# 个人助理的时代：更多的机器学习，更少的手工制作！

像Alexa和Siri这样的私人助理无处不在，但在处理任何比在Spotify上设置闹钟或播放歌曲更复杂的任务时都有困难。PolyAI是一家新的伦敦初创公司，拥有一个新一代的平台，用于构建基于语音的代理。我们使用机器学习来处理跨不同应用程序域的复杂任务，使用的语言种类繁多。

第一个电话是1876年3月10日打的。格雷厄姆·贝尔冲他的助手喊道：“沃森先生，过来，我要见你！“这改变了游戏规则。有史以来第一次，我们有了实时远程通信，并且有了它，能够远程自动化各种服务。在他的情况下，贝尔先生可以让他的助手跑腿，而不必亲自去见他。当然，贝尔先生本来可以发一封信、一封电报或一只鸽子，但这不会像拿起他的电话那样快或方便，也不会是实时的！



随着时间的推移，人们可以拨打越来越多的企业电话，远程和实时获取服务。然而，要发现这些服务并不容易。解决方案附带了黄页：列出当地企业及其电话号码的大号印刷目录。

经过长时间的等待，万维网将服务目录从灰蒙蒙的黄页转移到了时尚简约的搜索引擎中。用户可以阅读最新的新闻、订购书籍，甚至可以查看银行账户的余额，而无需等待电话接线员。然而，网站迫使用户与后端系统进行结构化交互，迫使公司在用户界面设计上花费数百万美元，以吸引用户使用其服务。由于没有统一的设计标准，用户必须为每项服务适应不同的界面，而不是单独使用他们的语音。



下一站，智能手机。由于其有限的屏幕空间，智能手机给软件和界面开发人员带来了新的挑战。反过来，苹果为如何创建极其简单直观的移动界面树立了一个令人惊叹的榜样。这些原则在即将到来的AppStore中大量涌现。尽管有太多可用的应用程序，但用户在选择下载哪个应用程序时还是要依靠自己的设备，这并不总是那么简单。事实上，大多数美国智能手机用户。

继智能手机之后，操作系统供应商现在为我们提供虚拟个人助理。机器学习的最新进展导致了语音识别的巨大改进，使谷歌和亚马逊等公司能够将语音驱动的个人助理带到每家每户、每部手机、每块手表或任何其他装有麦克风的硬件上。而不是适应第三方应用程序的界面，助理将它们抽象出来，允许用户单独使用他们的语音访问大量的服务。



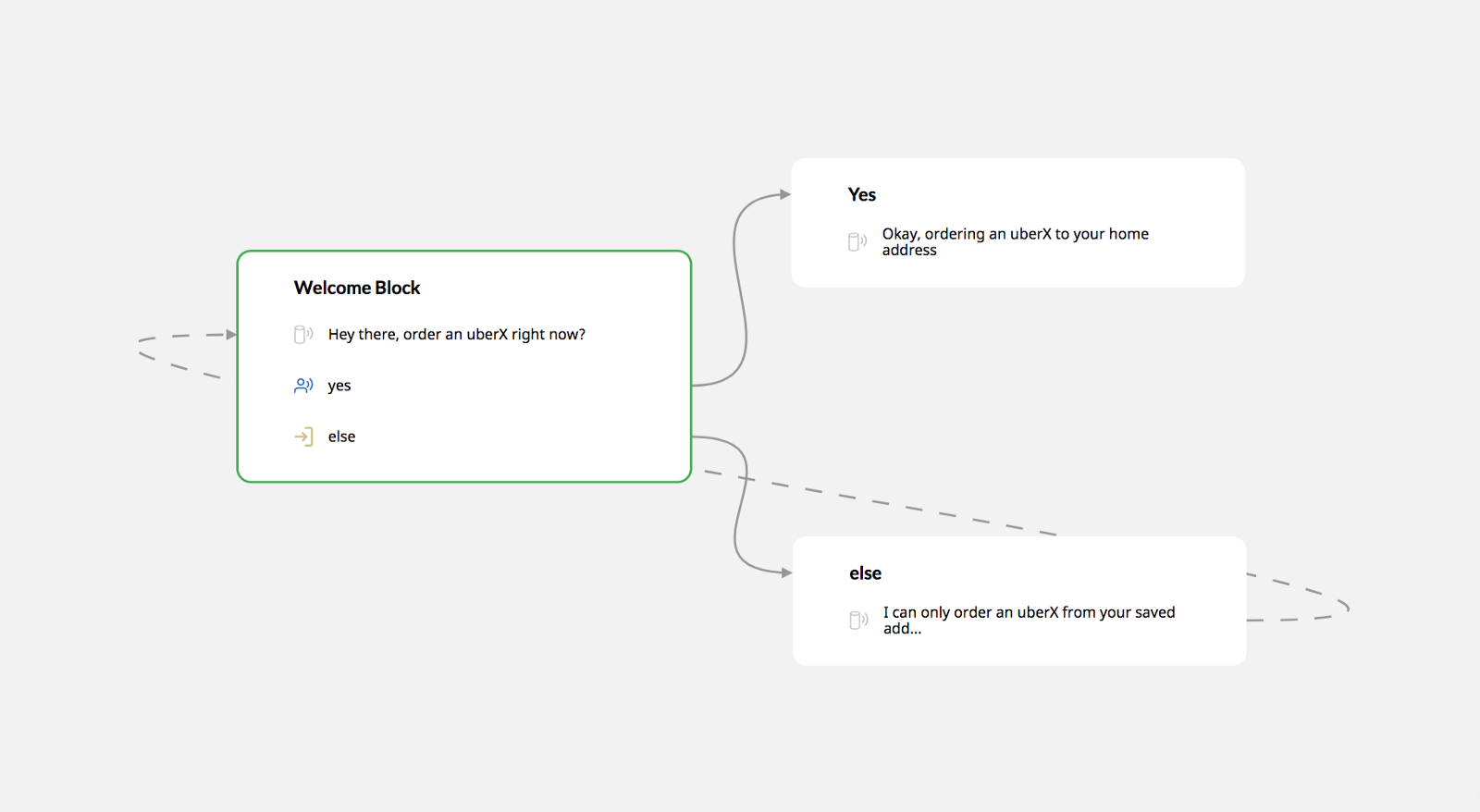
Siri、Alexa和Google Assistant等个人助理希望成为用户可能希望在智能手机、智能家居和其他支持助理功能的设备上执行的大多数操作的实际入口点。个人助理并没有强迫用户选择正确的应用程序，而是为访问第三方服务提供了一个自然的渠道。由于它们是基于语音的，因此允许用户完全绕过图形用户界面。这对于已经习惯了智能手机和其他联网设备的年轻一代来说尤其直观。

从搜索引擎到移动操作系统平台，各大科技公司之间的竞争围绕着平台的控制展开，因为平台是接触数十亿客户的中心点。要实现这些收入流，个人助理平台必须连接到第三方服务。如果不是的话，助理们是空脑袋的，就像没有搜索结果的谷歌或者没有AppStore的iPhone。事实上，亚马逊的Echo在很多方面都类似于2007年左右的iPhone。第一款iPhone的触摸屏风靡全球，同样，消费者对去年圣诞节亚马逊最畅销的产品Echo也无法满足。

平台提供商希望为其个人助理提供第三方应用程序。Alexa Skills、Google上的操作、Azure机器人……所有的巨头都在努力使他们的平台成为构建语音应用程序的最佳生态系统。然而，建立人们想要使用的会话应用程序被证明是非常困难的。尽管部署了数万个Alexa技能，其中62%的人，只有4个拥有超过1000个等级。其中，有14个播放环境声音，这表明我们距离使用第三方语音应用来完成复杂的日常任务还很远。

对话应用不会取代移动应用，就像移动应用不会取代网络一样。好的基于语音的应用程序将简单地为这些场景补充现有的接口。现有的BOT建筑服务只是一个开始。

像Google的Dialogflow或Amazon Lex，以及a这样的工具一直试图通过使用流程图和基本的机器学习模型来简化会话应用程序的设计。这些工具通过调节可编程系统对用户输入和先前对话上下文的响应，帮助开发人员手工制作对话流程。这些工具适用于简单的语音应用程序，例如可以向您的房子订购uberX的应用程序。然而，他们再也不能参与进来了。



这些工具有一个相当温和的学习曲线，在理论上工作良好。然而，开发人员往往发现，创建一个非常有用的应用程序意味着他们必须满足特定的用户偏好。在Uber的例子中，这意味着理解用户是喜欢UberX还是Uber Black，如果他们现在订购还是在15分钟内订购，以及到哪个目的地？在发言时，用户通常按照他们认为合适的顺序指定这些条件。为了处理用户输入的所有排列，流程图变成了非常难以管理或进一步改进的东西。

这种设计范式意味着开发人员需要为每个可能的用户输入及其前面的对话上下文指定响应。然而，域越复杂，可能出现的场景就越多。事实上，通向成功对话完成的路径数量随着应用程序域的复杂性而呈指数增长。这使得设计语音驱动的应用程序比在Spotify上设置闹钟或播放歌曲更难（如果不是完全不可能的话）。

是一家总部位于伦敦的技术公司，由三位剑桥实验室的同事创建，他们在过去的四年里一直在攻读博士学位，研究如何利用机器学习的最新发展，使口语对话系统更强大、更易于设计。

我们创建PolyAI是为了向全世界展示，口语对话系统设计不必涉及大量的手工制作或噱头来引导用户远离困难的用例。通过开发深度学习算法，我们允许会话代理学习如何解释用户想要的内容，以及如何在不必遵循脚本的情况下智能地响应。我们的算法以自然支持多种语言和应用程序域的方式提供这样的性能。所以我们叫波利亚！

虚拟个人助理的时代已经到来，但在我们有相当于支持助理的平台的AppStore之前，还有很多事情要做。当格雷厄姆·贝尔发明电话时，他几乎无法想象它的变革影响。同样，很难说功能齐全、自然的基于语音的应用程序会对我们的生活以及我们与技术互动的方式产生怎样的影响。如果你想在基于语音的会话代理的设计中使用适当的机器！

Special thanks to [Angus Bayley](https://medium.com/@angusbayley), [Nathan Benaich](https://medium.com/@NathanBenaich), [Matt Clifford](https://medium.com/@matthewclifford) and [Eileen Burbidge](https://medium.com/@eileentso) for their help with this post! To learn more about PolyAI, stay tuned.