

# 《人工智能导论》大作业

大作业以小组为单位完成。每个小组原则上3-4人，并由小组协商推选组长，组长负责任务统筹和确定成员贡献比例。

## 题目：谣言检测

### 1、目标

基于谣言检测数据集，构建一个检测模型。该模型可以对数据集中的推文进行谣言检测与识别。要求：

- ✓ 模型是2分类（0代表非谣言、1代表谣言），分类准确率越高越好；
- ✓ 模型具有一定的泛化能力；
- ✓ 有合理的运行时间。

### 2、检测模型

可选用各种分类模型。

### 3、数据集

由以下数据组成：

- 1) 正常推文，非谣言；
- 2) 谣言推文。

### 4、评分标准：满分100分

- ✓ 大作业报告叙述清楚（30分）

- ✓ 提交的代码是否能运行成功，代码书写清楚有条理（30分）
- ✓ 与训练集同源的数据作为测试集1，测试分类准确率（15分）
- ✓ 与训练集不同主题的数据作为测试集2，测试分类准确率（15分）
- ✓ 使用教师自有的谣言测试数据作为测试集3，测试分类准确率（10分）

## 5、实现要求

- ✓ 模型设计及其训练测试过程均需自己实现，可以参考公开的资源
- ✓ 对于雷同作业，经调查核实后，扣除一定分数
- ✓ 实现的模型具有一定的检测性能
- ✓ 数据集下载，建议运行环境和参考代码见 [《人工智能导论课程大作业支持文档》](#)

## 6、提交作业的具体内容

- 1) 大作业报告report.pdf。内容参考 [《人工智能导论大作业模板》](#)，突出重点内容，篇幅尽可能精简，不超过2000字，pdf格式文件。
- 2) 接口类文件classify.py。实现接口类RumourDetectClass，类的初始化函数中完成模型加载等初始化工作。接口类需提供一个接口函数classify，该函数的输入是一条字符串，输出是一个int值（值为对应的预测类别，即整数0或1，0代表非谣言、1代表谣言）。
- 3) 接口调用实例说明README.md（markdown格式）。

由组长到canvas上提交一个GitHub的地址或交大云盘的链接，包括如下文件：

--README.md：1) 注明组长姓名和组长学号；2) 接口调用实例说明

--report.pdf：大作业报告

--classify.py : 接口类文件

--其他支持文件和目录