AI 产品经理必备基础技术知识

对于 AI 产品经理或相关领域入门者、从业者，可以把本文看作需要阅读的入门文章，后续再根据自己情况进一步 dig 研究。

一、AI 通用技术划分

AI 的通用技术包括语音识别（ASR）、自然语言处理（NLP）、语音合成（TTS）、计算机视觉（CV）、知识图谱（KG）、即时定位与地图构建（SLAM）等。下面将介绍 AI 产品经理需要知道的几个主要的 AI 技术。

二、语音识别（ASR）

语音识别（speech recognition）技术，也被称为自动语音识别（英语：Automatic Speech Recognition，ASR）、电脑语音识别（英语：Computer Speech Recognition）或是语音转文本识别（英语：

Speech To Text,STT），其目标是以电脑自动将人类的语音内容转换为相应的文字。与说话人识别及说话人确认不同，后者尝试识别或确认发出语音的说话人而非其中所包含的词汇内容。

目前，主流的大词汇量语音识别系统多采用统计模式识别技术。典型的基于统计模式识别方法的语音识别系统由以下几个基本模块所构成：

信号处理及特征提取模块。该模块的主要任务是从输入信号中提取特征，供声学模型处理。同时，它一般也包括了一些信号处理技术，以尽可能降低环境噪声、信道、说话人等因素对特征造成的影响。

声学模型。典型系统多采用基于一阶隐马尔科夫模型进行建模。

发音词典。发音词典包含系统所能处理的词汇集及其发音。发音词典实际提供了声学模型建模单元与语言模型建模单元间的映射。

语言模型。语言模型对系统所针对的语言进行建模。理论上，包括正则语言，上下文无关文法在内的各种语言模型都可以作为语言模型，但目前各种系统普遍采用的还是基于统计的 N 元文法及其变体。

解码器。解码器是语音识别系统的核心之一，其任务是对输入的信号，根据声学、语言模型及词典，寻找能够以最大概率输出该信号的词串。

但在目前的 AI 语音产品中存在着一些语音识别的难点与瓶颈，比如，说话场景中主要发声源要靠近机器，发音要尽量标准以识别更准确，环境不能过于嘈杂，持续对话不能打断等等。因此，需要 AI

产品经理结合更加合理的麦克风阵列结构、更好的模型建模解决对话时的混响、噪声及回声环境问题。

三、自然语言处理（NLP）

自然语言处理(英语:Natural Language Processing,缩写作 NLP）是人工智能和语言学领域的分支学科。此领域探讨如何处理及运用自然语言；自然语言处理包括多方面和步骤，基本有认知、理解、生成等部分。

自然语言认知和理解是让电脑把输入的语言变成有意思的符号和关系，然后根据目的再处理。自然语言生成系统则是把计算机数据转化为自然语言。

当前 NLP 研究的难点有一下几点：

1.单词的边界界定

在口语中，词与词之间通常是连贯的，而界定字词边界通常使用的办法是取用能让给定的上下文最为通顺且在文法上无误的一种最佳组合。在书写上，汉语也没有词与词之间的边界。

2.词义的消歧

许多字词不单只有一个意思,因而我们必须选出使句意最为通顺的解释。

3.句法的模糊性

自然语言的文法通常是模棱两可的,针对一个句子通常可能会剖析（Parse）出多棵剖析树（Parse Tree），而我们必须要仰赖语义及前后文的信息才能在其中选择一棵最为适合的剖析树。

4.有瑕疵的或不规范的输入

例如语音处理时遇到外国口音或地方口音，或者在文本的处理中处理拼写，语法或者光学字符识别（OCR）的错误。

5.语言行为与计划

句子常常并不只是字面上的意思；例如，“你能把盐递过来吗”，一个好的回答应当是动手把盐递过去；在大多数上下文环境中，“能”将是糟糕的回答，虽说回答“不”或者“太远了我拿不到”也是可以接受的。再者，如果一门课程去年没开设，对于提问“这门课程去年有多少学生没通过？”回答“去年没开这门课”要比回答“没人没通过”好。

四、语音合成（TTS）

语音合成是将人类语音用人工的方式所产生。若是将电脑系统用在语音合成上，则称为语音合成器，而语音合成器可以用软/硬件所

实现。文字转语音（Text-To-Speech，TTS）系统则是将一般语言的文字转换为语音，其他的系统可以描绘语言符号的表示方式，就像音标转换至语音一样。

合成器的技术目前有串接合成、共振峰合成、发音合成、HMM 基础合成、正弦波合成、深度学习合成。目前来说，通用 TTS 基本满足商业化需求，但缺乏人声自然度，无法满足用户高体验预期。

五、计算机视觉（CV）

计算机视觉（Computer vision）是一门研究如何使机器“看”的科学，更进一步的说，就是指用摄影机和计算机代替人眼对目标进行识别、跟踪和测量等机器视觉，并进一步做图像处理，用计算机处理成为更适合人眼观察或传送给仪器检测的图像。

计算机视觉的主要研究内容包括：

识别--识别的几个具体应用方向：

基于内容的图像提取：在巨大的图像集合中寻找包含指定内容的所有图片。被指定的内容可以是多种形式，比如一个红色的大致是圆形的图案，或者一辆自行车。在这里对后一种内容的寻找显然要比前一种更复杂，因为前一种描述的是一个低级直观的视觉特征，而后者则涉及一个抽象概念（也可以说是高级的视觉特征），即‘自行车’，显然的一点就是自行车的外观并不是固定的。

姿态评估--对某一物体相对于摄像机的位置或者方向的评估。例如：对机器臂姿态和位置的评估。光学字符识别对图像中的印刷或手写文字进行识别鉴别，通常的输出是将之转化成易于编辑的文档形式。

运动--基于序列图像的对物体运动的监测包含多种类型，诸如：自体运动：监测摄像机的三维刚性运动。图像跟踪：跟踪运动的物体。

场景重建--给定一个场景的二或多幅图像或者一段录像，场景重建寻求为该场景创建一个三维模型。最简单的情况便是生成一组三维空间中的点。更复杂的情况下会创建起完整的三维表面模型。

图像恢复--图像恢复的目标在于移除图像中的噪声，例如仪器噪声、动态模糊等。

六、知识图谱（KG）

知识图谱（Knowledge Graph），是结构化的语义知识库，用于以符号形式描述物理世界中的概念及其相互关系。其基本组成单位是“实体-关系-实体”三元组，以及实体及其相关属性-值对，实体间通过关系相互联结，构成网状的知识结构。知识图谱可以实现 Web 从网页链接向概念链接转变，支持用户按主题而不是字符串检索，真正实现语义检索。基于知识图谱的搜索引擎，能够以图形方式向用户反馈结构化的知识，用户不必浏览大量网页即能准确定位和深度获取知识。

知识图谱可以应用在哪些方面呢？智能搜索，对查询分词之后，对查询的描述进行归一化，从而能够与知识库进行匹配。查询的返回结果，是搜索引擎在知识库中检索相应的实体之后，给出的完整知识体系。；深度问答，多数问答系统更倾向于将给定的问题分解为多个小的问题，然后逐一去知识库中抽取匹配的答案，并自动检测其在时间与空间上的吻合度等，最后将答案进行合并，以直观的方式展现给

用户。社交网络，Facebook 于 2013 年推出了 Graph Search 产品，其核心技术就是通过知识图谱将人、地点、事情等联系在一起，知识图谱会帮助用户在庞大的社交网络中找到与自己最具相关性的人、照片、地点和兴趣等。垂直行业应用，对于特定行业对整合性和关联性的资源需求迫切，知识图谱可以为其提供更加精确规范的行业数据以及丰富的表达，帮助用户更加便捷地获取行业知识。

七、即时定位与地图构建（SLAM）

即时定位与地图构建（英语：Simultaneous localization and mapping，一般直接称 SLAM）是一种概念：希望机器人从未知环境的未知地点出发，在运动过程中通过重复观测到的地图特征（比如，墙角，柱子等）定位自身位置和姿态，再根据自身位置增量式的构建地图，从而达到同时定位和地图构建的目的。

地图构建，SLAM 的地图构建通常指的是建立与环境几何一致的地图。传感，SLAM 研究中经常使用许多不同型号的传感器来获得地图数据。这些数据带有统计独立的误差。这个统计独立是解决度量偏差和检测中的噪声的强制需求。定位，传感器的结果会作为定位算法的输入。建模，以上结果对地图构建的贡献，可以在“2D 建模并分别表示”或者在“3D 建模并在 2D 上投影表示”中工作得一样出色。地图构建就是这样一个动态模型的最终运算结果。

结语

在笔者所希望深入研究的建筑机器人领域中，SLAM 技术和机器视觉技术我认为成为了当前阻碍产品有效落地和大量商业化应用的两大主要问题技术痛点。在建筑施工过程中，不论是材料零件的加工切割，还是在工地上进行现场施工，甚至到施工完成后，进入室内装修阶段。不同的场景面临着不同且及其复杂的场地情况，这就需要相应的建筑机器人能够在有限的空间和复杂的场地内进行高效率的施工作业。这对这种机器人的 SLAM 技术和 CV 技术提出了非常严苛的考验。