# 测试目的

本阶段测试的主要目的是：

1. 测试对于有解决方案的数据，对于将不同比例的数据以不同的方式放入训练集、知识库、测试集后得到的模型准确性。

2. 测试上述方法中的模型处理时间是否能满足要求。

# 2. 对于500条有标签数据的情况

## 2.1 对于微调数据，不将其放入知识库的方法

方案1：按照训练数据 : 测试数据 = 325 : 175的比例测试。

​ 方案2：按照训练数据 : 测试数据 = 300 : 200的比例测试。

​ 方案3：按照训练数据 : 测试数据 = 350 : 150的比例测试。

​ 方案4：按照训练数据 : 测试数据 = 375 : 125的比例测试。

​ 方案5：按照训练数据 : 测试数据 = 400 : 100的比例测试。

## 2.2 对于微调数据，将其全部放入知识库的方法

同2.1的五种方案中的训练数据与测试数据的比例，进行测试，与2.1的区别仅在于我们将微调使用到的数据全部放入到了外挂的知识库里面。

除了上面的区别，这一类中我们还可以控制以下两个变量：

1. 控制是否将测试数据放入知识库。
2. 控制大模型在知识库中查询的温度T（0, 0.25, 0.5, 0.75）

## 2.3 对于微调数据，将其部分放入知识库的方法

五种比例仍同2.1。对于五种比例的方案，分别再测试将微调数据的25%、50%、75%放入知识库中的模型精度。

除了上面的区别，这一类中我们还可以控制以下两个变量：

1. 控制是否将测试数据放入知识库。
2. 控制大模型在知识库中查询的温度T（0, 0.25, 0.5, 0.75）

# 对于2000条未标签的数据

## 数据说明

这一部分的2000条数据即excel表中没有解决方案的数据，这一类数据不能够作为训练以及知识库的数据，只能用来测试。

## 测试方法

对于这一部分数据，我们在训练好的大模型上输入故障现象，然后输出可能的故障原因，使用gpt4对其进行打分。

# 对于可能以故障代码形式输入的数据

对于系统，可能会以故障代码的形式输入数据，故障代码往往与故障现象有着强关联。因此在系统中有了将故障代码转换为故障现象字符串的函数之后，我们便可以对其进行统一的处理。