

# 市场新动态

2019 年 3 月

## · IBM “可解释的” AI — 让人工智能 变得透明、可信

作者：Charles King , Pund-IT, Inc.

Pund-IT, Inc.  
Hayward, CA  
U.S.A.94541

联系方式：  
charles@pund-it.com  
电话：510-909-0750  
www.pund-it.com

## IBM “可解释的” AI - 让人工智能变得透明、可信

作者：Charles King，Pund-IT, Inc.

### 引言

在有关现代计算系统的讨论中，信任问题鲜少出现。这并不是说我们的硬件和软件已经完美无瑕。事实上，出版物和在线论坛中包含数万篇文章，详细列出了各种个人电脑、工作站和服务器的相对优点。但这些产品存在时间已久且普及广泛，人们对它们基本的“正确性”和所提供的结果鲜少质疑。

然而，事情并不总是这样，大部分新兴技术和科学突破并未得到这样的“待遇”，包括专为企业和其他组织而设计的商业人工智能 (AI) 解决方案。考虑到机器学习、神经网络和其他人工智能相关流程的固有复杂性，客户对人工智能感到困惑不已也是预料之中的事情。但怎样才能打消他们的疑虑，让人工智能成为主流呢？

包括 IBM 在内的供应商自有应对之道，他们推出了一系列旨在让人工智能流程和结果更易于解释、理解且可信的解决方案。这应该能打消客户的疑虑，加速商业人工智能产品的采用。但可解释的 AI 也能带来其他重大好处。让我们考虑一下为何可解释的 AI 如此重要，以及 IBM 创新技术如何影响其客户。

### 人工智能的黑匣子问题

缺乏清晰的认知和理解往往会造成一系列问题。如果无法将人工智能解释清楚，可能会出现三个问题：

1. 首先，缺乏透明度会让用户无法确定结果的有效性和准确性。这就削弱了人工智能项目和流程的基本价值。
2. 其次，如果人工智能项目不可解释，其结果可能会出现偏差或不准确。这个问题可能您自己都意识不到。
3. 最后，如果人工智能流程不可解释，对异常结果进行诊断排除会很困难，甚至不可能。因为缺乏透明度，组织无法修复受损部分。

专注于人工智能的供应商如何解决这些问题？很遗憾，他们的解决方法往往会让问题更严重，其中包括“黑匣子”解决方案。这些供应商自称将交付人工智能的所有优势，但却无法提供足够的透明度，让客户了解其工作方式，知道如何确定结果是否准确或者如何解决问题。

这些解决方案还鼓动大家将人工智能视为神秘的存在，认为其功能不被一般人所理解。换句话说，不管人工智能“黑匣子”会带来怎样的好处，让客户“蒙在鼓里”不利于他们开展工作和实现目标。

## 可解释的 AI 的优势

是否有更好的办法？当然有。组织怎样才能成功地解释 AI？答案就是通过可应对人工智能生命周期的几个不同阶段的全面方法：

- 让人工智能项目变得更加可解释的关键就是让人工智能模型变得可解释，摆脱当前所处的黑匣子状态。
- 组织必须以清晰、透明的方式阐明人工智能在业务流程中的使用位置、方式以及目的。
- 他们还必须支持对人工智能结果进行分析，并在必要时提供更改和覆盖这些结果的办法。

为此，专业的用户和管理人员应该采用供应商设法用于增强人工智能透明度的技术和解决方案。这些方法可以让人们更快地理解人工智能，非常不错。这对于 IT、市场营销、销售和客户服务组织，尤其是高度管制行业（如银行和保险）中的组织来说，也是一个关键问题。

整个过程可以有机地发生，人们先是体验到人工智能，然后开始理解它给企业和个人所带来的影响，最后影响组织的文化。人们也可以利用当前可用的最佳工具和解决方案主动实现可解释的 AI。不管企业采用何种方式，人们都需要牢记人工智能的巨大潜力。为什么呢？因为一个时代即将来临，届时人工智能对于组织是否成功的重要性就等同于当前普遍应用的商业技术。

## 可解释的 AI 的商业优势

为何可解释的 AI 是一个如此重要的问题，一项如此重要的任务？这涉及到组织开展业务的真正起因。如果他们要采用并适应人工智能流程，就需要了解结果是否准确。否则，他们怎么向客户和合作伙伴保证受人工智能影响的决策是有效且可靠的？考虑下面两个例子：

1. 在金融服务领域，很显然，结果的准确性对于最大程度实现业务成果和开展客户交互十分关键。但是，像其他高度管制行业中的企业一样，银行和其他金融机构必须能够证明受人工智能影响的流程符合政府和行业规定，否则可能面临严重的制裁和处罚风险。这就够糟的了，但无法解释的 AI 还可能会破坏客户关系。如果客户寻求贷款，而信用卡或其他服务均被人工智能相关的系统拒绝，公司管理人员必须能够解释为何做出这一决定，并告知客户如何解决或纠正问题。
2. 全球供应链管理是另一个人工智能将大有作为的领域，因为供应链数据的复杂性、数量和多样性，让人们很难有效地跟踪需求的实时变化并做出相应调整。人工智能可以提高增值预测 (FVA) 指标 - 从过去成功和失败的预测中汲取经验，帮助规划者更好地做出调整。但除非供应链团队能够轻松监控人工智能模型的准确性，否则他们无法确保系统是否真正交付了这些模型所承诺的效益。

鉴于这些及其他一些观点，很难理解为何供应商会开发或者客户会考虑采用不可解释的 AI 解决方案。

## IBM 做了哪些工作来打开人工智能“黑匣子”

IBM 多管齐下，针对多个领域开发并交付了可解释的 AI 和高级分析解决方案。IBM Watson Data and AI 组织的副总裁 Ritika Gunnar 在最近的一篇博客中强调了公司前进的动力。“作为人类，我们习惯认为决策是基于任何人都可以理解的一系列证据和逻辑推理制定的。但如果人工智能系统基于不同或陌生的标准提出建议，它的结果很难令人信服，也很难解释。”

该公司的工作重心围绕 IBM 于 2018 年推出的 Watson OpenScale 平台。Watson OpenScale 旨在“打开人工智能模型中心的‘黑匣子’”，简化人工智能流程，包括详细阐述如何提出建议，以及自动检测并缓解偏差，确保产生公正、可信的结果。

IBM 目前正在利用现有的开源技术和 IBM 研究院开发的专有算法，探索、改善并解释人工智能的决策制定过程。

- LIME（局部可理解的与模型无关的解释）是一个广泛应用的开源算法，旨在通过对比对可轻松理解的模型的解释，解释人工智能系统做出的预测。
- 由 IBM 研究院开发的 MACEM（与模型无关的对比解释方法）超越了 LIME，可以识别一段数据中存在和不存在的功能，支持构建“对比解释”。

对比解释的一个应用场景是银行，它可用于分析贷款申请数据。该系统会就各种问题向银行发出警报，包括较差的信用评级，还可以发现并指出存在缺失的文件，比如不完整的信用报告。然后，银行可以通知客户其决定的原因，并提出建设性的建议。

从本质上讲，像 IBM Watson OpenScale 这样能交付更加准确、透明且可靠的人工智能结果的解决方案，可以帮助企业制定更加明智的决策，改善其客户服务，增进与客户的关系。

## 最终分析

人们常常会关注新技术，尤其是高度复杂或难以理解的那些技术。要让这些技术得到人们的广泛信任并取得在商业上的成功，关键是要解决人们的这些疑惑。事实上，如果不增进人们对新技术的了解，那么新的解决方案很难立足，也就无法让人们和组织从中受益。

如果能够清楚地解释诸如人工智能这样的技术以及基于人工智能的解决方案和服务，供应商就能更快地将新的服务产品投入主流市场。这就是可解释的 AI 服务产品（如 IBM Watson OpenScale）如此重要的原因。

通过打开“人工智能中心的黑匣子”，让流程和结果完全可解释，IBM 将为其客户和合作伙伴带来帮助，也将进一步推进自己的市场战略。更重要的是，IBM 在可解释的 AI 方面所做的努力将有助于建立这些解决方案的基本“正确性”，使其成为完全有效且有价值的企业技术。

总的来说，IBM 的努力将推动全球的大部分个人和组织理解、接受和采用人工智能。

*© 2019 Pund-IT, Inc. All rights reserved.*

关于 Pund-IT, Inc.

*Pund-IT™ ([www.pund-it.com](http://www.pund-it.com)) 强调要理解技术和产品的演进，并解释这些变化对于企业客户及更广泛的 IT 市场所带来的影响。尽管 Pund-IT 为技术供应商提供咨询和其他服务，但本文中所表述的仅仅是作者本人的观点。*