**问答记录**

以下均为组员向别组提出的问题：

* ××：
  1. 如何将联邦学习应用至大模型训练中？
* ×××：
  1. 是否应该考虑到项目成本对于投资选择的影响？
  2. 具体采用何种方法进行专家打分？
* ×××：
  1. 如何通过提高大模型的效率实现项目经济价值的提高？
* 高歌：
  1. 指标中“商业潜力”这项仅分配了5%的权重有何依据？
  2. 使用已有的大模型模拟专家评价是否在可信度上存在问题？

以下为部分回应其他小组提问的回答（均由汇报人××进行回应）：

1. 你们指标中的“潜在价值”一项是否有些模糊？它具体涵盖了哪些方面？

答：“潜在价值”一项是衡量一个模型在“语言准确率”、“响应速度”、“响应质量”、“知名度和用户粘性”、“训练及部署成本”这五项指标以外的其他潜在应用前景的指标。例如，对于Google Bard，该项指标可以结合谷歌自身的搜索业务进行打分；又如对于文心一言，可以结合其对中文理解的特化能力进行打分。某种意义上，该项指标也反映了投资人自身对产品应用前景的倾向性，如投资人倾向于投资某个模型在小型化应用上的前景，那么阿里的通义千问模型就可以在“潜在价值”一项中获得较高的分数。

1. 在通过AHP方法为各指标进行打分以确定其最终权重的过程中，你有哪些考量？

答：作为用户，我们可能更倾向于从使用者的角度出发进行考虑，因此在权重分配上总体上会更关注于“语言准确率”“响应质量”等对用户体验能够产生更直接影响的指标，而不太偏向于“知名度和用户粘性”、“训练及部署成本”这些更偏向投资方面的考量。毕竟是模拟打分，我们不太可能在有限的时间内聘请专家进行更严谨的打分。某种意义上，这的确反映了我们权重分配方式的局限性。但是，我们也通过AHP方法尽量降低了打分中的主观成分，使其尽可能的客观。