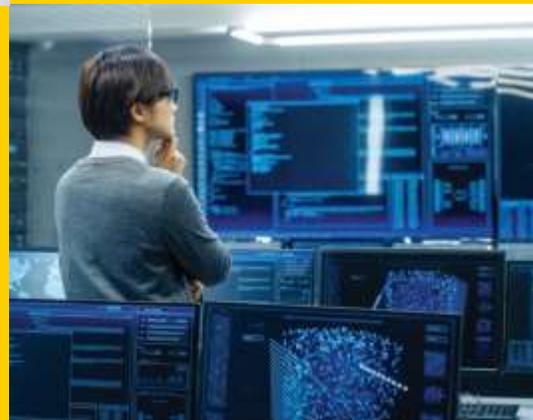
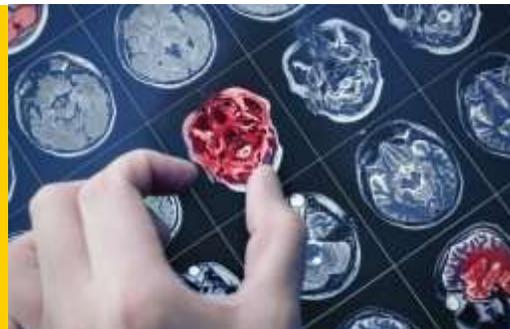


# AI职业发展途径：

步入职业正轨



Workera是一家deeplearning.ai公司，通过提供指导和优质的工作机会，帮助数据科学家，机器学习工程师和软件工程师实现其职业目标。我们的使命是确保每个人，无论背景如何，都可以发挥最大的潜力并实现其在AI方面的职业目标。



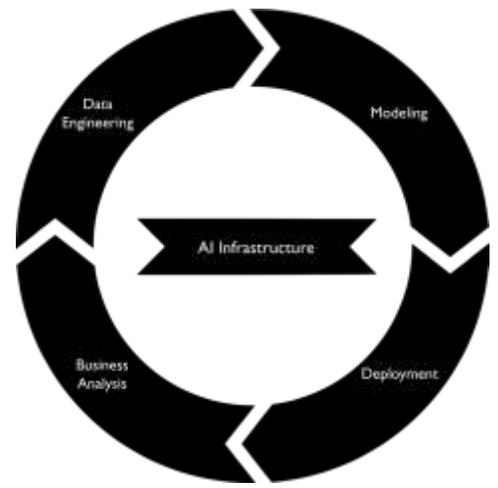
**WORKERA**

一家deeplearning.ai公司

# I 执行摘要

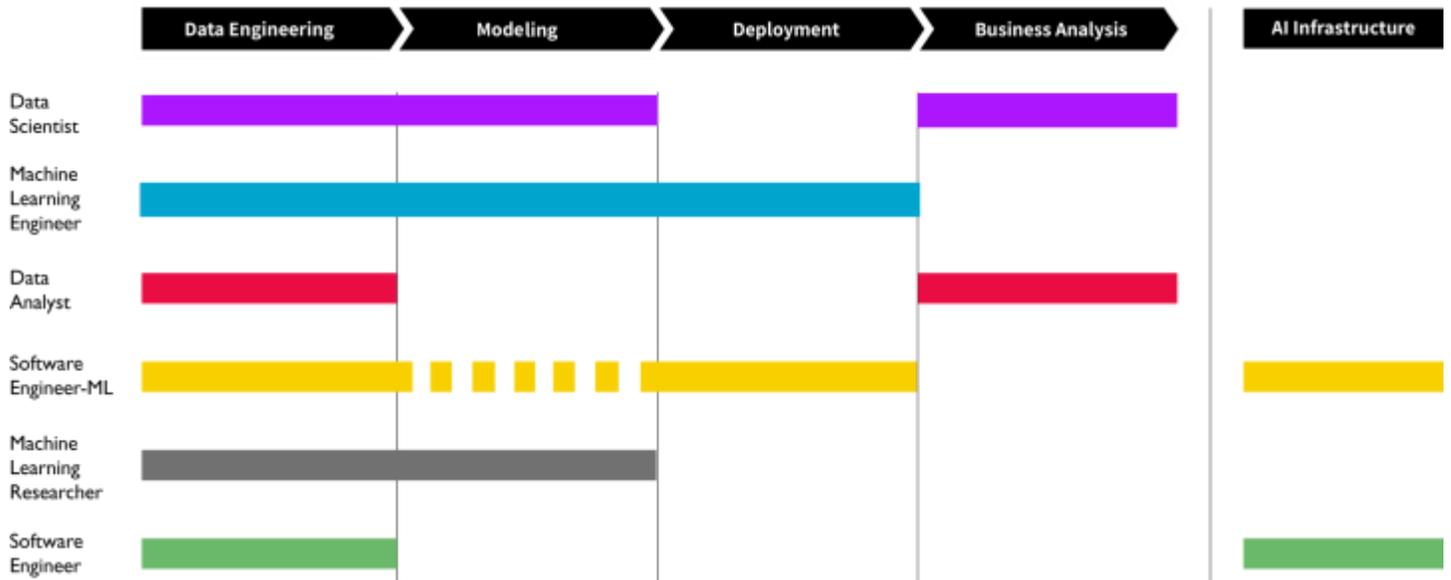
AI项目开发生命周期涉及五个不同的任务：

- **数据工程**：负责数据工程的人员准备数据并将数据转换为其他团队成员可以使用的格式。
- **建模**：负责建模的人员会在数据中寻找模式，这些模式可以帮助公司预测各种决策的结果，识别业务风险和机会或确定因果关系。
- **部署**：负责部署的人员在将模型投入生产之前，先获取数据流，并将其与模型组合，然后测试集成。
- **业务分析**：负责业务分析的团队成员评估已部署模型的绩效和业务价值，并进行相应调整以最大化收益或放弃无效模型。
- **AI基础架构**：从事AI基础架构的人们可以构建和维护可靠，快速，安全和可扩展的软件系统，以帮助从事数据工程，建模，部署和业务分析的人们。



六个基本角色执行这些任务。

没有一个人具有足够的技能来执行AI项目开发中的所有任务。因此，团队包括专注于周期某一部分的个人。这是六个技术角色及其与各种任务的关系的直观表示：



每个角色都需要特定的技能和知识。

负责**数据工程**的人员需要强大的编码和软件工程技能，最好与机器学习技能结合在一起，以帮助他们做出与数据相关的良好设计决策。大多数时候，数据工程都是使用数据库查询语言（例如SQL）和面向对象的编程语言（例如Python，C++和Java）来完成的。Hadoop和Hive等大数据工具也很常用。

**建模**通常使用Python，R，Matlab，C++，Java或其他语言进行编程。它要求在数学，数据科学和机器学习方面有坚实的基础。一些组织需要深度学习技能，尤其是那些专注于计算机视觉，自然语言处理或语音识别的组织。

从事**部署**工作的人员需要编写生产代码，拥有强大的后端工程技能（使用Python，Java，C++等），并且需要了解云技术（例如AWS，GCP和Azure）。

从事**业务分析**的团队成员需要了解用于分析的数学和数据科学知识，以及较强的沟通技巧和业务敏锐性。他们有时会使用R，Python和Tableau等编程语言，尽管许多任务可以在电子表格，PowerPoint或Keynote或A/B测试软件中完成。

在**AI基础架构**上工作需要广泛的软件工程技能来编写生产代码和理解云技术。

# 目录

第1部分：AI组织.....	4
数据科学与机器学习组织.....	5
第2部分：AI开发生命周期的任务和技能.....	6
AI项目开发生命周期概述.....	7
数据工程.....	8
建模.....	9
部署.....	10
商业分析.....	11
人工智能基础设施.....	12
第3部分：AI团队的角色.....	13
总览.....	14
数据科学家.....	15
数据分析师.....	16
机器学习工程师.....	17
机器学习研究员.....	18
软件工程师-机器学习.....	19
软件工程师.....	20
附录：定义技能.....	21
结论.....	22

第I部分

# AI 组织

# 数据科学与机器学习组织

我们确定了两种使用AI的组织：

- **数据科学组织**可以帮助公司的领导者做出科学或数据驱动的决策，从而更有效地开展业务。团队成员收集数据，分析数据集并提出假设和行动。
- **机器学习组织**使任务自动化以降低成本或扩展产品规模。输出是通过收集数据，训练模型并部署它们而实现的自动化本身。

尽管机器学习和数据科学组织不同，但是公司经常互换使用这些术语。您可以通过评估给定组织是否符合上述说明之一来区分彼此。一些公司拥有混合组织，既可以制定数据科学决策又可以自动执行任务。在本报告中，当提及数据科学，机器学习或混合组织时，我们将使用短语“AI组织”。

## 第2部分

# 人工智能开发生命周期的任务和技能

AI组织将其工作划分为数据工程，建模，部署，业务分析和AI基础架构，这些任务共同构成了AI项目开发生命周期。每个任务都需要特定的技能，并且可能是多个角色的重点。

您可以在附录中找到本报告中提到的技能的简要说明。

我们将讨论机器学习（ML）和数据科学（DS）的项目开发生命周期之间的差异，然后研究每个任务的目标，执行任务所需的技能以及组织中的哪些角色专注于哪些任务。

# AI项目开发生命周期概述

这是AI项目开发生命周期的工作方式。

首先，有人为建模准备数据。然后有人在此数据上训练模型。一旦训练完成，模型便交付给客户。然后，团队成员分析模型，以确定它是否为企业和/或用户带来了价值。如果一切顺利，该周期将重复进行新的数据，模型和分析。一直以来，从事AI基础架构工作的人们都在构建软件来提高周期的效率。

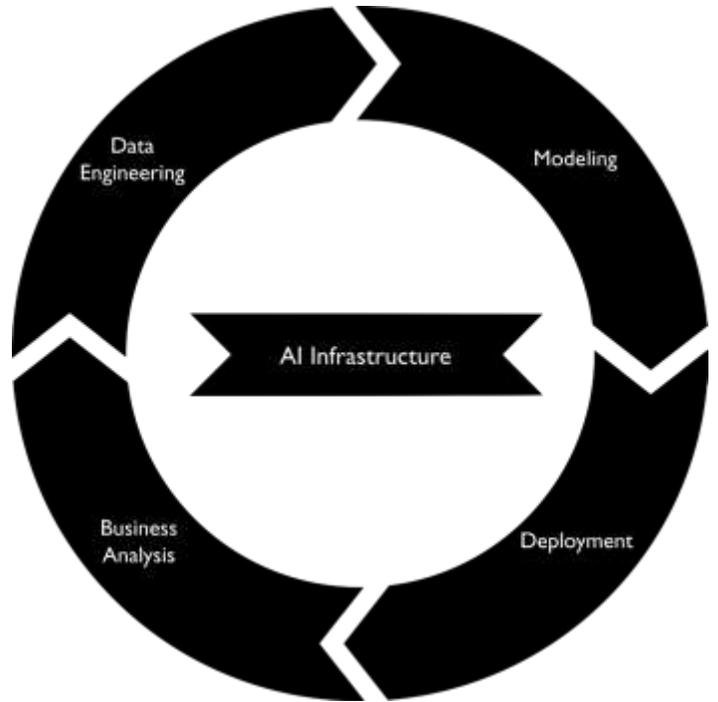
## 机器学习项目开发生命周期

机器学习项目从与服务或产品有关的数据开始，您可以在该数据或服务上拟合要部署到生产中的模型。这些模型需要进行监控并评估其性能。公司的AI基础架构支持这些任务。

## 数据科学项目开发生命周期

数据科学项目从数据开始，您可以在该数据上建立模型，该模型可以帮助公司或其客户做出有关产品或服务的可行业务决策。AI基础架构支持所有涉及的任务。

现在，让我们逐一考虑AI项目中的任务。我们将通过具体示例说明每个任务，并确定执行这些任务所需的技术技能。



# I 数据工程

公司收集数据并将其存储在各种数据库和文件中。负责数据工程的人员准备数据并将其转换为其他团队成员可以使用的格式。

从事这项工作的人需要强大的编码和软件工程技能，最好与机器学习技能结合起来，以帮助他们做出与数据相关的良好设计决策。大多数时候，数据工程都是使用数据库查询语言（例如SQL）和面向对象的编程语言（例如Python，C++或Java）来完成的。Hadoop和Hive等大数据工具也很常用。

数据工程工作包括：

子任务	例子	涉及技能
定义数据要求	<ul style="list-style-type: none"><li>- 创建数据模型</li><li>- 定义高质量数据的特征</li><li>- 定义要收集的协变量以实现所需功能</li><li>- 提供有关数据要求的清晰度和完整性的反馈</li></ul>	机器学习商业敏锐度 软件工程
收集数据	<ul style="list-style-type: none"><li>- 设置Mechanical Turk项目</li><li>- 通过手动拍摄猫的图像来收集数据</li><li>- 在网站上编码JavaScript跟踪器以收集用户数据</li><li>- 搜刮网络，并在必要时同步来自不同来源的数据</li></ul>	机器学习软件工程
给数据打标签	<ul style="list-style-type: none"><li>- 在图像上绘制边框</li><li>- 使用Mechanical Turk建立标签管道</li><li>- 为工人编写标签教程</li><li>- 重新贴错标签的数据</li><li>- 评估工人的标签表现</li></ul>	机器学习软件工程
检查和清洁数据	<ul style="list-style-type: none"><li>- 使用Python库（例如pandas）替换所有不可用的结构化数据记录为NaN</li><li>- 使用bucketing将连续特征转换为分类特征</li><li>- 重新格式化数据集（例如，将图像转换为jpeg并进行尺寸处理）</li><li>- 清理文本数据集（例如，删除特殊字符）</li></ul>	机器学习算法编程
扩充数据	<ul style="list-style-type: none"><li>- 使用skimage编写Python脚本以旋转，扭曲，平移或模糊图像</li><li>- 使用测试时扩充（test-time augmentation）来减少算法的方差</li><li>- 通过叠加不同的音频信号来合成语音</li></ul>	机器学习算法编程
移动数据和建立数据管道	<ul style="list-style-type: none"><li>- 编写脚本以允许在线学习模型</li><li>- 设计ETL系统</li><li>- 编写脚本以预处理训练数据并将其作为输入自动发送到模型</li><li>- 编写脚本以在数据库中记录模型预测</li></ul>	特定领域（例如，数据查询）语言
查询数据	<ul style="list-style-type: none"><li>- 从数据库中提取数据</li></ul>	特定领域（例如，数据查询）语言
追踪数据	<ul style="list-style-type: none"><li>- 跟踪数据源</li><li>- 设置数据版本控制系统</li></ul>	软件工程

# 建模

分配给建模人员查找数据中的模式，这些模式可以帮助公司预测各种决策的结果，识别风险和机会或确定因果关系。例如，房地产模型可能会采用房屋大小，卧室数量，位置和年龄来预测售价。

建模通常使用Python, R, Matlab, C

++, Java或其他语言进行编程。它要求在数学，数据科学和机器学习方面有坚实的基础。一些组织需要深度学习技能，尤其是那些专注于计算机视觉，自然语言处理或语音识别的组织。

建模工作包括：

子任务	例子	涉及技能
训练机器学习模型	- 使用以下方法之一：线性回归，逻辑回归，决策树，随机森林，XGBoost，支持向量机，K均值，K最近邻，神经网络，主成分分析，朴素贝叶斯分类器，Lasso/Ridge回归等	机器学习 算法编程 数学 数据科学
拟合概率模型或统计模型	- 拟合概率图形模型 - 通过数据实验检验假设 - 在数据集上应用降维以促进模型训练或收集见解	数据科学 算法编程 数学
训练深度学习模型	- 将深度学习用于特定领域的应用程序，例如欺诈检测，文本摘要，机器翻译，语音识别或对象分类，检测或细分 - 调整神经网络优化中涉及的超参数	深度学习 算法编程 数学 数据科学
加快训练	- 设置代码以在多台机器上并行训练模型	特定领域的语言（例如CUDA） 算法编程
定义评估指标（通常还涉及数据产品经理）	- 选择F1分数以评估模型在分类任务上的性能 - 实施评估指标，例如准确度，精确度，召回率，并集相交或平均均值精度（mAP）	机器学习 算法编程 数学
加快预测时间	- 应用修剪，量化或压缩等技术来减少内存需求 - 在模型上运行推理速度与准确性实验	机器学习 算法编程
遍历机器学习项目的良性循环：理念，代码，实验	- 将业务问题转换为机器学习问题。例如，根据可访问数据的质量和数量，端到端网络可能比流水线网络产生更好的结果 - 与团队一起实施构思的三个步骤，编码以设置实验，分析结果	机器学习 业务敏锐度
搜索超参数	- 组织实验以在最短的时间内获得结果 - 使用AutoML等工具设置超参数搜索实验	机器学习 算法编程
不断更新知识	- 阅读研究论文 - 观看会议讲座或参加会议	研究 数学 数据科学 机器学习

# I 部署

一些AI团队成员负责将项目提供给用户。他们获取数据流，将其与模型组合，并在将模型投入生产之前对其进行测试。部署人员需要能够编写可进行严格测试的生产代码。他们应具备Python, Java, C++或其他语言的强大后端工程技能，并了解AWS, GCP和Azure等云技术。

部署工作包括：

子任务	例子	涉及技能
将原型代码转换为生产代码	<ul style="list-style-type: none"><li>- 重构存储库的代码</li><li>- 减少重复代码</li><li>- 编写干净的代码以提高可读性和一致性，例如，通过遵循Python中的PEP8准则</li></ul>	软件工程
设置云环境以部署模型	<ul style="list-style-type: none"><li>- 掌握由AWS, GCP, Azure等提供的云工具和基础架构</li><li>- 准备文件（通常是模型架构和参数）以进行部署</li></ul>	软件工程
分支（版本控制）	<ul style="list-style-type: none"><li>- 设计分支工作流程，并使用开发，阶段和生产分支</li><li>- 参与或领导代码审查</li></ul>	软件工程
缩短响应时间并节省带宽	<ul style="list-style-type: none"><li>- 与负责AI基础架构的工程师一起制定负载平衡要求</li></ul>	软件工程
加密存储模型参数，体系结构和数据的文件	<ul style="list-style-type: none"><li>- 全面了解加密并利用现有功能</li></ul>	软件工程
构建供应用程序使用模型的API	<ul style="list-style-type: none"><li>- 设置HTTP RESTful API服务以促进软件组件之间的通信</li><li>- 设置授权和身份验证以访问API</li></ul>	软件工程
再训练机器学习模型（终身学习）	<ul style="list-style-type: none"><li>- 监视数据分布的变化并升级模型</li></ul>	软件工程 机器学习
在受限资源的设备上拟合模型	<ul style="list-style-type: none"><li>- 修剪或量化模型，使其适合内存需求</li><li>- 使用TensorFlow在移动设备上部署模型</li></ul>	软件工程 机器学习

# I 商业分析

一些团队成员评估已部署模型的性能和业务价值。他们建议或做出改变以增加收益或放弃无效模型。

假设部署了一个模型来提供电影推荐。用户首选项和等级将转换为数据。负责业务分析的人员将使用此数据来评估推荐系统的性能以及它为客户创造多少价值。

分配到业务分析的员工需要对数学和数据科学进行分析的理解，以及强大的沟通技巧和业务敏锐度。他们有时使用诸如R，Python和Tableau之类的编程语言，尽管可以在电子表格，PowerPoint或Keynote或A / B测试软件中执行许多任务。

业务分析工作包括：

子任务	例子	涉及技能
建立数据可视化	<ul style="list-style-type: none"><li>- 使用PCA或t-SNE等方法可视化低维的高维数据</li><li>- 建立和呈现使用Tableau，ggplot或matplotlib生成的图形</li><li>- 用JavaScript，HTML或CSS构建可视化</li></ul>	领域特定的编程语言 数据科学 数学 商务敏锐度
建立业务仪表盘情报	<ul style="list-style-type: none"><li>- 编写脚本，定期将数据趋势通知业务主管</li></ul>	特定领域的编程语言 业务敏锐度
向客户或同事介绍技术工作	<ul style="list-style-type: none"><li>- 准备演示文稿（例如PowerPoint）</li><li>- 与团队成员有效沟通</li><li>- 进行技术交流以介绍研究成果</li></ul>	沟通 业务敏锐度
将统计数据转化为可行的业务见解	<ul style="list-style-type: none"><li>- 根据各种来源做出营销决策</li></ul>	数据科学 业务敏锐度
分析数据集	<ul style="list-style-type: none"><li>- 绘制相关矩阵以分析协变量</li><li>- 计算统计变量，例如均值，方差和众数</li><li>- 将客户分成组</li></ul>	数据科学 算法编程 数学
运行 评估部署模型的实验	<ul style="list-style-type: none"><li>- 与部署团队一起评估已部署模型的业务绩效</li><li>- 帮助部署团队做出决策</li><li>- 将模型绩效转化为业务成果，例如收入</li></ul>	数据科学 算法编程
运行A / B测试	<ul style="list-style-type: none"><li>- 优化网页</li><li>- 评估生产中的系统</li></ul>	数据科学 算法编程 商业头脑

# I 人工智能基础设施

在AI基础架构中工作的员工可以构建和维护可靠，快速，安全和可扩展的软件系统，以帮助从事数据工程，建模，部署和业务分析的人员。他们建立了支持项目的基础架构。

继续以电影推荐器为例，人工智能基础设施中的某人将确保该推荐器系统对全球用户都是24/7可用，底层模型被安全地存储，并且可以可靠地跟踪用户与该模型在网站上的交互。

在AI基础架构上工作需要强大而广泛的软件工程技能，以编写生产代码并了解AWS，GCP和Azure等云技术。

人工智能基础设施工作包括：

子任务	例子	涉及技能
制定软件设计决策	-通过定位靠近数据的模型来减少延迟	软件工程
开发 分布式存储和数据库系统	- 建立数据库（SQL，NoSQL，MySQL，Cassandra等）以存储数据 并促进其他团队成员的访问	软件工程 领域特定语言
规模设计	-根据需要添加GPU计算或存储	软件工程
维护软件基础 架构	-通过自动监控和警报管理软件升级并提高稳定性	软件工程
网络	-控制对所有基础架构元素的访问	软件工程
保护数据和模型	- 建立安全功能，允许将生产部署到受监管的组织中，从而满足隐 私和安全性的需求	软件工程
编写测试	-编写跨AI项目生命周期任务的多个组件的单元和功能测试	软件工程
实施 各种软件任务	-建立标签程序，A/B测试框架或分析环境	软件工程

## 第三部分

# AI团队的角色

有时，想要在AI领域工作的人的技能与招聘经理的要求不匹配。为帮助您缩小这一差距，我们将解释AI团队的不同角色，他们的技能以及他们关注的任务。

# 人工智能团队的 六个角色

我们的研究突出了六个技术角色，每个角色都有不同的技能和重点领域。每个角色在AI开发周期中执行许多任务。

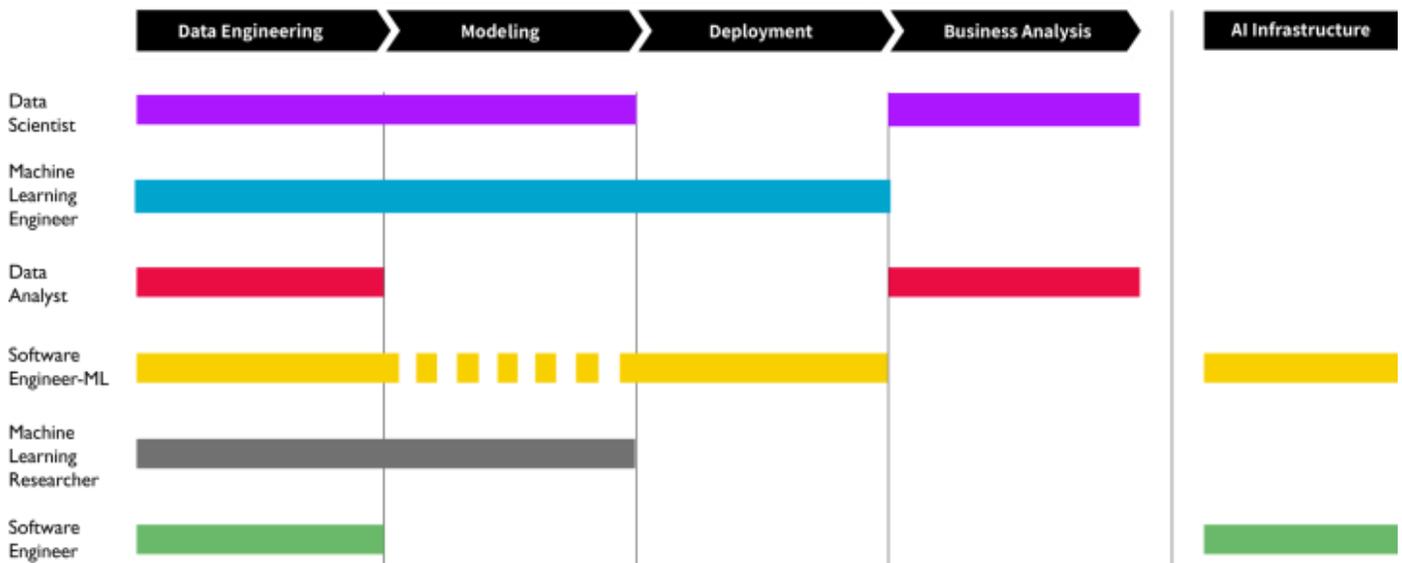
数据是所有这些任务的核心。从事AI项目的每个人都需要知道如何准备数据。如果不是这样，则该项目可能容易受到瓶颈的影响，从而可能减慢AI开发周期。了解如何访问和准备数据将使您对团队及其客户更有价值。

在我们研究的所有公司中，人们在周期内执行多项任务。这可能是非常有效的一种工作方式。

因为一个人可以解决潜在的瓶颈。例如，如果您正在建立电影推荐器，而该推荐器在青少年年龄段效果不佳，一个常见的解决方法是在来自青少年的更多数据上训练您的模型。如果您可以访问和准备数据以及训练模型，则开发周期将加快。

尽管如此，很少有人具备处理整个周期所需的全部技能。

对于每个角色，我们都列出了可能执行的任务以及完成这些任务所需的技能和工具。



# 数据科学家

数据科学家执行数据工程，建模和业务分析任务，如图1所示。他们的技能可以补充那些部署模型和构建软件基础结构的人员的技能。

数据科学家展示了扎实的科学基础以及商业头脑（见图2）。通常需要沟通技巧，但水平取决于团队。

公司可以将此职位称为数据科学家，数据分析师，机器学习工程师，研究科学家，统计学家，定量分析师，全栈数据科学家等职称。

## 任务

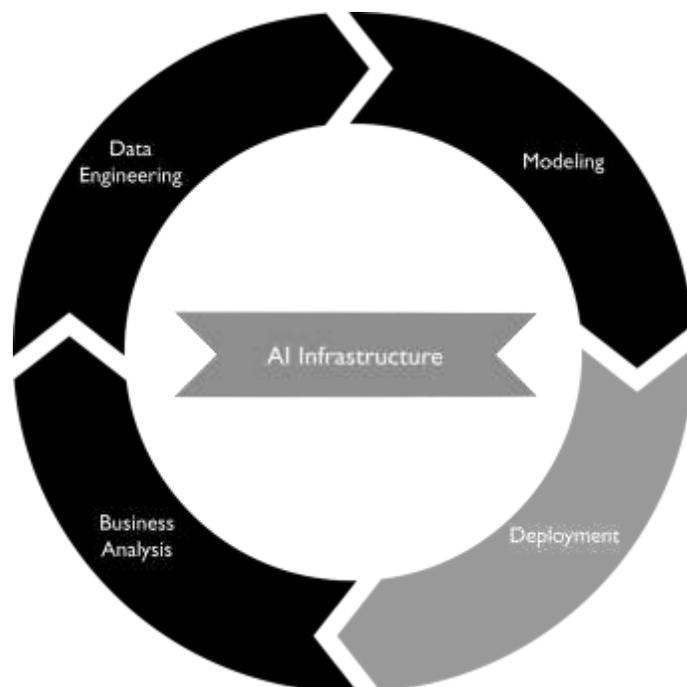


图1。可视化表示AI项目开发生命周期中数据科学家的关注点。

## 技能

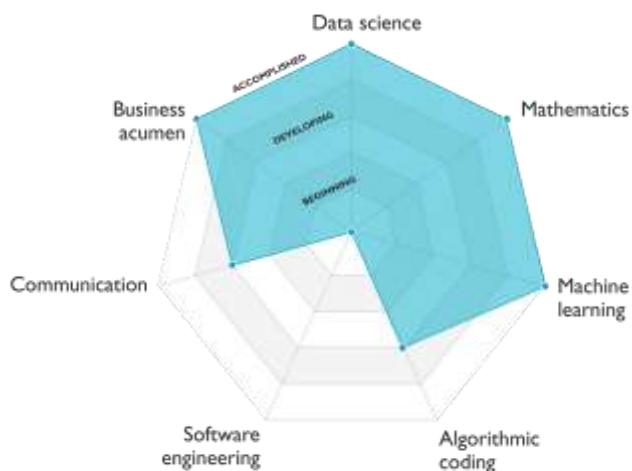


图2。数据科学家的技能和熟练程度的直观表示。

## 工具

- 使用numpy, scikit-learn, TensorFlow和PyTorch等软件包在Python中建模
- 使用Python和/或SQL或其他特定于域的查询语言进行数据工程
- 使用Python, R, 其他特定于域的工具（例如Tableau或Excel）或演示软件应用程序（例如PowerPoint或Keynote）进行业务分析
- 使用版本控制系统（例如Git, Subversion或Mercurial）以及命令行界面（CLI）（例如Unix）和集成开发环境（IDE）（例如Jupyter Notebook或Sublime）进行协作和工作流

# 数据分析师

数据分析师执行数据工程和业务分析任务，如图1所示。他们的技能与培训模型，部署模型和建立软件基础结构的人员相辅相成。

数据分析师展示了扎实的分析能力以及业务敏锐度（参见图2）。它们以查询语言（例如SQL）和常用的电子表格软件工具完成。但是，他们不需要算法编码技能。通常需要沟通技巧，但水平取决于团队。

公司可以将数据分析师的职位称为数据科学家，研究科学家，业务分析师，风险分析师，市场分析师等职称。

## 任务

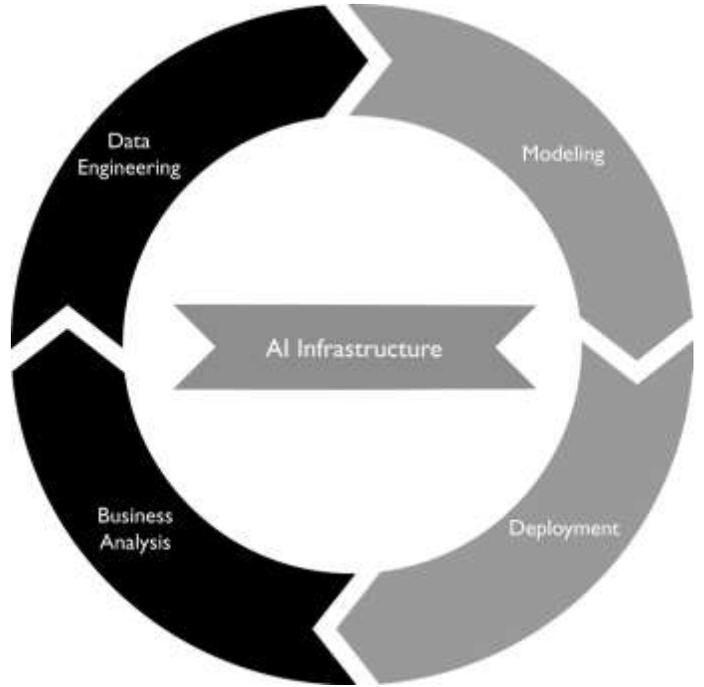


图1。可视化表示AI项目开发生命周期中数据分析师的关注点。

## 技能

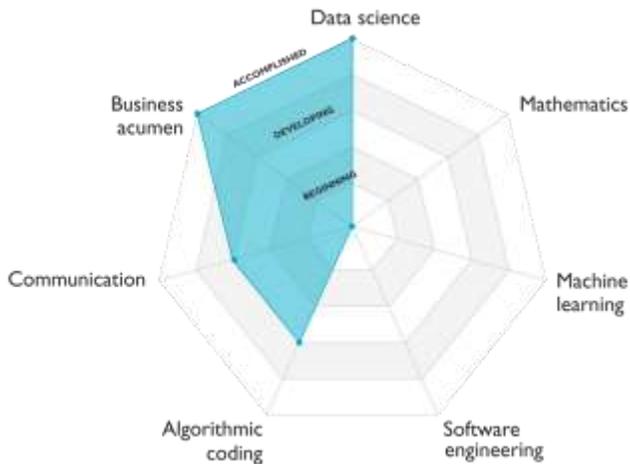


图2。数据分析师的技能和熟练程度的直观表示。

## 工具

- 使用Python和/或SQL或其他特定于域的查询语言进行数据工程
- 使用Python, R, 其他特定于域的工具（如Tableau和Excel），演示软件应用程序（如PowerPoint和Keynote）以及用于A / B测试的外部软件服务进行业务分析

# 机器学习工程师

机器学习工程师执行数据工程，建模和部署任务，如图1所示。他们在能够通过业务分析和AI基础架构为他们提供支持的团队中发挥最大潜力。

机器学习工程师具有扎实的科学和工程技能（见图2）。团队之间的沟通技巧要求各异。

公司可以将此职位称为机器学习工程师，软件工程师 - 机器学习，软件工程师，数据科学家，算法工程师，研究科学家，研究工程师，全栈数据科学家，以及许多其他职位。

机器学习工程师的一种变体称为深度学习工程师。除了图1所示的技能简介外，该角色还需要深度学习知识。它专注于通常由深度学习提供支持的应用程序，例如语音识别，自然语言处理和计算机视觉。因此，它需要特定于深度学习项目的技能，例如理解和使用各种神经网络体系结构，例如完全连接的网络，CNN和RNN。

## 任务

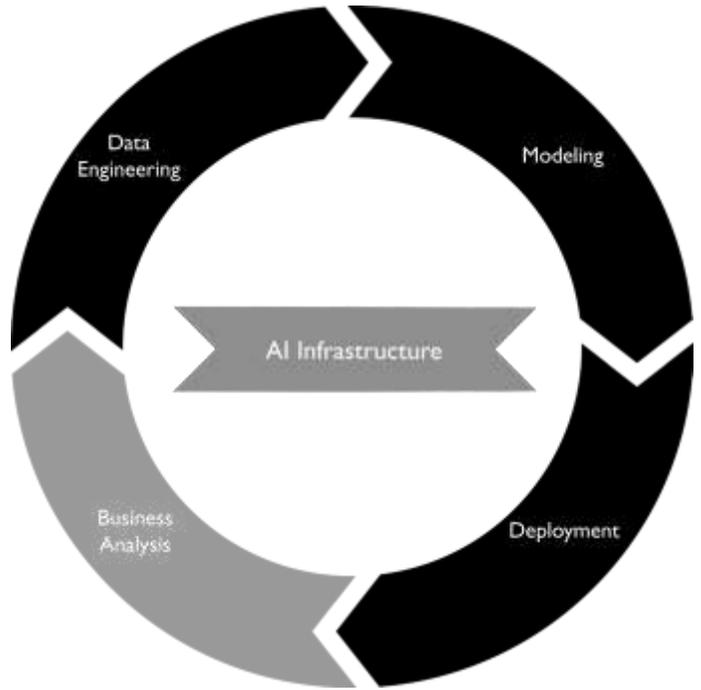


图1。视觉表示AI项目开发生命周期中机器学习工程师的重点。

## 技能

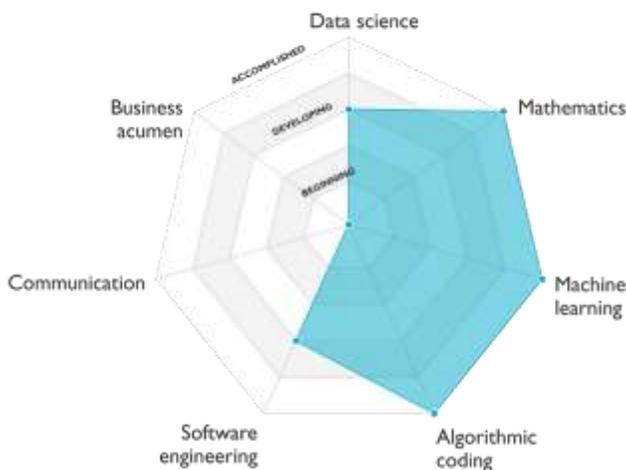


图2。机器学习工程师的技能和熟练程度的可视化表示。

## 工具

- 使用Python和/或SQL或其他特定于域的查询语言进行数据工程
- 使用numpy, scikit-learn, TensorFlow 和 PyTorch 等软件包在Python中建模
- 使用面向对象的编程语言（例如 Python 和 Java, C++）和云技术（例如 AWS, GCP 和 Azure）进行部署
- 使用版本控制系统（例如Git, Subversion和Mercurial），诸如Unix之类的命令行界面（CLI），诸如Jupyter Notebook 和 Sublime 之类的集成开发环境（IDE）以及问题跟踪产品的协作和工作流如JIRA

# 机器学习研究员

机器学习研究人员执行数据工程和建模任务，如图1所示。在负责部署，业务分析和AI基础架构的团队的支持下，他们在研究环境中发挥了最大的潜力。

机器学习研究人员展示了杰出的科学技能（见图2）。团队之间的沟通技巧要求各异。

公司可以将这个职位称为机器学习研究员，研究科学家，研究工程师，数据科学家以及许多其他职位。

机器学习研究人员的一种变体称为深度学习研究人员。除了图1所示的技能简介外，该角色还需要深度学习知识。它专注于通常由深度学习提供支持的应用程序，例如语音识别，自然语言处理和计算机视觉。因此，它需要特定于深度学习项目的技能，例如理解和使用各种神经网络体系结构，例如完全连接的网络，CNN和RNN。

尽管未在图1中显示，但一些机器学习研究人员专注于部署（例如终身学习，模型记忆或边缘部署优化）和AI基础架构（例如分布式培训，计划，实验和资源管理）。

## 任务

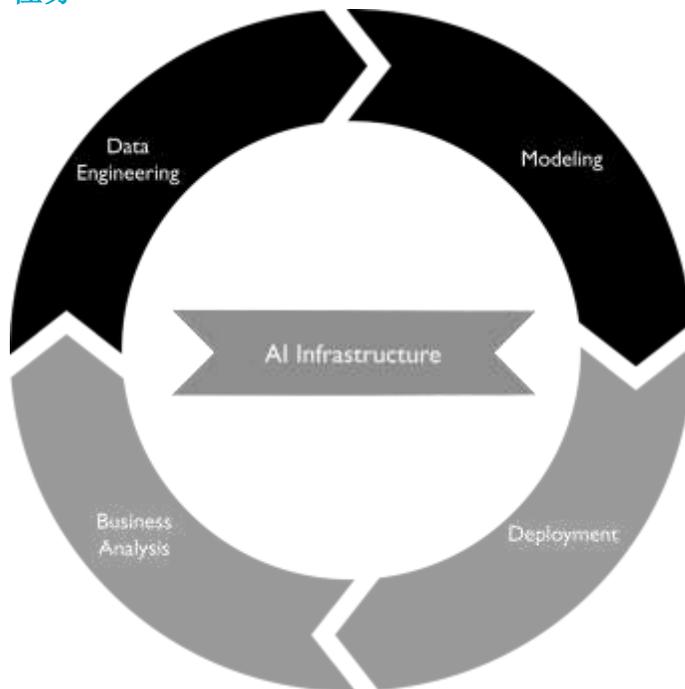


图1。视觉表示AI项目开发生命周期中机器学习研究员的重点。

## 技能

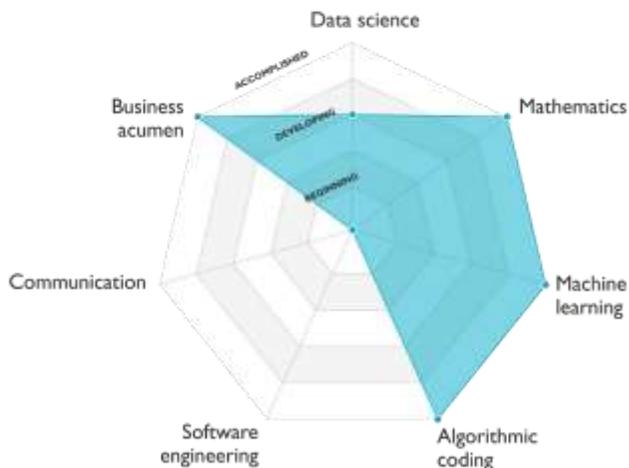


图2。机器学习研究员的技能和熟练程度的直观表示。

## 工具

- 使用Python和/或SQL（或其他特定于域的查询语言）进行数据工程
- 使用numpy, scikit-learn, TensorFlow, PyTorch 等软件包在Python中建模
- 使用版本控制系统（如Git, Subversion和Mercurial），命令行界面（CLI）（如Unix），集成开发环境（IDE）（如Jupyter Notebook 或 Sublime）以及问题跟踪产品（如JIRA）进行协作和工作流
- 通过Twitter, Reddit, Arxiv等渠道的更新以及NeurIPS, ICLR, ICML, CVPR和ACM等会议来跟踪研究

# 软件工程师-机器学习

拥有软件工程师-

机器学习职称的人员可以执行数据工程，建模，部署和AI基础架构任务，如图1所示。他们与负责业务分析和建模的科学家，分析师和研究人员合作良好。软件工程师的机器学习也是早期团队或初创企业的入门角色，旨在部署机器学习模型。如此多才多艺的人非常适合初创企业，因为工程师往往会执行各种任务。

这个单人团队是将软件工程师与数据科学家和/或机器学习工程师结合在一起的团队的替代方案。对于从事更成熟的机器学习项目的团队（例如改进生产中现有的机器学习模型。同样，找人来填补这个职位比找机器学习工程师更容易，并且填补一个职位的成本和雇用一名数据科学家加一名软件工程师的成本也更快。

软件工程师机器学习证明了扎实的工程技能并正在发展科学技能（参见图2）。团队之间的沟通技巧要求各异。

公司可以将这个职位称为机器学习工程师，软件工程师，全栈数据科学家等等。

软件工程师机器学习的一种变体称为深度学习软件工程师。除了图1所示的技能外，该角色还需要深度学习知识。它专注于通常由深度学习提供支持的应用程序，例如语音识别，自然语言处理和计算机视觉。

因此，它需要特定于深度学习项目的技能，例如理解和使用各种神经网络体系结构，例如完全连接的网络，CNN和RNN。

## 任务

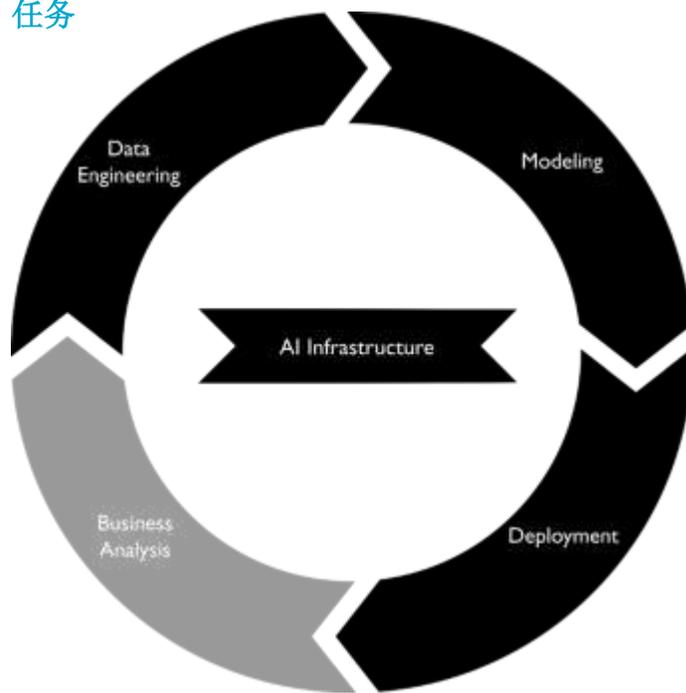


图1。AI项目开发生命周期中软件工程师-机器学习重点的可视化表示。

## 技能

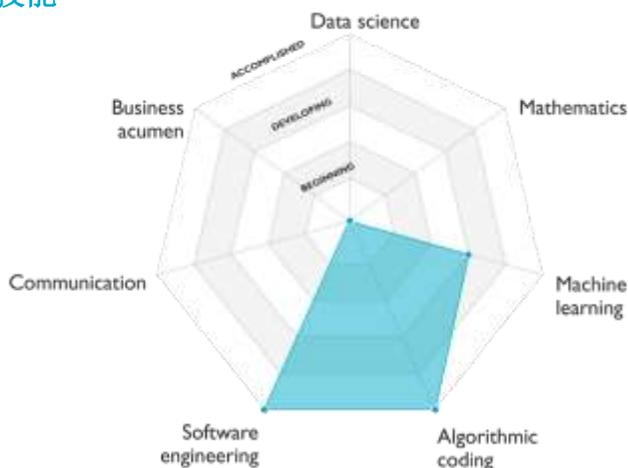


图2。软件工程师机器学习技能和熟练程度的直观表示。

## 工具

- 使用numpy, scikit-learn, TensorFlow 和 PyTorch 等软件包在Python中建模
- 使用Python和/或SQL（或其他特定于域的查询语言）进行数据工程
- 使用诸如 Python, Java 或 C++ 面向对象的编程语言以及AWS, GCP或Azure等云技术部署和设置AI基础架构
- 使用Git, Subversion或Mercurial等版本控制系统, Unix等命令行界面（CLI）, Jupyter Notebook或Sublime等集成开发环境（IDE）以及JIRA等问题跟踪产品进行协作和工作流

# 软件工程师

软件工程师执行数据工程和AI基础设施任务，如图1所示。他们与负责建模，部署和业务分析的人员合作良好。

软件工程师展示出杰出的编码和软件工程技能（见图2）。团队之间的沟通技巧要求各不相同。

公司可以将此职位称为数据工程师，软件工程师，软件开发工程师，AI基础设施软件工程师，软件工程师数据。

## 任务

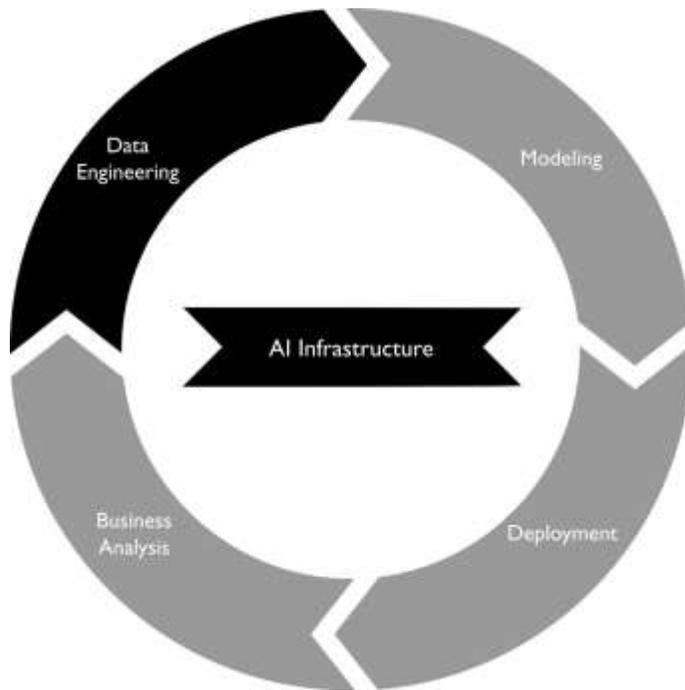


图1。在AI项目开发生命周期中可视化表示软件工程师的重点。

## 技能

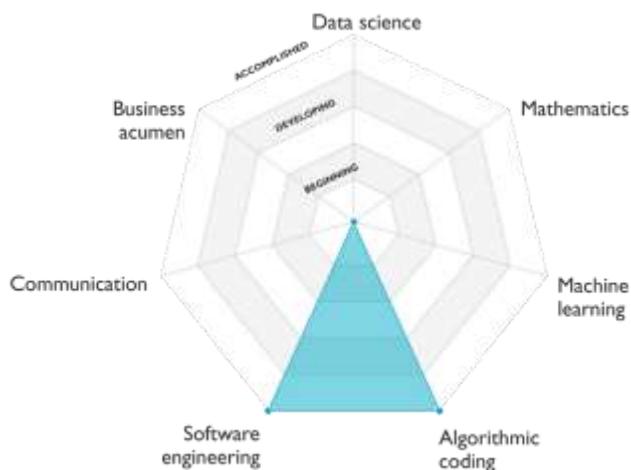


图2。可视化表示软件工程师的技能和熟练程度。

## 工具

- 使用Python和/或SQL（或其他特定于域的查询语言）进行数据工程
- 使用诸如 Python，Java 或 C++ 面向对象的编程语言以及AWS，GCP或Azure等云技术设置AI基础架构
- 使用Git，Subversion或Mercurial等版本控制系统，Unix等命令行界面（CLI），Jupyter Notebook 和 Sublime 等集成开发环境（IDE）以及 JIRA 等问题跟踪产品进行协作和工作流

# 附录：定义技能

这是对报告中提到的技能的简短描述。

## 机器学习

具有机器学习技能的人展示了使用经典机器学习模型（例如PCA，K-means，K-NN，SVM，Logistic回归，线性回归和决策树学习）的能力，训练方法（例如（例如初始化，优化，正则化和超参数调整），以及用于制定机器学习项目战略的技术。

## 深度学习

具有深度学习技能的人展示了使用经典深度学习模型（例如完全连接网络，卷积神经网络，递归神经网络和层）的能力，训练它们的方法（例如初始化，正则化，优化和转移学习）以及制定深度学习项目战略的技术。

## 数据科学

具有数据科学技能的人展示了使用概率（包括分布，条件概率，独立性，贝叶斯定理等），统计（包括假设检验，偏差/方差折衷，均值，方差和模式）和数据分析的能力。

（包括预处理，可视化和指标，例如准确性，R平方，残差，精度和召回率）。

## 数学

具有数学技能的人展示了使用线性代数解决问题的能力（例如，矩阵向量运算，特征值，特征向量和组合），演算（导数，积分等）和数学函数（简单函数，min / max / argmin / argmax等）。

## 算法编程

具有算法编程技能的人展示了理解用代码编写的算法，实现诸如排序和搜索之类的经典算法以及使用诸如树，字典和数组之类的经典数据结构的能力。

## 软件工程

具有软件工程技能的人展示了使用多种计算机科学和软件方法的能力，例如面向对象的编程，Internet协议，HTTP请求，敏捷/ scrum方法，数据库，版本控制（例如Git），容器和单元测试。

# 结论

世界需要工程师和科学家来构建未来。实际上，截至2020年，本报告中定义的所有角色都严重短缺。行动吧！

本报告旨在阐明什么是AI组织，您将从事哪些任务以及现有的职业发展轨迹。它可以帮助世界各地的学习者选择与其技能，背景和志向相匹配的职业道路，我们希望它也对您的学习过程和职业发展有所帮助。

人工智能组织不断发展，因此本报告仍在进行中。我们打算在我们的团队了解有关人工智能人才供求的更多信息后对其进行修订。

我们欢迎您的反馈。请将评论和/或问题发送到 Kian Katanforoosh (kian @ workera.ai)。

-Workera团队



**WORKERA**

一个deeplearning.ai公司