

## Default Question Block

此问卷调查的目标人群是机器学习项目的贡献者或者专业的软件开发人员，完成问卷大约需要10分钟。在这个问卷中，我们想要知道您对于机器学习软件维护方面的一些意见。您的参与是自愿性质的，您可以随时撤回。为了保障您的隐私，我们不会记录任何针对个人的识别信息。该问卷由美国加州大学尔湾分校（UCI）和浙江大学的计算机科学研究团队设计，问卷结果会被用于某个专业的科学出版物。如果您想要最终的研究结果或者出版物，请联系 吉日格斯（fjiriges@uci.edu）

- ☐ 同意并开始回答问卷
- ☐ 以后再说

### Block 1

你当前的工作单位是？

- ☐ 公司或者私人机构
- ☐ 大学（包括学生）
- ☐ 自由职业者
- ☐ 其他

你的主要工作领域是什么？

- ☐ 数据及应用科学。相关的职称或者头衔包括：应用科学家、数据科学家、数据科学技术负责人等。
- ☐ 项目管理。相关的职称或者头衔包括：项目经理、项目负责人等。
- ☐ 科研。相关的职称或者头衔包括：研究者、研究负责人等。
- ☐ 软件工程。相关的职称或者头衔包括：软件工程师、软件开发经理等。
- ☐ 其他

你工作的地点在哪里？

- ☐ 北美：美国
- ☐ 北美：加拿大、墨西哥

- ☐ 中南美洲
- ☐ 欧洲
- ☐ 亚洲：中国
- ☐ 亚洲：印度
- ☐ 亚洲：中东
- ☐ 亚洲：日本
- ☐ 亚洲：韩国
- ☐ 亚洲：其他\_\_\_\_\_
- ☐ 大洋洲（澳大利亚、新西兰）
- ☐ 非洲
- ☐ 其他\_\_\_\_\_

请估计一下在机器学习项目中你一般与多少人一起工作？

- ☐ 0（我不需要和别人合作）
- ☐ 1-2人
- ☐ 3-5人
- ☐ 6-14人
- ☐ 15-29人
- ☐ 30人以上

你至今参与了多少个机器学习项目？

这个问题需要以一个有效的数字

你现在有多少年的工作经验？

这个问题需要以一个有效的数字

你有多少年与机器学习项目相关的工作经验？

这个问题需要以一个有效的数字

你当前工作的项目组参与机器学习相关的工作有多少年了？

这个问题需要以一个有效的数字

你从哪里获取机器学习相关的知识和技能？（多选）

- ☐ 正规教育（例如本科、硕士或者PhD教育）
- ☐ 内部培训项目
- ☐ 工作中（例如构建机器学习或者人工智能相关的产品）
- ☐ 在线资料（网上公开课如慕课、Coursera、Edx；视频网站如YouTube）
- ☐ 自学
- ☐ 其他\_\_\_\_\_

Block 2

下面是一些机器学习软件项目可能遇到的开发挑战，你在机器学习软件开发项目中有没有遇到过类似的情况？如果选“是”，请评价该情况对项目维护造成的影响程度（1-5分）。如果选“否”，请跳到下一个情况

	是
	是
1. 你有没有遇到过，你本来可以使用一个第三方的机器学习库或者框架，但是你的项目组自己实现了功能	<input type="radio"/>
2. 你有没有遇到过你想要使用库中更加具体的函数，但是你的项目组已经使用了一个更泛化的函数，但是因为使用了更泛化的函数增加了更多的参数？	<input type="radio"/>
3. 你有没有遇到过你的项目组把所有函数都定义为公共访问（public access）但是后来发现了冲突，这是因为有相同名字的函数在其他类或者文件中已经被定义了，从而导致需要将相关的函数改为私有（private）？	<input type="radio"/>

是

是

4. 你有没有遇到过你的项目直接在模型的训练文件中测试训练出的机器学习模型，而不是创建新的函数或者文件去测试机器学习模型？ ☐
5. 你有没有遇到过你的项目中很多新定义的函数缺少默认参数从而使这些函数更加容易在之后出错？ ☐
6. 你有没有遇到过你的项目没有很好的将模型训练、测试及数据处理代码写到各自单独的文件？ ☐
7. 你有没有遇到过在你的项目中，机器学习模型的编写时是基于开发人员自己的理解，而不是基于模型的原来设计，从而造成代码变得难以理解？ ☐
8. 你有没有遇到过你的项目由于安装了不需要的第三方库或者依赖，从而使得项目的部署遇到困难？ ☐
9. 你有没有遇到过你的项目中的一些字符串存在硬编码（例如文件路径、模型名称等）而不是通过变量赋值？ ☐
10. 你有没有遇到过你的项目忘记删除一些不想要的调式代码？ ☐
11. 你有没有遇到过你的项目多次导入了一个相同的包或者模块（例如同时有import A 和 from A import B）？ ☐
12. 你有没有遇到过你的项目在使用一些第三方库和框架时（例如tensorflow、pytorch），因为这些第三方库和框架更新，导致你项目产生了错误从而迫使开发人员做出相应的修改？ ☐

除了以上在机器学习软件项目中遇到的代码维护问题，你还有没有遇到过其他的问题和挑战？如果有，请提供更多的信息：

## Block 4

以下是一些机器学习软件开发过程中可能遇到的场景，你是否在你的机器学习项目中也遇到过相似的场景，如果选“是”，请评价该场景对你的项目维护造成了多大的困难（1-5分），如果选“否”，请继续下一个问题。

	是否遇到 过?	我 不 明 白	如 身 评 化 造 成 非 常 大
1. 你有没有遇到过在一个机器学习系统中修改了一个模块（例如一个特征、超参数等），同时也需要修改其他多个模块？	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2. 你有没有遇到过在一个机器学习系统中需要依次的运行多个机器学习模型（例如一个模型的输出是下一个模型的输入）？	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
3. 你有没有遇到过在一个机器学习系统中你的模型结果已经被其他组，或者第三方使用在他们的系统里，但是没有告知你？	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4. 你有没有遇到过一个机器学习系统的输入的数据是不稳定的，例如随着时间会变化？	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
5. 你有没有遇到过一个机器学习系统的某个特征在一段时间后已经变成冗余特征了但是没有及时发现？	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
6. 你有没有遇到过一个机器学习系统必须要提升模型的精确度，即使提升非常小或者获得提升，需要极大的开销？	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
7. 你有没有遇到过一个机器学习系统中的一个支持包中的某些代码阻碍了模型的提升或者修改？	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
8. 你有没有遇到过一个机器学习系统中，数据管道随着开发需求的增长而增长，而和预先的计划不相符？	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
9. 你有没有遇到过一个机器学习系统使用了捆绑的组件，而其中并不是所有的组件都是需要的？	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
10. 你有没有遇到过一个机器学习系统包含了一些过时的代码路径，这些代码可能是实验性质的或者为了测试而加的？	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
11. 你有没有遇到过一个机器学习系统，因为同时使用了多种编程语言，从而导致整个系统效率不高？	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

除了以上在机器学习软件项目中遇到的高层次的代码维护问题，你还有没有遇到过其他的问题和挑战？如果有，请提供更多的信息

