SCOPE：评估并提升移动应用中的权限说明透明度

Liu Wang[1]，Tianshu Zhou*[2]，Haoyu Wang[1],Xiyuan Liu[3],Yi Wang†[2]1 中国华中科技大学 [2] 中国北京邮电大学[3] 中国盒马、阿里巴巴集团

arXiv:2611.32122v1 [cs.SE] 24 Nov 2026

摘要——**权限说明是随移动应用权限请求附带的解释性文本，对于用户隐私透明度和知情同意至关重要。尽管其重要性不言而喻，当前的权限说明实践仍常常未能达到监管要求，原因在于缺乏系统化的评估机制。通过对 600 款主流移动应用开展实证研究，我们发现其中普遍存在不足：15% 的权限请求没有提供说明，其他说明则使用含糊措辞或技术术语；更关键的是，许多说明未披露第三方 SDK 的数据访问情况，尽管这些组件会主动使用已授予的权限。为弥补这些透明度缺口，我们推出了 SCOPE，这是一种自动化多智能体框架，可系统评估权限说明合规性，并生成有针对性的优化建议。SCOPE 采用四个协同工作的专用智能体：基于多模态大语言模型的说明提取、基于少样本学习的语言分析、基于动态 API 的用途推断，以及自适应报告生成。全面评估表明，SCOPE 效果显著：说明提取准确率达到 98%，合规性分析一致性达到 93.5%，用途推断准确率达到 92%。一项由 30 名参与者参加的用户研究显示，84.6% 的参与者更青睐经 SCOPE 优化的说明，证实了其实用价值。本研究首次对权限说明实践进行了系统化分析，并提出了一种可扩展的解决方案，以提升移动应用的隐私透明度。**

索引词—**权限解释、可用隐私、数据透明度、移动应用**

I. 引言

移动应用通常需要访问大量敏感用户信息，例如位置数据、联系人列表和设备标识符，才能提供各种功能 [1,2]。应用权限是一种关键的访问控制机制，用于规范应用访问此类敏感资源的能力：应用需要请求相关权限，用户则需要明确授予这些权限，应用才能访问受限数据或执行受限操作 [3]。用户对权限授予的决定——允许还是拒绝这些请求——会显著影响其设备的隐私状况及相关风险。鉴于相当一部分安卓应用表现出过度请求权限的行为，申请超出其声明功能所需的权限，这尤其令人担忧 [4,5]。然而，用户往往难以做出此类决定，主要是因为现有做法通常未能充分提高权限使用的透明度，例如未能说明数据如何、为何以及由谁访问，导致用户无法确定是否授予访问权限 [6]。因此，用户往往只是为了“点掉”权限请求并继续执行主要任务而授予权限，从而增加了数据泄露和隐私侵犯的风险 [7]。

安卓的权限继承模型使情况变得更加复杂：嵌入式第三方库可以访问宿主应用已授予的所有权限。在现代应用开发中，集成多个第三方库（如广告、社交或分析 SDK）来委托处理特定任务和功能，进而简化开发流程，已是常见做法。由于这些库继承了应用的所有权限，它们往往会获得超出实际功能需求的权限。已有研究报告了“权限搭便车”问题 [8],：第三方库会探查已授予的权限，并在未明确自行请求权限的情况下使用或收集可获取的数据。此外，应用开发者往往不了解各个库的具体内部机制及其后台操作 [9]。因此，这些库可能会收集超出其主要功能所需范围的个人和敏感用户数据，在用户不知情且开发者无意为之的情况下造成严重的隐私风险。

欧盟《通用数据保护条例》（GDPR） [10],、美国《加州消费者隐私法案》（CCPA） [11]，和中国《个人信息保护法》（PIPL） [12]等隐私法规强调用户享有知情同意权和个人数据控制权，并要求
在收集数据之前提供清晰、易于理解的披露信息并获得明确同意。尽管应用通常依赖隐私政策来传达数据实践，但此类文件往往篇幅冗长、内容复杂，普通用户难以理解，从而降低了其及时且有效传达透明度的作用。此外，业界越来越多地倡导使用**“权限解释”**，即开发者主动在应用内提供简洁、易于用户理解的提示词，阐明权限请求的目的和必要性（如图 1所示），以此提升用户认知并促进更加 informed 的决策。鉴于大多数用户很少完整阅读隐私政策，而更多依赖实时权限提示词来进行决策， [13,14],精心设计的权限说明代表着提升透明度的一条可行路径。研究表明，当权限请求附有清晰的解释时，用户更有可能作出知情同意 [15,16]。

* Tianshu Zhou 是共同第一作者。† Yi Wang（yiwang@bupt.edu.cn）是通讯作者。

然而，目前权限说明的实施几乎不受监管。开发者可以完全自行决定解释内容和设计，也不存在系统级执行机制来确保遵守隐私法规。这一监管缺口为旨在诱导用户同意的操纵性做法创造了机会，例如以往研究揭示的暗黑模式 [17,18]。这些问题凸显了建立系统化评估机制的必要性，以确保权限说明真正践行其旨在促进的透明度原则并赋予用户更多自主权。

为了解当前状况，我们首先对 600 款主流移动应用中的权限解释实践开展了实证研究。调查发现，主要存在三类缺陷：(*i*) 许多应用完全没有提供任何解释；(*ii*) 另一些应用使用含糊或专业性过强的语言，掩盖了实际的数据使用情况；(*iii*) 最关键的是，尽管这些嵌入式库继承了所有已授予的权限，并可能加以滥用，但解释却普遍未披露第三方 SDK 访问。为弥补这些透明度缺口，并支持开发者编写合规、以用户为中心的解释，我们提出了 SCOPE，这是一种多智能体框架，可系统地评估权限解释合规性并生成有针对性的优化建议。该框架采用四个专用智能体，分别负责说明提取、语言适当性分析、用途推断和自适应报告生成，从而能够全面评估可见性、语言质量和实质准确性。我们开展了广泛评估，结果表明，SCOPE 在说明提取（使用多模态大语言模型时准确率为 98%）、语言合规性分析（评估一致性为 93.5%）和用途推断（结合上下文增强时准确率为 92%）方面均表现出较高的准确性。一项有 30 名参与者参加的用户研究显示，84.6% 的参与者更喜欢经过 SCOPE 优化的解释，而非原始解释，验证了我们优化方法的实际效用。该工具已公开提供 [19]。总之，本研究作出以下贡献：

- 我们首次对权限说明实践进行了系统性的实证分析，揭示了普遍存在的
  三个层面均存在透明度缺陷，包括普遍缺少说明、语言不透明性，以及系统性地未披露第三方 SDK 数据访问。
- 我们提出了一种新颖的多智能体框架，用于自动评估和优化说明。我们的方法融合了多模态大语言模型、动态API分析和法规合规性评估，可针对多个维度提供全面、可操作的反馈，以提升说明质量。
- 我们通过广泛的实验验证，证明了该方法在技术上的有效性以及用户对它的接受度。评估结果表明，该方法的表现优于传统方法，且用户显著偏好框架优化的解释，从而验证了其在现实场景中的适用性。



(a) 示例 1 (b) 示例 2 (c) 示例 3

图 1. 风格各异的权限解释界面示例。

II. 背景与动机

A. 权限解释

自 Android 6.0（Marshmallow）引入动态权限授权模型以来，应用必须在运行时请求权限，以便用户进行交互式授权。在该模型下，权限请求对话框由 Android 操作系统统一生成，通常会显示标准化提示，例如“允许此应用访问您的位置吗？”。开发者无法自定义这些系统生成对话框的内容。为了促进用户理解权限的用途，Android 官方建议开发者在调用系统对话框之前，通过应用内界面（例如自定义对话框或叠加提示）主动说明请求权限的原因和必要性 [20]。应用开发者为阐明所请求的敏感权限的用途和必要性而提供的此类说明信息或文本提示，称为权限解释。

由于开发者完全掌控这些解释的内容和设计，因此其格式和风格可能存在显著差异。如图 1 所示，常见的设计模式有几种：有些应用会在默认系统对话框旁增加一个解释性文本框（示例 1）；有些则会在系统提示之前显示自定义解释对话框（示例 2）；还有一些应用采用更具视觉吸引力的设计，加入插图或交互元素，以帮助用户理解（示例 3）。这些多样化的设计实践体现了开发者在信息透明度与用户体验之间寻求平衡的努力，同时也暴露出当前权限说明实践缺乏统一的设计标准和评估标准。

B. 光明面与阴暗面

与冗长且往往含义模糊的隐私政策相比，权限说明被广泛视为一种有助于提升用户理解能力并促进知情决策的有效机制 [13,14]。研究表明，当用户获知请求权限的理由时，他们授予权限的可能性会提高 12% [16]。

然而，权限说明的实际效果在很大程度上取决于实施质量，因为开发者在实际操作中对内容和设计拥有完全的自由。虽然监管指南提供了高层次的原则（例如数据最小化、透明度和情境化的权限请求），但其概念仍然宽泛，缺乏具体的实施说明或可直接使用的模板

供开发者遵循。此外，也没有系统级执行机制来阻止开发者违反这些法规。这些因素导致权限说明的质量参差不齐，并可能遭到滥用。例如，Mohamed 等人 [17] 发现，现实中的 iOS 应用跟踪透明度（ATT）提醒使用了误导性、含糊不清或带有诱导性的措辞，使用户感到困惑，并扭曲了他们对权限的理解。这些暗黑模式背离了权限说明的初衷，使其从帮助用户做出知情同意转变为暗中操纵用户的决定，最终降低透明度，削弱用户对其个人数据的掌控。

C. 研究动机

以往研究探讨了移动应用开发者可能采用的各种暗黑模式，以诱使用户采取非预期行动 [17,21]–[23]。然而，在权限请求这一情境下，现实中的安卓应用是否以及如何利用开发者撰写的解释来影响用户决策，尚未得到充分研究。几个关键问题仍未解决：权限解释中会出现哪些类型的暗黑模式？如何系统地识别这些做法？又有哪些机制能够遏制此类滥用？这些知识空白促使我们系统地调查当前的权限说明实践，并制定提升其透明度与完整性的策略。

III. 实证研究

为了解当前权限说明实践的现状并识别系统性缺陷，我们对 600 款移动应用开展了实证研究。这项实证研究有两个目的：建立有科学依据的评估框架，并提供现实应用中存在透明度缺口的实证证据。

A. 分层合规性审查框架 *rk*

为系统评估权限解释的合规性与质量，我们构建了一个三层分层审查框架。鉴于现有监管指南仅提供高层原则，未规定具体的评估标准，我们通过构建专家共识，将这些要求转化为实用的评估维度。

1）专家共识构建：我们首先审阅了《通用数据保护条例》、《个人信息保护法》等主要法规文件，以及中国行业指南《移动互联网应用程序系统权限使用指南 [24]》。这些法规强调目的清晰、用户可访问性以及避免误导性内容等原则，但缺乏用于系统化评估的具体操作指南。

为建立科学可靠且具有实践适用性的评估框架，我们采用德尔菲法 [25] 开展结构化专家咨询。这种方法尤其适用于缺乏统一标准的领域，并能通过多轮匿名专家反馈 [26,27]，系统地制定评估标准。我们邀请了六位领域专家：两位隐私学术界的研究人员、两位科技行业的移动应用开发者，以及两位监管机构的合规专业人员，他们均拥有三年以上相关经验。咨询过程分为两轮：

- **第1轮：**专家收到的全面材料包括主要隐私法规中的原则和要求摘要。在没有预设评估维度的情况下，专家独立确定评估权限解释合规性时应考察的关键方面。通过开放式问卷，专家提出了多个评估维度，包括可见性、可访问性、清晰度、具体性、可理解性、完整性、准确性、第三方披露、目的契合度以及避免误导性内容。我们综合了专家的反馈，并归纳出这些维度中反复出现的主题。

- **第2轮：**根据第1轮的综合结果，我们将确定的维度划分为三个层级：解释可见性（解释是否便于用户获取）、语言适切性（解释是否表达清晰且恰当）和实质准确性（解释是否如实反映实际权限使用情况）。我们向专家展示了这一三层框架。肯德尔W系数分析显示，框架所有组成部分的数值均超过0.75，表明专家共识度较高，并符合既定的德尔菲标准 [28]。

2）三层评估框架：根据专家反馈，我们最终确定了一个分层结构（如下所示），用于系统地评估权限说明，并以此作为实证分析和工具开发的基础。

- **L1：视觉可访问性**。本层考察在请求权限时，解释是否已清晰可见。解释不得隐藏在隐私政策等较难访问的资源或其他隐藏信息页面中。只有明确展示了解释的应用才会进入后续评估阶段。

- **L2：语言适切性**。本层评估解释文本的清晰度、具体性和适切性。解释应当：（1）明确说明请求权限的目的，避免使用含糊或过于宽泛的措辞；（2）使用通俗易懂的语言，避免晦涩的技术术语；（3）不得包含夸大权限必要性或向用户施压的误导性说法。本项评估主要依靠语义和文本分析。

- **L3：实质准确性**。本层评估解释是否如实反映了所请求权限的实际用途。它会检查解释是否完整、诚实，确保清楚披露所有相关用途，包括第三方 SDK 的用途或数据共享实践。此步骤需要结合动态分析技术，将应用的实际行为与解释内容进行比较，并识别潜在的不一致之处。

这一三层框架构建了一个渐进式、深入的合规性评估体系，涵盖可访问性、

语言适切性和行为对齐。该框架以法律要求、实践见解和专家共识为坚实基础，为本研究后续的评估和系统设计奠定了基础。

B. 权限说明的现状

基于这一三层框架，我们对来自中国应用市场、涵盖各类应用的 600 个真实世界移动应用进行了系统分析，旨在考察实践中权限说明的实施现状。

1) 分析过程：鉴于自动化用户界面探索工具难以实现全面的权限覆盖，我们采用手动测试，以确保最大程度地触发权限请求。经过培训的研究人员对每个应用进行了至少 5 分钟的系统探索，重点触发依赖权限的功能，例如定位服务、相机访问、联系人管理和媒体访问。在测试过程中，我们使用了 Frida [29]，这一动态插桩工具包，在运行时对 Android 系统 API 进行挂钩，并捕获敏感资源的调用堆栈跟踪，包括日历、通话记录、联系人、位置、设备标识符、身体传感器、短信、存储、相机和麦克风。我们对所有权限请求对话框进行屏幕截图，同时监控相应的敏感 API 调用。

两名具备相关专业知识的作者依据这三层层级框架，独立评估了收集到的每条权限说明：首先评估解释可见性（L1）。如果其被认为是可接受的，则进一步评估语言适切性（L2），并利用捕获的 API 调用信息考察实质准确性（L3）。如有任何分歧，则由全体作者通过共识讨论解决。

2）主要发现：通过对 600 款应用进行系统分析，我们在各种敏感资源中共发现 1,031 项权限请求。采用三层框架进行评估后，我们发现当前的权限说明实践存在显著不足：

- **缺少解释：**在 1,031 项权限请求中，有 158 项（占 15%，涉及 89 款应用）完全未提供解释。这些应用直接显示系统权限对话框，未提供任何关于权限用途的背景信息，使用户难以做出明智的决定。

- **语言适切性问题：**在已有的 873 条解释中，有 70 条（涉及 60 款应用）不符合语言适切性标准。主要问题包括：缺乏具体性（*43* 条）：解释使用“为了提供更好的服务”等含糊表述，却没有说明具体的使用场景。表达复杂、术语晦涩（*9* 条）：使用普通用户可能难以理解的技术术语，例如“SDK”或“基于位置的服务”。操纵性语言（*18* 条）使用夸大必要性或施加压力的误导性说法，例如毫无依据地声称“必须授权才能正常使用”。

- **实质准确性问题：**最为关键的是，246 条说明（占有说明的权限中 28%，涉及 159 款应用）存在信息披露不完整的问题，尤其是未说明第三方 SDK 使用情况。这只是保守估计，因为许多权限使用可能未在我们的测试过程中触发。我们的 API 分析发现，许多情况下，广告或分析 SDK 在未于随附说明中提及的情况下访问了权限。例如，某款导航应用称，位置访问“用于路线规划和查找附近车辆”，但我们的分析检测到，字节跳动的广告 SDK（com.bytedance.sdk.openadsdk）也在后台线程中访问位置数据——这一用途完全未向用户披露。

现代应用通常集成 5–10 个第三方 SDK（有些甚至多达 50 个） [8]，而这些组件会继承授予宿主应用的所有权限。然而，我们的分析表明，权限说明始终只关注应用的主要功能，却忽略了第三方 SDK 的使用场景。这造成了透明度缺口：用户同意了明确说明的用途，却不知道集成的 SDK 还会将数据用于其他次级用途。

C. 问题陈述

我们的实证分析揭示了当前权限说明实践中存在的重大缺陷，削弱了其预期的透明度效益。主要有三个问题：（1）大量权限请求缺少说明（占 15%）；（2）含糊的措辞和技术术语造成语言障碍，妨碍用户理解；（3）尽管我们的监测明确检测到这些组件调用了 API，但说明仍系统性地遗漏了第三方 SDK 的数据使用情况。

这些缺陷从根本上违反了隐私法规所要求的知情同意原则。当说明缺失、含糊或不完整时，用户就无法真正做出知情决定；而未披露的第三方访问则会带来用户无法预料或防范的隐私风险。法规意图与实际落实之间的差距表明，亟须采用系统化评估和优化方法，自动评估说明质量并提供有针对性的改进建议。

IV. SCOPE

为应对已识别的挑战，我们提出了 SCOPE（**S**mart **C**ompliance and **O**ptimization of **P**ermission **E**xplanations，权限解释的智能合规与优化），这是一种自动化多智能体框架，能够系统地评估权限解释合规性，并生成有针对性的优化建议。SCOPE 构建于 OpenAI Agents SDK [30]，之上，体现了智能体系统的关键特征，如自主规划、推理和行动，从而能够基于我们在实证研究中建立的三层评估框架开展系统性评估。

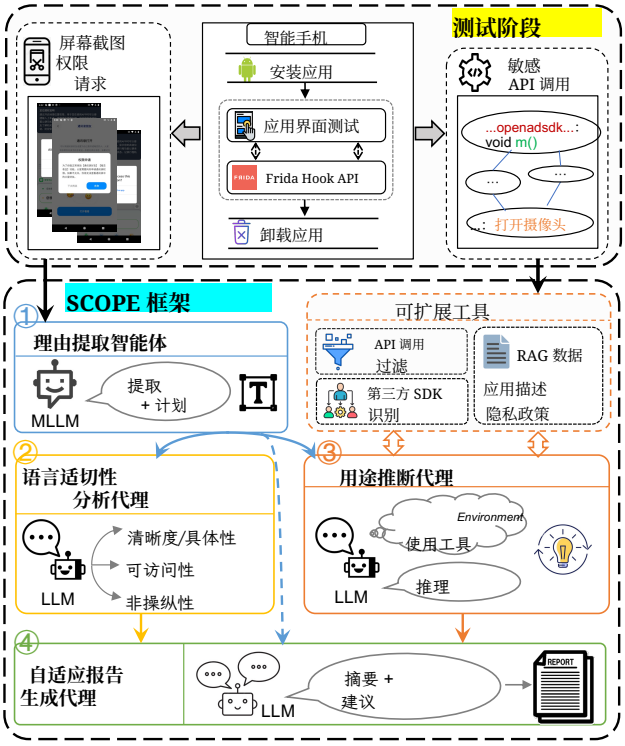


图 2. SCOPE 工作流程概览。

A. 总体架构

SCOPE 采用协作式多智能体架构，专用智能体通过结构化的规划、执行和通信机制协同工作。如 图 2所示，该框架编排四个专用智能体，并遵循一条基于中间结果进行条件分支的结构化流程：

解释提取智能体使用多模态大语言模型（MLLM）处理捕获的屏幕截图，以识别并提取权限说明文本。该代理利用规划能力来确定执行路径：如果成功提取出解释文本，协调器会将内容同时路由至语言适切性分析智能体和用途推断代理进行并行分析；如果未找到解释文本，工作流程将跳过语言适切性和用途分析，并将控制权交给自适应报告生成智能体，以记录未发现结果。

语言适切性分析智能体采用少样本学习方法，评估提取的解释文本是否符合法规合规要求。该代理维护语言模式和合规标准记忆，从而能够对不同解释样式进行一致评估。

用途推断代理分析运行时 API 调用信息和应用元数据，以确定实际的权限使用场景。该代理采用检索增强生成（RAG），结合应用描述和隐私政策信息，并通过上下文推理和第三方 SDK 识别，实现全面的用途推断和真实性评估。

自适应报告生成智能体综合结果汇总所有前序代理的结果，生成全面的评估报告。该代理汇总发现、识别合规差距，并根据检测到的具体问题组合生成针对性建议。

B. 解释提取智能体

解释提取智能体处理捕获的屏幕截图，以识别并提取权限说明文本。鉴于不同应用中权限说明的呈现形式多种多样，该代理必须适应各种用户界面样式、文本布局和视觉设计。

我们考虑了多种文本提取方法，包括传统的光学字符识别（OCR）技术（例如 Tesseract、PaddleOCR）和多模态大语言模型（例如 Gemini 2.0 Flash、gpt-4o）。虽然传统 OCR 方法能够提供基本的文本识别功能，但在处理复杂用户界面场景时存在显著局限，包括低对比度文本、复杂的用户界面元素以及多样的字体样式。此外，OCR 方法提取的是原始文本，缺乏上下文理解，因此需要额外的后处理来筛选相关的权限相关内容。相比之下，多模态大语言模型展现出更优异的性能，能够直接理解界面上下文，并利用其语言建模能力精准定位权限说明文本。我们的初步评估表明，多模态大语言模型在文本提取方面具有更高的准确性，能够更好地适应多样化的用户界面设计，并且由于无需大量后处理，能够实现更高效的端到端处理。因此，我们采用多模态大语言模型作为解释提取的核心技术。

代理使用精心设计的提示词，引导多模态大语言模型准确识别并提取说明内容。如下所示，提示词包含识别与权限相关说明的具体指示，同时过滤掉无关的界面元素。

提示词示例 1：说明提取

系统： 你是一名协助从图像中提取文本的助手。**用户：** 分析这张移动应用截图，并提取任何用于说明权限请求理由的解释文本。重点关注开发者提供的、用于说明为何需要特定权限的内容，而非系统生成的权限对话框。如果存在，请提取完整的解释文本。如果没有开发者提供的说明，请回答“未找到说明。”

输出格式：
- 权限类型： [所请求的具体权限]
- 解释文本： [权限的完整解释文本]

C. 语言适切性分析智能体

该代理会评估提取出的解释文本质量，以确保其符合监管要求，满足清晰度、具体性和非操纵性沟通这三个关键维度；这三个维度是在我们的实证研究中通过专家共识确立的。为提高文本质量评估的准确性和适应性，该代理采用了基于大语言模型的少样本学习技术。不同于依赖预定义关键词库或

可读性指标的传统基于规则的方法，基于大语言模型的方法能够提供更细致的语义理解，并且在多样化的表达风格之间具有更好的泛化能力。该代理从以下三个关键维度评估解释文本：

(1) 清晰性与具体性：代理评估解释是否提供了关于权限使用的具体、明确说明，而非含糊、笼统的表述。它会识别含糊术语，并评估解释的整体可理解性。

(2) 语言可及性：该分析考察所使用的语言是否便于普通用户理解，避免过多使用可能妨碍用户理解的技术术语或复杂法律术语。

(3) 非操纵性沟通：代理会检测潜在的操纵性或胁迫性语言模式，包括施压策略、强调益处却淡化隐私影响，或使用旨在不当影响用户决策的说服技巧。

少样本学习方法在提示中融入多个带注释示例，以指导大语言模型进行准确评估，如以下示例所示。

提示示例 2：语言适切性分析

系统： 你是一名文本语义学和语言学专家。**用户：** 请从以下三个维度评估下面的权限解释文本：

评估标准：

1) **清晰度与具体性：**目的是否具体明确？ 2) **语言可及性：**不具备技术专长的人能否理解？ 3) **非操纵性沟通：**是否避免使用强迫性或过度劝说的语言？

示例：

- **清晰度与具体性：**
  – 合规：“我们会访问您的照片相册以上传个人资料图片。”– 不合规：“需要照片相册权限以优化您的使用体验。”（过于含糊）
- **语言可及性：**
  – 合规：“我们使用位置数据来显示附近的商店。”– 不合规：“基于位置的服务（LBS）提供兴趣点信息。”（过于技术化）
- **非操纵性沟通：**
  – 合规：“我们访问您的联系人列表，以实现联系人同步功能。”– 不合规：“允许使用联系人权限即可解锁高级服务。”（操纵性激励）

待评估的解释：”[解释文本]”**响应格式：** _

- 总体合规性：[合规/不合规]
- 发现的问题：[列出具体问题]

D. 用途推断代理

用途推断代理用于确定所请求权限的实际使用场景，从而能够验证说明的真实性和完整性。既有研究表明，从敏感 API 调用和应用元数据（如应用描述、隐私政策）中提取上下文线索，可以有效推断应用实际的数据访问用途。 [31,32]受这些研究发现的启发，我们的代理通过系统化工作流协调多个分析组件。首先，代理调用 API 调用筛选工具，获取与各项权限请求在时间上相关的 API 跟踪记录。同时，它调用第三方 SDK 识别工具，确定各 API 调用的归属，区分主应用功能与第三方 SDK 使用情况。在此基础上，代理利用 RAG增强的上下文整合，从应用描述和隐私政策中检索相关应用元数据。通过分析多个来源中的命名模式、执行上下文和语义线索，代理推断权限使用的实际用途，并将其与声明的说明进行比较，从而识别完整性方面的差异（例如，未披露第三方使用情况）和准确性方面的差异（例如，所述用途与 API 使用情况不符）。这一工作流能够系统地检测说明真实性方面的违规行为，并为改进权限透明度提供详细见解。

1) API 调用筛选工具： 该工具从动态分析期间采集的完整运行时监控数据中提取相关的 API 调用轨迹。该工具采用时间相关性分析：当发生权限请求时，以请求时间戳为中心定义一个时间窗口，并捕获该窗口内的所有 API 调用。随后，该工具执行去重以移除冗余轨迹，并过滤掉无关的系统调用，仅保留与权限请求在时间上相关联的权限敏感型 API 调用。这个经过过滤的 API 调用集合为后续的用途推断分析提供基础数据。

2) 第三方 *SDK* 识别工具： 该工具通过分析调用包名并确定其归属，识别权限使用源自主应用还是集成的第三方 SDK。该工具采用结合参考数据库与基于大语言模型的推断的混合方法：(1)参考数据库： 我们首先从多个来源汇编了一个已知第三方 SDK 的综合数据库，包括谷歌官方文档、开源学术资料和行业报告。该数据库将包名映射至相应的服务类别（例如广告、分析和社交功能）。(2)大语言模型增强扩展：鉴于参考数据库不可能涵盖所有情况，我们集成了一个大语言模型组件，用于处理未知包名。当数据库中找不到某个包名时，该工具会查询大语言模型，根据命名模式、语义线索以及大语言模型自身的知识，判断该包名是否可能代表第三方 SDK。经过验证的预测结果随后会被写回数据库，从而持续改进并扩展我们的 SDK 识别能力。

*3) RAG*增强的上下文整合：为提高用途推断的准确性，代理采用检索增强生成，将相关应用元数据与 API 调用分析相结合。代理维护一个知识库，其中包含应用商店描述和隐私政策片段，以便在 API 调用语义不足时提供上下文补充。

在分析权限使用时，代理首先从应用描述中检索相关上下文，以了解其主要功能，然后搜索隐私政策，查找现有的权限使用说明。代理将这些上下文信息与 API 调用分析相结合，从而全面推断用途。

提示示例 3：用途推断

系统： 你是一名安卓安全分析专家。**用户：** 请分析以下 API 调用信息、应用上下文和隐私政策声明，以推断实际的

p权限使用情况 ge:

API 调用信息：

- 权限：位置访问
- API 方法：getLastKnownLocation()
- 调用类：LocationTracker
- 调用栈：com.example...AdNetwork → ... → android.location...getLastKnownLocation
- 调用方类型：第三方 SDK（广告）

应用上下文： 该应用提供路线规划和交通更新.....

隐私政策声明： “我们使用位置数据来提供导航服务和路线规划。”

任务：

1) 根据提供的上下文推断此 API 调用的实际用途 2) 确定负责权限使用的实体 3) 提供支持推断的证据

响应格式：

- 推断用途：[实际使用场景]
- 调用方实体：[主应用/第三方 SDK]
- 证据：[支持该推断的 API 调用和上下文中的关键信息]

E. 自适应报告生成智能体

最终代理综合所有前序代理的分析结果，生成全面的评估报告，并提供针对性的优化建议。该代理是多智能体框架的集成节点，为开发者提供可行的见解，帮助他们改进权限说明实践。

1) 合规性评估：该代理根据前三个代理的结果评估整体合规状态：

- **可见性合规：** 在权限请求期间，解释内容是否存在且用户能否访问；
- **语言合规：** 解释文本是否符合清晰度、具体性和非操纵性方面的标准；
- **真实性合规：** 所述目的是否准确反映实际权限使用情况，包括是否披露第三方数据访问。

2) 针对性建议生成：根据已发现的合规缺口，该代理针对每个应用的具体问题生成相应建议。建议按合规层分类，包括：

- **内容改进：** 关于提升解释文本清晰度、替换含糊术语以及提供更具体使用说明的具体建议；
- **披露增强：** 关于解决第三方 SDK 数据使用披露不完整问题的建议，包括用于透明沟通数据共享情况的措辞建议；

该代理具有自适应特性，可确保建议与具体情境相关且切实可行，帮助开发者打造更加透明、符合监管要求的权限实践，最终增强用户信任并促进知情同意。

提示示例 4：自适应报告生成

系统： 你是一名合规分析专家。**用户：** 请根据以下分析结果生成一份合规性评估报告：**A**

nalysis Results:

- **解释提取结果：** [提取结果]
- **语言分析结果：** [语言分析]
- **目的推断结果：** [用途推断]

应用信息：

- 应用名称：[APP NAME]
- 权限类型：[PERMISSION TYPE]

任务：

1) 评估三层框架的整体合规状态 2) 针对识别出的每个问题生成针对性建议 3) 提供实施指南和优先级排序

响应格式：

- **合规摘要：** [可见性、语言和真实性各层的状态]
- **建议：** [按合规层分类的可执行改进措施]
- **实施优先级：** [建议改进措施的排序列表]

V. 评估

为评估我们提出的用于评估和优化权限说明的多智能体框架的有效性，我们使用实证研究中的 600 个真实世界移动应用数据集开展了全面实验。这些应用由领域专家进行人工标注，为评估提供了真实值。我们的实验考察了框架中各代理的准确性和性能，以及所生成优化建议的实际效用。

- **研究问题1：** 解释提取智能体在从应用截图中识别和提取权限说明文本方面的效果如何?
- **研究问题2：** 语言适切性智能体在语义分析和合规性评估方面的准确度如何?
- **研究问题3：** 用途推断代理在确定实际权限使用目的方面的精确度如何?
- **RQ4：** 我们的框架生成的优化建议是否会对用户的隐私感知和理解产生积极影响?

A. 研究问题*1*：解释文本提取效果

1）实验设置：我们评估了解释文本提取代理从应用程序界面中自动识别和提取权限说明文本的能力。实验比较了传统光学字符识别工具（Tesseract [33]）与多种热门多模态大语言模型（gpt-4-vision-preview、gpt-4o、gemini 2.5 flash、claude-3-sonnet）。

我们基于从数据集中收集的 100 张权限请求屏幕截图，对提取结果进行了评估。评估衡量了两个关键维度：(1) 使用权限匹配率（PMR）识别权限类型，该指标表示正确识别出的权限类型所占的比例；(2) 使用说明提取准确率（UEA）提取解释文本，该指标表示准确提取且与人工标注完全匹配的解释文本所占的比例。

2）结果：表 I 总结了不同方法的实验结果。在权限识别方面，所有方法都取得了近乎完美的表现，PMR 分数接近 100%，表明从应用截图中识别权限类型对于不同方法而言都是相对直接的任务。然而，在解释文本提取方面，各方法的性能出现了显著差异。Tesseract OCR 的提取能力存在明显限制，使用说明提取准确率仅为 50%，尤其难以处理深色背景和复杂的用户界面布局。相比之下，多模态大语言模型取得了显著更高的提取准确性，范围为 65% 至 98%。在多模态大语言模型中，gpt-4-vision-preview 的表现最佳，使用说明提取准确率达到 98%，其次是 gpt-4o（92%）、gemini 2.5 flash（82%）和 claude-3-sonnet（65%）。与其他模型相比，gpt-4 系列在处理复杂用户界面布局、多样化文本格式和低对比度场景方面表现出更优的性能。此外，多模态大语言模型相较于传统 OCR 的一个关键优势在于，它们能够通过整合视觉上下文与语义理解，直接识别与权限相关的解释性内容；而 OCR 则会不加区分地提取所有文本，并需要大量后处理才能筛选出相关内容。

表 I 解释文本提取方法的性能

方法	PMR	UEA	备注
Tesseract OCR	98%	50%	难以处理深色背景、复杂布局
gpt-4-vision-preview	100%	98%	难以处理深色背景、复杂布局 在各种用户界面设计
gpt-4o	100%	92%	表现出色，仅有轻微差异
gemini 2.5 flash	100%	82%	整体表现良好，偶尔上下文理解错误
claude-3-sonnet	100%	65%	难以应对 with 复杂的场景

研究问题1的答案：说明提取智能体展现出很高的有效性，多模态大语言模型的提取准确性最高可达98%。多模态大语言模型在各种用户界面设计中均表现出卓越性能，并且无需大量后处理即可直接识别权限相关内容，性能优于传统光学字符识别。

B. 研究问题*2*：语言适当性分析准确率

1) 评估设置：本实验评估了语言适当性分析代理在语义理解和合规违规检测方面的性能。我们将提出的基于大语言模型的少样本学习方法，与现有研究中常用的传统关键词匹配方法和基于规则的方法进行了比较。

关键词匹配和基于规则的基线采用了以下检测规则：模糊表达（例如“可能”“一些”“相关”）、胁迫或威胁性语言（例如“必须”“否则”“强制”）以及技术术语（例如“SDK”“国际移动设备识别码”“MAC 地址”）。我们通过加入结构化语义规则（包括句子分割策略和否定语境分析）来增强这一基线，以提高准确性并减少误分类。这种全面的比较有助于系统化评估不同的语言得体性评估方法。

我们使用包含多个示例及相应标签的少样本学习提示，对多种主流大语言模型（gpt-4o、claude-3-sonnet、Gemini 2.5 Pro、Grok 3、DeepSeek V3、doubao-1.5-pro-256k、qwen3-32b）进行了测试。评估数据集由我们实证研究中的100条权限说明文本构成，并经人工标注为“合规”（50条）或“不合规”（50条），同时注明具体违规类型。我们测量了三个关键指标：（1）评估一致性（EC）：模型判断与专家标注相符的比例；（2）假阳性率（FPR）：被错误标记为违规的合规文本所占比例；（3）假阴性率（FNR）：系统漏检的实际违规所占比例。

2) 结果：表 II 展示了不同方法的性能比较。结果显示，基于大语言模型的少样本学习与传统的基于规则的方法之间存在显著的性能差距。关键词匹配基线的整体效果较差，评估一致性仅为 57.5%，且 FNR（70%）和 FPR（33.3%）均较高。这表明，基于规则的方法在捕捉语言适切性违规的语义细微差别方面存在根本性局限，因为它们依赖表层模式匹配，而非语境理解。

在大语言模型中，qwen3-32b 的表现最佳（评估一致性为 93.5%），且错误率较为均衡（FPR 为 8.5%，FNR 为 4%）。不同模型呈现出不同的错误模式：gpt-4o 的召回率达到完美水平（FNR 为 0%），但假阳性率较高（23%）；而 gemini-2.5-pro 和 deepseek-v3 的精确率达到完美水平（FPR 为 0%），但假阴性率较高（30%–35%）。一项值得注意的发现是，不同模型呈现出系统性的权衡模式：精确率达到完美水平（FPR 为 0%）的模型，假阴性率始终较高；而召回率达到完美水平（FNR 为 0%）的模型，假阳性率则较高。这一模式反映了模型在处理边界案例时所采用的不同决策边界。通过针对特定应用需求优化精确率与召回率的平衡，对决策边界进行任务特定校准，或可提升模型在语言得体性评估中的表现。

表 II 语言适切性分析方法的性能

方法	EC (%)	FPR (%)	FNR (%)
关键词匹配与规则	57.5	33.3	70
qwen3-32b	93.5	8.5	4
gpt-4o	85	23	0
Gemini 2.5 Pro	85	0	30
DeepSeek V3	82.5	0	35
doubao-1.5-pro-256k	82.5	12	25
Grok 3	70	25	27.8
claude-3-sonnet	67.5	23	50

研究问题2的回答：语言适切性分析代理表现出较高的有效性，其中 qwen3-32b 的评估一致性达到 93.5%，gpt-4o 则达到 85%。基于大语言模型的少样本学习显著优于传统的关键词匹配；根据精确率与召回率权衡的不同，模型会呈现出不同的保守或激进评估策略。

C. 研究问题3：用途推断准确率

1）实验设置： 本实验通过分析结合应用上下文的API调用栈信息，评估用途推断代理识别应用实际权限使用意图的准确性。我们从数据集中选取了 50 个权限使用案例，涵盖不同的权限类型（位置、联系人、存储）和使用场景。领域专家对每个案例进行人工标注，记录实际使用目的和调用实体（主应用或第三方 SDK）。我们使用 gpt-4o 作为基础推断模型，将调用栈上下文与应用背景信息相结合。

我们开展了两组实验，以评估不同输入下的推断性能：（1）对照组：仅提供 API调用栈信息；（2）实验组：提供 API 调用信息以及包括应用商店描述和隐私政策在内的元数据。我们将用途推断准确率（UIA）定义为正确推断权限用途的比例。

2）结果：结果表明，我们的方法在权限用途推断方面有效。实验组的 UIA 达到 92%（46/50），而对照组为 76% （ 38/50 ），这说明，虽然仅凭 API 调用栈信息也能较为合理地识别用途，但加入应用上下文信息后，推断准确性会显著提升。当 API 调用栈提供的语义线索有限，或类名/方法名称含义模糊时，这一改进尤为重要。例如，在一个案例中，某应用通过通用的 BaseService.doTask() 方法调用了 getLastKnownLocation()，模型仅根据调用栈也能作出较为合理的推断；而在获得该应用的商店描述（“提供城市路线规划和实时公交导航”）后，则成功明确了具体用途。这些结果验证了基于 API 的用途推断的可行性，也证明了我们采用的上下文增强推理方法有助于实现高精确率的权限用途识别。

对研究问题3的回答：用途推断代理表现出较高的有效性：基于 API 调用的分析准确性达到 76%，而上下文增强推理则达到 92%。该方法成功实现了自动化用途识别，应用上下文信息进一步显著提升了识别效果。

D. RQ4：优化效果的用户研究

1）实验设置： 为评估我们框架提出的优化建议能否积极改善用户隐私感知，我们对 30 名参与者开展了用户研究，对比原始权限说明与优化后的权限说明。参与者看到 5 组权限说明（原始版本和优化版本），并被要求选择他们认为更清晰、更易理解的版本。我们采用被试内设计，每位参与者都评估所有组别，共进行了 150 次比较。研究既包含定量偏好测量，也通过开放式问题收集参与者对选择理由的定性反馈。

2）结果： 结果显示，在 150 次比较中，有 127 次（84.6%）更倾向于优化后的解释版本，而非原始版本。在个体层面，18 名参与者（60%）在全部 5 项任务中都选择了优化版本，而 27 名参与者（90%）在超过一半的任务中选择了优化版本。定性反馈也印证了这些定量研究结果。大多数参与者指出，优化后的解释措辞更加具体、语言更加简单，减少了技术术语和含糊表述，也更有助于理解权限用途和识别隐私风险。一些用户还提到，优化版本增强了他们对应用的信任，使他们更愿意做出知情授权决策。这些结果验证了我们的智能评估与优化框架在提升用户隐私感知和理解方面的有效性。

RQ4 的回答：优化建议显著提升了用户隐私感知，84.6% 的用户更偏好优化后的解释，而非原始解释。用户认为优化后的版本更加清晰、具体且值得信赖。

VI. 讨论

A. 可扩展性

虽然我们的研究采用手动测试以实现全面覆盖，但该框架只需将权限截图和 API 跟踪记录作为输入，即可无缝集成 DroidBot [34] 和 Fastbot [35] 等自动化用户界面探索工具，因此不受数据收集方法的限制。模块化架构支持对代理进行独立扩展和定制，以满足特定组织的需求。SCOPE 可集成到 CI/CD流水线中，以便在企业环境的开发过程中自动评估说明。

B. 启示

1）对于应用开发者： SCOPE 为开发者提供自动化评估和针对性建议，支持基于证据改进说明清晰度和法规合规性。该框架能够检测未披露的第三方 SDK 使用情况，这一点尤其有价值，可帮助开发者识别隐藏的合规违规问题。将其集成到开发工作流中，可以及早发现透明度问题，降低部署后风险，并支持敏捷开发流程中的持续合规。

2) 致隐私监管机构：我们的研究首次提供了系统性的实证证据，表明权限说明实践存在不足。SCOPE 展示了自动化合规性评估的可行性，为大规模应用组合提供了可扩展的监测能力。三层评估框架为标准化合规性测试和未来监管指南提供了结构化基础。我们发现第三方 SDK 披露失败现象普遍存在，这凸显了亟待通过更具体的监管要求解决的关键缺口。

C. 局限性与未来工作

1) 样本代表性与主观标注：我们对来自中国市场的 600 款应用进行的分析，只能代表更广泛移动生态系统中的有限样本。未来的研究应扩大范围，纳入长尾应用、特定地区应用以及不同开发者类别的应用。此外，所有标注和评估任务均由作者团队完成。尽管团队经验丰富，但由于缺少外部评审人员，仍可能无意中引入主观偏差。虽然我们通过作者间共识解决了分歧，但未来研究应通过纳入更多元的标注人员来解决这一局限，包括外部专家或众包工作者，以提升结果有效性和可推广性。

2）混淆与推断挑战：我们的用途推断依赖于对 API 调用进行语义分析，而代码混淆技术会将有意义的方法名称替换为无意义的字符串，因此这种推断容易受到影响。未来的工作应探索行为分析技术，重点关注 API 调用模式、数据流分析、周围的代码函数以及注释等辅助线索，从而更好地解释权限用途。动态加载和基于反射的 API 访问也需要更为复杂的分析技术。

3）框架优化与可扩展性：语言适切性和合规性评估等组成部分涉及细致且依赖语境的标准，难以完全操作化。未来的工作可通过采用自洽性或多步推理，或集成多个模型以提升稳定性、降低方差，从而缓解这一问题。当前的评估结果可能受到特定提示设计和大语言模型选择的影响，进而影响该框架在实践中的有效性。考虑到实际部署中的经济限制，未来将探索较小的开源模型，以及针对权限说明分析任务定制的领域专用微调模型。

VII. 相关工作

移动应用的数据透明度机制。数据透明度对于建立用户信任和促进知情决策至关重要。多个利益相关方推广了隐私政策以外的各种隐私工具，包括苹果的隐私营养标签 [36], 应用跟踪透明度 [37], 应用隐私报告 [38],谷歌的数据安全板块 [39], 隐私仪表板 [40],以及权限说明。然而，越来越多的研究指出，这些工具在支持知情隐私决策方面存在显著局限 [17,32,41]–[46]。研究表明，现有机制往往无法提供可操作信息，实施实践不佳，或缺乏系统化评估。我们的工作通过揭示权限说明中普遍存在的缺陷，并提供首个全面评估其实际实施质量的框架，补充了这些透明度机制。

权限请求中的暗黑模式。先前的研究发现，权限界面中存在被称为“暗黑模式”的操纵性设计策略。Gray 等人 [47] 将暗黑模式分为纠缠、阻碍和强制操作，这些模式通过重复出现的模态窗口或隐藏的退出选项表现出来。其他例子包括功能捆绑，即除非用户授予权限，否则核心功能会受到限制 [22],；以及隐私诱导，即同意信息被刻意隐藏 [21]。与本文最相关的是 Mohamed 等人 [17],的研究，他们发现现实中的权限提示采用了误导性、含糊不清或带有诱导性的语言，使用户感到困惑，并扭曲了他们对权限影响的理解。我们的工作将这一暗黑模式研究方向进一步拓展至权限说明，开发了一种自动化框架，能够系统地识别操纵性语言模式，并提供针对性建议，以减轻此类欺骗性做法。

VIII. 结论

我们对 600 款主流移动应用进行的实证分析揭示了权限说明实践中普遍存在的不足，包括缺少解释、语言不透明性，以及系统性地遗漏第三方 SDK 数据访问信息。这些问题违反了监管透明度要求，并损害了知情同意。我们提出了 SCOPE，这是一种用于自动化解释评估与优化的多智能体框架，融合了多模态大语言模型、少样本学习、推理和规划。评估结果表明，各个代理均表现良好。用户研究显示，84.6% 的用户更偏好经 SCOPE 优化的解释，验证了其实际效用。SCOPE 弥合了监管意图与实际实施之间的差距，为提升移动应用隐私透明度提供了可扩展的解决方案。

致谢

本研究部分由中国国家自然科学基金（项目编号：62572209、62172049）和湖北省重点研发计划（项目编号：2025BAB057）资助。

参考文献

[1] 美国网络安全和基础设施安全局（CISA），“隐私与移动设备应用程序”，https://www.cisa.gov/news- events/news/privacy-and-mobile-device-apps，2022年。[2] R. Bateman，“安卓数据收集与敏感数据”，https://www.termsfeed.com/blog/android-sensitive-data-collection/，2025年。[3] A. Developers，“安卓上的权限”，https://developer.android.com/guide/topics/permissions/overview，2025年。[4] “贪婪应用程序收集的信息超出应有范围”，https://betanews.com/2023/10/27/greedy-apps-collect-more-information-than-they-should/，2023年。[5] “安卓金融应用程序对权限的索求过于贪婪”，https://cybernews.com/security/android-financial-apps-greedy-for-permissions/，2023年。[6] K. Bongard-Blanchy、J.-L. Sterckx、A. Rossi、V. Distler、S. Rivas 和 V. Koenig，“一种（不）必要之恶——用户对智能手机应用权限的（不）确定性及其对隐私工程的影响”，载于《2022年IEEE欧洲安全与隐私研讨会专题讨论会》（EuroS&PW）。IEEE，2022年，第01–08页。[7] P. Wijesekera、A. Baokar、A. Hosseini、S. Egelman、D. Wagner 和 K. Beznosov，“重新审视安卓权限：一项关于情境完整性的实地研究”，载于《第24届USENIX安全研讨会》（USENIX Security 15）。华盛顿哥伦比亚特区：USENIX协会，2015年，第499–514页。 [在线]。获取地址：https://www.usenix.org/conference/usenixsecurity15/technical-sessions/presentation/wijesekera [8] “安卓应用中的第三方库权限搭便车”，https://blog.appicaptor.com/2025/02/27/third-party-library-permission-piggybacking-in-android-apps/，2025年。[9] K. Heid、E. J. Sonntag 和 J. Heider，“重新审视安卓应用中第三方库的权限搭便车问题”，2025年。 [在线]。获取地址：https://publica.fraunhofer.de/handle/publica/486121

[10] 欧盟委员会，“通用数据保护条例”，https://gdpr.eu/tag/ gdpr/，2025年。[11] 加利福尼亚州司法部长办公室（OAG），“加利福尼亚消费者隐私法”，https://oag.ca.gov/privacy/ ccpa，2025年。[12] DigiChina，“中华人民共和国个人信息保护法”，https://digichina.stanford.edu/work/translation-personal-information-protection-law-of-the-peoples-republic-of-china-effective-nov-1-2021/，2025年。[13] 皮尤研究中心，“美国人对隐私政策与法律的态度和体验”，https://www.pewresearch.org/internet/2019/11/15/americans-attitudes-and-experiences-with-privacy-policies-and-laws/，2019年。[14] Syrenis，“隐私政策：有人在读吗？”，https://syrenis.com/resources/blog/privacy-policies-is-anyone-reading-them/，2024年。[15] W. Cao、C. Xia、S. T. Peddinti、D. Lie、N. Taft 和 L. M. Austin，“安卓权限下用户行为、预期和参与度的大规模研究”，载于《第30届USENIX安全研讨会》（USENIX Security 21）。USENIX协会，2021年，第803–820页。 [在线]。获取地址：https://www.usenix.org/conference/usenixsecurity21/presentation/cao-weicheng [16] M. Rosala，“有效移动应用程序权限请求的三项设计考量”，https://www.nngroup.com/articles/permission-requests/，2019年。[17] R. Mohamed、A. Arunasalam、H. Farrukh、J. Tong、A. Bianchi 和 Z. B. Celik，“请注意！对应用跟踪透明度权限的调查”，载于《第33届USENIX安全研讨会》（USENIX Security 2024），2024年，第5017–5034页。[18] S. Hoober，“移动端暗黑模式”，https://www.uxmatters.com/mt/archives/2019/11/mobile-dark-patterns.php，2019年。[19] “制品可用性”，https://github.com/dangyssss/SCOPEAgent，2025年。[20] A. Developers，“请求运行时权限”，https://developer.android.com/training/permissions/requesting?utm source=chatgpt. com&hl=en，2025年。[21] C. B¨osch、B. Erb、F. Kargl、H. Kopp 和 S. Pfattheicher，“暗黑面的故事：隐私暗黑策略与隐私暗黑模式”，《隐私增强技术论文集》，第2016卷，第237–254页，2016年7月。[22] L. Di Geronimo、L. Braz、E. Fregnan、F. Palomba 和 A. Bacchelli，“用户界面暗黑模式及其踪迹：一项关于移动应用程序与用户感知的研究”，载于《2020年计算机系统中的人为因素CHI会议论文集》，2020年，第1–14页。[23] J. Gunawan、A. Pradeep、D. Choffnes、W. Hartzog 和 C. Wilson，“网页与移动端形态中暗黑模式的比较研究”，《ACM人机交互论文集》，第5卷，第CSCW2期，2021年10月。[在线]。获取地址：https://doi.org/10.1145/3479521

[24] 中国国家标准化管理委员会网络安全技术委员会，“移动互联网应用程序系统权限使用指南”，https://www.tc260.org.cn/front/postDetail.html?id = 20200918163359 2020 ，。 –

[25] 维基百科，“德尔菲法”，https://en.wikipedia.org/wiki/Delphi method，2025年。[26] C. Okoli 和 S. Pawlowski，“德尔菲法作为研究工具：一个示例、设计考量与应用”，《信息与管理》，第42卷，第15–29页，2004年12月。[27] D. Varona 和 L. F. Capretz，“使用德尔菲法构建软件行业角色分配模型”，载于 IECON 2021——IEEE工业电子学会第47届年度会议。IEEE，2021年10月，第1–7页。 [在线]。可访问：http: //dx.doi.org/10 1109 48115 2021 9589957 。/IECON。。E.

[28] Głuszek，“使用电子德尔菲法验证企业声誉管理成熟度模型（cr3m）”，《可持续发展》，第13卷，第21期，2021年。 [在线]。可访问：https://www.mdpi.com/2071-1050/ 13/21/12019[29] Frida，“Frida：面向开发者、逆向工程师和安全研究人员的动态插桩工具包。”https://frida.re/docs/home/，2025年。[30] OpenAI，“OpenAI Agents SDK”，https://openai.github.io/openai-agents- python/，2025年。[31] S. Yang、Y. Wang、Y. Yao、H. Wang、Y. Ye 和 X. Xiao，“DescribeCtx：移动应用中敏感行为的上下文感知描述合成”，载于第44届国际软件工程大会（ICSE 2022）论文集，2022年，第685–697页。[32] L. Wang、D. Wang、S. Pan、Z. Jiang、H. Wang 和 Y. Wang，“向前迈出一大步？以用户为中心考察iOS应用隐私报告及其改进”，载于2025年IEEE安全与隐私研讨会（SP）。美国加利福尼亚州洛斯阿拉米托斯：IEEE计算机学会，2025年，第4210–4228页。 [在线]。可访问：https: //doi.ieeecomputersociety.org/10 1109 61157 2025 00223 。/SP。。“

[33] Tesseract OCR”，https://github.com/tesseract-ocr/tesseract，2025年。[34] DroidBot，“面向安卓的轻量级测试输入生成器。”https://github.com/honeynet/droidbot，2023年。[35] 字节跳动，“Fastbot-Android开源手册”，https://github. com/bytedance/Fastbot Android，2023年。[36] A. Inc.，“透明是最佳政策”，https://www.apple.com/hk/en/privacy/labels/，2020年。[37] ——，“如果应用请求追踪你的活动”，https://support.apple.com/ en-hk/102420，2021年。[38] ——，“关于应用隐私报告”，https://support.apple.com/en-sg/ 102188，2021年。[39] G. Play，“为Google Play的数据安全板块提供信息”，https://support.google.com/googleplay/android-developer/answer/10787469?hl=en，2022年。[40] ——，“通过隐私仪表板管理权限”，https://support.google.com/android/answer/13530434?visit id 638526274467900320 3285927032 = - &p privacy dashboard&rd=1，2022年。[41] S. Zhang、Y. Feng、Y. Yao、L. F. Cranor 和 N. Sadeh，“iOS应用隐私标签的可用性如何？”，《隐私增强技术论文集》，2022年。[42] S. Koch、M. Wessels、B. Altpeter、M. Olvermann 和 M. Johns，“保持隐私标签诚实”，《隐私增强技术论文集》，2022年。[43] A. DeGiulio，“‘请求应用不要追踪’：主动选择追踪授权对移动隐私的影响”。[44] X. Wu、L. Hu、E. Zeng、H. Habib 和 L. Bauer，“透明度还是信息过载？评估用户对iOS应用隐私报告的理解和看法”，载于网络与分布式系统安全研讨会，2025年。[45] R. Khandelwal、A. Nayak、P. Chung 和 K. Fawaz，“解读隐私标签：从测量和开发者视角看谷歌的数据安全板块”，载于第33届USENIX安全研讨会（USENIX Security 24），2024年，第2831–2848页。[46] R. Baalous、A. Althobaiti、D. Alyoubi、R. Alzahrani 和 M. Aljohani，“使用静态分析检测安卓应用数据收集与Google Play数据安全板块之间的不一致”，《控制论与信息技术》，第25卷，第1期，2025年。[47] C. M. Gray、Y. Kou、B. Battles、J. Hoggatt 和 A. L. Toombs，“用户体验设计中黑暗（模式）的一面”，载于2018年计算机系统人因国际会议论文集，系列：CHI ’18。美国纽约州纽约：计算机协会，2018年，第1–14页。 [在线]。可访问：https://doi.org/10.1145/3173574.3174108