

# 暴力检测系统报告

任昱菡 2022201533

在人工智能的浪潮中，多模态大模型以其独特的魅力吸引了众多研究者的目光。GLM-4V，作为智谱 AI 的杰出之作，更是将多模态处理能力推向了新的高度。

## 一、研究问题及意义

在此次项目中，我计划的选题是：做未成年“数字守门人”——多模态检测暴力信息。此次项目中我首先接入 GLM-4V 的 API，采用自监督学习，利用大规模多模态数据集进行预训练，然后设计专门的融合层，将不同模态的特征进行联合训练。注意力机制和特征融合技术增强了跨模态信息交互，使得模型能够更好地理解和处理多模态数据。

随着网络的普及和发展，未成年人在网络上受到各种形式的暴力信息的侵害与干扰，例如暴力游戏、恶意言论等，给其心理健康和成长带来了严重影响。

《中华人民共和国未成年人保护法》第十二条规定了国家对未成年人保护方面科学研究的鼓励和支持，因此，加强对未成年人网络保护的研究与实践显得尤为迫切。

当今社会，互联网的普及和社交媒体的盛行使得网络暴力言论问题日益突出，尤其对未成年人的影响更为严重。网络暴力言论不仅会对未成年人的心理健康造成负面影响，还可能导致他们受到欺凌、伤害甚至自杀。

根据研究表明，全国超六成网民遭受过网络暴力言论的侵扰，而未成年人正是网络暴力言论的主要受害者之一。他们在社交媒体上受到攻击、侮辱、威胁等形式的暴力言论，这种言论不仅会影响他们的自尊心和自信心，还可能引发焦虑、抑郁等心理问题。

此外，网络暴力言论也会影响未成年人的社交关系，导致被孤立或隔离。为了保护未成年人免受网络暴力言论的侵害，有必要进行相关研究，探讨如何有效检测和防范网络暴力言论。

## 二、数据集：多模态的宝藏

GLM-4V 的强大能力，离不开其背后丰富多样的数据集。这些数据集如同宝藏，为模型提供了源源不断的营养。

### 1. 文本数据集：语言的海洋

文本数据集是 GLM-4V 的语言基础。从新闻、书籍到学术论文、社交媒体，这些文本来源覆盖了多种语言和领域。经过分词、去噪、标注等预处理工作，这些数据变得井然有序，为模型提供了丰富的语言知识。

## 2. 图像数据集：视觉的盛宴

图像数据集则展现了 GLM-4V 的视觉理解能力。公共图像数据集、专业图像库以及自采集图像，构成了一个庞大的视觉世界。缩放、裁剪、增强等预处理手段，使得这些图像更加符合模型的学习需求。

## 3. 音频数据集：声音的旋律

音频数据集为 GLM-4V 带来了声音的感知能力。语音识别数据集、音乐库和自采集音频，构成了一个充满声音的世界。降噪、归一化、特征提取等预处理步骤，确保了音频数据的质量和可用性。

## 4. 视频数据集：动态的画卷

视频数据集则展现了 GLM-4V 对动态场景的理解能力。影视作品、短视频平台和自采集视频，构成了一个生动的动态世界。帧提取、标注、剪辑等预处理工作，使得这些视频数据更加适合模型的学习。

## 5. 多模态融合数据集：融合的艺术

多模态融合数据集是 GLM-4V 的独特之处。将文本、图像、音频和视频数据按照特定任务进行融合，形成多模态数据对。详细的人工标注确保了数据的一致性和准确性，为模型提供了丰富的多模态学习素材。

# 二、训练过程：精雕细琢的艺术

有了丰富的数据集，接下来就是训练过程的精雕细琢。GLM-4V 的训练过程分为多个阶段，每个阶段都有其独特的目标和策略。

## 1. 预训练：打牢基础

预训练阶段是打牢模型基础的关键环节。采用自监督学习，利用大规模多模态数据集进行预训练。大规模 Transformer 模型、动态调整的学习率以及 Adam 优化器，共同构成了这一阶段的优化策略。

## 2. 任务特定微调：精益求精

任务特定微调阶段则是针对特定任务进行精细化调整。使用小规模标注数据进行微调，针对图像分类、语音识别、视频理解、多模态问答等任务类型，采用精细的调参和正则化技术，确保模型在特定任务上达到最佳性能。

## 3. 多模态融合训练：融会贯通

多模态融合训练阶段是实现多模态信息有效融合的关键。设计专门的融合层，将不同模态的特征进行联合训练。注意力机制和特征融合技术增强了跨模态信息

交互，使得模型能够更好地理解和处理多模态数据。

#### 4. 工具集成与扩展：功能丰富

工具集成与扩展阶段为模型增添了更多的功能和灵活性。集成 Python 执行环境、实现浏览器功能以及集成图像生成和编辑工具，使得模型能够执行复杂逻辑、访问互联网以及进行数据可视化。

#### 5. 用户自定义智能体训练：个性定制

用户自定义智能体训练阶段则允许用户根据需求配置智能体的行为、知识和技能，进行个性化训练。这一阶段提供了开放的智能体框架，使得用户能够定制属于自己的智能体。

### 三、训练效果评估：量化的证明

训练效果评估是验证模型性能的重要环节。GLM-4V 采用了多种评估指标和方法，对其在文本、图像、音频、视频以及多模态任务上的性能进行了全面评估。内部评估和外部评估相结合，确保了评估的客观性和准确性。结果显示，GLM-4V 在各项任务上均取得了显著提升，尤其是跨模态处理能力更是突出。

### 四、挑战与解决方案：不断前行

在训练过程中，GLM-4V 也面临着数据不平衡、模型复杂度高等挑战。然而，通过数据增强、重采样、模型简化等解决方案，这些挑战被逐一克服。正是这些挑战和解决方案，推动了 GLM-4V 的不断前行和进步。

### 五、结语

GLM-4V 多模态大模型以其丰富的数据集、精细的训练过程和卓越的性能，展现了多模态处理的无限可能。随着技术的不断进步和应用场景的拓展，相信 GLM-4V 将在未来的人工智能领域绽放更加耀眼的光芒。