

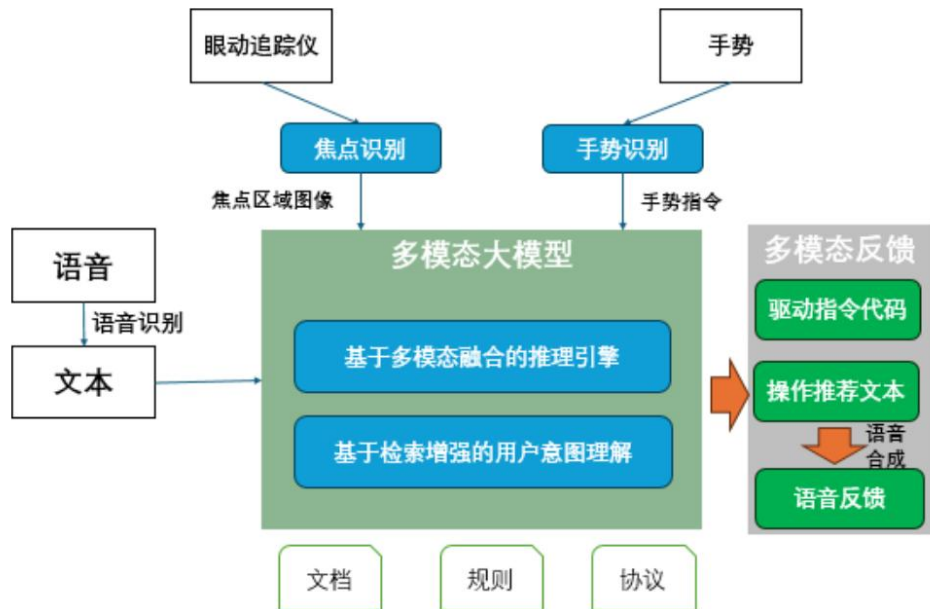
# 题目："车载多模态交互系统"

## ➤ 用户角色

• 驾驶员（主交互对象）、乘客、车辆维护人员、系统管理员

基础功能项：要求软件至少需要完成的功能

1. 语音控制：实现基本的语音指令识别，比如“打开空调”、“播放音乐”等。
2. 手势交互：定义并实现至少3种手势控制功能：
  - 握拳暂停音乐
  - 竖起大拇指确认
  - 摇手表示拒绝
3. 视觉交互
  - 实现简单的图像识别功能，能够识别用户的眼光集中区域。
  - 基本的头部姿态识别，能够识别点头（确认）和摇头（拒绝）。
4. 系统管理功能
  - 用户个性化配置，能够保存驾驶员的常用指令和交互习惯。
  - 记录和分析多模态交互日志，帮助优化用户体验。
  - 实现系统权限管理，区分驾驶员与乘客的操作权限，确保安全。



## 开发要求

1. 实现一个可运行演示的原型系统（需包含至少三种模态的完整交互流程）

2. 编程语言与框架不限

3. 组队要求

• 团队规模：不多于8 人（需指定组长）

• 角色建议：

- 项目经理（制定开发计划、协调进度）
- 算法工程师（多模态数据融合与模型部署）
- 交互设计师（设计车载界面与交互逻辑）
- 系统开发工程师（搭建通信框架与硬件模拟环境）
- 测试工程师（设计测试用例）

## 交付物

1. 可运行Demo

• 演示至少1个典型场景（三个模态的交互场景）

2. 软件文档

- 需求分析（包含用户场景流程图）
- 系统架构设计（多模态数据融合方案）
- 模块详细说明（语音/手势/视觉子系统的实现逻辑）
- 测试报告
- 软件开发文档

## 典型场景举例：

场景：异常状态反馈（整合语音、手势、视觉三模态）

### 1. 场景描述

驾驶员在行驶过程中分心，系统通过视觉检测分心状态，触发语音警告，并要求驾驶员通过手势或语音主动确认安全状态。

### 2. 输入

• 眼动数据（视觉模态）

视线偏离道路超过3秒

• 语音指令（新增语音模态）

驾驶员说“已注意道路”解除警告

• 手势动作（新增触觉模态）

竖起大拇指（确认安全）/摇手（拒绝警告）

### 3. 输出

• 文本反馈

状态栏显示：“警告！请目视前方”（红色闪烁）

• 语音反馈

系统播报：“请注意行车安全”，若未响应则升级为“请立即目视前方！”

• 视觉提示

仪表盘警告灯高频闪烁（频率随危险等级提升）

### 典型场景举例:

#### 3. 数据来源 (仿真实现方案)

##### 3.1 视觉模式仿真

###### 1. 摄像头数据

###### 1. 使用开源数据集:

- GazeCapture: 手机摄像头采集的眼动焦点数据集

###### 2. 或者自制数据、网络爬取:

- 用手机拍摄同学模拟驾驶场景 (示例: 注视前方/低头看手机/转头说话等), 并且具备标签



(a) 注视前方



(b) 低头看手机



(c) 转头说话

###### 2. 手势动作

###### • 使用Unity手势仿真工具:

- 导出握拳/竖拇指/摇手的动作视频 (帧率30fps)

###### • 或者自制数据、网络爬取:

- 用手机录制手势视频 (背景需纯色墙面), 握拳/竖拇指/摇手的动作。

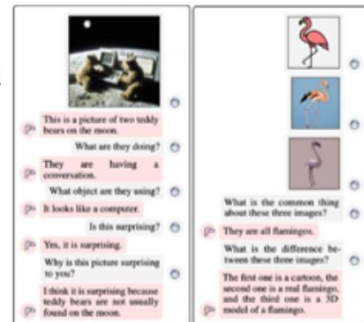
##### 3.2 语音模式仿真

###### 1. 从公开数据集截取片段:

- LibriSpeech (英文)
- THCHS-30 (中文)

###### 2. 自定义录音:

- 手机录制指令 (示例: “已注意道路”/“打开空调”)



后续可以利用已有的多模态视觉语言模型识别多个模态的数据, 例如:

Flamingo: a Visual Language Model for Few-Shot

## 软件开发文档

### • 一、项目概述

- 开发背景
- 项目目标
- 开发环境
- 可行性分析
- 项目计划

### • 二、需求分析与系统设计

- 需求分析
- 系统设计
- 详细设计
- 数据库设计
- UI设计

### • 三、系统测试

- 测试环境
- 功能测试
- 性能测试

### • 四、项目管理

- 参与人员及分工
- 项目进展记录
- 项目管理工具

### • 五、用户手册