

## 056 | 成为“香饽饽”的数据科学家，如何衡量他们的工作呢？

2018-02-09 洪亮劼

AI技术内参

[进入课程 >](#)



讲述：初明明

时长 08:24 大小 3.85M



本周我们聊了在构建一个数据科学家团队时，从筛选简历入手到电话面试，再到 Onsite 面试这一系列的流程。从无到有，建立一个数据科学家或者人工智能团队的确是一件煞费苦心的事情。

那么今天，我们来聊一聊数据科学家团队管理的下一个重要的步骤，那就是**如何来衡量数据科学家或者人工智能工程师在团队中的业绩，有时候也被称为是绩效评定**（Performance Review）。绩效评定的种种规则必须在团队建立的初期就明确，否则就会出现一些不定因素，对于招聘、培训以及留住人才都有着不可估量的影响。

### 数据科学家的价值

如何对数据科学家团队进行绩效评定呢？这个问题的核心其实是要回答，数据科学家或者人工智能工程师究竟应该（以及实际）为你的公司或者组织带来什么核心价值？只有梳理清楚这个核心问题，才能真正建立起衡量数据科学家团队的价值体系，从而达到为公司和组织赋能的目的。

那么，数据科学家团队或者人工智能工程师团队应该为企业或组织带来什么样的价值呢？

对于这类相对来说比较抽象的问题，其实很难有一个标准答案。每一个组织或者公司都有自己一套衡量价值的方式。这里我们并不追求一个统一的答案，而是希望能够为这个问题提供一些参考。

关于这个问题，在我们前面的一些分享中，其实已经提到过，那就是**数据科学团队最重要的一部分价值来自于为企业或者团队引入数据驱动的决策过程**，这是数据科学团队或者人工智能团队的一个核心价值。很多企业或者组织，在没有这些团队之前是无法真正做到数据驱动、持续决策的。

也就是说，“**数据驱动**”和“**持续决策**”这两点可以看作是数据科学家团队的主要价值体现。那么，怎么衡量数据科学家团队这个问题，也就变成了如何来衡量这些团队在围绕这两方面的工作中做得怎么样。

注意，我在这里其实并没有明确提及数据科学团队和人工智能团队对产品直接带来的效果，比如点击率升高了多少，用户存留增加了多少，什么产品又上线了等等。主要是出于以下两点考虑。

第一，每一个公司、每一个组织甚至是每一个产品在这些价值上都有不一样的需求，没有一个统一的模式。

第二，如果数据科学团队为组织建立起了数据驱动的持续决策过程，那么很多产品级别的性能提高或者核心功能的实现就会成为顺理成章、水到渠成的结果。相反，如果仅仅强调某一个产品性能的提升或者某一个单点技术的突破，很可能无法真正建立有效的人工智能团队，并且团队的“战斗力”也无法真正得到最大程度的发挥。

## 数据科学家团队的评价误区

刚才我们从一个比较大的概念上做了一个讨论，看数据科学家团队的价值应该如何来评价。但在实际操作中，往往存在两种比较明显的误区。

**第一种误区是“唯技术论”**。那就是觉得人工智能团队能够快速帮助公司、组织甚至是项目很快打开突破口，希望人工智能技术能够给公司业务带来突破性的发展，从而对人工智能团队有过高的预期。

在这种思路的指导下，在前期往往可能有一个比较大的热情，能够招聘到不少的人才或者能够拉起团队开始一些不错的项目。但很快，由于急功近利的心态和不切合实际的需求，常常又让人工智能团队身陷绝境。而这个时候，最容易产生的一种情绪是走另外一个极端，那就是从“唯技术论”到“技术无用论”。在这样的背景下，产品遇到的任何困难、任何失败都有可能归因到人工智能团队上。

**第二种误区是对人工智能团队或者数据科学家团队心存怀疑，本质上觉得这些团队都无法真正能够帮助到团队**。因此从一开始就不信任这些团队，蹑手蹑脚，在政策和发展上限制这些团队。由于这种不信任，使得人工智能团队不能真正发挥作用，因此催生了进一步的不信任，恶性循环，最终得出这样的结论，“人工智能是花瓶，没有用”。

这两种错误的核心其实是一种行为，那就是忽略了**人工智能团队需要一个“生态环境”**。什么生态环境？比如产品部门、数据部门以及其他的工程部门，必须协调发展。

绝大多数数据科学团队和人工智能团队都需要依赖一个比较强有力的数据部门的支持。同时，产品上，如果人工智能的算法或者模型并不能和产品有机结合，那无论如何，都是无法真正帮助产品，为产品赋能的。我们之前也提过，其实在很多时候，人工智能都是锦上添花的部分，而产品的整体呈现才是最为重要、做需要认真思考的问题。

## 数据科学家的评定

有了前面这些思路作为基础，我们现在来看一看数据科学家的评定的问题。

**第一，对于数据科学家或者人工智能工程师来说，我们需要看他们是否对于建立、完善、和推动“数据驱动的持续决策”这一长期任务有不间断的贡献。**

具体来说，那就是数据科学家或者工程师是不是在帮助建立和推动数据驱动的链路，是不是在思考如何能够更快、更好地解决数据的问题。这里的数据包括获取数据、整理数据、分析数据以及利用数据的整个流程。我们的数据科学家或者人工智能工程师应该持续在这几个方面有所贡献。

你可能会疑问，这不是数据工程师的责任吗？没错，这确实是数据工程师的职责。但是，如果我们的科学家并不清楚数据的情况，并不了解如何进一步推动数据链路的进步，那产品线将来肯定会出问题。

同理，数据科学家也需要在帮助“持续决策”上不断做出贡献。这里主要指的是实验的平台，以及围绕着实验平台进行决策的工具，比如图表，比如更加复杂的假设检验工具，比如因果推断的工具等等。数据科学家和人工智能工程师必须要具备这样的敏感度。

**第二，那就是考察数据科学家和人工智能工程师本身的职责和专长，针对某一个产品能否提出切实可行的机器学习解决方案，能否和产品部门以及其他工程部门一起，让解决方案落地。**

这一点检验的就是解决方案的落地能力。当然，这里不仅仅依赖于解决方案本身，还依赖于其他的因素，比如数据，比如产品。这里面有一部分是第一点的内容，主要是评定数据科学家或者人工智能工程师对于跨部门合作以及共同构建一个人工智能生态系统的能力。这一条的重点是评定在一个较小范围内落地解决方案的能力。

**第三，那就是数据科学家和人工智能工程师必须能够不断提高自我修养，能够持续学习不断进步。**

这一点，可能是在所有的工程团队和产品团队里面都比较突出的。虽然所有的团队都需要不断进步，然而人工智能这个领域实在是变化太快。因此，在这个方面，人工智能相关的工作都必须要有比较不一样的评价标准。

在一些企业中，数据科学家的持续学习主要体现为参加会议、发表论文、参与学术讨论、发表开源软件等形式。如果在一些初创公司或者是暂时没有这些能力的组织中，我们也要思考如何来评价员工是不是在积极地持续学习。

## 小结

今天我们分析了组建一个数据科学家或人工智能团队后，你怎样来认识这个团队的价值，怎么来评价员工的工作。

简单地做个总结：第一，我们讲了数据科学家团队对于推动“数据驱动持续决策”这一目标的作用；第二，我们梳理了面对人工智能团队上可能存在的两个误区；第三，我们简单聊了聊如何在大的方面来评定数据科学家的工作。希望这些内容能给你带来一些启发和借鉴。



最后，给你留一个思考题，需不需要把人工智能团队的工作和企业的 KPI 挂钩？如果需要，该怎么挂；如果你觉得不需要，又是什么理由呢？

欢迎你给我留言，和我一起讨论。



# AI 技术内参

你的360度人工智能信息助理

洪亮劼

Etsy 数据科学主管  
前雅虎研究院资深科学家



新版升级：点击「 请朋友读」，10位好友免费读，邀请订阅更有**现金**奖励。

© 版权归极客邦科技所有，未经许可不得传播售卖。页面已增加防盗追踪，如有侵权极客邦将依法追究其法律责任。

上一篇 055 | 数据科学团队养成：Onsite面试面面观

下一篇 057 | 人工智能领域知识体系更新周期只有5~6年，数据科学家如何培养？

## 精选留言 (1)

写留言



林彦

2018-02-09



我之前的公司里曾经空降了新的部门领导及其嫡系的数据科学家并建立了相应团队后。我当时的感受更多是在提升数据报告的制作效率，可视化的丰富程度和互动操作便利程度上发力，其中涉及数据获取和清洗的工作，数据源是GA, Facebook API，内部和外部的数据库。

...

展开 ✓