# 基于人工智能的模型检测平台研究需求

本项目将以全面提升人工智能的模型检测业务技术支撑能力为目标，完成基于人工智能的模型检测平台研究的开发工作，需求如下：

（1）人工智能数据与模型

平台集成了多模态数据处理和多种高级模型，为用户提供了一个全面、高效和精确的数据分析工具。无论是企业、研究机构还是个人用户，都可以从中受益，实现各种数据驱动的应用和服务。平台具有图像模态数据、语音模态数据、文本模态数据、涉黄分析模型、情感分类模型、目标识别模型等模型。本平台的分析模型通过深度学习和其他机器学习技术，对图像、视频和文本内容进行深入的分析，从而实现对图像、文本、语音进行识别分类、命名、情感分析、目标提取等操作。

（2）人工智能模型增强

平台具有多种核心算法，可对图像数据集进行数据扩样增强功能，包括在训练时对加载数据进行翻转，旋转，裁剪，缩放，平移等预处理以及利用梯度、置信度等进行扰动增强样本；对语音数据集进行数据扩样增强功能，包括在训练前对数据集进行多目标扰动、普适扰动、语音合成扰动等处理，得到多倍的数据集；对文本数据集进行数据扩样增强功能，包括包括在训练时对加载数据进行替换、增词，混淆同形等预处理。

（3）样本智能分析生成

构建数据集自动生成和智能管理工具，支持整合主流开源数据集及定制化数据集的自动生成，并对数据集的样本稀疏性和类别不平衡性进行智能分析，通过数据增强等技术手段实现数据集的智能扩充，生成测试样本。开发技术建立网页测评展示平台，将本项成果集成部署到测试平台中，在网页上进行测评展示。

（4）模型检测

本项目针对不同的深度学习框架之间存在兼容性问题，给定任意框架下的一种或者多种模态数据，能够自动确定合适的模型结构，实现鲁棒/可解释/可靠的模型训练，是现实有用的。因此，本项目研究多模态数据和全覆盖模型在跨平台框架下的可靠性测试与鲁棒增强集成研究技术，为相关领域科研人员提供一套具有良好交互性的可靠性测试与鲁棒增强平台原型系统，具有数据和模型的动态可扩展功能，集成评估、检测、鲁棒加固技术，具备从鲁棒性、透明性、隐私保护等多角度对系统进行分析的能力，以及定制个性化的评分系统和测试报告的功能。

（5）数据分析管理

通过引入数据集管理、模型库管理、算法实现与调用管理（包括脆弱性评估算法、检测与防御加固算法、鲁棒性测试算法）、结果可视化报告等模块，智能协调地驱动测试与攻防算法的运行与结果呈现，有效协同大规模图像、语音、文本等数据和多类测试与攻防算法，实现体化平台集成建设。平台集成智能系统的内生脆弱性评估分析、恶意数据检测、全周期鲁棒加固、鲁棒性测评等技术，具备从鲁棒性、透明性、可解释等多角度对系统进行分析的能力，以及全面的评分系统和测试报告的功能。