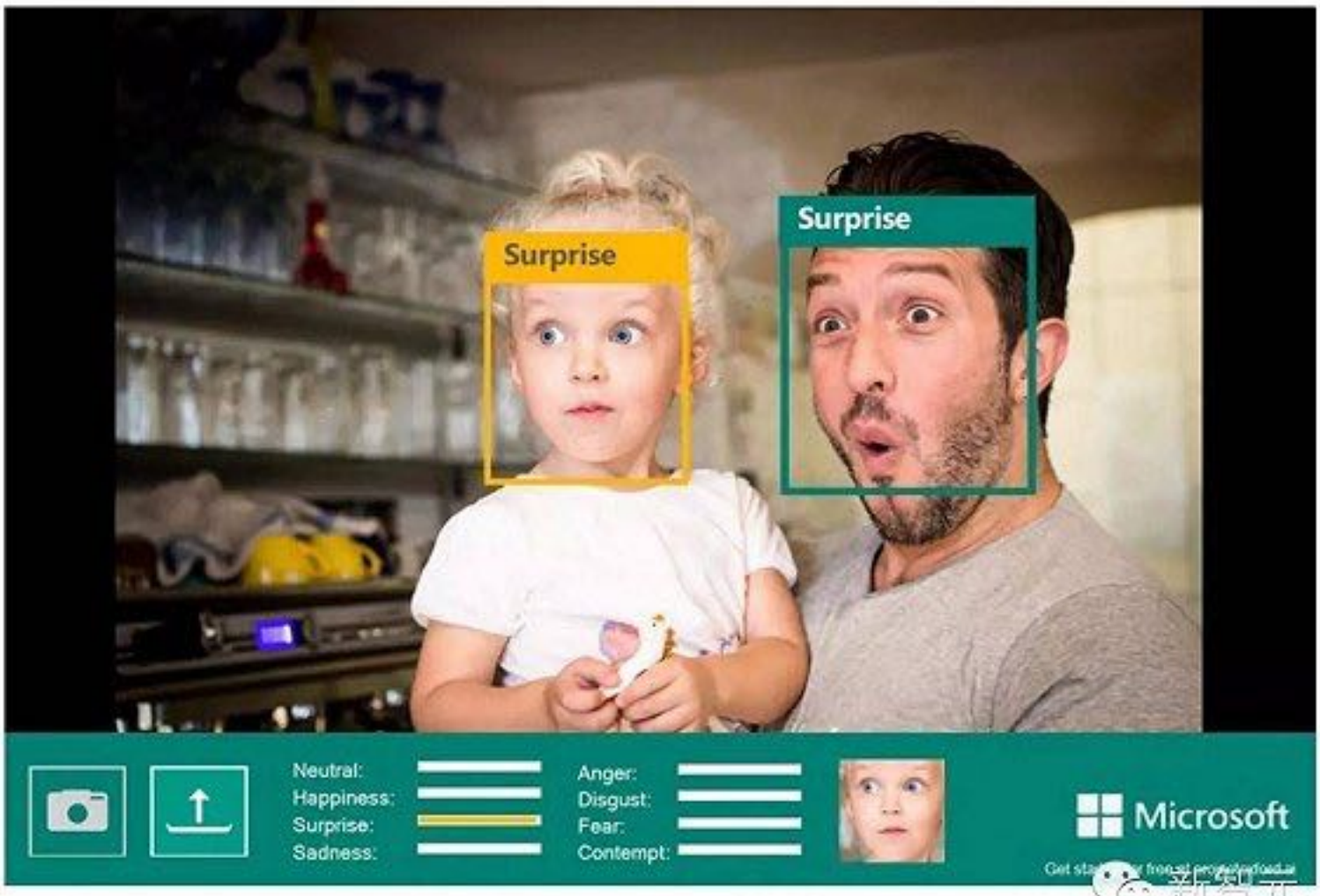
# How-Old.net的魔力

```
static async void MakeRequest()
{
    var client = new HttpClient();
    var queryString = HttpUtility.ParseQueryString(string.Empty);
 // Specify values for optional parameters, as needed
 // queryString["analyzesFaceLandmarks"] = "false";
 // queryString["analyzesAge"] = "false";
 // queryString["analyzesGender"] = "false";
 // queryString["analyzesHeadPose"] = "false";
 // Specify your subscription key
    queryString["subscription-key"] = "";
 // Specify values for path parameters (shown as {...})
 var uri = "https://api.projectoxford.ai/face/v0/detections?" + queryString;
 HttpResponseMessage response;
 // Specify request body
 byte[] byteData = Encoding.UTF8.GetBytes("");
    using (var content = new ByteArrayContent(byteData))
    {
    response = await client.PostAsync(uri, content);
    }
    if (response.Content != null)
    {
        var responseString = await response.Content.ReadAsStringAsync();
        Console.WriteLine(responseString);
    }
}
```




也有很多人上传了很有意思的相片，比如，奥巴马的老婆测出来只有36岁。

还有，大家都知道，微软刚创立时只有11个人，照片左下角长得比较清秀的就是比尔盖茨。微软两个合伙人建立的，另一位是艾伦，在照片右下角，他比盖茨大两岁，但是留了那么长的胡子，所以机器认为他是50多岁。这就是估算年龄的这么一个APP，20行代码调用API就写出来了。




2015年12月份我们发布了这个APP的第二版，能识别人的表情：人是高兴，是吃惊，是悲伤，还是愤怒，都可以辨别。我们想让计算机像人类一样，通过它的眼睛去看这个世界。有太多的东西要去看了，其中很重要的一个技术是图像的理解和分类。在座的大部分都是证券投资业的朋友，对计算机视觉不是很了解，计算机视觉领域有一个很大的比赛，每年全球最高等的研究机构和学府都会派出最强的队伍去参加这个比赛，这个比赛中，有1000张图片，让计算机进行分类，让它分辨没有见过的图片是属于这一千类的哪一类。

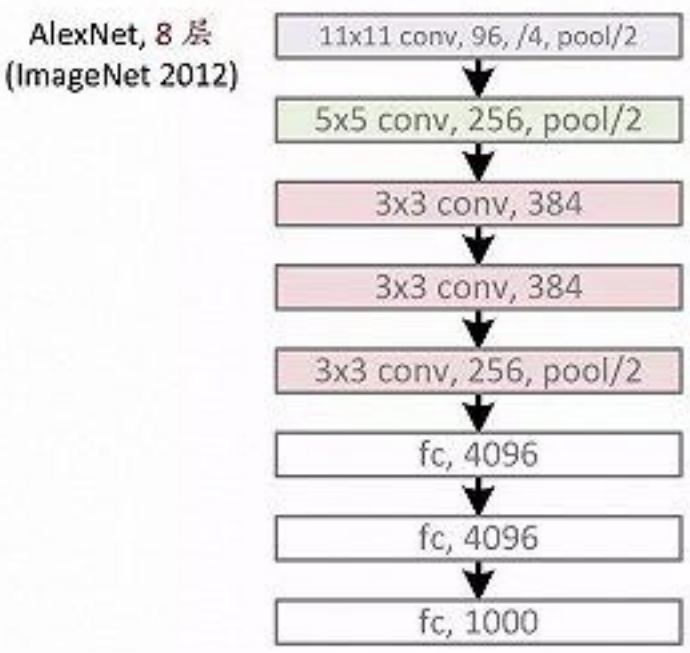
说到这里就要说到深度学习。深度学习其实就是原来人工神经网络，80年代末和90年代的时候只有一层，层数够深的话，就可以去模拟人类脑子里面神经元之间的连接。

# 图像分类

- 最大的图像数据集 – **ImageNet**
  - 120万训练样本，10万测试图像
  - 1000个类别


如果人类把一千类物体去分类，人去做，错误率是多少？斯坦福在读的一个博士生，他说来看一看，来分一分没有见过这一类的图片，这一千类里面有67种狗，我可能就懂4、5种狗。还有好几十种不同的植物和树，这位在读博士生好好训练了一下自己，之后去分类，错误率达到了5.1%，非常厉害。

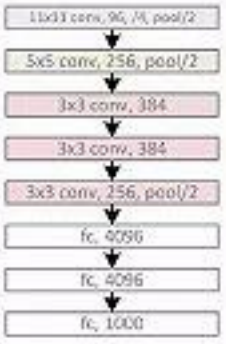

2011年以前，深度学习还没有进入计算机视觉的时候，1000类物体计算机做分类的错误率是20%，到2011年、2012年深度学习被引入计算机视觉领域，错误率一下子降到了10%，之后在2013年、14年，错误率越来越低，2014年已经低到了6.7%。去年年底的时候，深度学习算法有了进一步的发展，错误率降到了3.57%。可以说，比很聪明的斯坦福博士错误率更加低。



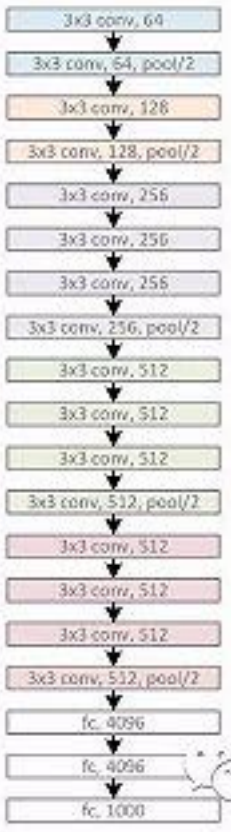

深度学习发展的一个标志就是人工神经网络有多深，做得越深，学习的能力就越强。2012年深度学习这个词出来，当时是8层，到2014年的，19层就出来了，关键是算法要对，训练算法不对，你是训练不下去的，因为算法不对，在GPU上跑不起来，错误率降不下去。

# 很深层网络 (2014)

AlexNet, 8 层
(ImageNet 2012)




VGG, 19 层
(ImageNet 2014)





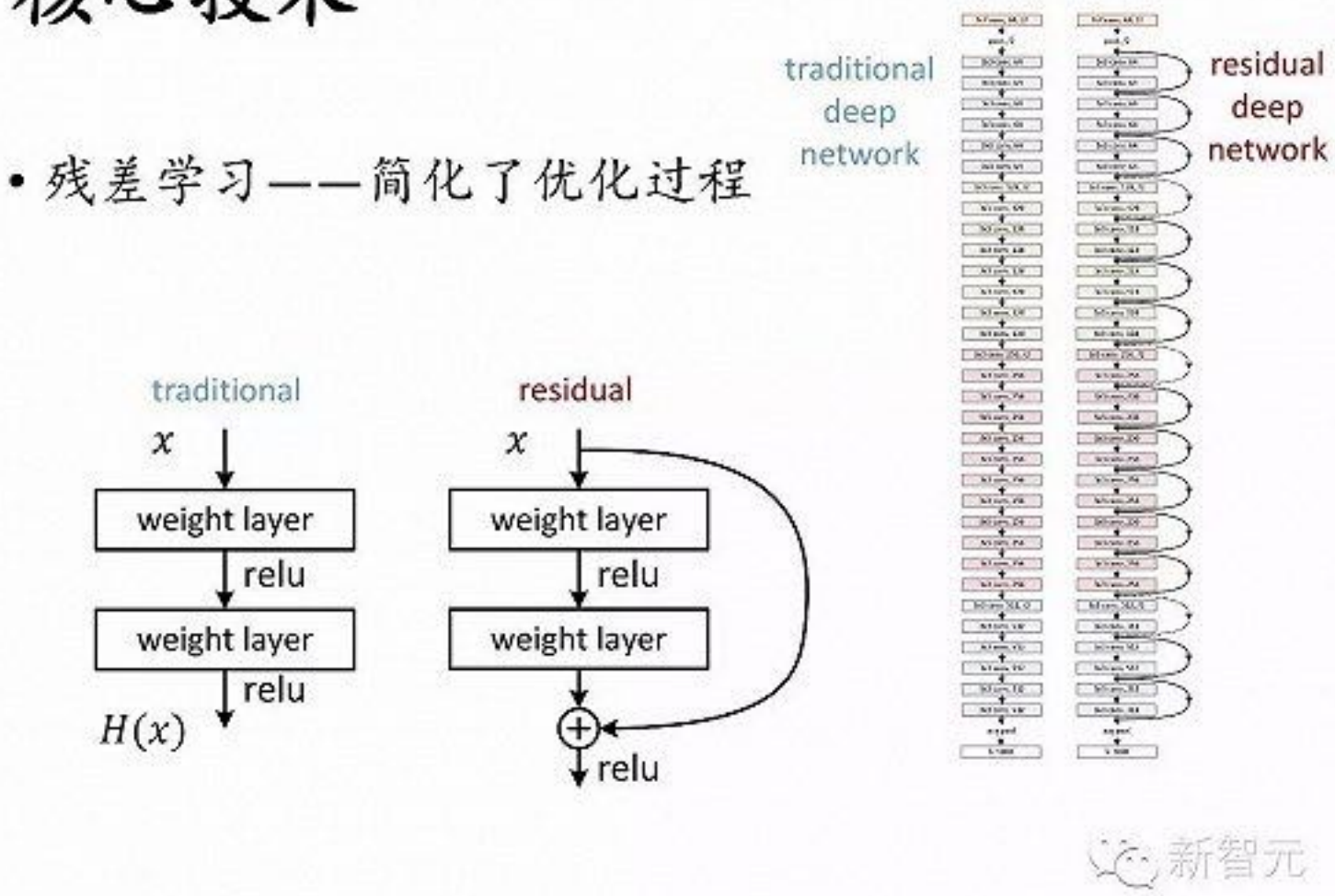

# 我们的极深层网络 (2015)

AlexNet, 8 层
(ImageNet 2012)

VGG, 19 层
(ImageNet 2014)

**ResNet, 152 层**
**(ImageNet 2015)**

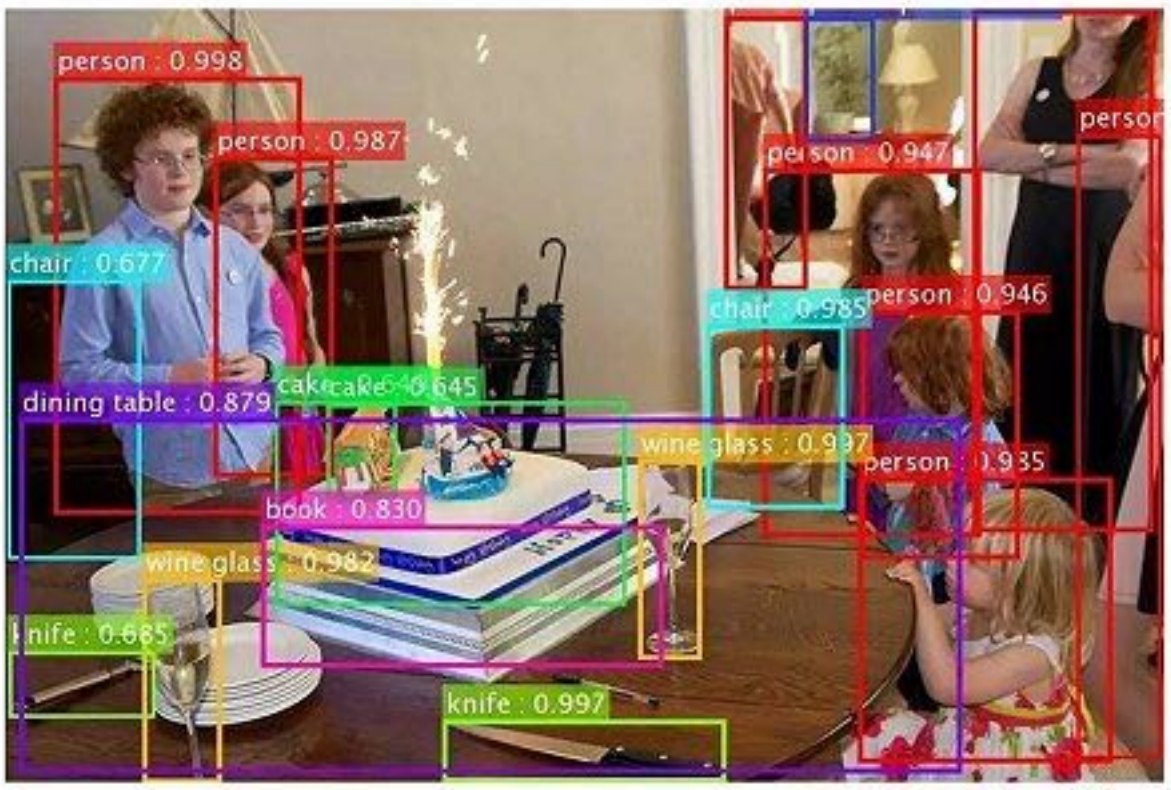

在15年底的时候，微软研究院的小伙伴，把它深度到152层，达到人类不可以企及的高度，非常的深，可以去挖掘大数据原来没有发现的一些东西。去年年底的时候，我们的错误率降低到了3.57%，比人类还好。




当然，这不是说一层一层加进去，算法上一定要有创新，一个很重要的创新就是残差学习，除了原来传统经过非线性的传输，还要跳跃性的直接过去，这个就相当于有好几层不同的神经在一起学，正是因为这样一个很新颖的算法结构，使得错误率大幅度的降低，微软在2015年的一千类的物体识别比赛中三个项目都拿到了第一名，并且比12年的准确率要高出很多。

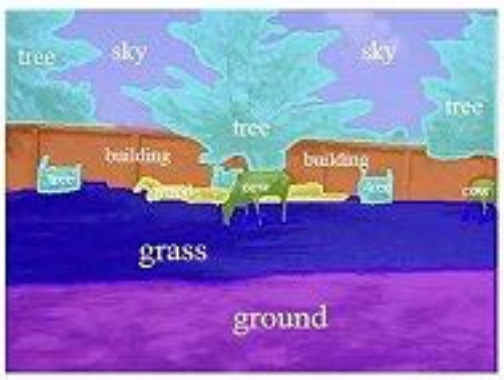

比图片分类更难的是把图片里面的物体检测出来，也是深度学习的方法，下一步，叫做物体的检测。现在物体的检测已经可以做得非常精准，有人，有书，都可以检测到。

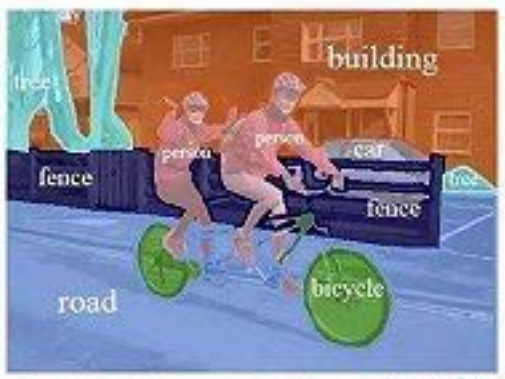
# 物体检测



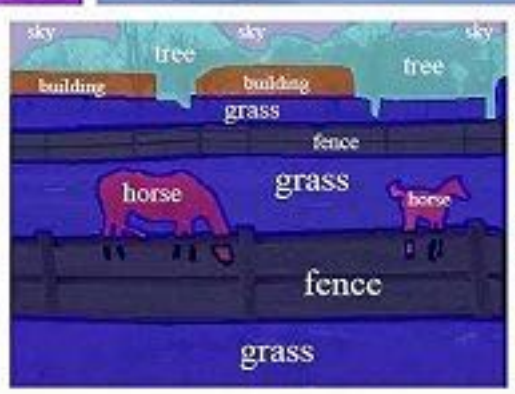

比物体检测再难的，就是像素级的精确识别，计算机看到一个图片，说这个图片里有一只狗，我让计算机告诉我狗大概是哪个位置，这个算难了，比这个还要难的，就是图片中某个像素属于这个狗身体的一部分，还是属于旁边这个草地，还是那个房屋，每个像素都要分类，要达到这样的效果，对人来说没有什么特别的，一生出来就会，但是对计算机非常难。计算机看东西就是两个，一个0，一个1，它没有物体的概念。

# 视频中的物体分割




像素级的
精准识别





Source: Pascal VOC Challenges, http://host.robots.ox.ac.uk/pascal/VOC/

计算机如果能在每个像素级都做到这么精准，就有很多的应用，比如说无人车，还有更加重要的在精密工业制造上面的一些应用，这就是跟大家分享的第一个人工智能属性，叫聚合智能。

微软实时翻译技术揭秘


手机自拍，在座都拍过自拍，尤其是女士们，有时候你拍自拍的时候，是不是窗户太亮了，脸会比较黑，是不是光线不太好，我们有没有办法做出一个APP，让它来帮我们把这些很困难的事情都考虑进去，我们写了一个APP，我们没有做任何的推广，完全是口口相传的，一夜之内用户全球就达到了100万，它能做什么?

# 微软自拍

- 一键提高照片质量
- 发布1个月，全球用户达百万

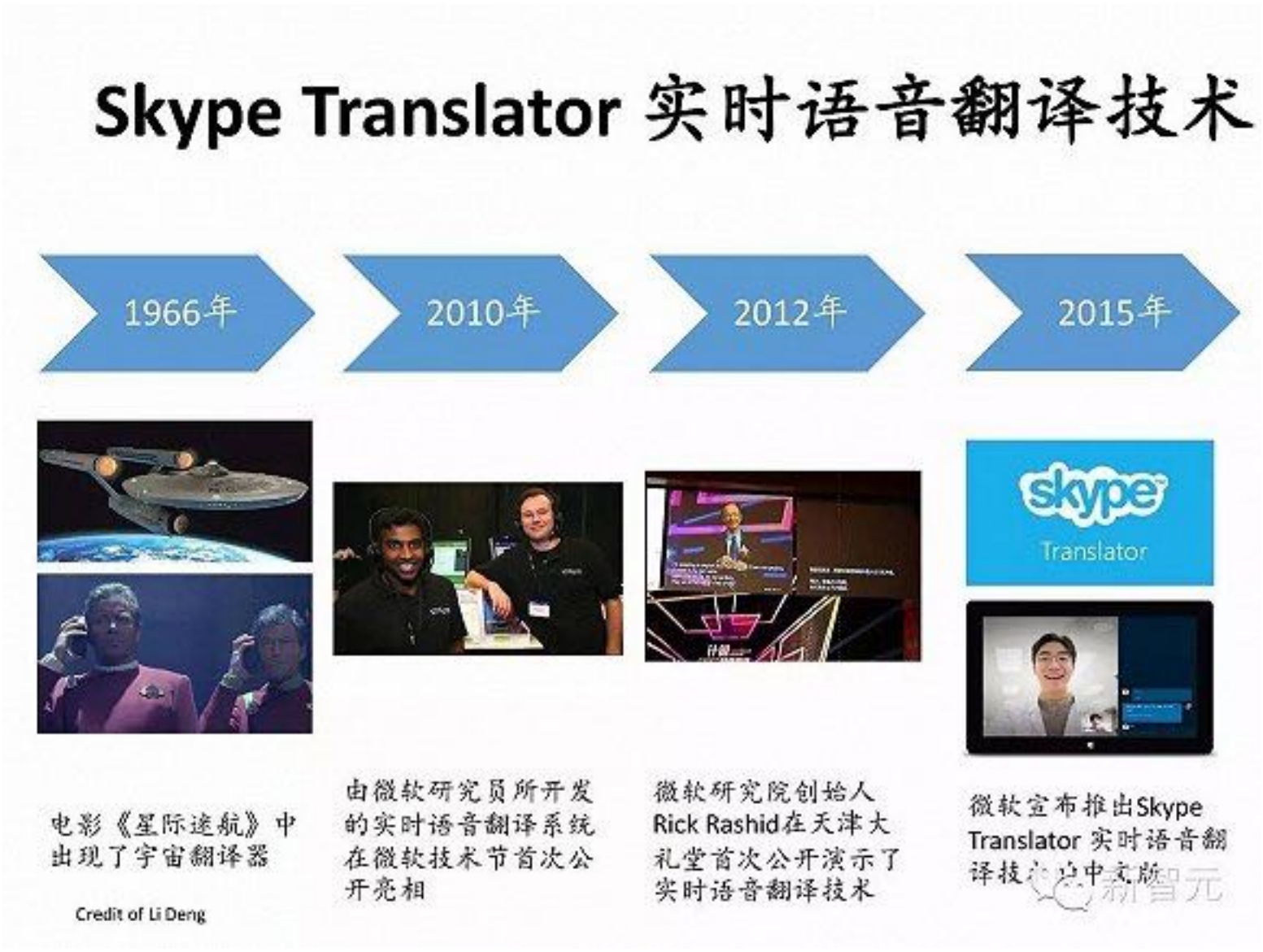
自然美颜

智能降噪

曝光增强


做美颜！其实是可以知道你的年龄，你的性别，你的肤色，你的种族，如果是新智元创始人杨静女士拍一张自拍，我们希望把她美颜得非常漂亮，甚至

可以加上口红，如果是我拍一个自拍，加上口红，那是非常不好看的。你要知道这个人是什么性别，什么种族，这些智能都在APP里面，不用人类去考虑，所以叫自适应的。

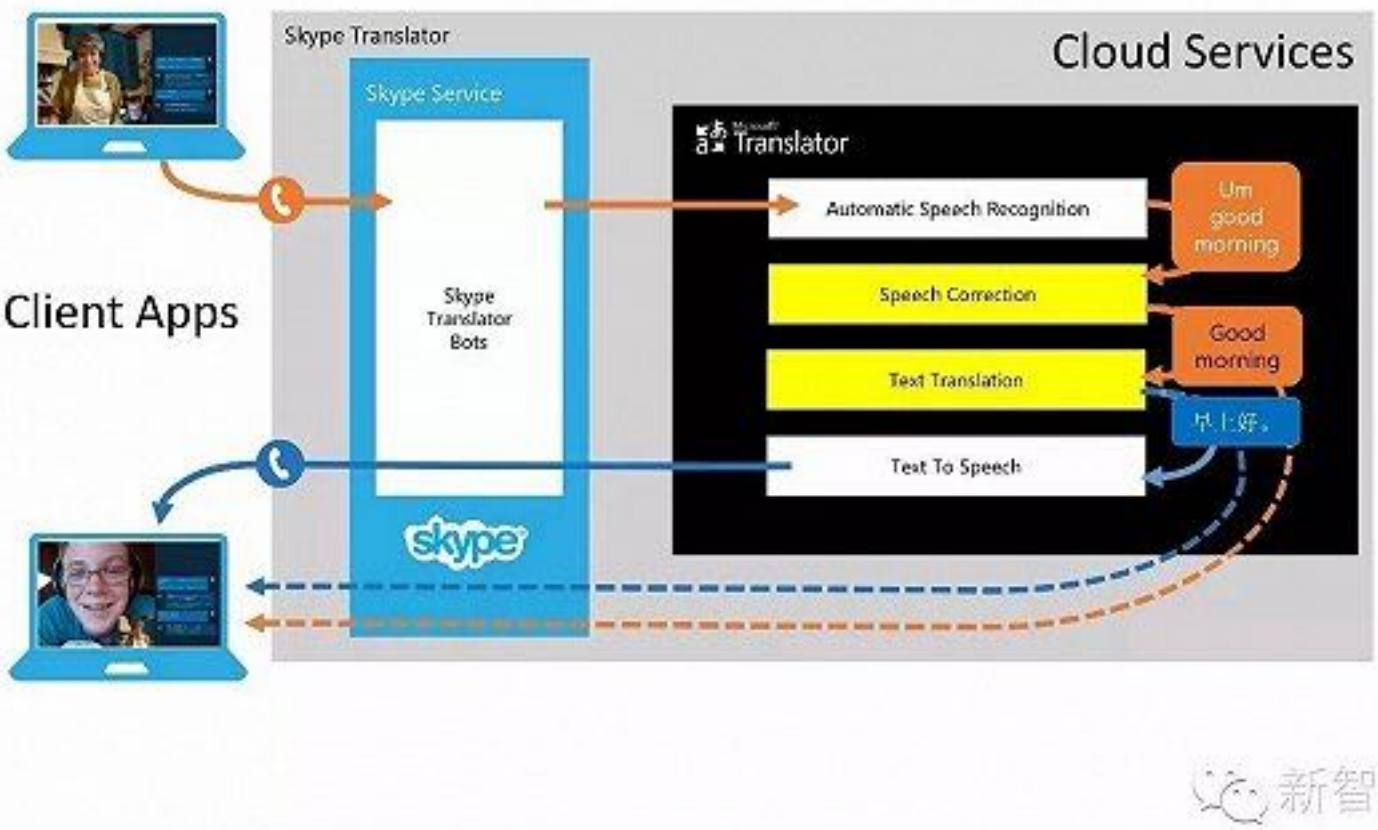



还有一个例子跟大家分享的，就是实时的语音翻译技术。我们人类一直有这么一个梦想，大家看过60年代一个很火的电视连续剧，叫《星际迷航》，有一群人在星际各个空间，各个星球里面去探索。如果抓起一个电话，这个电话就让我跟另外一个星球的人可以实时交互，在66年的时候这还是一个科幻，但是做科研的总是希望把科幻变成现实。2010年前后我们就开始组织团队，希望把这个事情能够做出来。这个之前有很多的积累——语音识别在92年就已经成立了。我们在5、6年前集中发力做这个事情，现在已经做到中文和英文实时的翻译。2015年我们出了一款产品，通过Skype Translator可以享受到全球八种语言的实时翻译。

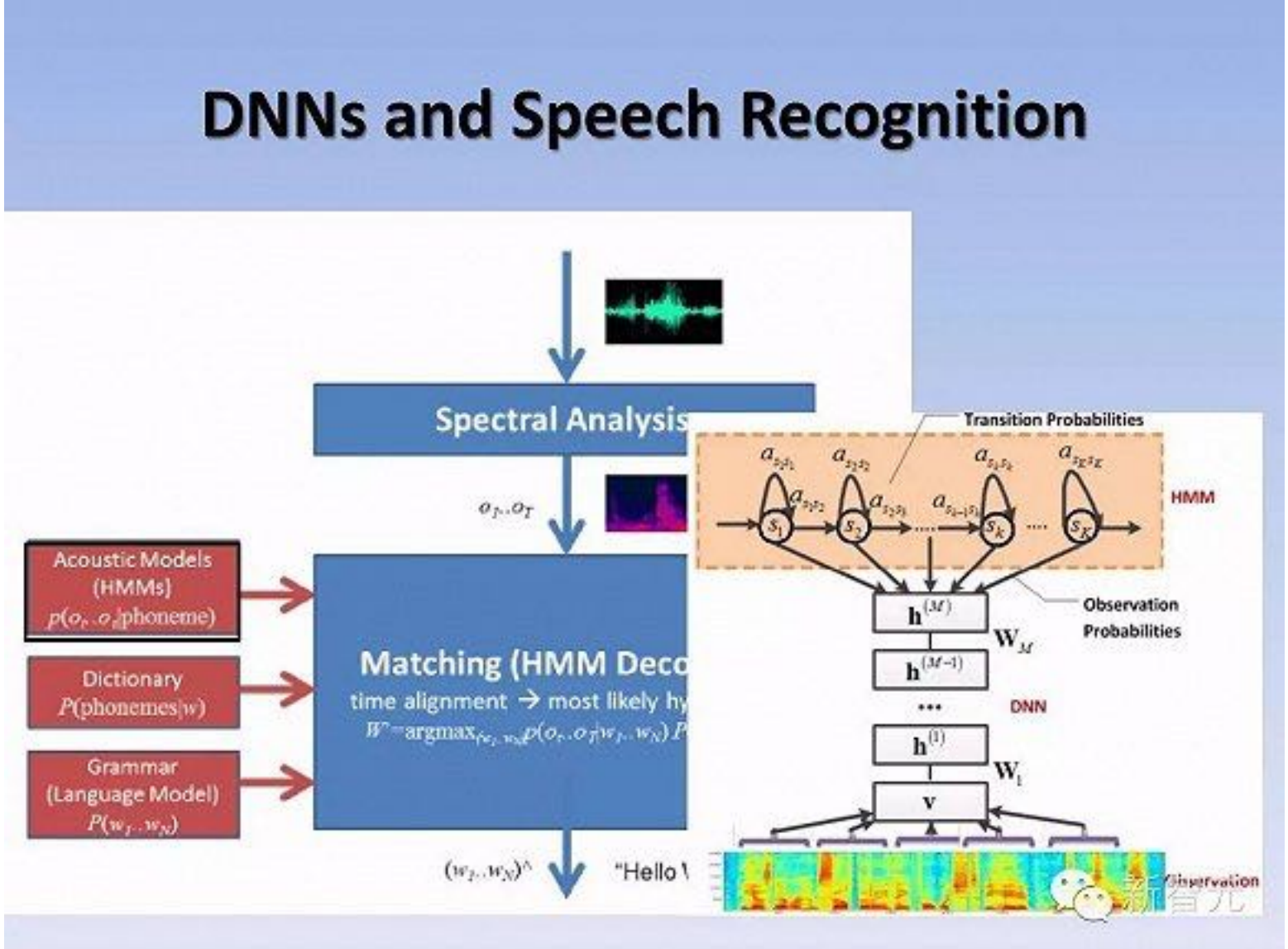
# Skype Translator 实时语音翻译技术



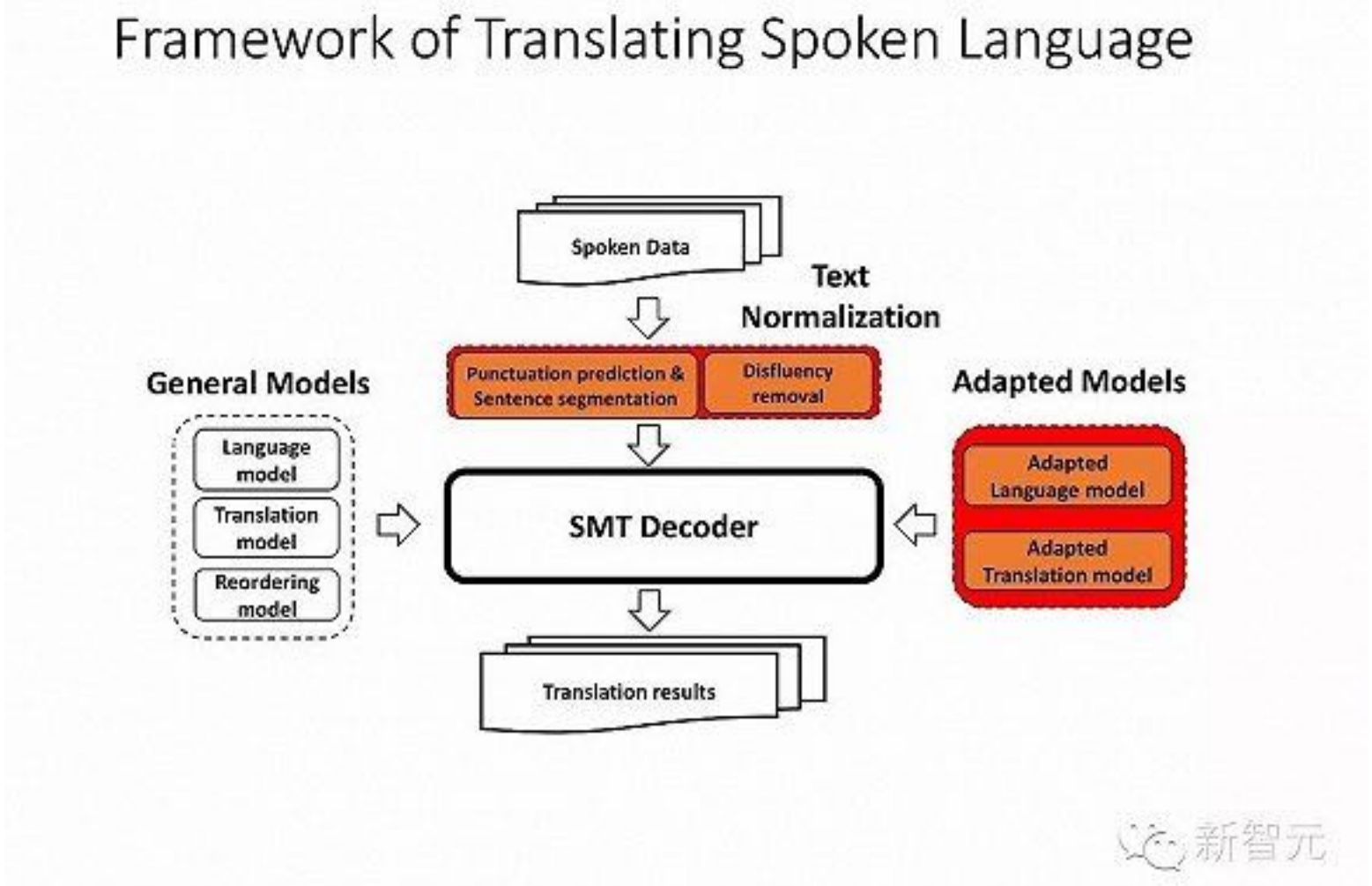

上图这个是美国的背包客来中国已经很久了，喜欢摄影，总共走了35000英里的路，每次去不同的地方去拍摄，他不会说中文，都是一些很痛苦的事情，我们看一下有了Skype Translator之后对他有什么样的帮助?



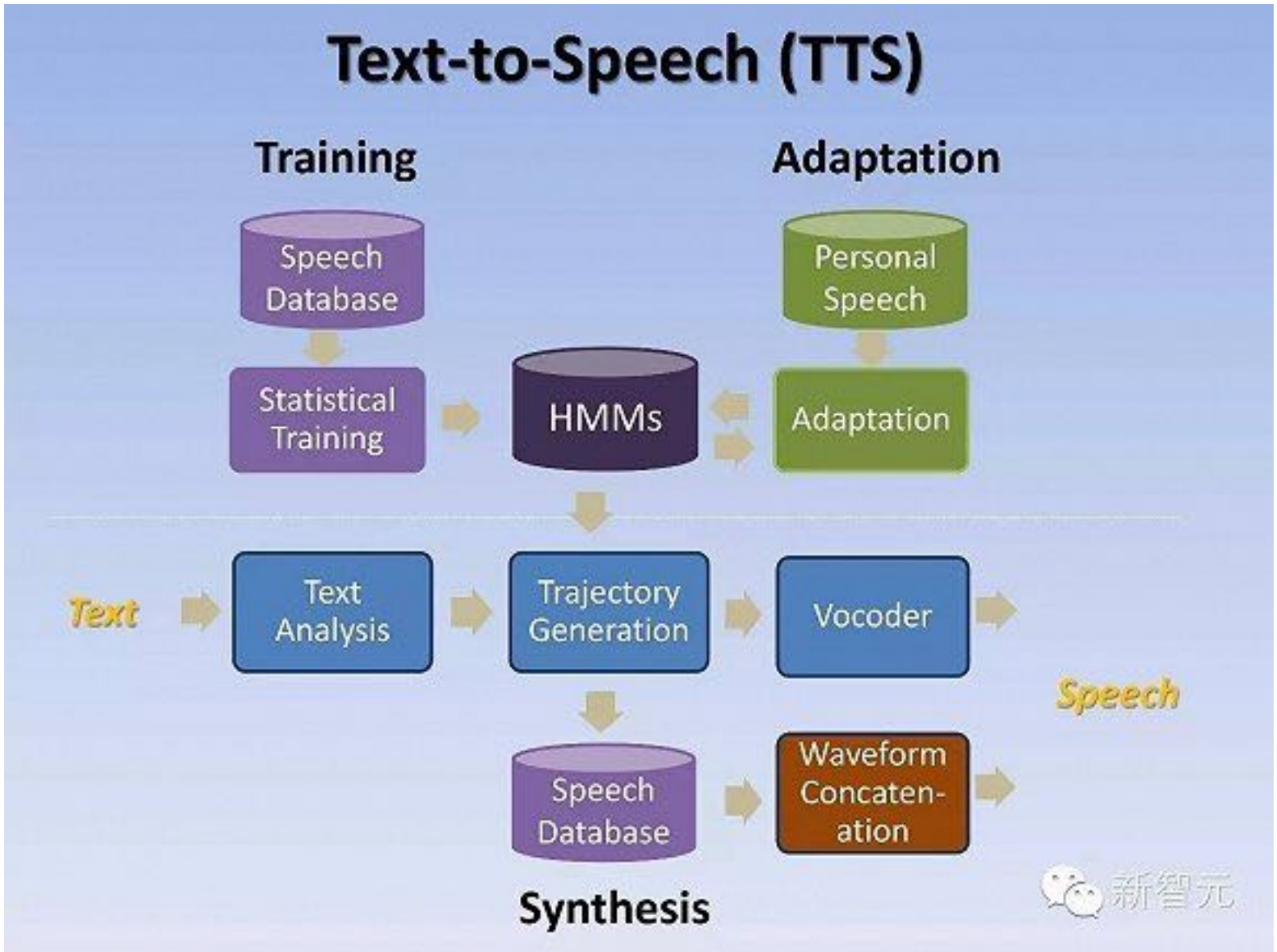

为了能做到这个，其实有很多技术必须去实现的。比如，一个说中文的人和一个说法语的人要实现实时交流。首先，在说中文的时候，计算机需要把音频信号能够实时识别成文字，中文音频信号变成中文的文字。



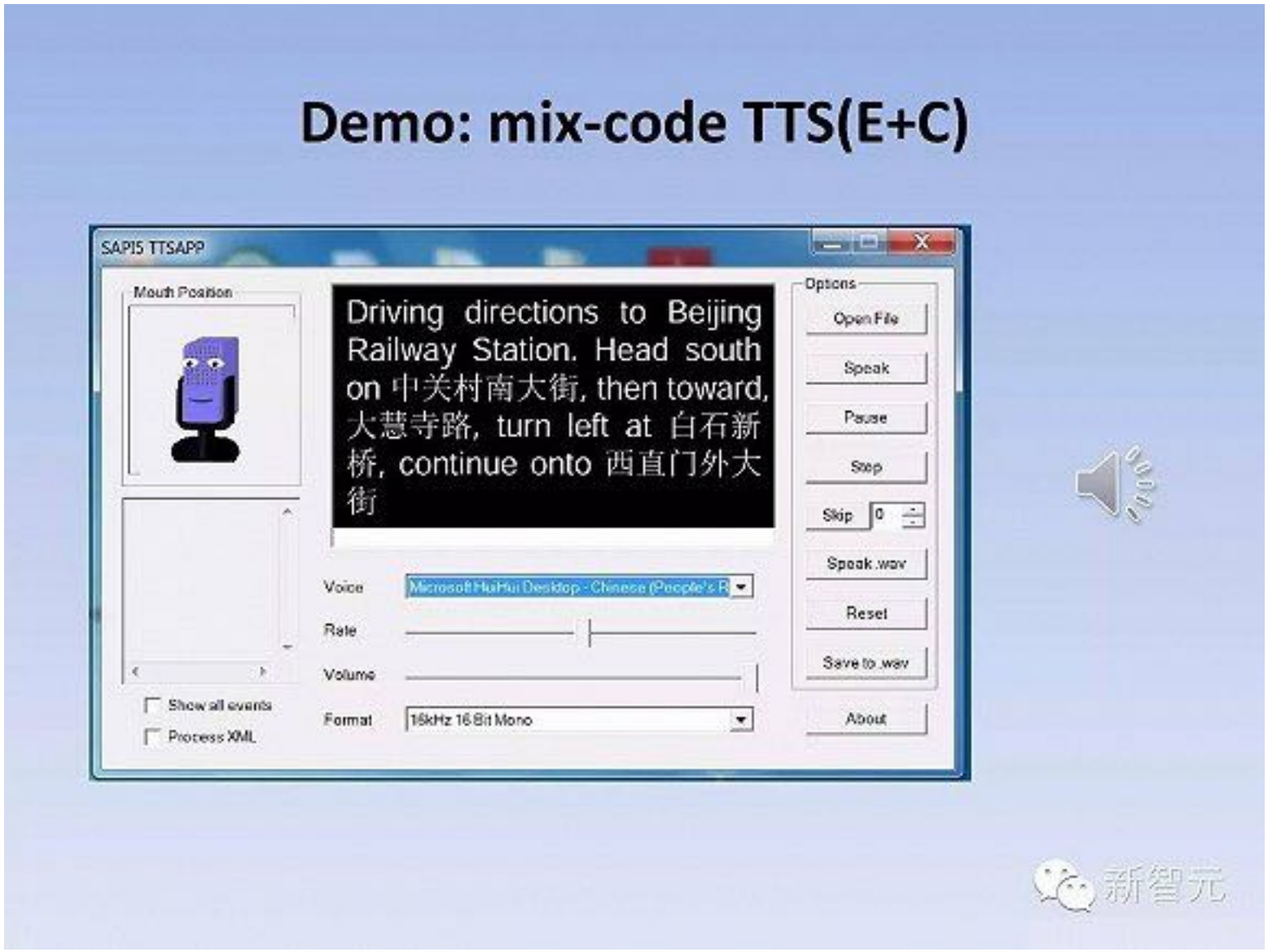

第二，因为不是在读报纸，是在说话，说话的时候人会经常下意识加入一些“恩、啊”的词，要给计算机加这些东西，它一下子就傻了。机器要把你在说话中间毫无意义的东西弄掉，这是必须要做到的事情。




第三，我们有了改良以后更加准确的中文文字，要翻译成法文。



第四步，法文的文字变成法文的声音说出来，以我说话的方式说出这种法文。这四步都要做到，并且是一个串联。这四步中间每一步错一点，乘在一起，错误率就高得没法用了，所以每一步几乎不可能有任何的错误，这就是整个为了实现这种实时不同语言之间的交流，需要做到的一些方法。




这至少有四个技术，第一个技术是语音识别，这几年与深度学习相结合，使原来语音识别的错误率从30%，一下子降到百分之几，这是一个非常大的飞跃，也是得益于深度学习。

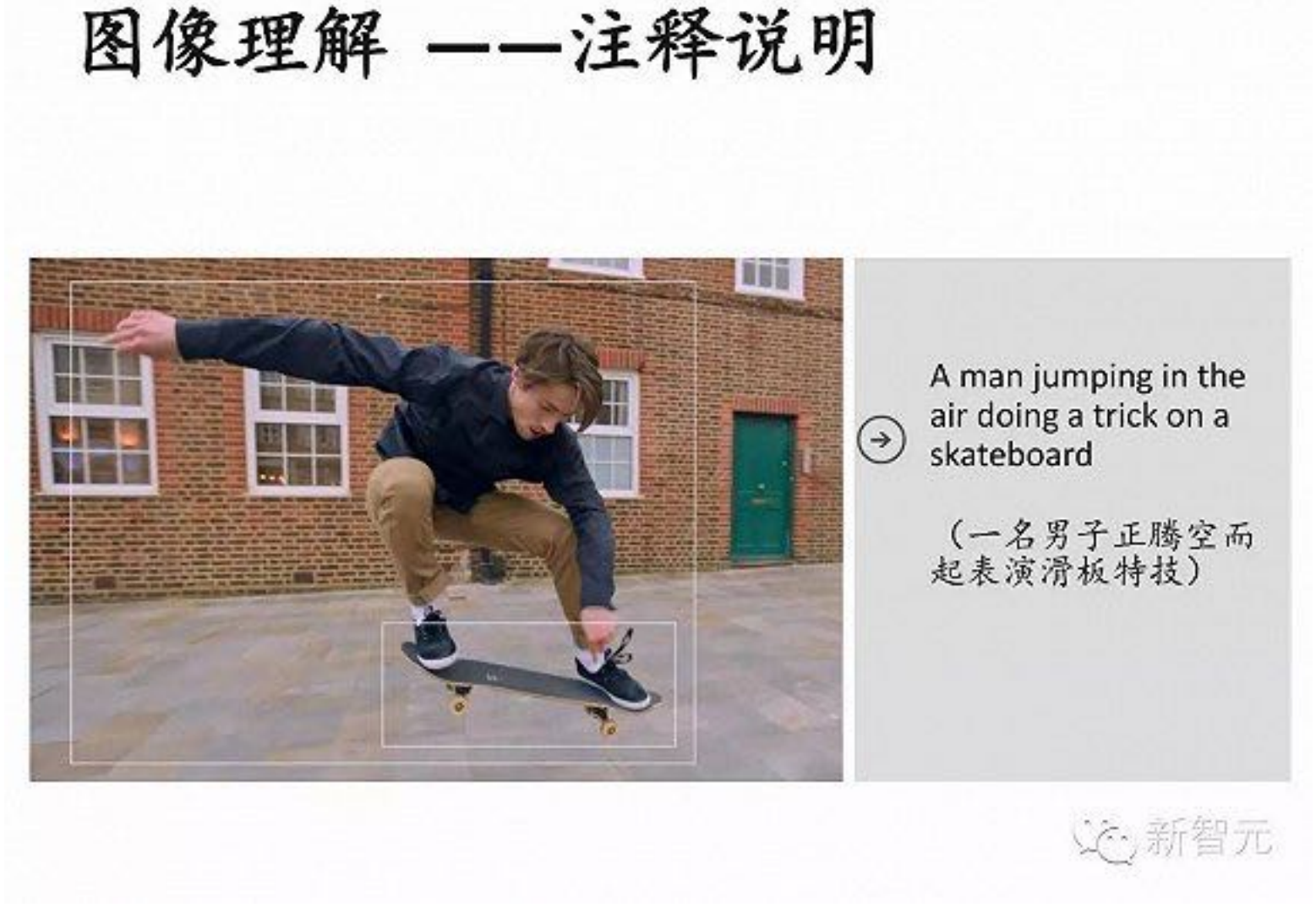
第二、三步就是实时的机器翻译，你本人说话的方式是什么样的，怎么样去调整你语言的模型。

第四步就是把文字再变回到声音，用法文的方式说出来。这个东西很难做，但是也有一些算法可以做。

一触而发的隐形革命


隐形智能，这个就更加难一点。不管是智能的会场，智能的家居，智能的家电，其实我们有很多的传感器在环境里，这些传感器能够实时感知我们在做什么，然后给我们提供更好的服务。这个时候也可能是在楼宇里面，在家庭里面，也可能是我们身上的穿戴设备，比如说我们戴着的眼镜，戴的手表，我们穿的一些衣服，为了使得这些智能的设备，智能的穿戴，智能的楼宇有智能，很重要的一点，就是说他们有没有这种能力看到世界，理解世界的能力，这个就包括了计算机视觉等好几个人工智能技术。






如果我让计算机看这么一幅场景，它能够说出来这名男子在表演滑板。我觉得它真是理解了它所看到的世界是怎么回事。确实今天也有这个技术，我们

可以想象有很多的应用。

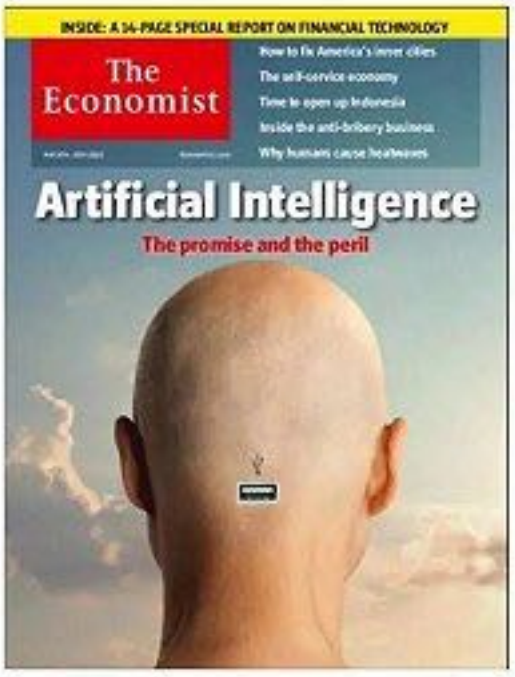

其中的一个应用对我们一些盲人朋友很有帮助。他如果看不见，他可以通过机器的眼睛看到世界，通过声音知道外面在发生什么事情。我在微软有一个同事，七岁时眼睛就失明了。我们一起开发了一个程序，使得看不见的他能听到外面是怎么回事。



还有全息计算的未来，微软Hololens在很多领域都得到了应用：在医疗行业和人体分析的教学，以及在室内设计，在工业设计等。在没有做出模型之前，用Hololens能看到预览效果，可以有效节约成本。

增强智能：机器负责推理，人类负责抽象

# “人工智能”备受关注




Stanford Report, December 16, 2014

## Stanford to host 100-year study on artificial intelligence

Stanford University will lead a 100-year effort to study the long-term implications of artificial intelligence in all aspects of life.

BY CHRIS CESARE

Stanford University has invited leading thinkers from several institutions to begin a 100-year effort to study and anticipate how the effects of artificial intelligence will ripple through every aspect of how people work, live and play.

This effort, called the One Hundred Year Study on Artificial Intelligence, or AI100, is the brainchild of computer scientist and Stanford alumnus Eric Horvitz, who, among other credits, is a former president of the Association for the Advancement of Artificial Intelligence.

In that capacity, Horvitz convened a conference in 2009 at which top researchers

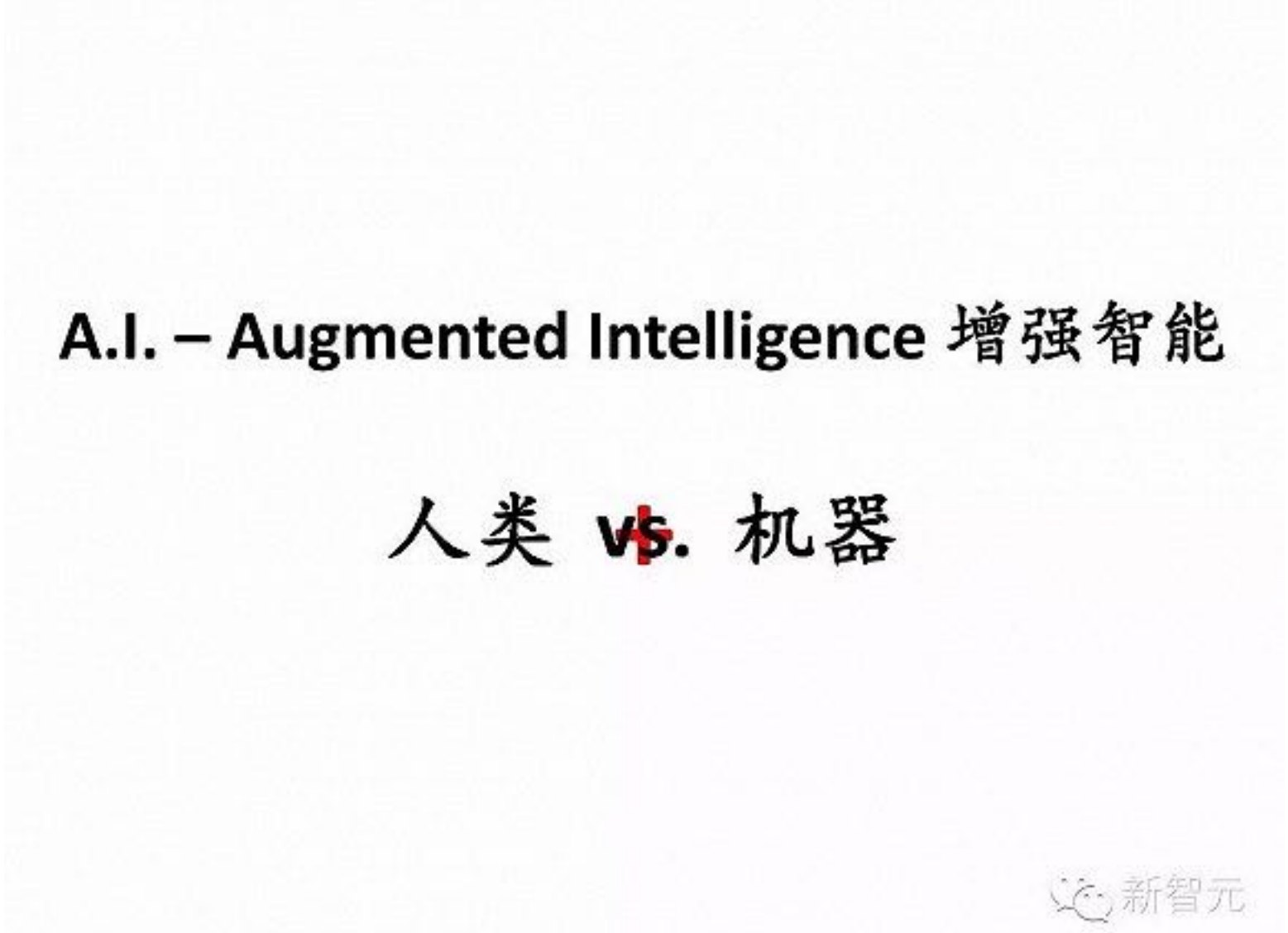
Rod Searcey



Russ Altman, a professor of bioengineering and of computer science at Stanford, will serve as faculty director of the One Hundred Year Study on Artificial Intelligence.



过去七八年，人工智能在全球各个国家都很热，包括在中国、美国，以及欧洲和日本。不仅仅是一些科研院所高校在看，我们做投资的朋友也在看，老百姓也在谈论，但是绝大多数的人工智能还是弱的人工智能，就是一个我们人类规定好的，有规则的事情，机器可以去做了。比如说下国际象棋和围棋，以及其他任何有规则的事情。




人脑有左右脑，左脑是记忆和计算，涉及有规则的事情的推理。右脑是想象力，抽象能力，经常神来之笔都是从右脑来的。弱人工智能是在模拟人的左脑，所以我个人的感觉今后不是说人要和机器去PK，更是说人类和机器各有所长，但是怎么去把对方强的地方结合在一起，成为一个AI，也就是增强智能。要知道，计算机存储计算根本就不是事。大家能把圆周率背到100位之后吗，计算机没有问题的。还有围棋的棋谱记在脑子里，计算机也没有问题的。人的抽象思维和想象力是计算机是很难做到的，今后60年更是一个人类和人工智能相互增强，各取所长，让人类能够更有效的去处理各种事情的时代。

谢谢大家！

*本文整理自芮勇在中信证券与新智元合办的人工智能研讨会上的演讲，芮勇授权新智元刊发*