

想入行 AI，怎么选择有效的技术培训

引子

IT 行业发展迅速，各种新名词此起彼伏。身处这样一个热点行业，**学习是必须的**。不打算成为终身学习者的程序员，失业就在明天。

可是，**怎么学呢？**

都已经毕业了，每天要上班，不能像以前读书的时候，整天只是学习，学什么都有老师教，坐在那儿听就可以了。

自己从头看书太辛苦了，网上的文章又太碎片化——是不是报一个培训班，交点钱听人讲更容易学会？

培训班有那么多，这个是证书，那个是优惠，再一个给提供工作机会，到底选哪个好呢？

今天我们就来说说这件事。

NOTE：我们此处说得培训班指需要较长时间（一般以月为计）的付费培训。线上线下都包括，但一定能够和培训教师直接交流，有答疑过程。

自学的困境

如今，网络资源那么发达，IT 类技术都不难找到各式各样免费或费用很低的书籍、资料、课件、讲座、代码。从理论到实践都有讲解。

好在读本文的同学，已经共同拥有了一个优势：目标明确。

我们学习的目的是为了入行AI！

在明确目标的指引下，比较容易找到什么“**有用**”——虽然有点功利，但不得不承认，“有用”与否是大多数人产生驱动力的重要所在。

我们可以以“在AI领域找到技术岗位工作”为基准，进行学习。避免东一下，西一下的“乱学”。

什么样的目标才算明确

但是要注意，这个目标如果仅仅停留在：“我要做AI”，是不能称之为一个目标的。**真正的目标必须具有可实施性，并最终体现为实施计划。**

想要入行AI，先搞清楚几件事：

- 目前真正落地的AI领域都有哪些？
- 每个领域有什么样的代表企业？
- 这些企业中，都有什么样的技术岗位？
- 每种岗位需要什么入门条件？

至少要能够定位到你的目标岗位，明确了这个/类岗位的技术要求，才有可能确定需要学习的内容有哪些，从而选择到正确的培训课程。

如何划定明确的目标

但是怎么能知道这些呢？笔者个人推荐如下调研步骤：

Step 1. 先确定一个领域（图像、语音、NLP等等）

可能你本来就有特别感兴趣的领域；或者你对某一种技术，比如人脸识别，特别兴趣，那么可以直接由此入手。

真的能认真读进去10篇比较新的论文，哪怕是普通博士生发表的，也能让你对一个学术领域有最基本的理解了。

Step 3. 了解本领域理论的落地技术以及相关企业

AI作为一个新兴方向，很多领域还处在研究阶段，真正能够应用到现实产品中的领域相当有限。

有代表性不过就是：语音识别/合成，图片/人脸识别，和NLP的一些分散应用。

当然并不是说尚未投入使用的技术就不值得去了解或者投身，比如现今热点中的热点——自动驾驶——尚处于研究性质远超实用的探索阶段，虽然很多公司都在做，但其实并没有实际的投入真实世界使用。

此处只是说，落地技术的范围并不算太广，了解起来投入也有限。

有了目标技术再找企业就相对容易多了。虽然大公司掌控了当前AI领域的绝大多数人才和资源，但是也有越来越多的小企业在具体技术点上发力。

普遍来说，进大公司是为了公司，而进小公司则是为了跟人。而AI行业又是一个强学术背景的行业，一个公司也好，团队也罢，如果连一个有一些最起码学术建树的博士都没有，那能走多远真的不好说。

从这一点来看，step 2的调研过程也可以应用到此处。如果有感兴趣的小公司，尤其是刚刚创业不久的startup，不妨先评估一下技术合伙人的学术水平。

Step 4. 了解具体岗位的招聘需求

这里的具体岗位，到并不一定指xxx公司的xxxx岗位，而是指同一类型公司同一技术角色的相对普遍要求。

AI行业的技术岗位，按角色可以简单地分为三类：

角色1：科学家——研究理论，开发/改进算法；

角色2：工程师——结合业务，训练模型；

角色3：产品经理——负责产品需求，协调资源，推动落地；

领域、企业和角色共同定义了岗位之后，再根据岗位需求来反推需要学习的内容，就是有的放矢了。

容易被误导的“捷径”

虽然推荐上述路径，但是笔者确实知道，很多人喜欢走“捷径”——去招聘网站用AI、人工智能等关键词搜索一堆职位，看看那些职位要求的工具和语言是什么，直接去学就好了。相当于从step1直接跳到了最后。

反正现在大多数职位都要求Python，Tensorflow，直接报个班学学怎么用Python调用现成的算法，或者怎么用tensorflow处理数据不就好了？何必那么麻烦，还要看什么论文，学什么理论。

这种想法，属于典型的被“**捷径**”**误导**。在AI行业从事技术工作，哪怕是做角色3的工程辅助工作，如果想要做得长久，有所发展，理论学习是必不可少的。

要详细解释这一点，完全可以单独开个chat了。此处且举个直观的例子：

工具就像是武器，学会使用一种工具只是学会了使用这种武器的最基本的招式和套路。而理论学习则是学习策略，决定了未来在真实对战中，遇到对手攻击时，你选取哪些招式套路，如何组合起来去迎敌。

不排除现在有些公司跟风慕名，想做AI，自己没有人才，就直接招聘，要求会用XX工具就可以了。只学会用工具做一些基本操作，也许就可以应聘这样的职位。但是这样的职位能长久吗？能解决真正的问题，产生价值吗？做这样的工作，能有长进提高个人价值吗？

为了个人长远的职业发展，我们还是扎实打牢基础。

制定学习计划

有了明确的目标，也就有了明确的范畴（scope）。同时，经过上节step2-4，还搞明白

3. 填充知识模块；
4. 列举针对具体模块的主要知识点。

可以借鉴做作文列提纲的办法：在勾勒出轮廓之后，先把知识结构的骨干勾勒出来，分为篇章，列出大标题，再在其中填注小标题（知识点）。

这里很**关键**的一点是，计划中最细粒度的“小标题”到底有多大。

个人意见：学习一个计划中的最小单位，如果是不脱产，仅在业余时间学习，掌握它的时间不应该超过一周；如果是全脱产学习，最好不要超过2天。

我们下面用一个例子来说明一下具体从目标到学习计划的制定过程。

【举例说明】

从需求出发：笔者要学习基于机器学习的自然语言处理，具体的应用是开发聊天机器人的语言理解模块。

进行调研：通过 i) 向有类似经验的同事请教；ii) 到网上搜索综述性文章和NLP领域论文；iii) 查找实践类的文章、类似开源项目……等一系列手段，笔者了解到，最起码有两件事情必须要做：意图分类和实体提取。

目前，要做这两件事情，有基于规则和基于机器学习/深度学习模型两类方法。

基于规则虽然直接、初始代价小，但是可扩展性差，所有规则都需要人工添加。

而基于模型的方法有较强的可扩展性，而且随着聊天机器人用户和语料的增多，还可以通过反馈持续增强模型。在聊天机器人中应用机器学习/深度学习不仅符合产品发展的需求，也是当前业界的发展方向。

机器学习和深度学习的区别在于：前者适合相对数据量、运算资源较小，而开发者对业务理解较深的场景；后者则更加“自动化”，但对于数据量和运算量需求巨大，尤其是对人工标注的数据要求很高，先期投入太大。

笔者决定在工作应用机器学习模型，因此就要先从机器学习学起。

构建体系：通过对比多本机器学习著作，不难发现，机器学习理论是以一个个模型为主要内容的。

之前工作中要用到的意图识别和实体提取，正好对应分类和seq2seq预测模型。而为了获取意图，可能还需要对原始语料做文本聚类。

因此我们构建的体系可以以模型为主要节点。

填注内容：确定了主干节点，下一步就是填注课程的“血肉”。

归根到底，模型是用来解决问题的。比如chat bot语言理解所用到的分类，序列预测，聚类等，都是典型的机器学习问题，每一个问题都对应多个模型。每一个模型都有其适用的具体情形。

模型的获得包括算法和数据两个方面。需要一个训练过程，训练过程一般迭代进行，期间要做多次验证，根据验证结果调优，最终通过测试来检验模型质量。

再通过进一步查询资料得知，要了解这些模型的运行原理，就得读公式，那么就要求导、求微分、求积分、矩阵运算、概率统计等方面的知识。

还有就是，所有模型处理的都是数值，我们要把现实当中的文本等人类可读信息都转化成向量。这就需要掌握将文本转化为向量空间模型的能力，和数据清洗整理的能力。这些能力又都是以编码能力为支撑的。

制定提纲：由此，我们学习计划体系就已经有了眉目。

我们需要学习（复习）下列这些知识：

- 1) 高数、线代和概率论的知识；
- 2) 选择一批在实践中应用较多的典型模型，学习其原理、数学推导过程和训练模型的算法；
- 3) 模型的训练、验证、测试过程和评判模型的指标；
- 4) 将人类可读信息转化为数值的方法；
- 5) 训练、测试模型的工具、框架和编程语言。

- c. Naive Bayes
- d. Decision Tree(ID3, C4.5)
- e. SVM
- f. CRF
- g. KMeans
- h. Spectral Clustering
- i. LDA

C. 模型的构建和验证

- a. Binary-Classification vs Multi-Classification
- b. Normalization & Regularization
- c. Validation & Test
- d. Cross Validation Methods
- e. Precision, Recall, F1Score f. ROC, AUC

D. 数据处理和向量模型空间的构建

- a. 文本标注
- b. Bagging & Boosting
- c. 中文分词方法及原理
- d. n-gram模型原理
- e. bi-gram文本特征提取
- f. 计算文档tf-idf 及其信息熵

E. 工具及语言

- a. Python,Java,C#类比及对比
- b. Python 库 , Java 库 , C# 库
- c. 分词工具 (jieba分词) 和词库
- d. word2vec

F. 实践

如上只是第一个版本，可以先依据它制定一个为期两到三个月的学习计划。在执行过程中，根据新的认识和具体需要可随时调整提纲和计划。

看看星级，还有优惠券拿，直接选一个不就行啦。何必还要自己定制什么学习计划，多此一举。

此处笔者想说的是：**选择培训课程容易，选择有效的培训课程可就难了。**

如果选择了错的培训课程，浪费钱还是小事，浪费自己的时间才可惜。

尤其是像AI这种热点、风口，如果因为总也学不进去东西，错过了发展最迅速阶段的黄金入门期，可能错过的是大好的职场机遇！

培训的用处

AI行业的知识、技术，包括各种工程实践，best practice，都是纯粹的书面知识。

不像乐器、舞蹈之类，文字不足以描述，需要老师面对面指导和反复多角度演示才能掌握到位的技能；书面知识，肯定是可以自学的。

对于这样的知识，培训课程的作用何在？

上培训课的**原因**很简单：为了省事——自己学太累了，不如听老师讲。用耳朵听，比用眼睛看书轻松，而且老师讲得，总比书上写的丰富细致吧。

用自己的金钱和时间换取他人的知识和经验当然没问题，甚至可能是高效的办法。但如果觉得，只要自己花了钱，就可以坐在那里，等着老师把知识塞到自己脑子里，可就大错特错了。

学习的过程，必须学习者主动吸收理解，才能达到掌握知识的目的。

而培训课程的作用，无外乎两点：

1. 讲解细致

相对于力求清晰、简洁、逻辑性强，具有“攻击性”的书面资料，合格的课堂讲解应该是丰富、细腻、包容性强，具有“防守性”的。

学习的过程中，疑问是难免的，如果疑问解决不了，则学习很难继续下去。

无论文字还是影音资料，都是只有传达过程，没有针对每个读者个人的答疑解惑。培训课程恰恰可以弥补这个不足。

让培训有效

可能有些同学上培训班的目的之一是直接获得系统性知识的传达，因此，也就觉得不必做之前说的，自己构建知识体系，定制学习计划的事情了。

自己就算定了个计划，也不太可能找到完全一致的培训班。随便报个班，老师自然有教学大纲，听着就是了，为什么要自己费力气？

此处需要提示这些同学一点：**如果你自己都不知道想要学习什么东西，又怎么能够正确判断一门培训课程的质量呢？**

毕竟，**培训（不止是培训，其实所有教学都是如此）要有效，一定是教师和学生双方配合的结果。**

教学双方能够配合好至少有三个前提：

- 1) 教的人有足够的个人积累和授课技巧用以传达知识；
- 2) 学的人主动吸收传授内容并同步思考、理解；
- 3) 教的人教的内容与学的人想学的内容相匹配，且张力适度。

学生在一无所知的情况下，看广告报个班，对于学习本身缺乏动力和认识，自己都不知道想从课程中获得什么；或者教的内容与学生的背景相差太远，学生一句都听不懂——这种情况下，即使有幸碰到了好的老师和课程，能吸收的恐怕也非常有限。

通过制定学习计划，学生不仅能够明确自己的需求，具体到每一节课想学什么，而且能够了解到这些授课内容之间的联系和优先级排序。

拿着自己的学习计划去找培训，能在大概率上保证培训课程的有效性。因为：

1、制定计划的过程，是一个带着问题探索的过程，带着问题，但高目标生驱动上

对于课程深度的了解，还可以借助事先咨询授课教师或者培训学校来完成，对自己的了解则无人可以替代。

搜寻有效培训的途径

现在我们已经有了了一份自己定制的详细学习计划，而且也明白了按图索骥的重要性。下一步就要寻找具体的培训课程了。

对找寻过程，笔者的建议是：**先从免费资源开始。**

1) 首先，AI领域有少量非常经典的现在培训课程，是完全免费的，而且非常容易找到。

如果你和前面例子一样，已经选定了机器学习作为学习目标，Andrew Ng的Machine Learning是一定要从头到尾学一遍的。每节课后的习题都比较简单，要逼着自己做完。

2) 其次，通过书籍、文章、论坛、微信/QQ群、免费讲座和分享等渠道，去搜集大家的学习经验和各种培训班的信息。

和有同样目标的人交流，经常能在不经意间得到优质推荐。

3) 然后，在尽量广泛地收集了大量培训信息后，对照课程大纲和自己的学习计划进行选择。

教师声誉、培训机构声誉、网友评价等可以作为参考和背书，但**核心是课程大纲与个人计划的匹配度！**

4) 再次，按优先级裁剪个人计划。

当然，想找到和最初版本计划100%匹配的培训课程基本是不可能的（除非你的计划就是照着某个培训课抄的）。

这就不能不涉及到对于个人学习计划的裁剪。裁剪的原则应该是：**确保核心知识和大体结构，在细节上做出妥协。**

比方说 你自我判断数学且必须要补习的部分 那么就不要找空余一半数学基础都不

不仅内容大多是笔者想听的，而且老师的风格与笔者颇为合拍——做计划做得类似的人，想必有不少共同的特质，交流起来也比较顺畅。这就是自己定制学习计划更深层次的好处。

为了避免软文之嫌，这个课程具体为何在此就不说了。不过，对别人合适的课程未必对自己合适。这门课我曾经推荐给几位有兴趣的朋友，有人真的买了课，但是却听不下去，浪费了钱。原因就是教学不相匹配。

本文的目的，并不是推荐具体的课程或者讲师，而是希望大家掌握去寻找适合自己的有效培训课程的方法。

有效学习的几点提示

分享几个笔者认为行之有效的日常学习Tips。

关联

把新学到的东西和日常的工作、生活联系起来，进行思考。比如：将理论代入现实，不同领域间事物的类比，相似内容的对比等。

以身边的实例来检测理论。不仅能够加深理论知识的理解，而且有助于改进日常事物的处理方法。

记录

准备一个笔记本，纸的或者电子的。有什么发现、感想、疑问、经验等等，全都记下来。

如果是对某个话题、题目有比较完整的想法，最好能够及时整理成文，至少记录下要点。

这一点算是针对AI方面学习的专门建议吧。

AI涉及到的数学特别多。很多数学问题，之所以让人头大，其实并不是真的有多难，而是符号系统比较复杂，运算繁复，或者运算所表达的物理意义多样。

很多时候造成困扰是因为想不起来这里用到什么定理，哪个公式，或者这样操作表达的含义是什么了。

如果把常用的细小知识点都记录下来，按主题整理在一起，做成速查手册（小字典），需要用的时候迅速查找一下对应点，效果往往意想不到地好。能让我们的学习“机器学习”之路顺畅不少。

下面是两个笔者自己制作的例子，大家可以参考一下：

- [机器学习常用微积分知识速查手册](#)
- [机器学习常用线性代数知识速查手册](#)

GitChat