

广东创智杯 AI 应用赛

AI 手机垃圾短信识别智能体挑战赛竞赛规则

一、赛程介绍

（一）算法原型验证

参赛团队需利用主办方提供的 200 条脱敏的标注短信数据，离线下载到本地，构建一个端到端的垃圾短信智能分类模型。提交分类模型详细构建方案 PPT 作为应答材料，无需进行线下答辩。

（二）复赛阶段：对抗环境生存

参赛选手需使用主办方提供的线上训练平台和 10 万条短信数据，其中包含 9 万条带标签的训练数据和 1 万条无标签测试数据。选手需基于训练数据构建模型，并对测试数据进行预测，在线提交预测结果，线上训练平台评判结果。

（三）决赛阶段：模拟攻防

决赛期间，参赛选手每 48 小时将获得一批带标签的训练数据及无标签测试数据，需依托主办方提供的线上平台持续优化垃圾短信分类模型，并在每轮 48 小时内完成预测并提交结果，逾期或缺交记为 0 分。计算所有结果的平均值，作为决赛的最终成绩。全部轮次结束后，选手还需参与线下答辩，最终成绩由模型检测结果和答辩评分共同构成。

二、数据集描述

（一）数据来源与特点

1. 数据来源：开源短信数据集，包含真实短信数据和大模型生成的对抗变体（如 GPT 生成的钓鱼文本），确保数据真实性和攻击场景的多样性。参赛者可自行划分训练集与测试集，用于模型训练、调优和评估。

2. 任务类型：二分类任务

Label 0：正常短信（普通通信内容）

Label 1：垃圾短信（推广、诈骗、钓鱼、恶意链接等）

数据格式：数据集格式为 csv 文件，具体格式为：Id：编号、Label：短信类别,(0 或 1)、message：短信内容

（二）文件格式规范

文件为 csv 格式，后缀名也必须以.csv 作为文件后缀名

id,label

0,1

1,0

2,1

...

（三）结果数据格式要求

CSV 格式要求

编码: UTF-8

分隔符: 逗号(,)

列数: 固定 2 列

第一列: 短信 id(字符串), 请按 id 从 0 开始, 从小到大进行排序

第二列: 分类标签(字符串)

无标题行: 直接从数据行开始

无空行: 文件末尾不应有空行

三、比赛评分规则

（一）初赛评分

由评审专家根据 PPT 内容评估项目进行打分。

（二）复赛评分

结果提交后系统将计算 F1 指标。

（三）决赛评分

最终成绩由两部分组成: 所有提交结果的 F1 值平均分 (占 60%) 和答辩成绩 (占 40%) , 按 6:4 比例加权计算。