|  |  |
| --- | --- |
|  | **第十七届全国大学生软件创新大赛** |
| **文档编号：SWC2024-Hi安卓** |



**按键精灵**

**AutoDroid**

**作品创新性分析报告**

**Version: 1.0.0**



**Hi安卓**

**2024-03-19**

**All Rights Reserved**

目录

[1 痛点分析 1](#_Toc56720942)

[1.1 痛点概述 1](#_Toc56720943)

[1.2 相关工作 1](#_Toc56720944)

[2 项目创新点 3](#_Toc56720945)

[2.1 技术性创新点 3](#_Toc56720946)

[2.2 功能性创新点 3](#_Toc56720947)

[2.3 其他创新点 4](#_Toc56720948)

[3 竞品分析 5](#_Toc56720949)

文档修订历史

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **序号** | **修订原因** | **版本号** | **作者** | **修订日期** | **备注** |
| **1** | **初稿** | 0.5.0 | **Hi安卓队** | 2023.12.10 |  |
| **2** | **完善了相关工作**  **增加项目创新点** | 0.7.0 | **Hi安卓队** | 2024.01.28 |  |
| **3** | **对项目创新点进行了提炼** | 0.7.0 | **Hi安卓队** | 2024.02.15 |  |
| **4** | **对项目创新点及竞品分析进行修改** | 0.9.0 | **Hi安卓队** | 2024.02.29 |  |
| **5** | **增加了相关工作中的内容对项目创新点进行进一步细化和增加** | 1.0.0 | **Hi安卓队** | 2024.03.14 |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

# 痛点分析

## 痛点概述

基于项目背景，如今的智能手机助手应用面临着以下4个方面的问题：

1. 界面操作复杂性与用户熟悉度不足：

对于新用户或不常使用某些应用的用户来说，面对繁杂的操作流程或不直观的用户界面(UI)，完成简单任务都可能变得困难。例如，一个购物应用可能要求用户经过多个步骤才能完成购买，或者一个社交应用的隐私设置可能隐藏在多层菜单之下。这种复杂性不仅增加了用户的认知负担，也大大降低了操作的效率。

1. 针对残障人士的可访问性挑战：

残障人士，尤其是视障用户，使用触控屏幕设备和导航复杂的移动应用界面时面临巨大挑战。传统的触控操作和基于视觉的界面设计不满足他们的需要，导致这部分用户无法充分利用智能手机应用提供的服务和功能，增强他们的日常生活质量。

1. 隐私保护与数据安全问题：

在解析用户指令和执行自动化任务的过程中，需要处理敏感的用户数据和隐私信息。如何确保这些信息在自动化过程中得到妥善处理，避免泄露给第三方或上传到云端，是项目实施中必须严格考虑的问题。用户对于隐私保护的担忧可能会阻碍他们使用自动化助手，尤其是当涉及到需要访问个人信息的任务时。

1. 跨应用操作的复杂性：

用户经常需要在不同的应用程序之间完成任务，例如，从电子邮件中复制会议链接到日历应用中创建事件。这类跨应用操作涉及不同界面和操作逻辑的切换，对于自动化系统来说是一个极大的挑战。不仅需要精确理解每个应用的UI和功能，还需要在不同应用间高效准确地传递信息和执行操作，这对于设计高度灵活和智能的自动化助手提出了更高要求。

## 相关工作

1. 用户界面（UI）任务自动化：

用户界面任务自动化是指建立能根据用户界面自动执行任务的系统。最相关的工作可分为以下几类：1.理解图形用户界面的语义。这些方法可以帮助总结图形用户界面的主要功能，理解每个控件的目的，并解决与图形用户界面相关的用户问题。这对实现用户界面自动化和无障碍等各种交互任务非常重要。2.将用户指令映射到用户界面元素。这些方法旨在选择与用户指令最相关的图形用户界面元素。最近，一些研究正在使用 LLM 将指令映射到图形用户界面元素。3. 移动用户界面任务自动化。这些方法建立了一个代理，通过在图形用户界面上执行操作来完成用户的任务。另一方面，AutoDroid利用自动生成的用户界面过渡记忆来完成高级任务，包括智能手机上的多个步骤。该记忆可以帮助代理理解信息丰富的用户界面和应用程序的使用，并在推理和规划中增强 LLM。

1. 增强型 LLM：

虽然 LLM 可以执行问题解答和文本生成等各种任务，但它们仍然受到固定权重集和最大上下文长度所能存储信息的限制。因此，研究人员正在尝试用各种工具来增强 LLM，如网络浏览器、各种 API，甚至其他模型。这些研究通常依靠巧妙设计的提示来提高 LLM 的推理能力，或在特定领域的数据集上对 LLM 进行微调。此外，它们还经常假设有公共可用的应用程序接口来解决任务。而我们的方法并不依赖人工设计的应用程序接口，因为这些应用程序接口在移动应用程序中并不常见。

1. 自动应用程序分析：

自动应用程序分析是使用软件工具自动分析 Android 应用程序的功能、性能和可用性的过程。典型的技术包括静态分析（分析移动应用程序的静态组件，如代码、资源文件和配置）和动态分析（侧重于运行时行为）。在移动任务自动化中，关键是要了解应用程序的功能。先前的工作已经探索了基于代码、图形用户界面、网络流量和元数据来提取此类功能信息，但它们大多过于粗粒度，无法促进任务的完成。我们的方法使用 LLM 处理动态应用程序分析器收集的原始痕迹，以获得每个图形用户界面元素的细粒度描述。

# 项目创新点

## 技术性创新点

1. 将APP的UI界面转换为HTML文本。

使用Android的无障碍服务功能，动态监测用户手机APP界面的变化（如页面跳转、点击、输入），当界面状态发生变化时，就会调用无障碍服务捕获界面上的所有组件信息。根据组件的属性，将所有组件转化为HTML形式的文本描述（比如可点击的组件转换为HTML中的<button></button>标签）。组件的名字作为HTML标签的属性，并为转换后的HTML标签添加id属性供LLM选择。

1. 基于离线探索APP的Memory生成技术。

AutoDroid系统在离线阶段会对APP进行探索，生成APP的UI转换图，并对UI组件进行剪枝操作，对于不同界面中出现的重复的组件，仅保留离初始界面最近的一个。然后会见之后的所有组件通过query LLM的方式获取其功能，我们称之为模拟任务。基于此，我们生成了APP Memory表，每一个表项中存储：一个组件、到达该组件的路径（从初始界面到达该组件所在界面所需点击的组件）、该组件模拟任务。APP Memory会在之后用于对Prompt的增强，以及Memory Shrortcut快速跳转的操作。

1. 基于APP Memory的指令快速跳转。

所有的APP Memory中的模拟任务会通过一个特征提取模型Instructor-XL映射成固定维度的特征向量。在用户输入指令以及执行任务的APP后，系统首先会通过Instructor-XL将用户输入指令做映射，然后和该APP Memory中的模拟任务计算相似度。挑选相似度最高的模拟任务对应的组件，在任务执行开始的时候，先行跳转到该组件对应的界面并点击该组件。这样的设计减少了执行指令需要query LLM的次数，减少了开销，也减少了完成指令的所花费的时间。

## 功能性创新点

1. LLM赋能，能够对用户提出的任何移动任务进行自动化。

传统的手机助手（Siri、华为小艺、Google Assistant等等）通过调用API的方式来对移动任务进行自动化，整个助手所支持的功能需要开发者逐一实现，只能在部分APP上处理部分任务。AutoDroid使用了具有强大文本理解、推理能力的LLM（如：ChatGPT、Vicuna、OPPO AndesGPT等）来进行移动任务的自动化。由于LLM具有海量的先验知识，能够灵活得根据用户界面的文本描述做出正确的推理，得到完成任务所需的动作，因此该系统不局限于某个特定的APP，能够在用户手机的任何APP上完成任务。

1. 将LLM部署到移动手机，保障用户隐私安全。

基于大语言模型的高性能部署解决方案MLC LLM，将具有强大性能的开源大语言模型Vicuna部署到手机上，从而调用本机上的LLM来进行移动任务动作的推理。由于系统需要将用户界面上的信息转化为文本形式的UI描述（HTML的格式）传输给LLM，因此这种文本形式的UI描述中可能会包含用户界面中的隐私信息。而将LLM部署到本机上就避免了向云端大模型传输用户界面的信息，使得系统可以在离线状态运行，从而保护用户的隐私安全。

## 其他创新点

1. 无障碍服务，模拟点击、输入等操作。

与传统的智能助手如 Siri 不同，AutoDroid不依赖于系统后端API的调用，而是通过Android提供的无障碍功能，通过模拟人类的点击和输入等操作，直接与手机应用的图形用户界面（GUI）互动。这种方法不仅提高了安全性和隐私性，还确保了agent能够适应应用界面的变化和更新。

1. 隐私过滤器，保障用户隐私安全

基于Microsoft Presidio算法，在线处理过程中，AutoDroid运行一个人身份信息（PII）扫描仪，检测UI信息中的用户隐私信息，包括姓名、电话号码、电子邮件地址等。将用户隐私信息用非专用词取代（例如“＜姓名＞”→“张三”），然后将Prompt发送到云。在收到LLM的响应后，AutoDroid在解析操作之前将特殊单词映射回原始单词。

# 竞品分析

竞品如Siri、Google Assistant和华为小艺等传统智能助手，在智能手机应用自动化领域已有广泛应用。这些助手依赖于预定义的指令集和API调用来执行任务，优势在于能够快速响应系统级操作和常见查询任务，如设置闹钟、查天气等。它们也通过第三方应用的集成提供有限的跨应用操作能力，但这通常需要应用开发者主动适配和提供支持。对于新或少见的应用场景，这些传统助手往往因缺乏深入的应用内操作支持而受限。此外，这些系统在隐私保护方面虽有措施，但用户数据处理和云端交互的方式仍然可能引起用户对隐私安全的担忧。

相比之下，AutoDroid在以下方面展示了明显的优势和创新点：AutoDroid采用将APP的UI界面转换为HTML文本的方法，配合大型语言模型（LLM）的综合理解和推理能力，实现对任意APP内部任务的自动化操作。这种方法突破了依赖特定API和开发者适配的限制，为用户提供了在任何APP上完成任务的能力，显著扩展了自动化助手的应用范围和深度。特别关注残障人士的使用需求，AutoDroid通过语音输入指令和高度自动化的UI操作，为视障等残障用户提供了更便捷的手机使用体验。这不仅展现了技术的社会价值，也为特殊群体带来了实际的便利和自立能力的提升。在隐私保护方面，AutoDroid通过先进的隐私过滤技术，确保用户数据在本地处理，避免敏感信息传输至云端，有效保护了用户的隐私安全。这一点对于越来越关注数据安全的用户群体具有重要吸引力。