

AI 产品经理入门要求

AI 产品经理需要熟悉 AI 基础知识，包括 AI 行业现状，数学统计学，AI 模型构建和模型基本概念。之所以需要具备这些知识，是因为实现 AI 产品必然会涉及相应的 AI 技术，如果 AI 产品经理不了解相应技术基础，就不能很好的和研发人员沟通，完成 AI 项目的管理。当然，AI 产品经理并不需要像 AI 算法工程师那样，懂很底层的技术细节、数学公式的逻辑推导，但其中涉及的基本概念和行业现状应有所了解。



一. AI 行业现状

首先需要了解 AI 行业现状。AI 的产业架构可以分成**基础技术层**、**算法层**、**应用层**和**解决方案层**。

1) 基础技术层：

硬件设备：包括用于 AI 计算的芯片、服务器和设备。例如云计算、GPU 等。

基础软件：包括用于数据处理、模型训练和部署的开发工具和框架，例如 Tensorflow、Pytorch 等框架。

2) **算法层：**包括机器学习、深度学习、增强学习等算法。

3) **应用层：**主要是垂直领域应用，包括人工智能在医疗、金融、零售、交通等行业的具体应用，例如金融风控识别系统、智能客服系统等。

4) **解决方案层：**包括 AI 技术在智能制造、智慧城市、智慧医疗等场景的解决方案，主要关注如何将 AI 技术与特定行业的业务需求相结合，提供定制化的解决方案，以推动该行业的数字化转型和智能化发展。

其中基础技术层主要负责人员是软件开发，算法层负责人是算法工程师，而应用层和解决方案层是 AI 产品经理主要工作方向。

AI 应用层和解决方案层中，涉及到的最新技术就是大模型，目前也是各科技企业竞相追逐的风口。CHATGPT、Gemini、Sora、文心一言、通义千问等大模型产品层出不穷，基于大模型的 AI 程序员 Delvin 已经能实现独立开发，最新的 GPT4o 已具有实时视频和语音功能；掌握大模型工具、紧跟 AI 应用前沿能帮助 AI 产品经理在竞争中脱颖而出。



二. 数学统计学基本概念

数学统计学是人工智能的基础，AI 产品经理应了解并掌握，包括线性代数、概率论和统计学的基本概念。

1) 线性代数

线性代数是人工智能和机器学习中的基础数学概念，涉及向量、矩阵、线性方程组等内容。

需要理解常量、向量、矩阵、张量的概念。

常量 (Scalar)：常量是一个单独的数值，比如一个用户的年龄数据。

向量 (Vector)：向量是一个有序的数值集合，具有大小和方向。比如多个用户的年龄数据集合。

矩阵 (Matrix)：矩阵是一个二维的数值集合，由行和列组成。矩阵可以看作是向量的推广，其中每个元素都有一个行索引和列索引。

在机器学习中，矩阵常用于表示数据集或模型的参数，例如多个用户的年龄和收入数据、灰度图像的像素值均为 2 维矩阵。

张量 (Tensor)：张量是多维的数值集合，张量可以有任意数量的维度。在深度学习和神经网络中，张量是数据在神经网络中传播和处理的基本单位。常量是 0 阶张量，向量是 1 阶张量，矩阵是 2 阶张量，而彩色图片因为有 RGB 三通道，是 3 阶张量。

2) 概率统计

需要重点掌握随机变量和概率分布，了解业务场景下的特征数据和模型结果概率分布情况，有助于产品经理对 AI 模型的验收（例如已知身高是正态分布，但模型输出的身高预测结果却不是正态分布的，则需要质疑模型效果）。

随机变量 (Random Variable)：

随机变量是描述随机现象结果的数学变量。它可以取多个值，分为离散和连续随机变量两类。

离散随机变量：只能取有限个或可数无限个值的随机变量，如抛硬币的结果（正面或反面）。

连续随机变量：可以取任意实数值的随机变量，如身高、体重等。

概率分布 (Probability Distribution)：概率分布描述了随机变量可能取值的概率分布情况，分成离散和连续概率分布两类。

离散概率分布主要有：

二项分布：描述了在一系列独立重复的是/非试验中成功的次数的概率分布。

泊松分布：用于描述单位时间或空间内随机事件发生次数的概率分布。

超几何分布：描述了从有限总体中抽取不放回样本的概率分布。

贝努力分布：描述了只有两种可能结果的单次随机试验的概率分布。

多项式分布：描述了多项试验中每个类别出现次数的概率分布。

连续概率分布主要有：

正态分布：也称为高斯分布，是最常见的连续概率分布，具有钟形曲线。

指数分布：描述了独立随机事件发生时间间隔的概率分布。

均匀分布：所有数值在一个区间内具有相同的概率密度的分布。

t 分布：用于小样本情况下对总体均值的推断。

三. AI 模型构建和模型基本概念

还应熟悉 **AI 模型构建流程**和**模型基本概念**，有助于更好的和研发协作，管理整个 AI 项目的研发周期。

1) AI 模型构建：主要包括模型设计、数据准备和特征选择、模型训练、模型验证 4 步，最后模型才会作为产品交付。

其中，在模型设计阶段，产品经理需要明确当前的场景适用的算法有哪些、每种算法适合解决什么问题；在数据准备阶段，产品经理需结合业务判断什么数据更具有代表性，提供更高质量的数据；在模型验证阶段，需要评估模型是否达到了上线的标准。

2) 模型基本概念：

监督学习 (Supervised Learning): 指训练数据有标注，按照预测的结果类型，可以进一步分成分类（预测类别）和回归（预测数值）

无监督学习 (Unsupervised Learning): 指的训练数据中没有标注。主要用于聚类。

半监督学习 (Semi-supervised Learning): 介于监督学习和无监督学习之间，训练数据只有部分有标注，模型需要利用有标注和无标注的数据一起训练。

训练集、测试集、泛化能力: 用于给模型学习规律的数据称为“**训练集**”，用于检验机器学习效果的数据称为“**测试集**”。模型在未知数据上或测试集上的表现称为“**泛化能力**”。

欠拟合和过拟合: 如果模型在所有数据上表现效果不好，预测结果和实际结果偏离程度大（高偏差），则是欠拟合；如果模型只在一部分数据上表现效果好，在其他数据上表现差，模型输出结果不稳定（高方差），则是过拟合。

模型表现的衡量指标:

对于分类模型，有准确率 (Accuracy)、精确率 (Precision)、召回率 (Recall)、F1 分数 (F1 Score)、ROC 曲线和 AUC 值等。

对于回归模型，有均方误差 (Mean Squared Error, MSE)、平均绝对误差 (Mean Absolute Error, MAE)、对数损失 (Log Loss) 等。

3) 常见的算法:

监督学习: 逻辑回归、支持向量机、随机森林、神经网络等

无监督学习：K 均值聚类、主成分分析、层次聚类等

半监督学习：自训练、半监督支持向量机、标签传播、生成式半监督学习等方法。

对于上述的概念和名词，需要有印象，遇到了具体业务场景再去详细了解。