UNIVERSIDAD POLITÉCNICA DE MADRID
Escuela Técnica Superior de Ingenieros de Telecomunicación

Contributions to the Automated Assessment of Mobile Applications' Compliance with Privacy and Data Protection Requirements

D O C T O R A L T H E S I S

Submitted for the degree of Doctor by:

David Rodríguez Torrado
Master's Degree in Cybersecurity

Madrid, 2025


UNIVERSIDAD POLITÉCNICA DE MADRID
Escuela Técnica Superior de Ingenieros de Telecomunicación

Doctoral Degree in Telematics Systems Engineering

Contributions to the Automated Assessment of Mobile Applications' Compliance with Privacy and Data Protection Requirements

D O C T O R A L T H E S I S

Submitted for the degree of Doctor by:

David Rodríguez Torrado
Master's Degree in Cybersecurity

Under the supervision of:
Dr. José María del Álamo Ramiro

Madrid, 2025

Title: Contributions to the Automated Assessment of Mobile Applications' Compliance with Privacy and Data Protection Requirements

Author: David Rodríguez Torrado

Doctoral Programme: Doctoral degree in Telematics Systems Engineering

Thesis Supervision:

Dr. José María del Álamo Ramiro, Associate professor at Escuela Técnica Superior de Ingenieros de Telecomunicación - UPM (Supervisor)

External Reviewers:

Thesis Defense Committee:

Thesis Defense Date:

This thesis has been partially supported by the **AutoGDPR project** (TED2021-130455A-I00) funded by MCIN/AEI/10.13039/501100011033 and by the European Union "NextGenerationEU"/PRTR. It has also been supported by the **CEDAR project**, funded by the Horizon Europe research program (2021-2027) under grant agreement no. 101135577, and by the **PRESECREL project**, funded by the Plan Estatal de Investigación Científica y Técnica y de Innovación 2017-2020 (Ministerio de Ciencia e Investigación (Spain) - MCIN/AEI/10.13039/501100011033) under Grant agreement PID2021124502OB-C43. Reseach conducted in collaboration with international institutions during this thesis have been partially supported by **Programa Propio UPM** under the call "Convocatoria de ayudas dirigidas al personal investigador en formación predoctoral para realizar una estancia de investigación internacional igual o superior a tres meses", as well as supported by **Escuela Técnica Superior de Ingenieros de Telecomunicación (ETSIT)** under the call "Ayudas

de la ETSIT-UPM para la publicación de artículos en revistas, congresos, primeros proyectos, estancias predoctorales y movilidad".

Dedicado a todos aquellos que me habéis enseñado lo que sé, a los que habéis hecho que sea quien soy hoy. Y en especial a ese chico cabezota que confió en sí mismo y siempre tuvo fe.

Acknowledgement

A Chema, por haber sido mi mentor y guía profesional, por permitirme desarrollarme hasta llegar donde estoy hoy. Aunque existen muchos libros de *desarrollo personal*, ninguno explica con tanto ahínco el Principio de Pareto como has hecho tú durante estos años: "*El 80% de los resultados proceden del 20% del esfuerzo*". Esta frase siempre ha entrado en conflicto con mi naturaleza perfeccionista, pero ha sido clave para estar aquí hoy. Por ello, y con el objetivo de no menoscabar el principio en el último momento, advierto al lector que esta tesis podrá contener hasta un 20% de erratas y/o errores; de lo cual me enorgullezco. Gracias, Chema, por haberme dado la oportunidad y por tu confianza en todo momento.

A mi pareja, por haberme dado una de las lecciones de vida más importantes. Por enseñarme que la actitud lo es todo. Sin ti, este viaje no habría sido lo mismo. No puedo resumir en tan pocas palabras tu contribución, simplemente, gracias. A mi familia, por el apoyo y ayuda incondicional, por haberos preocupado y cuidado siempre de mí, y por enseñarme con el ejemplo. Esto incluye a mi sobrina, la nueva integrante de la familia, gracias por habernos traído tanta felicidad. Os quiero a todos.

A mis amigos que, lo sepan o no, todos han contribuido en esta tesis. Por privacidad, no pondré nombres, pero quien se dé por aludido con estas líneas, no se equivoca. Cada uno ha aportado de diferente manera, y entre todos me han hecho mejor. Pocos o ninguno leerán esto, pero les haré saber que me importan, como vengo haciendo desde hace unos años. Espero estar ahí para vosotros de la misma forma que habéis estado. Gracias en especial a los que habéis compartido conmigo momentos difíciles; habéis sido un apoyo clave.

A todos mis compañeros de STRAST, de la ETSIT y de la UPM. Los que están y los que estuvieron. De todos ellos he aprendido mucho y he compartido momentos muy variados, algunos de agobio, estrés o cansancio y otros de risas. Gracias por los grandes consejos. Sois una fuente inagotable de conocimiento y ejemplo. Romperé la privacidad de Celia (lo siento). Tu actitud altruista, desinteresada y siempre colaborativa es digna de mención. Me has enseñado que el trabajo debe ir de la mano del impacto, y en eso eres una experta.

A Norman y a William (y a Jose), que aunque leyeran esto, probablemente no lo entenderían. Gracias por abrirme las puertas a vuestra casa y por haberme guiado. Me habéis permitido crecer y vivir experiencias únicas. Gracias, Joe, por tener siempre buenas palabras, y por haberme enseñado tanto en tan poco tiempo trabajando contigo.

Gracias a todos los que habéis contribuido, a sabiendas o no, durante el transcurso de esta tesis. Espero haberos aportado algo de vuelta. Y por último, gracias a mí mismo, por el esfuerzo y dedicación, por el trabajo duro y la resiliencia durante estos años.

Abstract

Smartphones have seen widespread adoption in society, largely due to their Internet connectivity and their ability to perform a wide range of activities—from work-related tasks like document creation and editing to multimedia consumption. However, while their capabilities have been significantly enhanced by the advent and proliferation of mobile applications, they also introduce risks for users. The numerous sensors integrated into these devices and their extensive connectivity enable the collection and transfer of vast amounts of personal data, making it possible to identify users and track their behavior, movements, and usage patterns. This practice, mainly driven by business models based on hyper-targeted advertising, poses a considerable risk to privacy. In response to this and other challenges in our digital society, the European Union enacted the General Data Protection Regulation (GDPR) to promote greater protection and handling of personal data.

Evaluating the regulatory compliance of mobile applications with the GDPR is a formidable challenge, particularly in fast-paced environments like the Google Play Store, where thousands of apps are published and updated daily. The sheer volume and frequency of these changes make manual inspection impractical, and the widespread use of third-party libraries further compounds this challenge. Although these libraries allow developers to rapidly integrate new functionalities, they often collect and transfer user data without the developers' awareness.

Within the framework of the AutoGDPR project, which the Spanish government funded with European funds, this thesis contributes a set of methods and artifacts that allow for the comprehensive analysis of mobile app behavior. The approach involves the use of both static and dynamic techniques—employing tools such as Mitmproxy and Frida—to assess app behavior, as well as the automation of privacy policy and privacy label processing through natural language processing techniques and large language models. Moreover, GDPR requirements have necessarily been translated into programmable rules that enable automatic evaluations, with frequent collaboration with legal experts to ensure accuracy.

The developed methods and artifacts have been integrated into a modular platform based on technologies like Docker and RabbitMQ, facilitating large-scale studies and the extraction of empirical evidence. The results from these studies reveal widespread non-compliance with transparency obligations: more than 80% of the analyzed apps potentially fail to meet GDPR requirements regarding the disclosure of data transfers to third parties, and significant shortcomings exist in the privacy configurations of third-party libraries, which are responsible for over 70% of undeclared data transfers. Additionally, about 50% of privacy policies do not correctly state the data retention periods, and 48% of those transferring data do so to servers outside the European Union without proper declaration, as required by the GDPR. Furthermore, studies have shown that privacy labels often do not match the actual behavior of the apps or the information provided in their privacy policies. All this demonstrates that the current mobile application ecosystem is misaligned with data protection regulations, underscoring the need for authorities to employ automated monitoring and review mechanisms and for developers to have tools that help them comply with these regulations.

This research has resulted in a total of 12 scientific publications. Seven articles are a direct

outcome of this thesis—four published in journals indexed in the Journal Citation Report (with one in the first quartile Q1 and three in the second Q2), and two in international peer-reviewed events, including a notable publication at the *Privacy Enhancing Technologies Symposium* (PETS), one of the most prestigious congresses in privacy research. Additionally, the developed methods and artifacts have indirectly contributed to five further publications—two in international peer-reviewed events, two in JCR-listed Q1 journals, and one in a Q2 journal currently under review.

International collaborations have played a significant role in this work. Research stays of three months each were carried out at renowned institutions such as Carnegie Mellon University and King's College London, along with additional collaborations with research groups at ETH Zurich. In the industrial realm, the artifacts have been employed in regulatory compliance audits, demonstrating their practical utility. Regulatory bodies, including members from the Federal Trade Commission (FTC) in the United States and the Spanish Data Protection Agency (AEPD), have expressed interest in these findings and tools, highlighting their potential to enhance supervision and enforcement of data protection laws.

Moreover, some of the research findings have reached a wider audience through major Spanish media outlets such as La Vanguardia, Computer Hoy, La COPE, and TreceTV, raising public awareness about the importance of privacy in our digital age. The thesis also includes the direct supervision of five final degree projects and the provision of technical support to other research initiatives, which have helped advance the automation of regulatory compliance evaluation and contributed to the education of new talent in the field of data protection.

Overall, this doctoral thesis lays the groundwork for the automated evaluation of regulatory compliance in mobile applications by providing tools that foster a more transparent digital ecosystem, aligned with data protection laws. Future research will extend this approach to other platforms such as iOS, enable multilingual assessments of privacy policies, and apply these methods and knowledge to analyzing usage policies in customized chatbots, thereby addressing emerging challenges in an ever-evolving digital landscape.

Resumen

Los teléfonos móviles, y particularmente los inteligentes, han tenido una gran adopción en la sociedad, en especial gracias a su conectividad a internet y capacidad para realizar un amplio abanico de actividades, desde tareas orientadas al trabajo, como la creación y modificación de documentos, hasta el consumo multimedia. Sin embargo, aunque sus capacidades se han visto especialmente aumentadas tras la aparición y proliferación de las aplicaciones, también conllevan riesgos para los usuarios. La cantidad de sensores que integran estos dispositivos y su gran conectividad, permite que se recolecte y envíe un gran volumen de datos personales, posibilitando identificar a los usuarios, conocer su comportamiento, movimientos y patrones de uso. Esta práctica, impulsada en gran medida por los modelos de negocio basados en la publicidad hipersegmentada, supone un riesgo considerable para la privacidad. En respuesta a este problema ya presente y en auge en nuestra sociedad digital, la Unión Europea presentó el Reglamento General de Protección de Datos (RGPD) con el fin de promover una mayor protección y un adecuado tratamiento de los datos personales.

No obstante, la evaluación del cumplimiento normativo de aplicaciones móviles conforme al RGPD representa un desafío significativo, especialmente en un ecosistema marcado por el dinamismo de sus plataformas de distribución, como Google Play Store, donde miles de ellas son publicadas y actualizadas a diario. Este gran volumen y frecuencia de cambios hacen impracticable la inspección manual de todas las aplicaciones, subrayando la necesidad de métodos, técnicas y herramientas automatizados que permitan abordar esta tarea a escala. Además, este problema se ve agravado por el uso de bibliotecas de código de terceros, que permiten integrar funcionalidades de forma rápida y efectiva, pero que a menudo recopilan y transfieren datos de los usuarios, hecho que a menudo ocurre inadvertido para los desarrolladores y responsables de las aplicaciones.

En respuesta a esta problemática, y en el marco del proyecto AutoGDPR—financiado por el Gobierno de España y centrado en la automatización de la evaluación del cumplimiento del RGPD—, esta tesis contribuye con el diseño y desarrollo de métodos y artefactos. Estos permiten, de forma conjunta, 1) analizar el comportamiento de las aplicaciones, 2) automatizar el procesamiento de políticas y etiquetas de privacidad mediante técnicas de procesamiento de lenguaje natural, y 3) traducir los requisitos del RGPD en reglas programables que posibilitan evaluaciones automáticas. Para abordar este último punto se ha requerido, además, una frecuente colaboración con abogados expertos en protección de datos, destacando la faceta multidisciplinar de este trabajo.

El análisis del comportamiento de las aplicaciones se ha llevado a cabo mediante técnicas de análisis estático y dinámico de aplicaciones, apoyado parcialmente por herramientas de código abierto como Mitmproxy o Frida. Otras tecnologías, como los modelos de lenguaje de gran tamaño, han permitido identificar y extraer prácticas más complejas descritas en los textos legales, permitiendo evaluar el cumplimiento de requisitos de RGPD como las transferencias internacionales de datos personales, o transferencias a terceras organizaciones. Finalmente, los artefactos y métodos desarrollados se han integrado en una plataforma modular basada en tecnologías como Docker y RabbitMQ, que han propiciado realizar estudios con un gran volumen de aplicaciones y extraer conclusiones basadas en evidencia empírica.

Los resultados obtenidos con la plataforma y los artefactos desarrollados muestran incumplimientos generalizados en las obligaciones de transparencia de los responsables de las aplicaciones. Más del 80% de las aplicaciones analizadas potencialmente incumplen con los requisitos de transparencia del RGPD respecto a la cesión de datos a terceros y se identificaron deficiencias significativas en la configuración de privacidad de bibliotecas de terceros, responsables de más del 70% de las transferencias no declaradas. La falta de transparencia en las políticas de privacidad también se observó en otras prácticas evaluadas; por ejemplo, el 50% de las políticas no declaran correctamente el periodo de retención de datos personales, y el 48% de las que envían datos personales lo hacen a servidores en países fuera de la Unión Europea sin declararlo como exige el RGPD. Aunque otros mecanismos como las etiquetas de privacidad han surgido como alternativa para declarar el comportamiento de las aplicaciones relativas a la recolección y tratamiento de datos personales, en los estudios realizados durante esta tesis se observó que estos a menudo no coincidían con el comportamiento real de las aplicaciones ni con las declaraciones de las políticas de privacidad. Todo ello evidencia que el ecosistema de aplicaciones móviles actual está desalineado con las normativas de protección de datos y la necesidad de que las autoridades cuenten con mecanismos automatizados de control y de revisión y los desarrolladores cuenten con herramientas que les ayuden a cumplir con la regulación.

En consecuencia, los hallazgos expuestos y otros obtenidos a lo largo de esta investigación han culminado en la elaboración de 12 artículos científicos. De ellos, 7 artículos son resultado directo de esta tesis y han sido publicados en revistas, conferencias de alto impacto y otros eventos internacionales, incluyendo 4 en revistas indexadas en el Journal Citation Report (JCR). Una de ellas se encuentra dentro del primer cuartil (Q1) del JCR, tres en el segundo cuartil (Q2), y dos en eventos internacionales con revisión por pares. Además, destaca también una publicación en *Privacy Enhancing Technologies Symposium* (PETS), una de las conferencias más prestigiosas en el campo de la privacidad. Como resultado indirecto de los métodos y artefactos desarrollados, esta tesis ha contribuido a producir otros 5 artículos; de los cuales dos han sido publicados en eventos internacionales con revisión por pares, otros dos en revistas JCR pertenecientes a Q1 y un artículo Q2 en proceso de revisión.

Durante la realización de la tesis también se han llevado a cabo colaboraciones internacionales respaldadas por dos estancias de investigación de tres meses cada una en instituciones de prestigio como Carnegie Mellon University y King's College London, y dos colaboraciones adicionales con dos grupos distintos de ETH Zurich. En el ámbito industrial, los artefactos desarrollados han sido utilizados para auditorías de cumplimiento normativo, demostrando su aplicabilidad práctica. Asimismo, miembros de la Federal Trade Commission (FTC) de Estados Unidos (agencia de protección al consumidor y la competencia) y la Agencia Española de Protección de Datos (AEPD) han mostrado interés en estos artefactos y en los hallazgos de la tesis, subrayando su potencial para mejorar la supervisión y el cumplimiento de la normativa de protección de datos.

Los hallazgos también han sido divulgados a un público más general en algunos medios de comunicación relevantes en España, como el periódico La Vanguardia, Computer Hoy, La COPE y TreceTV, sensibilizando a la sociedad sobre la importancia de la privacidad en el entorno digital. Además, en el ámbito académico, se ha llevado a cabo la supervisión directa

de cinco trabajos de fin de titulación en esta misma línea de investigación, así como se ha dado soporte técnico a otros trabajos. Esto ha permitido avanzar en la automatización de la evaluación de cumplimiento normativo, mientras se forma a estudiantes en la temática de protección de datos.

Esta tesis doctoral sienta las bases para la evaluación automatizada del cumplimiento normativo en aplicaciones móviles, aportando herramientas que fomentan un ecosistema digital más transparente y alineado con la ley de protección de datos. Las líneas futuras de investigación incluyen la ampliación del enfoque a ecosistemas como iOS, la evaluación multilingüe de políticas de privacidad y la aplicación de los métodos desarrollados al análisis de políticas de uso en chatbots personalizados, abordando nuevos retos en un contexto digital en constante evolución.

Table of Contents

List of Figures

Abbreviations and acronyms

AEPD Agencia Española de Protección de Datos (Spanish Data Protection Agency)

ACM Association for Computing Machinery

ADB Android Debug Bridge

AdID Advertising Identifier

AI Artificial Intelligence

APK Android Package (Application Package)

API Application Programming Interface

AutoGDPR Automated GDPR Compliance Project

BCR Binding Corporate Rules

CCPA California Consumer Privacy Act

CNIL Commission Nationale de l'Informatique et des Libertés (French Data Protection Authority)

CPRA California Privacy Rights Act

CSV Comma-Separated Values

DNT Do Not Track

DPAs Data Protection Authorities

DSR Design Science Research

EEA European Economic Area

EECTI Estrategia Española de Ciencia, Tecnología e Innovación (Spanish Strategy for Science, Technology, and Innovation)

EIC European Innovation Council

EM2i Estrategia Madrileña de Investigación e innovación (Madrid 2030 Research and Innovation Strategy)

ERA European Research Area

EU European Union

FTC Federal Trade Commission

FWCI Field-Weighted Citation Impact

GDPR General Data Protection Regulation

HTTP HyperText Transfer Protocol

HTTPS Hypertext Transfer Protocol Secure

ICT Information and Communication Technologies

IEEE Institute of Electrical and Electronics Engineers

IP Internet Protocol

IWPE International Workshop on Privacy Engineering

JCR Journal Citation Report

LGPD Lei Geral de Proteção de Dados (Brazilian General Data Protection Law)

LLMs Large Language Models

MITM Man-in-the-Middle (attack or proxy)

MiTM Man-in-the-Middle

ML Machine Learning

MSCA Marie Skłodowska-Curie Actions

NLP Natural Language Processing

OS Operating System

PEICTI Plan Estatal de Investigación Científica y Técnica y de Innovación (State Plan for Scientific, Technical, and Innovation Research)

PETS Privacy Enhancing Technologies Symposium

PII Personally Identifiable Information

PIPL Personal Information Protection Law (China)

PoPETs Proceedings on Privacy Enhancing Technologies

PRICIT Plan Regional de Investigación Científica e Innovación Tecnológica (Regional Plan for Scientific Research and Technological Innovation)

R+D+i Research, Development, and Innovation

ROI Receiver Organization Identifier

S3 Estrategia de Especialización Inteligente (Smart Specialization Strategy)

SCC Standard Contractual Clauses

SDK Software Development Kit

SDGs Sustainable Development Goals

SECRYPT International Conference on Security and Cryptography

SECTI Sistema Español de Ciencia, Tecnología e Innovación (Spanish Science, Technology, and Innovation System)

SSL Secure Sockets Layer

SSIM Structural Similarity Index

SSRN Social Science Research Network

SVM Support Vector Machine (a type of ML classifier)

UPM Universidad Politécnica de Madrid

USENIX Advanced Computing Systems Association

Chapter 1

Introduction

1.1 Context and Motivation

The exponential adoption of mobile phones (Al-Sharafi et al., 2024), initially driven by their ability to combine basic communication functions, has evolved dramatically through the incorporation of sensors, internet access, and the development of mobile applications distributed via digital marketplaces. This ecosystem led to their rebranding as "smartphones"—devices that fundamentally changed how users interact with technology by offering personalized services and constant connectivity. Operating systems such as Google's Android and Apple's iOS have come to dominate this market (GlobalStats, 2023), establishing a global infrastructure that enables users to access a wide range of applications for everyday tasks, entertainment, and communication.

A significant factor driving the proliferation of mobile applications is the prevalent business model that offers apps free of charge to users, monetizing instead through the collection and utilization of user data. This model often involves in-app advertising, where developers generate revenue by displaying targeted ads based on user information. To accumulate a sizeable user base and gather information on the people interacting with their app, developers often sort and sell this data to app publishers who pay to place targeted ads within the app. While this approach allows users to access a multitude of applications without direct financial cost, it necessitates the extensive collection and processing of personal data (Sipior et al., 2014). This data is often shared with third-party advertisers and analytics companies, creating intricate and complex networks of data exchange that can operate beyond the user's and, oftentimes, the developers' awareness and control.

The extensive collection and sharing of personal data inherent in these business models have led to significant privacy risks (Ketelaar & van Balen, 2017). High-profile data breaches have exposed the personal information of millions of users, leading to financial losses, identity theft, and erosion of trust in digital services. For instance, the 2013 Yahoo data breach compromised over one billion accounts, making it one of the largest breaches in history (Trautman & Ormerod, 2017). Similarly, the 2018 Marriott International breach affected more than 300 million guests, exposing sensitive information such as passport numbers and payment card

details (Federal Trade Commission, 2024). A more recent example is Amazon, currently facing a class-action lawsuit alleging that it covertly tracks user movements via its Amazon Ads SDK (Stempel, 2025). Filed in January 2025 in a federal court in San Francisco, the suit argues that this practice facilitates the unauthorized collection of geolocation and other sensitive data without explicit consent.

These incidents underscore the potential consequences of inadequate data protection measures and highlight the critical importance of strong privacy practices. In response to these—and other digital—challenges, the European Union enacted the General Data Protection Regulation (GDPR) in 2016 (European Union, 2016a), providing a regulatory framework designed to promote transparency, user control, and the protection of personal information. The GDPR has influenced similar regulations around the world, prompting the adoption of similar measures in jurisdictions such as Brazil (LGPD), California (CCPA/CPRA), and China (PIPL) (Carrillo & Jackson, 2022; Eskhita & Stamhuis, 2024; Islam et al., 2022; Mahieu et al., 2021).

Beyond individual regulations, privacy and data protection have become strategic priorities for governments and institutions, aiming to foster digital trust and secure fundamental rights in an increasingly data-driven world. This is reflected in regional, national, and international policy agendas, where digitalization, artificial intelligence, and cybersecurity are positioned as significant roles for sustainable technological development. Within this broader regulatory and strategic landscape, initiatives such as the Madrid 2030 Research and Innovation Strategy (EM2i) (Comunidad de Madrid, 2020), the Spanish Strategy for Science, Technology, and Innovation (EECTI) (Ministerio de Ciencia e Innovación, 2020), and the Horizon Europe program (European Commission, 2020) highlight the importance of ensuring that digital transformation aligns with privacy protection principles. This direction also resonates with the United Nations Sustainable Development Goals (SDGs) (United Nations, 2015), particularly those focused on ensuring inclusive technologies, promoting peace, justice, and strong institutions (SDG 16) so that users can fully exercise their privacy and data protection rights, reinforcing the need for continued research and innovation in this domain.

This doctoral thesis is in line with these strategic priorities, addressing the pressing need for scalable, automated methods to evaluate privacy compliance in mobile applications. By developing innovative techniques for analyzing both the observed behavior of applications and their declared data practices, this research contributes to advancing transparency, accountability, and user protection in digital ecosystems. Ultimately, it seeks to bridge the gap between regulatory frameworks and their practical enforcement, ensuring that privacy regulations such as the GDPR are effectively upheld in a digital landscape.

1.2 Research Problem

One of the primary challenges in ensuring compliance with data protection regulations, such as the GDPR, is the scale and complexity of the mobile application ecosystem. Marketplaces like Google Play Store and the Apple App Store host millions of applications (Memon et al., 2024), each with unique data collection practices and privacy policies. The dynamic nature of these platforms—where applications are continuously updated and new ones are introduced

daily—makes manual compliance assessments impractical. This creates a pressing need for automated methodologies capable of evaluating their privacy practices at scale.

Article 13 of the GDPR (European Union, 2016b), anchored in the principle of transparency—one of the Regulation's seven core principles—plays a particular role in this context. It requires data controllers, the organizations deciding the data to be collected and the purposes for it, to clearly inform users, aka data subjects in the RGPD terms, about their specific privacy practices. In the mobile application domain, this allows users to make informed decisions when downloading and using apps. However, achieving this level of transparency in the mobile domain is especially challenging. Privacy policies, traditionally the primary mechanism for communicating privacy practices, are often lengthy, ambiguous (Sailaja & Jones, 2017), and incomplete (Pollach, 2006). Indeed, previous research (Story et al., 2019) has shown that the statements in these documents frequently do not align with the actual practices observed in application behavior. As an alternative, privacy labels were proposed (Kelley et al., 2013) and adopted by mobile app marketplaces as a structured-format alternative to inform users about apps' privacy practices, such as data collection and data sharing. Nonetheless, this new means also showed misalignments with privacy policies and app behavior (Khandelwal et al., 2024).

A contributing factor to these inconsistencies is the widespread reliance on third-party libraries (Akash et al., 2022; Hao et al., 2023), which constitute over 60% of an app's code on average (Wang et al., 2015). These libraries, intrinsic to the current mobile app economy, enable data controllers (often named developers, as they sometimes hold this responsibility) to rapidly integrate features such as analytics, advertising, authentication, and geolocation services. However, their use is closely tied to the predominant business model of mobile applications, where personal data is monetized through targeted advertising and other data-driven revenue streams. Many third-party libraries operate as intermediaries in this data economy, collecting user information and sharing it with a network of advertisers, data brokers, and analytics firms. Consequently, developers often incorporate these libraries without a complete understanding of the scope of data collection or the legal responsibilities associated with them (Balebako et al., 2014). This lack of transparency complicates compliance efforts for developers (Lu et al., 2024) and exposes end-users to significant privacy risks.

Furthermore, the technical documentation provided by third-party libraries is often inadequate or overly complex (Horstmann et al., 2024), making it challenging for developers to configure privacy settings correctly or fully comprehend the data flows within their applications. As a result, even well-intentioned developers may struggle to provide accurate disclosures in privacy policies and privacy labels, further exacerbating regulatory misalignment and undermining user trust.

At the intersection of technology and law, addressing these challenges requires a multidisciplinary approach. Ensuring compliance is not merely a technical problem of detecting data flows or analyzing privacy policies; it also demands an understanding of legal requirements, regulatory interpretations, and the broader implications of data protection principles. This thesis has, therefore, benefited from collaborations with legal experts—including data protection lawyers—to accurately translate GDPR requirements into programmatically verifiable rules, ensuring that the automated assessments align with both regulatory expectations

and real-world enforcement criteria.

In summary, the mobile application ecosystem—with its continuous influx of new and updated apps, prevalent reliance on third-party libraries, and often opaque privacy practices—presents significant challenges for achieving compliance with data protection regulations such as the GDPR. Privacy policies and labels frequently fail to accurately reflect actual data collection and sharing practices, thereby creating a substantial gap between regulatory requirements and real-world application disclosure. This discrepancy exposes developers to potential noncompliance and end-users to privacy risks and uninformed decisions when downloading and using apps.

1.3 Research Questions

To address the challenges associated with assessing GDPR compliance in mobile applications, this thesis is guided by the following research questions. Each question is preceded by a brief explanation to contextualize the specific problem it addresses:

Preamble RQ1. Assessing GDPR compliance of mobile apps manually at scale is impractical and, for some GDPR requirements, impossible without data controller cooperation.

RQ1: Which GDPR-required privacy practices can be evaluated by analyzing the behavior of mobile applications? How can they be systematically assessed?

Preamble RQ2. Privacy policies and labels are the primary mechanisms for informing users about data handling practices, but their analysis at scale is challenging.

RQ2: How can the search, extraction, and analysis of privacy policies and privacy labels be automated to identify specific personal data handling practices?

Preamble RQ3. Discrepancies between declared privacy practices and actual app behavior undermine transparency and regulatory compliance.

RQ3: How effective are automated methods in comparing the observed applications' behavior with the statements declared in their privacy policies?

Preamble RQ4. Identifying common patterns of noncompliance is essential for improving privacy practices in the mobile ecosystem.

RQ4: What are the most prevalent behavioral patterns and compliance violations in the mobile ecosystem, and how can they be documented to provide concrete recommendations for developers, regulators, and marketplaces?

1.4 Thesis Objectives

Given the rapid evolution of technology—and particularly the dynamic nature of the mobile application ecosystem—the automation of compliance evaluation arises as the only viable solution. To address this challenge, the thesis is organized around a primary objective and several specific objectives outlined in this section. Moreover, these objectives are closely

aligned with those of the AutoGDPR project, which has financially supported this thesis.

1.4.1 Main Objective

The main objective of this thesis is to develop automated methods and artifacts for assessing the regulatory compliance of mobile applications under the GDPR. This goal involves examining apps' actual behavior, evaluating the practices disclosed in privacy policies and labels, and comparing both dimensions with the GDPR requirements to determine their alignment.

1.4.2 Specific Objectives

1. **Analyze the personal data practices of mobile applications:**
    - Design, develop, refine, or utilize advanced dynamic analysis methods capable of identifying personal data transfers made by mobile apps on real devices.
    - Validate the developed methods to ensure accuracy and robustness in capturing and examining personal data practices.
    - Infer potential developer practices from intercepted connections and observed behavior that may impact regulatory compliance.
2. **Evaluate the practices disclosed in privacy policies and labels:**
    - Develop automated methods for searching, extracting, and analyzing privacy policies and labels.
    - Implement Natural Language Processing (NLP)-based techniques to identify specific data handling practices in these privacy statements.
    - Generate or find annotated datasets to validate and measure the performance of the proposed methods.
3. **Automate regulatory compliance assessment:**
    - Translate GDPR requirements into verifiable rules that can be automatically evaluated using the developed artifacts.
    - Create automated methods that compare app behavior analysis with the practices declared in privacy policies, enabling a systematic approach to compliance assessment.
    - Apply the proposed methods to large-scale evaluations, providing empirical evidence of application compliance levels within the Android ecosystem.
4. **Contribute to transparency and compliance in the mobile ecosystem:**
    - Document recurring patterns and trends of noncompliance, along with potential underlying causes.
    - Develop recommendations—based on the findings—directed toward developers, regulators, and marketplace owners to foster more transparent practices that adhere

to data protection regulations.

Chapter 2

Methodology

This section describes the methodological approach that underpins this research, which focuses on automating regulatory compliance assessments for mobile applications. To pursue this objective, a methodology was adopted incorporating descriptive and quantitative elements within the Design Science Research (DSR) framework. This comprehensive approach not only supports the analysis of existing phenomena but also guides the creation of specific artifacts that facilitate analysis and evaluation. The following pages detail the methods employed, the tools used, and the validation strategies adopted, illustrating how they collectively support the development of methods that automate legal compliance assessments.

2.1 Introduction to the Methodology

Automating legal compliance assessments in the mobile ecosystem requires concurrent progress in three key areas. First, it is essential to analyze the technical behavior of applications to identify the personal data they collect and how they process it. This analysis is essential for determining whether these practices comply with existing regulations. Second, the level of transparency on the part of developers or application owners must be evaluated through documents such as privacy policies and privacy labels. These documents serve as transparency tools mandated by regulators and marketplaces. Finally, understanding the legal obligations imposed by regulations like the GDPR is indispensable. Compliance assessment relies on detailed comparisons across these three pillars, and automation requires translating such evaluations into programmable, verifiable rules.

To address these areas, a methodological strategy that combines descriptive and quantitative techniques with the DSR framework was adopted. This approach enabled the identification and delineation of key problems, as well as the design, implementation, and evaluation of targeted artifacts in each area. Below are the primary accomplishments achieved in each of these domains.

2.1.1 Behavior Analysis

Reconstruction of a Dynamic Analysis Module: A complete overhaul of an existing module was carried out, rewritten in Python to enhance extensibility. This new version intercepts and analyzes connections generated by mobile applications, capturing and decrypting network traffic on real devices. It was further optimized to support parallel execution on multiple devices, thereby increasing efficiency and analytical capacity.

Method for Identifying Domain Controllers: An innovative method was developed to link a domain with its accountable organization, facilitating the identification of personal data recipients and highlighting patterns of data aggregation by key entities.

Identification of Connections Initiated by Third-Party Libraries: A method was created to discern whether personal data connections originate in the native app code or in third-party libraries.

Detection of Third-Party Libraries: A dynamic–analysis–based method was created to detect the presence and versions of third-party libraries within an app's code.

Evaluation of Library Privacy Settings: A dynamic method was devised to analyze the privacy settings of third-party libraries and to monitor their real-time behavior, capturing any changes in their configuration.

2.1.2 Privacy Statement Analysis

Classification of Texts as Privacy Policies: A method was developed to determine whether a given text represents a privacy policy.

Identification of Recipients in Privacy Policies: A technique was designed to extract the companies receiving personal data, as declared in privacy policies.

Identification of Data Controllers in Privacy Policies: A method was implemented to locate and extract the identity of the data controller as stated in such policies.

LLM-Based Method for Analyzing Privacy Policies: An automated approach was developed using LLMs—specifically ChatGPT—to identify and extract privacy practices from policies. This method sets a new benchmark for privacy policy analysis, offering higher efficiency and superior performance than traditional pre-trained machine learning classifiers.

2.1.3 Automation of Compliance Evaluation

Translating Regulations into Programmable Rules: In collaboration with legal experts, GDPR requirements were converted into specific, testable rules and applied in most of this thesis's studies. For example, the automated evaluation of international data transfers integrated behavioral data from apps with the declared practices in policies, automatically determining compliance or lack (Guamán et al., 2023). Similar methods were implemented to evaluate additional requirements defined under GDPR Article 13.

The methodological approach adopted in this thesis ensures that the artifacts developed are technically solid and applicable in real-world scenarios. In addition, the work addresses

the leading legal and regulatory challenges of the mobile ecosystem by proposing tools and methods that facilitate automated compliance evaluations in this field.

2.2 Methods and Techniques

2.2.1 Research Design

The research design employed in this thesis combines descriptive and quantitative approaches within the DSR framework. This methodological blend allows for structured examination of the three central study areas (application behavior, declared privacy practices, and regulatory compliance), aiming to design, implement, and evaluate artifacts that respond to identified challenges.

Descriptive and Quantitative Approach

All studies in this thesis adopt a descriptive research design, focusing on analyzing and appraising the current regulatory compliance status of mobile applications with GDPR. This design aims not only to document observed issues but also to explore their underlying causes. In particular, it:

- **Systematically characterizes** aspects of privacy policies, data flows, and third-party library configurations, providing a detailed overview of current practices in the mobile ecosystem.
- **Identifies recurring patterns and problems** and investigates potential root causes, such as developers' lack of awareness of third-party libraries they integrate or how default settings frequently prioritize functionality over privacy.

The quantitative component ensures objective, scalable, and data-driven evaluations. Through large-scale datasets and statistical analyses, the research offers solid, generalizable insights into noncompliance prevalence, transparency gaps, and practical hurdles that developers face in meeting regulatory requirements.

Integration of Design Science Research

The Design Science Research (DSR) served as the methodological framework guiding the creation of the artifacts and methods described in this thesis. This approach addresses complex problems through an iterative process that includes:

1. **Identifying Key Issues:** Examining transparency, legal compliance, and data handling challenges in mobile applications.
2. **Defining Solution Objectives:** Setting clear targets for the developed artifacts, such as detecting the organization responsible for a domain that receives personal data or identifying privacy settings in third-party libraries.
3. **Design and Development of Artifacts:** Building practical solutions through dynamic analysis, natural language processing (NLP), and machine learning (ML and LLMs).

4. **Validation:** Testing these artifacts in real-world scenarios using data from Android applications to evaluate their performance, scalability, and practical utility.
5. **Dissemination of Results:** Publishing the developed methods and tools in international journals and conferences, contributing to both practice and the advancement of the state of the art.

This structured, methodologically strong design ensures coherence across the three research areas while producing outcomes that are both relevant and implementable in practical environments. It also establishes a firm foundation for generalizing and applying the methods to other systems or regulatory frameworks.

2.2.2 Data Collection Methods

Conducting the studies in this thesis required collecting two principal types of data: mobile applications' code (APK file) and metadata and the information in their policies and privacy labels. The methods used to obtain these data are described below, emphasizing the reliability and reproducibility of the process.

Mobile Application Collection

Android applications code forms the basis of the analyses performed in this thesis, as they are installed and run on physical devices for subsequent examination. A Google Play Store API (marty0678, 2023) was utilized to download the APK files. This API simulates a real device's connection and authentication, enabling automated access to publicly available app metadata and the download of APK files. This method guarantees that collected data authentically mirrors the publicly released app versions at the time of download, ensuring their integrity for subsequent analyses.

Privacy Policy Collection

Privacy policies constitute a crucial component for assessing developers' claims regarding personal data processing. To gather these policies, a Selenium-based method was elaborated to automate the scraping of Google Play webpages. This process visits each app's page, locates the privacy policy link provided by the developer, and downloads its content. The retrieved policies are stored in plain text or HTML format, enabling subsequent analysis.

Privacy Label Collection

Introduced by Google Play Store as an additional transparency mechanism, privacy labels supply structured information on how apps manage personal data. An automated scraping method was designed to access the relevant section on each app's page. The extracted information is organized and saved in CSV files, streamlining data handling and analysis due to the structured format.

Ethical and Technical Considerations

Finally, APKs, along with their associated privacy policies and labels, were maintained on a secure local server for retrieval by analysis modules at a later stage. Throughout the data collection process, platform usage policies were respected, ensuring that all retrieved data were publicly accessible and used exclusively for research. Further, the implemented methods were designed to be flexible and adaptable, allowing for updates in response to potential changes in page structures or Google Play's policies. This strategy maintains continuity and reliability in the collection process over time.

2.2.3 Tools and Technologies Used

A range of tools and technologies were used in this thesis, selected and configured to serve specific objectives. They fall under two primary categories: those employed for analyzing mobile app behavior and those aimed at identifying and extracting declared practices in privacy policies. All utilized technologies are open-source or rely on public APIs, ensuring the reproducibility of the studies performed.

Application Behavior Analysis

Mitmproxy: A traffic interception tool that employs the Man in The Middle (MiTM) technique to capture and analyze network connections established by mobile apps. It records network traffic and decrypts data transmissions to identify personal data collection and transfer.

Frida: A dynamic instrumentation tool that bypasses certificate pinning[1] to inspect HTTPS connections. Frida is also used to detect integrated third-party libraries and analyze their privacy configurations.

LibScout: A static analysis tool that locates and characterizes third-party libraries within apps, complementing dynamic analysis with Frida.

ADB (Android Debug Bridge): Enables communication with mobile devices, assisting in app installations, command executions, and necessary data extraction for analysis.

Android Monkey: Generates pseudo-random events to interact with apps, simulating user behavior that is otherwise impractical to replicate manually.

Processing Privacy Policies

Selenium: Used to access dynamic webpages containing privacy policies. It executes JavaScript on these pages and extracts complete content, even when the policies are not directly available in plain text.

Machine-Learning Classifiers (ML): Initially, traditional classifiers—such as Support Vector Machines (SVM)—identified and categorized privacy practices, later complemented by

[1] Certificate pinning is a protection measure that mobile apps typically employ to avoid trusting unknown SSL certificates, thereby being susceptible to MiTM attacks.

LLMs due to their superior performance.

LLMs: Models like ChatGPT automate the analysis of privacy policies, identifying and extracting declared practices (e.g., data transfers and third-party recipients).

Method Development and Automation

Python: The programming language used to implement analysis and automation methods. Its flexibility enabled integrating Selenium, Frida, and Mitmproxy tools, as well as the processing of data gathered through classifiers and LLMs.

2.2.4 Validation and Evaluation of Results

Validation and evaluation of the methods developed in this thesis vary according to their nature and purpose. Some methods rely on empirical evidence, while others require more rigorous statistical and experimental validation. Below is an overview of the strategies adopted for each type of method.

Validation of Methods for Analyzing App Behavior

The methods designed to analyze app behavior do not necessitate statistical validation processes, as they aim to provide empirical evidence of actual data-handling practices. These methods focus on:

- Recording observable data such as network traffic, third-party library privacy settings, and the connections established by apps.
- Presenting findings based on directly and reproducibly collected data.

Their value lies in documenting the genuine practices of mobile apps, without involving predictive models or statistical inference.

Validation of Privacy Policy Processing Methods

On the other hand, NLP-based methods for analyzing privacy policies do require statistical validation, given their inferential nature. The validation of these methods was conducted through a rigorous DSR framework:

1. **Creation of a Ground Truth:** A manually annotated dataset was established as a reference, encompassing:
    - **Codebook Creation:** Detailed annotation guidelines were developed to standardize how privacy practices were identified.
    - **Discrepancy Resolution:** Specific cases were discussed and documented to maintain annotation consistency.
2. **Method Validation:** The ground truth dataset was then used to evaluate the performance of developed methods, employing key metrics such as:

- **Precision:** Proportion of correct predictions among all the method's identified cases.
- **Accuracy:** Percentage of correct predictions relative to the total number of cases.
- **Recall:** Proportion of correct identifications among all valid cases in the ground truth.
- **F1 Score:** The harmonic mean of precision and recall, balancing both aspects in a single metric.

Significance of Recall: In the context of regulatory evaluations, recall is particularly important. A low recall indicates the existence of many false negatives, which can result in overlooking valid cases and thus incorrectly assessing possible non-compliance. Given the potential legal and regulatory implications of these findings, minimizing false negatives is essential to prevent wrongly exonerating apps that should be flagged for potential non-compliance.

App Selection for Experiments

Conducting experiments to evaluate methods and explore hypotheses about the mobile ecosystem required strategically selecting sets of applications. Depending on the study's specific objective:

- **Large-Scale Studies:** Randomly chosen apps maximize sample representativeness to gauge the overall compliance environment in the Google Play Store.
- **Popularity-Based Comparisons:** In analyzing how compliance differs between popular and less-known apps, sets of applications were grouped by download count.
- **Third-Party Library Privacy Configurations:** Apps with a substantial number of downloads were selected to determine whether developers modified third-party library privacy settings, as such apps are more likely to integrate advanced SDKs and explicitly configure them.

Taken together, the validation and evaluation strategies adopted in this thesis have been essential to ensuring that the resulting methods are both reliable and applicable. From empirical documentation of app behavior to the statistical validation of privacy policy analysis methods, these approaches were tailored to test the research hypotheses and to drive the development of methods suitable for real-world contexts in the mobile ecosystem.

2.2.5 Methodology Limitations

Despite the advances made in this work, the developed methods present certain inherent limitations that must be considered when interpreting results and envisioning future studies.

Analyzing Application Behavior

Network interception in this research employs tools such as Frida and Mitmproxy, based on dynamic analysis techniques. This choice stems from the need for highly reliable evidence

regarding actual app behavior and data-handling practices. Unlike static analysis—inspecting code without running it and potentially detecting hypothetical connections that may never be activated (e.g., "dead code")—dynamic analysis focuses exclusively on connections genuinely established during an app's execution. This orientation ensures that gathered evidence reflects directly observed activities, a vital aspect in evaluating regulatory compliance.

However, this reliance on dynamic analysis constrains connection detection to the code segments triggered during the analysis process. While it ensures high reliability of the recorded data, it also reduces the number of identified connections, thus providing a conservative lower bound of possible practices. In other words, there may be additional connections and potential violations not uncovered by the analysis.

To increase coverage, Android Monkey is used to generate pseudo-random app interactions, which raise the likelihood of triggering extra connections. Despite this, its random nature introduces minor variations in reproducibility. Moreover, the inability to log into apps can restrict access to functionalities that might generate additional connections. Finally, the need to use rooted devices poses a technical hurdle, as certain applications—particularly banking apps—detect this condition and block their installation or operation.

Analyzing Privacy Policies

Methods used to process and analyze privacy policies encounter specific hurdles, whether with traditional machine-learning techniques or with LLMs. Traditional techniques like ML classifiers demand carefully prepared, representative training datasets to achieve solid performance, requiring substantial effort in both time and resources. Additionally, these models are less flexible when it comes to generalizing to new cases compared to LLMs.

Although LLMs offer notable advantages—such as handling tasks that require flexibility and complexity—they also present limitations. Thorough validation remains crucial to verify their performance in extracting specific practices. Moreover, their costs, especially with proprietary models, can limit availability in academic settings. Because of their "black-box" design, interpreting LLM outputs is challenging, and they may exhibit unpredictable behaviors that necessitate iterative testing and fine-tuning (i.e., A/B testing).

Impact on Results and Future Directions

These limitations mainly affect the coverage and reproducibility of the outcomes. In dynamic analysis, the findings represent a reliable but incomplete sample of observed practices, whereas, in privacy policy analysis, validation challenges and operating costs can influence scalability. Recognizing these constraints enhances clarity in interpreting results and highlights key areas for future research focused on overcoming them.

Chapter 3

Compendium of Publications

This section presents the articles that constitute the main contributions of this thesis. The following is a list of them in order:

1. D. S. Guamán, D. Rodriguez, J. M. del Alamo, and J. Such, "Automated GDPR compliance assessment for cross-border personal data transfers in android applications," *Computers & Security*, vol. 130, 103262, 2023. doi: 10.1016/j.cose.2023.103262
2. D. Rodriguez, J. M. del Alamo, M. Cozar, and B. García, "ROI: A method for identifying organizations receiving personal data," *Computing*, vol. 106, no. 1, pp. 163–184, 2024. doi: 10.1007/s00607-023-01209-2.
3. D. Rodriguez, A. Jain, J. M. Del Alamo, and N. Sadeh, "Comparing Privacy Label Disclosures of Apps Published in both the App Store and Google Play Stores," in *2023 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, 2023, pp. 150–157. doi: 10.1109/EuroSPW59978.2023.00022.
4. D. Rodriguez, J. M. Del Alamo, C. Fernández-Aller, and N. Sadeh, "Sharing is Not Always Caring: Delving Into Personal Data Transfer Compliance in Android Apps," *IEEE Access*, vol. 12, pp. 5256–5269, 2024. doi: 10.1109/ACCESS.2024.3349425
5. D. Rodriguez, I. Yang, J. M. Del Alamo, and N. Sadeh, "Large language models: a new approach for privacy policy analysis at scale," *Computing*, vol. 106, no. 12, pp. 3879–3903, 2024. doi: 10.1007/s00607-024-01331-9.
6. D. Rodríguez, C. Fernández-Aller, J. M. Del Alamo, and N. Sadeh, "Data Retention Disclosures in the Google Play Store: Opacity Remains the Norm," in *2024 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, 2024, pp. 19–23. doi: 10.1109/EuroSPW61312.2024.00009.
7. D. Rodriguez, J. A. Calandrino, J. M. Del Alamo, and N. Sadeh, "Privacy Settings of Third-Party Libraries in Android Apps: A Study of Facebook SDKs," *Proc. Privacy Enhancing Technologies*, Forthcoming, 2025.

The contribution of each article to the thesis will be discussed in Section 4.

Computers & Security 130 (2023) 103262

Contents lists available at ScienceDirect

Computers & Security

journal homepage: www.elsevier.com/locate/cose

Automated GDPR compliance assessment for cross-border personal data transfers in android applications

Danny S. Guamán[a], David Rodriguez[b], Jose M. del Alamo[b,*], Jose Such[c]

[a] Escuela Politécnica Nacional, Ecuador
[b] Universidad Politécnica de Madrid, Spain
[c] Department of Informatics, King's College London, United Kingdom

ARTICLE INFO

Article history:
Received 3 November 2022
Revised 12 March 2023
Accepted 14 April 2023
Available online 19 April 2023

Keywords:
D.4.6 security and privacy protection
J.9 mobile applications
K.4.1.f privacy
K.4.1.g regulation
k.4.1.h transborder data flow

ABSTRACT

The General Data Protection Regulation (GDPR) aims to ensure that all personal data processing activities are fair and transparent for the European Union (EU) citizens, regardless of whether these are carried out within the EU or anywhere else. To this end, it sets strict requirements to transfer personal data outside the EU. However, checking these requirements is a daunting task for supervisory authorities, particularly in the mobile app domain due to the huge number of apps available and their dynamic nature. In this paper, we propose a fully automated method for assessing the compliance of Android apps with the GDPR requirements for cross-border personal data transfers. We have applied the method to 4593 apps from the Google Play Store discovering that nearly half of the ones sending personal data are potentially non-compliant with GDPR requirements. These results reveal that there is still a very significant gap between what app providers do in practice and what is intended by the GDPR.

1. Introduction

THE distributed nature of today's digital systems and services not only facilitates the collection of personal data from individuals anywhere but also their transfer to different countries around the world (Razaghpanah et al., 2018). This raises potential risks to the privacy of individuals, as the organizations sending and receiving personal data can be subject to different data protection laws and, therefore, may not offer an equivalent level of protection. In some regions, such as China, privacy may be less valued or valued differently, when compared to order and governance (Weber et al., 2020). While in other regions, particularly in the European Union (EU), privacy is strenuously protected and is conceived as a Human Right (European Union, 2010). As a result, the General Data Protection Regulation (GDPR) (European Commission, 2016) constrains cross-border transfers (*also named international transfers*) outside the European Economic Area (EEA)[1] and recognizes only fourteen non-EU countries as providing protection equivalent to the GDPR.

Mobile applications, or just "apps", exacerbate the data protection compliance issues for organizations, notably with requirements related to cross-border transfers. The particularities of the app development and distribution ecosystems are major factors underlying these issues (European Union Agency For Network and Information Security, 2017). First, apps collect a great amount of personal data, which may be transmitted from the device to data processors across the world or shared between chains of third-party service providers (Razaghpanah et al., 2018), even without the app developer's knowledge (Balebako et al., 2014). Second, apps are distributed through global stores, enabling app providers to easily reach markets and users beyond their country of residence. In this context, there is a need for constant vigilance by the various stakeholders, including app developers, supervisory authorities, and app distribution platforms, to ensure that appropriate requirements have been met and to avoid potential data protection compliance breaches.

Nevertheless, testing or auditing mobile apps against legal data protection requirements is challenging. First, it is necessary to simplify the high-level legal requirements into concrete step-by-step technical criteria and indicators to be assessed in mobile apps. Second, parties responsible for checking compliance with data protection requirements, such as supervisory authorities, require automated assessment methods and tools to cope with the vast and

* Corresponding author.
E-mail address: jm.delalamo@upm.es (J.M. del Alamo).

[1] The EEA includes all the EU Member States plus Norway, Iceland, and Liechtenstein. For the sake of clarity, we will use the term EU from now on to refer to all of these countries.

https://doi.org/10.1016/j.cose.2023.103262

ever-changing mobile ecosystem (European Union Agency For Network and Information Security, 2017). Even distribution platforms like Google Play Store can benefit from these automated assessment methods and thus extend their privacy mechanisms, such as Google Play Data Safety, which currently allows developers to declare their apps' data collection and sharing practices but does not enforce them.

Relevant research efforts have been directed toward elaborating methods and tools for assessing whether mobile apps fulfill privacy and data protection requirements. In the past five years, these research efforts have mainly focused on two privacy-related requirements, i.e., disclosure of personal data and non-compliance (del Alamo et al., 2021). Disclosure of personal data is aligned with the *privacy as confidentiality* paradigm (Gurses, 2014). It is based on the binary criterion that any exposure of personal data outside a mobile app may result in a privacy violation. To assess this criterion, researchers rely on static analysis techniques performed on app models (e.g., Zhang et al., 2021), dynamic analysis techniques on the actual app behavior (e.g., He et al., 2018), or hybrid analysis techniques that combine both static and dynamic analysis (e.g., Ali-Gombe et al., 2018).

On the other hand, non-compliance is roughly aligned with privacy as contextual integrity. It does not equal the exposure of personal data to a privacy violation. Instead, privacy violation or preservation is determined by the appropriateness of the personal data flow within a particular context (Nissenbaum, 2004). These personal data flows are defined based on five key elements: the type of personal data, the data subject to whom the personal data refers, the sender and recipients of personal data acting in a particular capacity or role, and the transmission principles that constrain the flow (e.g., user consent). To identify appropriate or inappropriate personal data flows in terms of the aforementioned elements, some researchers rely on relevant regulations such as the GDPR (e.g., Jia et al., 2019), while others use privacy policies (e.g., Lin et al., 2022). Ultimately, the extracted (in)appropriate personal data flows are compared to the actual apps' behavior to check compliance.

Data protection regulations like the GDPR lay down multiple (in)appropriate personal data flows, where each contextual integrity element is instantiated differently. For instance, the GDPR distinguishes between regular categories of personal data and special categories of personal data (i.e., sensitive data), such as political opinions, which are subject to stricter constraints. Similarly, the GDPR recognizes that data subjects can be adults or children, which also implies different levels of constraint. GDPR also specifies the roles and capabilities of senders and recipients of personal data, which could be data subjects (e.g. app users), data controllers (e.g. app providers), data processors (e.g. cloud storage providers), or third parties (e.g. ads providers). These actors have different sets of responsibilities and capabilities depending on their roles. Finally, the transmission principle can be related to transparency (e.g., requiring informing a user that certain data flows will take place), related to legitimacy (e.g., requiring to obtain user consent before a data flow takes place), among others.

Various methods and tools for assessing personal data flow compliance with certain GDPR requirements have been proposed in the state-of-the-art, as detailed in Section 2. Yet, a prior work (del Alamo et al., 2022) found the need to address particular data flows, such as cross-border personal transfers, and consider contextual integrity elements to improve the effectiveness and utility of these methods and tools.

To address these challenges, we have developed a novel fully automated method for assessing the compliance of Android apps with the GDPR requirements for cross-border personal data transfers. To this end, similar to related work (Baskerville et al., 2022), we followed the Design Science research paradigm, which focuses on iteratively creating and evaluating innovative solutions to a practical problem. In a realm where assessing compliance with privacy requirements is highly required but manually doing it is not feasible, we iteratively implemented and evaluated a fully automated artifact (i.e., a method) to assess the compliance of Android apps with GDPR cross-border transfer requirements.

This method enhances the state-of-the-art related methods by allowing (i) more accurate detection of non-compliant apps, and (ii) a broader scope not only considering international transfers but other types of cross-border transfers. This method leverages our prior work on a compliance assessment process (Guaman et al., 2021) and further extends it with an automated approach to identify cross-border transfer statements from natural language privacy policies with an F-measure ranging from 85.7% to 100%.

Our results contribute to the practice of privacy engineering for different stakeholders. First and foremost, our approach can be exploited by supervisory authorities and even app distribution platforms to assess these privacy practices with a high degree of certainty. Indeed, the utility of the automated method was leveraged to assess Google Play Store apps' compliance with the GDPR cross-border transfer requirements. A set of 4593 apps in Spain was evaluated, providing relevant findings concerning compliance with these requirements. Privacy researchers and scholars can also benefit from the resources generated in this work e.g., we have released as open data a corpus of annotated privacy policies that can be used to train new assessment methods. Finally, we are transferring the method to an online platform to support developers in assessing their apps and thus being aware of the potential non-compliance issues they may incur.

The rest of the paper breaks down as follows: in Section 2, we review the related work on GDPR compliance assessment and highlight the unique features of our approach. Section 3 provides an overview of the requirements set on cross-border transfers by GDPR. Section 4 details our method, including the methodology we have followed and the method design. Section 5 reports the evaluation results and demonstrates the utility of the developed method by assessing Google Play Store apps. Section 6 discussed our research findings and how we address potential threats to the validity of our approach. Finally, in Section 7, we conclude and outline directions for future work.

2. Related work

The automated compliance assessment of Android apps requires the analysis of both the privacy policy text and the app's behavior.

For the privacy policy analysis, del Alamo et al. (2022) recently surveyed the different approaches available in the state of the art. They mostly rely on the codification or annotation method (Saldana, 2015), where one or multiple domain analysts generate structured annotations of privacy practices (i.e., a corpus) by systematically assigning a label to the policy statements. Useful corpora have been released in the privacy domain (Zimmeck et al., 2019; Wilson et al., 2016), which ultimately are used as ground truth for building automatic classification models. For example, Zimmeck et al. (2019) automated the extraction of data collection practices from privacy policies, while Andow et al. (Andow et al., 2020) distinguished the entity (i.e. first-party vs. third-party) to which personal data is sent.

Focusing on GDPR, Fan et al. (2020) empirically assessed transparency, data minimization, and confidentiality requirements in Android mHealth apps, checking whether six different practices are informed through privacy policies. Mangset (2018) also checked GDPR requirements related to transparency, data minimization (collection practices), confidentiality (data at rest in transit), and some user rights (particularly, consent and objection automatically individual decision-making). Similar in nature to our work,

Lin et al. (2022) proposed a semi-supervised methodology to audit website compliance. The study focused on checking the transparency and readability of 663 Chinese websites according to GDPR.

Unfortunately, none of the previous works addressed cross-border transfer practices nor proposed a fully automated method to check GDPR compliance by contrasting privacy policies to the actual system behavior in the mobile ecosystem.

As for app behavior analysis, researchers have leveraged static, dynamic, or hybrid techniques (del Alamo et al., 2021). For example, Ferrara and Spoto (2018) relied on static code analysis to detect disclosures of personal data so that data protection officers could spot potential GDPR infringements. Jia et al. (2019) leveraged dynamic techniques to detect personal data disclosures in network packets lacking user consent. While our work focuses on different GDPR requirements, Jia's work could be seen as complementary to our work as the app behavior analysis method could minimize the false-negative rate.

Finally, we consider Eskandari et al. (2017) as the closest related work regarding the GDPR requirements covered in this paper. They propose PDTLoc, an analysis tool that employs static analysis to detect violations of article 25.1 of the EU Data Protection Directive (European data protection law replaced by the GDPR). This Directive set requirements for international transfers like those laid down in the GDPR. However, this prior work presumes any transfer outside the EU to be a regulatory infringement, thus this approach would have incorrectly identified potential compliance issues. The authors did not consider the privacy policies as a means of disclosing the intention to perform cross-border transfers and the appropriate safeguards that do enable these transfers.

To fill this gap, in prior work (Guaman et al. (2021)) we defined an earlier method for the compliance assessment of Android apps with GDPR cross-border personal data transfer requirements. This work supported the app behavior analysis through dynamic testing techniques. It also identified the specific requirements for the transparency elements to be included in the privacy policies for the lawful disclosure of the international transfer. However, the compliance assessment process was not automated, as the interpretation of the privacy policies required human analysis, and thus did not scale. We have extended this prior work with an automated approach to identify cross-border transfer statements from natural language policies. As a result, we have been able to carry out an extensive assessment of cross-border transfers in Google Play Store apps.

3. GDPR cross-border transfers

As illustrated in Fig. 1, there are specific criteria for determining what is considered a cross-border transfer according to GDPR, and the information to be provided to data subjects in that case. Next, we briefly summarize the criteria and refer the interested reader to Guaman et al. (2021) for details.

Criterion **C1.1** determines whether personal data, in the meaning of GDPR (European Commission, 2013), are sent to remote recipients (See DT enumeration in Fig. 1).

Criterion **C1.2** determines whether an app targets EU citizens. We fairly assume that mobile apps available in the Google Play Store reachable from an EU country are indeed targeting EU users.

Criterion **C1.3** determines to which country personal data are sent. Data transfers between EU countries do not add further constraints, but those targeting non-EU countries must meet specific requirements.

Criterion **C1.4** distinguishes whether the servers located outside the EU belong to the app provider itself (i.e., *first-party recipient* or data controller in GDPR terms), or another organization (i.e., *third-party recipient*). In the former case, if the first-party recipient is also located outside the EU then it must disclose the contact details of its representative in the EU[2] (T1 in Fig. 1). The latter is considered an international data transfer.

Criterion **C1.5** seeks to determine whether the destination country provides protection equivalent to the GDPR and therefore has received an adequacy decision from the EU. Fourteen non-EU countries maintain an *adequacy decision* (See ADC enumeration in Fig. 1). International transfers to these countries can take place without any further safeguards but the app provider must disclose[3] (1) the *intention* to transfer personal data to a non-EU country, (2) the names of targeted countries, and (3) the *existence* of an adequacy decision by the Commission (European Commission, 2018) (T2 in Fig. 1).

International transfers to countries not covered by an adequacy decision require the app providers to adopt other *"appropriate safeguards"*. GDPR defines the following ones[4]: *Standard Data Protection Clauses, Binding Corporate Rules, Approved Codes of Conduct,* and *Approved Certification Schemes.* These assurance mechanisms must be approved by the EU and, in general, allow the app provider to ensure that third-party recipients have implemented appropriate safeguards to guarantee a protection level equivalent to GDPR. To ensure transparency, the app provider must inform the data subjects about (1) the *intention* to transfer personal data to a non-EU country, (2) the names of targeted countries, (3) a reference to the *appropriate safeguard(s)* according to the aforementioned options, and (4) the means to obtain a copy of the safeguard(s) (European Commission, 2018) (T3 in Fig. 1).

In the absence of an *adequacy decision* or any *appropriate safeguards*, some exceptions[5] allow for international transfers in specific situations. We highlight the *explicit consent,* which requires consent through an affirmative action of the data subjects, e.g., ticking a box, to be obtained after providing precise details of the international transfers.

4. Compliance assessment method

Our method combines different techniques to solve an actual problem in the data protection auditing field for developers, app distribution platforms, and data protection authorities. In this section, first we describe our research methodology, and then we detail the method design.

4.1. Methodology

We designed our automated assessment method following the Design Science Research (DSR) methodology (Hevner et al., 2004). DSR guides the iterative development of artifacts of various forms (e.g., models, methods, etc.) through a set of phases. Fig. 2 illustrates how our research approach aligns with DSR. We further elaborate on each phase in the following paragraphs.

4.1.1. Phase 1

As a result of (i) analyzing reports and recommendations from international bodies, and (ii) observing the recent changes of major stakeholders, we identified the need for automated approaches to assessing privacy and data protection requirements. The European Cyber Security Organization (ECSO) highlights the need to assess and certify aspects of privacy in information systems (ECSO, 2017), while the EU Agency for Cybersecurity (ENISA) highlights the need to build supporting methods and tools to assist regulators in their supervisory duties, and thus cope with the vast and ever-changing

[2] GDPR Art. 27(1)
[3] GDPR Art. 13(1)(f) and Art. 14(1)(f)
[4] GDPR Chapter V
[5] GDPR Art. 49

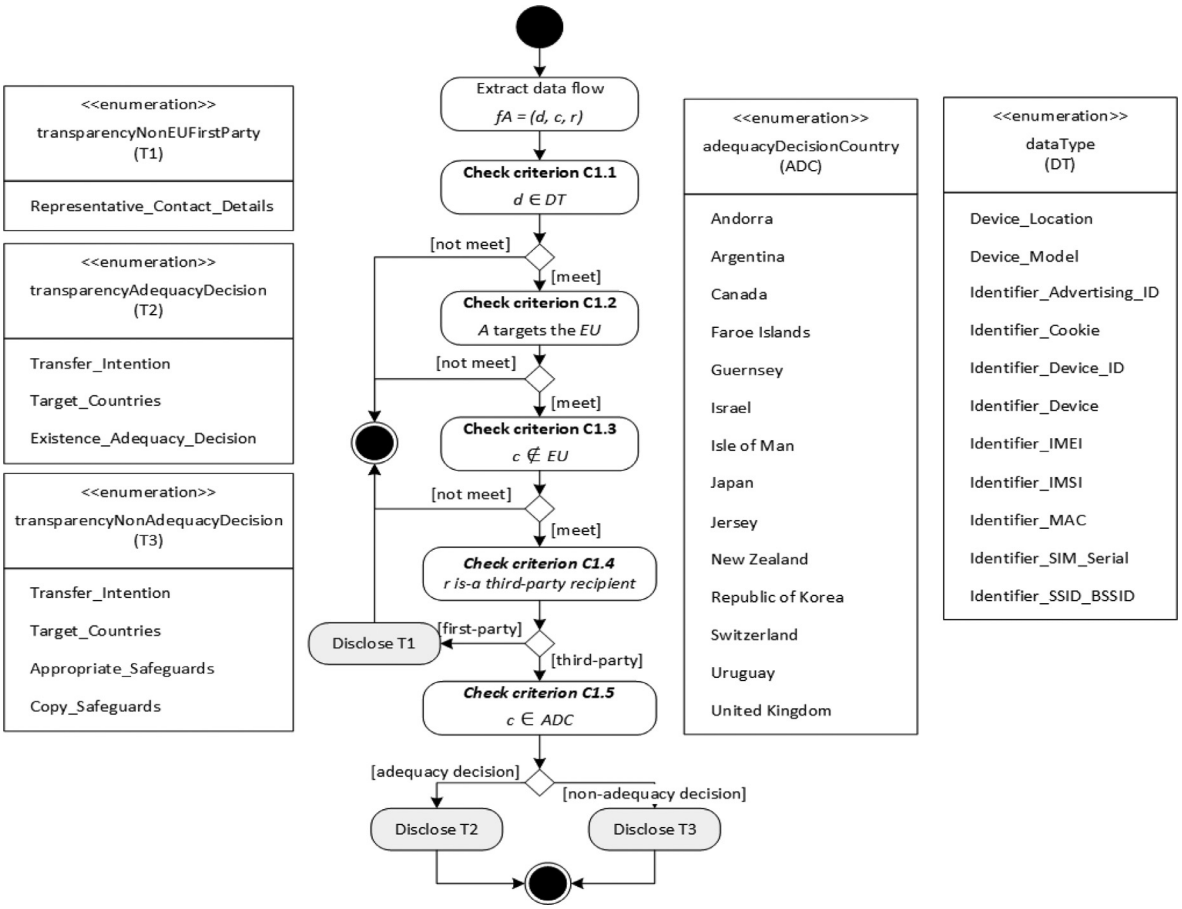


Fig. 1. Criteria for distinguishing the type of cross-border transfer performed by each app.

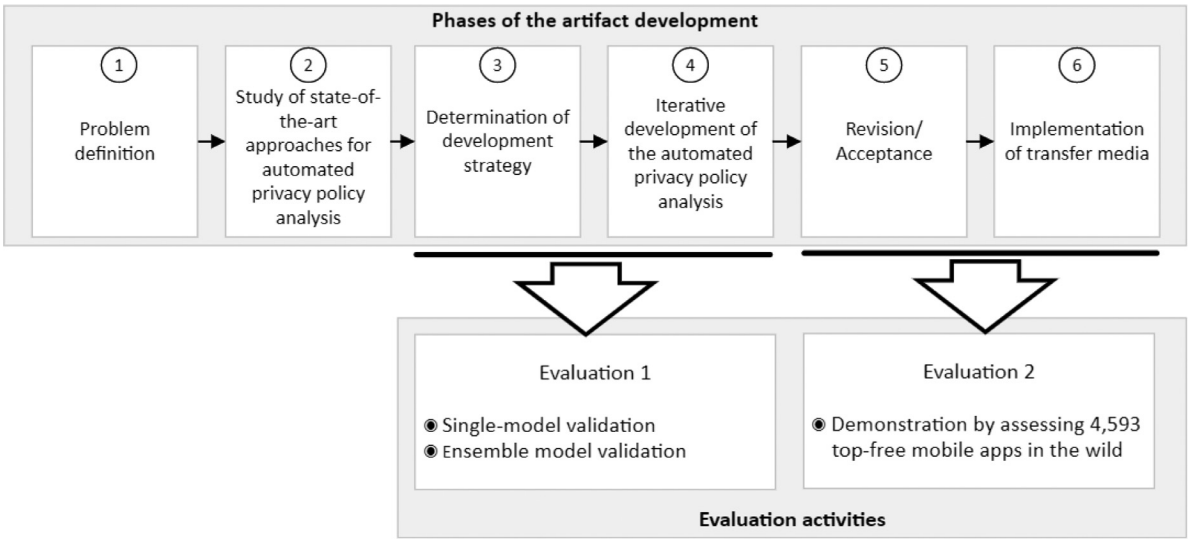


Fig. 2. Overall research approach.

mobile ecosystem (European Union Agency For Network and Information Security, 2017).

This need is also observed in recent changes in major app marketplaces, such as those of Google and Apple, which are increasingly interested in ensuring that their apps meet privacy and data protection requirements. For example, Google Play Protect (GPP) is a security service that scans apps submitted to Google Play Store for potentially harmful behavior. Although it mainly focuses on security aspects, it has just started also to analyze disclosures of pieces of personal information (Android Developers, 2020). Aligned to this endeavor, Apple already provides "nutrition labeling" of personal data disclosures based on information mandatorily provided by developers as a mechanism to improve transparency (Apple Developers, 2020).

4.1.2. Phase 2

After identifying that key stakeholders recognize the need for automated approaches for assessing privacy and data protection requirements in the mobile app ecosystem, we wondered to what extent this need had already been covered by researchers. To answer this question, we conducted a systematic literature mapping (SLM) to analyze existing approaches for automated privacy analysis. Scopus and Web of Science (WoS) databases were used to find high-quality peer-reviewed literature. This SLM, whose details can be found in (del Alamo et al., 2022), analyzed 39 papers from the 1097 publications found on the topic. Among the SLM findings, we identified two key requirements [R] that an automated privacy requirements assessment method needs to consider:

[R-1] automated assessment approaches should still consider requirements mandated only by specific privacy laws. Organizations that offer online products and services worldwide must comply with the requirements established by applicable privacy legislation in the places where these products and services are consumed. A prime case study is the transfer of personal data to organizations located in other countries. While some legislations (e.g. CCPA) do not impose restrictions, others such as the GDPR and PIPL set out detailed requirements. Therefore, automated assessment methods should further recognize more specific requirements, so that compliance can be automatically assessed against the various applicable laws.

[R-2] context must be considered to improve the performance of state-of-the-art automated assessment approaches. Several related approaches focus on assessing incomplete criteria of a privacy requirement, which may lead to a high false positive rate and jeopardize its utility. For example, Eskandari et al. (2017) proposed an approach that assumes a single criterion for assessing compliance with data transfer requirements, i.e., any transfer outside the EU constitutes a regulatory infringement. However, it is more useful to contextualize a transfer, including the country of destination (this allows the type of cross-border transfer to be identified), the type of recipient organization, the declared enabling measures, and other criteria. In this scenario, automated assessment approaches are expected to contextualize data processing practices.

4.1.3. Phase 3

Guided by the aforementioned findings and as motivated in Section I, we focused our efforts on GDPR cross-border transfer requirements. A prior work Guaman et al. (2021) provided the basis for the specific activities that integrate the developed artifact. It requires fundamentally three activities (i.e., *privacy policy analysis, app behavior analysis,* and *compliance checking*) and two inputs (i.e., the privacy policy used for the *privacy policy analysis* and the application package used for the *App analysis).* This prior work supported the *app behavior analysis* by using dynamic testing techniques, as detailed in Section 4.2.2. However, the overall assessment method was not automated, as the *privacy policy analysis* still required human analysis. Accordingly, we extended our previous work by automating the *privacy policy analysis*.

The automation of privacy policy analysis was carried out by building a set of machine learning (ML) and rule-based classifiers to extract cross-border transfer practices declared in privacy policies. The development strategy followed the standard process for building an ML model, i.e., preprocessing, feature engineering, model selection, hyperparameter tuning, and evaluation. We relied on the IT-100 Corpus[6] which consists of one hundred privacy policies manually annotated by two privacy experts and contains 3715 policy segments of which 281 segments contain transparency elements of cross-border transfers.

Details of the privacy policy analysis, app behavior analysis, and compliance checking activities are presented in Sections 4.2.1, 4.2.2 y 4.2.3, respectively.

4.1.4. Phase 4 and evaluation 1

As an iterative approach, each design cycle includes a design phase to improve the artifact design, as well as an evaluation phase to evaluate its performance. We used two iterations to build and evaluate a two-layer classification pipeline: a cross-border transfer intention classifier to identify entire policy segments disclosing the intention to perform a cross-border transfer, followed by a set of transparency element classifiers to identify the individual transparency elements disclosed, as per Table 1. Each iteration advanced the artifact's complexity. First, the individual elements of our model were designed and validated. Then, the individual elements were combined into an ensemble model (pipeline), which was validated again. Details of the process of building and evaluating the ML and rule-based classifiers are presented in Section 4.2.1.

4.1.5. Phase 5 and evaluation 2

The best performance classifiers were integrated into the overall automated assessment method, which was then used to conduct a controlled experiment for assessing 4593 apps from the Google Play Store in Spain. This experiment was also used to evaluate the two-layer classification pipeline by comparing it with the closest related work, which was used as a benchmark. The details of experiment settings and results of this evaluation are presented in Section 5.

4.1.6. Phase 6

Finally, the implementation of transfer media involves transferring the overall automated assessment method and its results into an appropriate form for the relevant audience, such as researchers, practitioners, and supervisory authorities. First, privacy researchers and scholars could leverage the method, individual resources, and results generated in this work. For instance, we released the IT-100 corpus at https://github.com/PrivApp/IT100-Corpus, which contains 281 segments with transparency elements of cross-border transfers that have been manually annotated by two privacy experts. It can be used for advancing research around GDPR compliance. Second, we are transferring the method and results into a work-in-progress online platform[7] to provide practitioners with a straightforward assessment tool for their new mobile apps, and the visualization of results. Finally, the overall results of this work will be presented to the Spanish Data Protection Authority, which was also involved in the early stages of the problem definition.

4.2. Design

Assessing compliance of an app cross-border transfers requires fundamentally three activities (*privacy policy analysis, app behavior analysis,* and *compliance checking*) and two inputs (the privacy policy used for the *privacy policy analysis* and the Android application package APK used for the *App analysis)* as shown in Fig. 3.

The *privacy policy analysis* parses the privacy policy to extract the cross-border transfer practices disclosed. In parallel, *the app analysis* installs and executes the application (APK) to observe its real behavior, particularly the type of personal data it leaks, the type of recipient who receives the data (i.e., first-party or third-party recipient), and the country in which the recipient servers are located. Finally, based on the practices disclosed through the app

[6] The corpus has been released at https://github.com/PrivApp/IT100-Corpus

[7] The alpha online platform can be found at https://www.aei.gob.es/

Table 1
Cross-border transfer annotation scheme.

Cross-border transfer type	Required transparency elements
T1. Transfer to non-EU data controller	EU Representative contact information
T2. International transfer (with adequacy decision)	Transfer intention Existence of EU adequacy decision Target country
T3. International transfer (without adequacy decision)	Transfer intention Target country Appropriate safeguards: - Standard Data Protection Clauses - Binding Corporate Rules - Approved Codes of Conduct - Approved Certification Schemes - Explicit consent Copy means

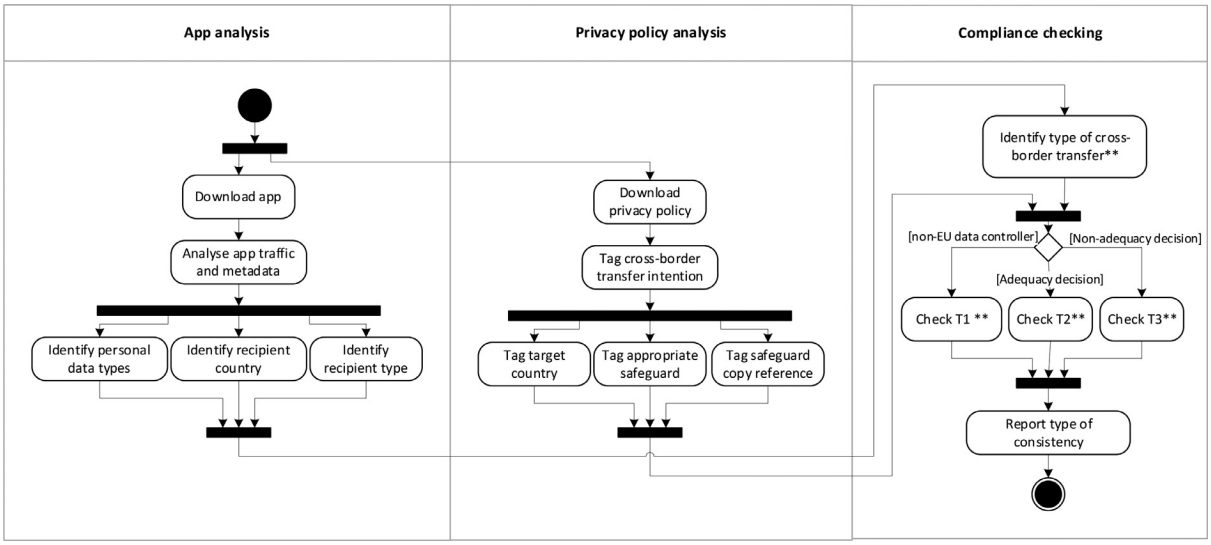


Fig. 3. The automated pipeline detects the app's cross-border transfers and automatically checks them against the corresponding privacy practices in its privacy policy. (**) These activities have been detailed in Fig. 1.

privacy policy and the cross-border transfers it performs, a *compliance check* alerts of potential non-compliant behavior.

4.2.1. Automated privacy policy analysis

In this section, we present our automated approach for classifying cross-border transfer practices declared in privacy policies (Table 1). We relied on the IT-100 Corpus[8] which consists of one hundred privacy policies manually annotated by two privacy experts and contains 3715 policy segments of which 281 segments contain transparency elements of cross-border transfers.

The policy segment disclosing a cross-border transfer practice may include different transparency elements (Fig. 4). Therefore, we composed a two-layer classification pipeline: a cross-border transfer intention classifier to identify entire policy segments disclosing the intention to perform a cross-border transfer (Section 4.2.1.1), followed by a set of transparency element classifiers to identify the individual transparency elements disclosed (Section 4.2.1.2). The validation of the two-layer classification pipeline is presented in Section 4.2.1.3.

Cross-border transfer intention classifier. This classifier tags each policy segment, roughly a paragraph, to indicate whether it discloses (1) or not (0) the intention to perform a cross-border transfer. We use the IT-100 corpus to train a supervised machine learning (ML) algorithm and generate a binary classifier. We have followed the systematic process shown in Fig. 5 to determine the best performance classification model.

> Our business may require us to transfer your Personal Data to countries outside of the European Economic Area (EEA), including to countries such as the Peoples Republic of China or Singapore. We take appropriate steps to ensure that recipients of your Personal Data are bound to duties of confidentiality and we implement measures such as standard contractual clauses. A copy of those clauses can be obtained by contacting our Help Center.

Fig. 4. Segments of the privacy policy of the net.manga.geek.mangamaster app. The segment discloses a typical cross-border transfer practice, including the transfer intention (blue), the target country (red), the appropriate safeguards (purple), and the means to get a copy of such safeguards (orange). (For interpretation of the references to color in the text, the reader is referred to the web version of this article.)

Feature vector composition. We relied on the *bag-of-words* model to define a set of candidate features based on all distinct terms in the IT-100 corpus policy segments. The rationale for this key assumption is that there is a distribution of individual terms in cross-border transfer practices that is distinct from the distribu-

[8] The corpus has been released at https://github.com/PrivApp/IT100-Corpus

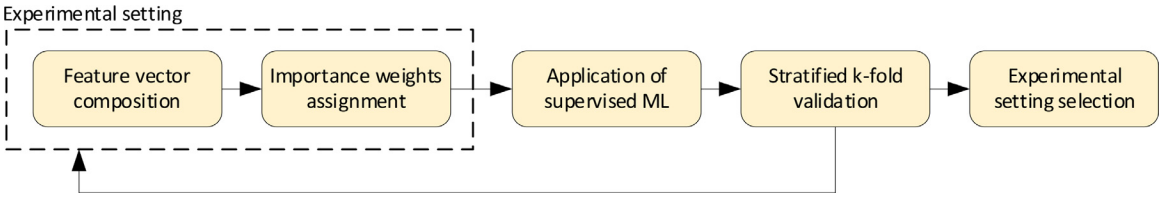


Fig. 5. The overall process to generate the cross-border transfer classification model.

tion of unrelated practices. For example, terms such as *"transfer"*, *"country"* or *"outside"* appear mostly in the segments disclosing cross-border transfers and only marginally in unrelated segments e.g. the term *"transfer"* appears in 87% of the cross-border transfer statements but only in 4% of the other statements. Furthermore, apart from the individual terms, we relied on the *n*-gram model to parse each policy segment into new composite features. We experimented empirically to determine the value of *n*, selecting the range of *n*-grams that provides the highest performance metrics.

Importance weight assignment. We experimented with three different approaches to assigning weights to the selected features: a binary counter (BC), the term frequency (TF), and the Term Frequency-Inverse Document Frequency (*TF.IDF*). The BC encodes the presence or absence of each feature. The *TF* encodes the number of times that each feature occurs in a policy segment. The *TF.IDF* relies on the *TF* to encode the number of times that each feature occurs in a policy segment, but the *IDF* penalizes (decreases) it as the feature x_i occurs across many policy segments. Formally, the *TF.IDF* for the feature x_i is computed as $TF_i . log\frac{N}{n_i}$, where N is the total number of IT-100 corpus policy segments, and n is the number of policy segments that contain the feature x_i.

Application of an ML-supervised algorithm. The above-mentioned feature vectors and their corresponding class labels can be denoted as $S = \{(x_1, y_1), \ldots, (x_n, y_n)\}$, where x_i is the feature vector of the policy segment i, and $y_i \in \{0, 1\}$ indicates the class label of the policy segment. In our case, 1 implies the policy segment i discloses the intention to perform a cross-border transfer, and 0 its absence. By using the training sample S, stemmed from the IT-100 corpus, we used the Support Vector Machine (SVM) technique to find the optimal separation hyperplane that best divides the dataset into the two classes mentioned. SVM has empirically demonstrated better performance over a variety of other ML techniques in high-dimensional spaces, being still effective in cases where the number of dimensions is greater than the number of samples (Moguerza and Muñoz, 2006). Also, prior work (Zimmeck et al., 2017) (Wilson et al., 2018) demonstrated that SVM can reach higher performance than Logistic Regression and Convolutional Neural Networks for privacy practices classification.

Stratified k-fold validation. The IT-100 corpus is an imbalanced dataset i.e. the number of samples tagged as disclosing cross-border transfers is much lower than the number of samples not disclosing it. Accordingly, stratified *k*-fold cross-validation was carried out, establishing $k = 5$. The advantage of the stratified *k*-fold cross-validation is that the entire dataset is used for both training and testing while ensuring that policy segments tagged as disclosing a cross-border transfer practice are represented consistently among all training and validation folds.

Results and experimental setting selection. We performed an empirical evaluation of the effect of iteratively applying the different experimental settings explained above. We relied on Scikit-learn (Pedregosa et al., 2012) and the Natural Language Toolkit (Bird and Loper, 2004) to carry out all our experiments. In each case, we computed the standard performance metrics to compare the models generated. Since the IT-100 corpus is an imbalanced dataset, we primarily used the F-measure instead of the accuracy metric. For models with comparable F-measures, we have favored the model with the highest recall for the sake of more conservative analysis. That is, we seek to prevent a privacy policy that does disclose the intention to perform a cross-border transfer from being tagged as not doing so.

The experimental setting that consistently provided better performance (F-measure of 90.9%) was built on top of a feature vector of stemmed uni- and bigrams and TF as the weighting assignment approach. Also, the best SVM parameters are the *Modifier-Huber loss function* and *SVM alpha* of 10^{-3}. This setting has been used to build the definitive binary classification model to identify the entire policy segments disclosing the intention to perform a cross-border transfer. This is then placed at the entry to the transparency element classifiers explained in the next section.

Transparency elements classifiers. The policy segments disclosing the intention to perform cross-border transfers are further processed to identify specific transparency elements, as per Table 1.

Target country classification. Target countries are disclosed in three different ways in the IT-100 corpus. First, explicit country names or abbreviations (e.g. U.S.) are mostly disclosed in privacy policies. Second, some domain-specific terms were also used to implicitly disclose the target countries. For example, *Privacy Shield* was a certification framework that, until 16 July 2020, ensured an adequacy decision to perform international transfers to the United States. Third, city names rather than country names are also used by a minor number of privacy policies. Our approach, therefore, involves a dictionary of countries, cities, and aliases, whose occurrences are sought in the pertinent policy segments. More specifically, we relied on the CountryInfo dataset[9] which provides details on all countries, including their canonical names, country codes, as well as their states and provinces. We extended it by adding an *'alias'* field to register domain-specific terms implicating a particular country. For example, "*Privacy Shield*" was added as an alias for the United States. Both policy segments disclosing a cross-border transfer and the country dictionary values were first normalized to lowercase. Then, if a non-EU country, state, province, or alias from the dictionary occurs in a policy segment, it is labeled with the identified country/s.

The approach provides high performance in detecting target countries in policy segments with an F-measure of 100% for all countries but for the U.S., which achieved 99%. A couple of policy segments were misclassified due to typos in the country name and compound ways of referring to countries (e.g. *California-based*), which scape the proposed approach.

Appropriate safeguard and copy means. We built one binary SVM classifier (*Adequacy Decision*) and four keyword-based rule classifiers (*Standard Data Protection Clauses, Binding Corporate Rules, Explicit Consent,* and *Means to get a copy of safeguards*) to identify the other individual transparency elements besides the *target country*. We have not generated classifiers for *Approved Certification* and *Approved Code of Conduct* as they are not disclosed in any IT-100 Corpus privacy policy. This makes sense since, so far, the

[9] https://pypi.org/project/countryinfo

EU has adopted only two Codes of Conduct, which focus on cloud providers and thus are out of our scope. The only adopted Certification Scheme for GDPR was approved on 10th October 2022, after our sample was collected.

The binary SVM classifier to identify an *Adequacy Decision* was built by following the same procedure explained in Section 4.2.1.1. The best performance was achieved by a binary SVM classifier built on top of uni- and bigram-based features, and using a TF.IDF weighting approach. The details on the performance achieved by other classifiers we tried are available in the accompanying replication package.[10]

On the other hand, due to the limited number of positive ground truths and the higher performance compared to binary classifiers, we generated keyword-based rule classifiers to identify the remaining four transparency elements. Our approach involves developing a set of rules leveraging that a scoped domain-specific vocabulary is used to refer to them. The set of 117 IT-100-Corpus policy segments disclosing a cross-border transfer were analyzed by a privacy expert, who selected minimum phrases (2–5 terms) that captured the key terms of each transparency element. These phrases were first normalized to lowercase and their grammatical roots by using Porter Stemmer and then turned into a set of rules. For example, the rule ('contract'|' standard') w/4 ('model'|'clause') implies that the normalized form 'contract' or 'standard' must occur before or after the normalized form 'model' or 'clause' by no more than 4 terms in the same sentence. If that rule is satisfied, a policy segment is labeled as disclosing *Standard Data Protection Clauses*. The same procedure was followed for the other transparency elements.

The binary *adequacy decision* classifier achieves high performance (F-measure of 94%), only misclassifying policy segments ambiguously stated. The keyword-based rule classifiers also allowed to correctly identify (F-measure of 100%) the transparency elements related to *Standard Data Protection Clauses, Binding Corporate Rules*, and *Means to get a copy of safeguards*. The *Explicit consent* classifier achieved an F-measure of 75%. Admittedly, their generalization may be hindered since their rules were built based on relatively small positive ground truths. We, therefore, performed a further evaluation on a subset of unseen privacy policies as explained next.

Pipeline validation. We validated the two-layer classification pipeline, i.e., the cross-border transfer intention classifier and the individual transparency element classifiers. To this end, we took advantage of the large-scale compliance assessment method presented in Section 5, which automatically tagged the privacy policies of 4593 apps using the aforementioned classifiers. A cluster sampling was conducted on these tagged privacy policies to randomly select a subset of 30 privacy policies while ensuring a balanced number of each transparency element. These 30 policies were manually annotated and used as ground truth for the evaluation.[11] Since the actual compliance checking is based on an entire privacy policy, we say that a privacy policy discloses a cross-border transfer practice or a transparency element if at least one policy segment contains them.

With F-measures ranging from 85.7% to 100% in the target country classifiers, from 94.4% to 100% in three out of the four *appropriate safeguard classifiers*, and 90.9% in the *copy means classifiers*, we believe that our approach can be exploited to extract these privacy practices at the level of transparency elements with a high degree of certainty.

We examined the misclassification of the appropriate safeguard classifier that showed the lowest performance (F-measure of 54.5%), that is, the explicit consent classifier. The reason was it did not distinguish between tacit and explicit consent. This is an area that could be improved, perhaps through an effort to extract more positive ground truths and then build a robust ML-based classifier. Nevertheless, this classifier achieves perfect recall, thus avoiding pointing out a wrong compliance issue. We further discuss the (mis)use of consent as an appropriate safeguard enabling cross-border transfers in Section 6.

4.2.2. App behavior analysis

This process aims to analyze the behavior of an Android app and then extract the personal data flows in terms of (i) the type of personal data; (ii) the type of recipient who receives the personal data (i.e., first-party or third-party recipient); and, (iii) the country in which the recipient servers are located. This information feeds the compliance checking process to be compared with the practices extracted from the app privacy policy.

We rely on dynamic analysis to observe the app's behavior and extract its personal data flows. Personal data flows can also be inferred from the app's representations or models by using static analysis (del Alamo et al., 2021). However, we favor dynamic analysis to prioritize soundness over completeness, as our goal is to extract actual evidence of cross-border transfers carried out by an app. Furthermore, we rely on app network interfaces as sources of behavior. Previous studies (Lindorfer et al., 2016) have shown the prevalent usage of network interfaces over SMSs or short-range interfaces such as Bluetooth or NFC by mobile apps to communicate externally. Thus, it is fair to assume that most cross-border transfers occur naturally through the network. Fig. 6 sets out the overall data flow extraction process, which is summarized below. Details can be found in Guaman et al. (2021).

Configuration. Based on the Google Play API, we automatically crawl and download the target mobile app (APK) from Google Play Store. Once downloaded, we extract the metadata from the APK digital certificate.

Stimulation. Automated stimulation is based on a random strategy provided by the U/I Exerciser Monkey (UI/Application Exerciser Monkey | Android Developers, n.d.), which provides better performance in terms of code coverage compared to other approaches (Choudhary et al., 2016).

Interception. This component is responsible for capturing the app network traffic and storing it for further analysis. Traffic capture from the device's network interface is built around a man-in-the-middle (MITM) proxy,[12] which requires installing a self-signed CA certificate in the device to become trusted. We leveraged Frida[13] to further bypass the most common countermeasures to HTTPS interception e.g. certificate pinning. After configuring the mobile device to connect to the Internet through the MITM proxy, it is possible to capture both HTTP and HTTPS traffic from any app. We further implemented a flow-to-app mapping component to filter the traffic belonging to the target app since each app is analyzed independently.

Analysis. This component analyses each app traffic to determine (i) the type of personal data transferred by an app, (ii) the country where the personal data recipient is located, and (iii) the type of

[10] The full relative frequency distribution can be found in the replication package available at https://short.upm.es/tqidi

[11] These annotated privacy policies can be found in the sheet *assembled_30_validation.csv* available at the replication package.

[12] https://mitmproxy.org/

[13] https://frida.re/

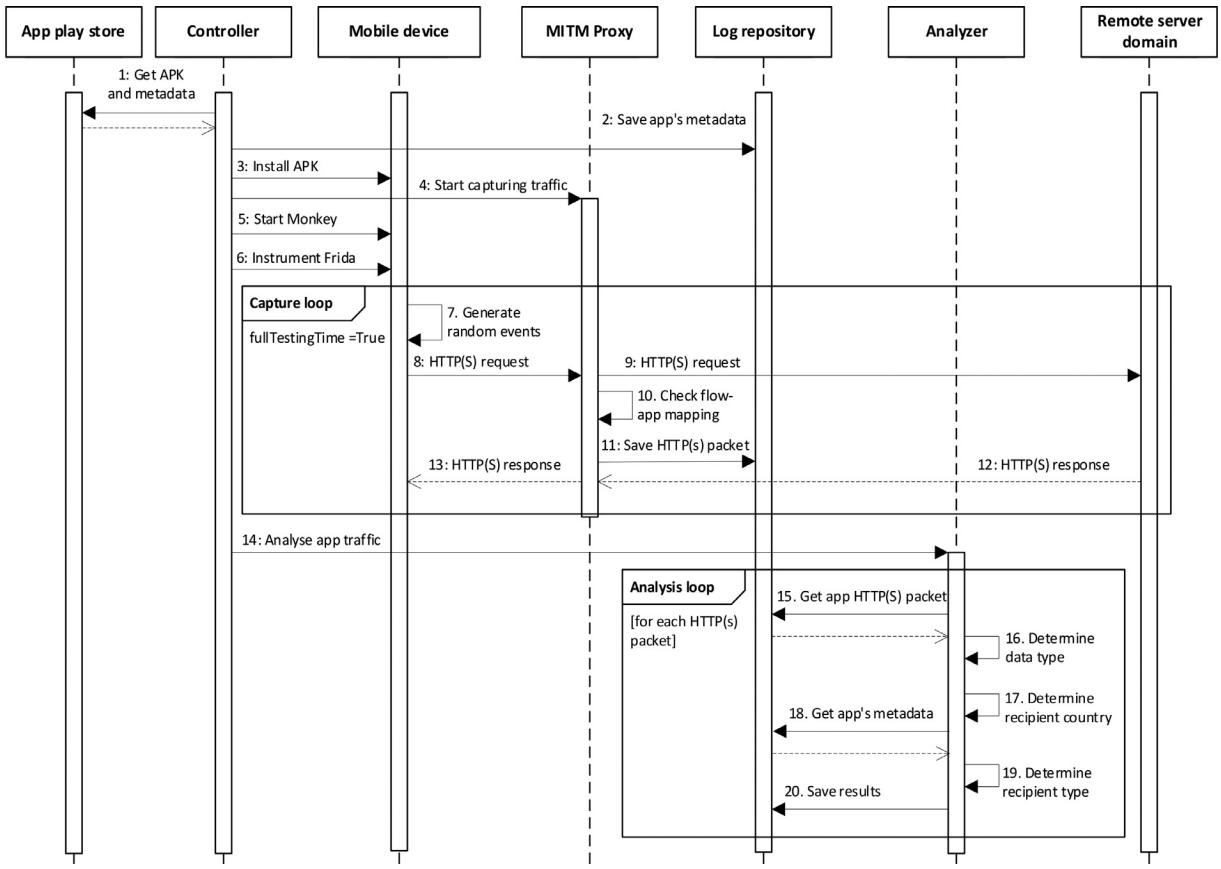


Fig. 6. *The overall* process to extract cross-border personal data flows from apps.

recipient who receives the personal data (i.e., first-party or third-party recipient). For (i), we use string searching in the packet payload. It also searches for data encoded in Base64, MD5, SHA1, and SHA256. For (ii), we relied on ipinfo[14] to determine the location of the servers receiving the app's connections. This IP geolocation service has been recently assessed as achieving almost 95% reliability (Cozar et al., 2022). For (iii), we first performed a token matching between a bag of tokens representing the app and a bag of tokens representing the target domain. The former consists of the second-level domain (SLD) and subdomains from the APK name, the organization name extracted from the digital certificate used to sign the app, and the app name retrieved from the Play Store. The latter consists of the SLD and subdomains of the domain targeted by the traffic. A token matching is then made between the two bags, classifying the domain as a *first-party recipient* if there is at least one token match. Domains not classified as *first-party recipients* were searched in webXray[15]; if found, they were classified as *third-party recipients*. This dataset[16] has been created in the specific context of disclosing personal data to third parties in the web and mobile ecosystem. It maps individual target domains to the owner company and even to parent companies, including the country in which the headquarters are located and the service category. Domains not classified as *first-* or *third-party recipients* were classified as *unknown* and excluded from further analysis.

4.2.3. Compliance checking

This final process checks whether the apps performing cross-border transfers properly disclose them through their privacy policies. There are four possible outcomes, which are explained below.

Full cross-border transfer disclosure. It implies that a privacy policy discloses all *transparency elements* according to the type of cross-border transfer carried out by an app. For illustrative purposes, consider the "Alibaba.com" com.alibaba.intl.android.apps.poseidon app, owned by Alibaba Mobile. We found that it transfers the Google Advertisement Identifier to, inter alia, www.googleadservices.com, whose servers are located in the United States (US). The domain googleadservices.com is owned by the third-party recipient Google LLC, which is based in the US.

The privacy policy of the app includes the statement shown in Fig. 7, which fully discloses the required transparency elements: the transfer intention (green); target countries (yellow); appropriate safeguard - Standard Data Protection Clause (blue); and the means to get a copy of these clauses (gray). Therefore, in this specific case, we classify this app as a full cross-border transfer disclosure.

Ambiguous cross-border transfer disclosure. It implies that a privacy policy includes only a subset of the *transparency elements* required by the GDPR according to the type of cross-border transfer carried out by the app. The missing *transparency elements* are either included in an ambivalent manner or not included at all. For example, consider the "Pou" (me.pou.app) app, owned by Zakeh Ltd. We found it transfers the Google Advertisement Identifier to, inter alia, adc3-launch.adcolony.com, whose servers are located in the United

[14] https:// ipinfo.io/

[15] https://webxray.org/

[16] As a further contribution of this study, we added 234 new domains to the dataset maintained by a research community, available at https://github.com/PrivApp/webXray_Domain_Owner_List

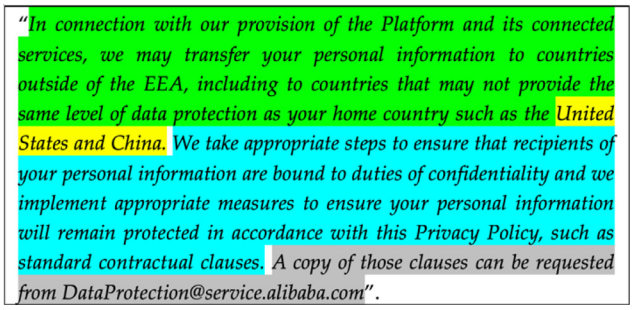

"In connection with our provision of the Platform and its connected services, we may transfer your personal information to countries outside of the EEA, including to countries that may not provide the same level of data protection as your home country such as the United States and China. We take appropriate steps to ensure that recipients of your personal information are bound to duties of confidentiality and we implement appropriate measures to ensure your personal information will remain protected in accordance with this Privacy Policy, such as standard contractual clauses. A copy of those clauses can be requested from DataProtection@service.alibaba.com".

Fig. 7. International transfer statement of the *com.alibaba.intl.android.apps.poseidon* app.

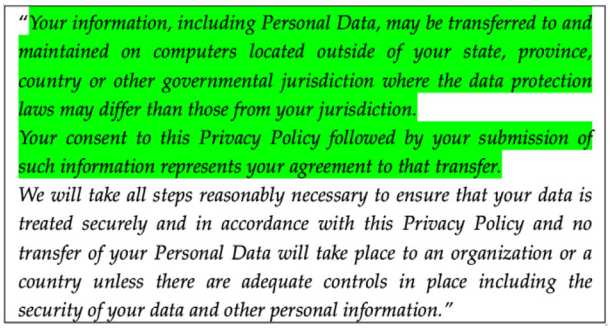

"Your information, including Personal Data, may be transferred to and maintained on computers located outside of your state, province, country or other governmental jurisdiction where the data protection laws may differ than those from your jurisdiction.
Your consent to this Privacy Policy followed by your submission of such information represents your agreement to that transfer.
We will take all steps reasonably necessary to ensure that your data is treated securely and in accordance with this Privacy Policy and no transfer of your Personal Data will take place to an organization or a country unless there are adequate controls in place including the security of your data and other personal information."

Fig. 8. International transfer statement of *me.pou.*app.

States. This domain is owned by the third-party recipient AdColony which is based in the United States.

The privacy policy of the app includes the statement shown in Fig. 8, which discloses the *intention* to transfer personal data (green). However, it does not reveal the target countries nor *appropriate safeguards* and the *means to get a copy* of such safeguards. Despite this app provider appeals to the consent, the identifier is transferred to me.pou.app before the user interacts with the app for the first time, nullifying any attempt to underpin the transfer by *explicit consent*, as explained in Section 3.

Inconsistent cross-border transfer disclosure. It implies that a privacy policy includes statements that contradict the cross-border transfers carried out by an app. For illustrative purposes, consider the "CATS: Crash Arena Turbo Stars" (com.zeptolab.cats.google) app, owned by ZeptoLab. We found it transfers the Device Model and the Kernel Build Number to, inter alia, the domain in.appcenter.ms, whose servers are located in the United States. This domain is owned by the third-party recipient Microsoft Corporation based in the US.

The privacy policy of the app includes the statement shown in Fig. 9. It properly discloses all three transparency elements, i.e., the *intention* to transfer personal data (green) to a country (yellow) covered by an *adequacy decision* (blue). However, an international transfer is made to the United States, which does not hold an adequacy decision. Therefore, in this specific case, we classify this app as *inconsistent cross-border transfer disclosure*.

Omitted cross-border transfer disclosure. It implies that a privacy policy does not include any *transparency element* when an app performs a cross-border transfer. For example, consider the "Plague Inc." (com.miniclip.plagueinc) app, which transfers the Device Model and Kernel Build Number to 4 different third-party recipients engaged with advertisement. Those recipients are mainly US-based companies, including Meta, Unity, InMobi, and Supersonic Studio. Meta was the only company whose servers were located in Spain and therefore the only one not involved in cross-border transfers. We couldn't find a policy statement describing at least the intention to make a cross-border transfer. We, therefore, classify this app as *omitted cross-border transfer disclosure*.

5. Evaluation

To demonstrate the utility and assess the efficacy of the developed artifact, as well as to advance the fundamental understanding of GDPR compliance in the Android ecosystem, we carried out a compliance assessment of cross-border transfers of apps from the Google Play Store in Spain. In particular, after defining the experimental environment setup (Section 5.1), we examined how many apps conducted cross-border transfers and how many of them disclosed properly such practices through privacy policies according to the type of cross-border transfer (Section 5.2). Finally, we compare our approach with the related work (Section 5.3).

5.1. Experimental environment

We conducted a controlled experiment by using the automated assessment method presented in Section 4. Both the Android mobile apps and their privacy policies were downloaded and tested between 2nd September and 19th September 2022 from Spain. The apps were installed and tested on five mobile devices Xiaomi Redmi 10 (API 30). Each app was run for 5 min considering two phases: idle stage (i.e., 2 min without user interaction) and active stage (i.e., 3 min with automated stimulation). Before starting the active stage, all permissions requested by the app were automatically granted.

We downloaded plain-text privacy policies from Google Play Store using the Selenium WebDriver[17] into a headless Chrome browser, allowing the execution of dynamic content such as JavaScript. We relied on these privacy policies as the European Data Protection Board (EDPB) explicitly states that, for apps, the necessary should be made available from an online store prior to download (European Commission, 2018). Non-English privacy policies[18] and their apps were excluded from the analysis. Since the classification models presented in Section 4.2.1 operate based on policy segments, each privacy policy was broken down accordingly, using the full stop as a paragraph separator. Each policy segment was finally pre-processed and then fed into the classification pipeline for extracting the cross-border transfer practices.

5.1.1. Apps selection

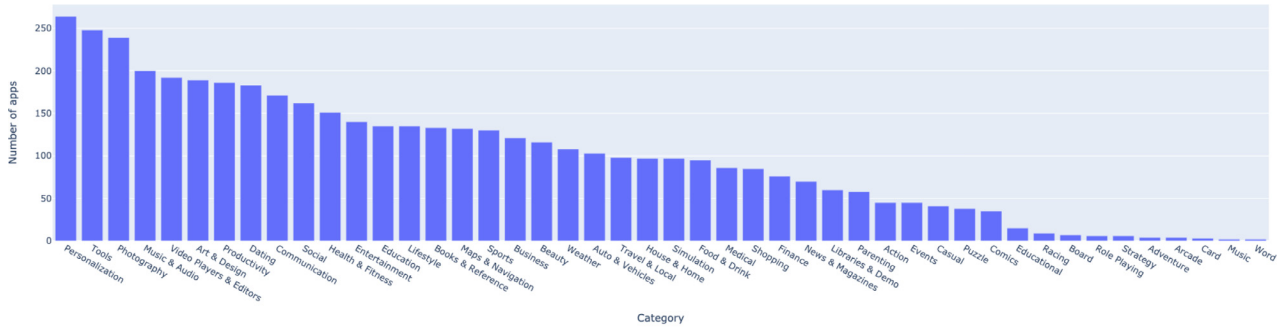
To have a representative sample of the apps mostly used by EU data subjects, the main criterion guiding the selection of apps was their popularity within an EU country. We relied on Google Play Store's categorization of the top-free apps in Spain to download a set of 4593 apps, which have been highly downloaded as shown in Table 2. These set of apps are distributed across the different categories available on Google Play Store (Fig. 10a). Furthermore, based on the *Issuer Locality* field of digital certificates used to sign the apps, we observe that apps have been signed by app providers coming from 94 different countries, thus ensuring diversity in the geographical location of app providers. Interestingly, a vast majority of apps (65%) have been signed by non-EU app providers, in particular by providers based in the United States, while only the

[17] Selenium Chrome WebDriver available at https://www.selenium.dev/documentation/en/webdriver [Accessed: 17-Feb-2023]

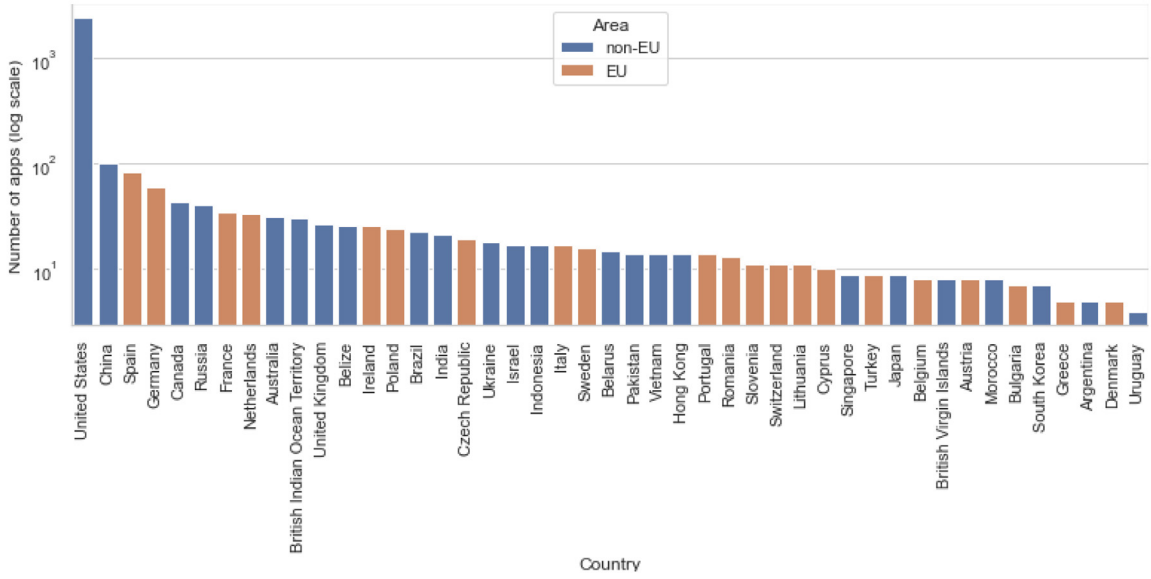
[18] LangDetect was used to determine whether the majority of a privacy policy was written in English. Available at https://pypi.org/project/langdetect/ [Accessed: 17-Feb-2023]

> "Your information, including Personal Data, may be transferred to — and maintained on — computers located outside of your state, province, country or other governmental jurisdiction where the data protection laws may differ than those from your jurisdiction.
>
> If you are located outside Viet Nam and choose to provide information to us, please note that we transfer the data, including Personal Data, to Viet Nam and process it there.
>
> Your consent to this Privacy Policy followed by your submission of such information represents your agreement to that transfer.
>
> VideoMaker will take all steps reasonably necessary to ensure that your data is treated securely and in accordance with this Privacy Policy and no transfer of your Personal Data will take place to an organization or a country unless there are adequate controls in place including the security of your data and other personal information."

Fig. 9. International transfer statement of the *com.forqan.tech.Jobs* app.



(a)



(b)

Fig. 10. Distribution of (a) categories and (b) providers' locality for the apps analyzed.

9% of apps were signed by EU app providers (Fig. 10b). The remaining 26% of apps do not provide *Locality Issuer* information in digital certificates.

5.1.2. App's data flow dataset

A total of 112,746 data flows generated by the 4593 apps have been logged. Each log includes the app name, app version, capture stage (idle or active), target domain, target country, and personal data type disclosed (if any). From these flows, 28,788 flows (25.53%) have disclosed at least one of the personal data types to 684 unique fully-qualified domain names with 459 unique second-level domains (SLDs). These SLDs are hosted on servers located across 23 different countries, 7 EU countries, and 16 non-EU countries hosting 220 (43.7%) and 283 (56.3%) SLDs, respectively. Note that the sum of both exceeds the aforementioned 459 unique SLDs because 44 of them are hosted on servers located in both EU and non-EU countries.

5.1.3. Privacy policy dataset

The 4593 privacy policies were fed into the classification models irrespective of whether cross-border transfers were detected during the testing time of their corresponding apps. A total of 142,753 policy segments have been analyzed. Of these, 3224 contain transparency elements of cross-border transfer practices.

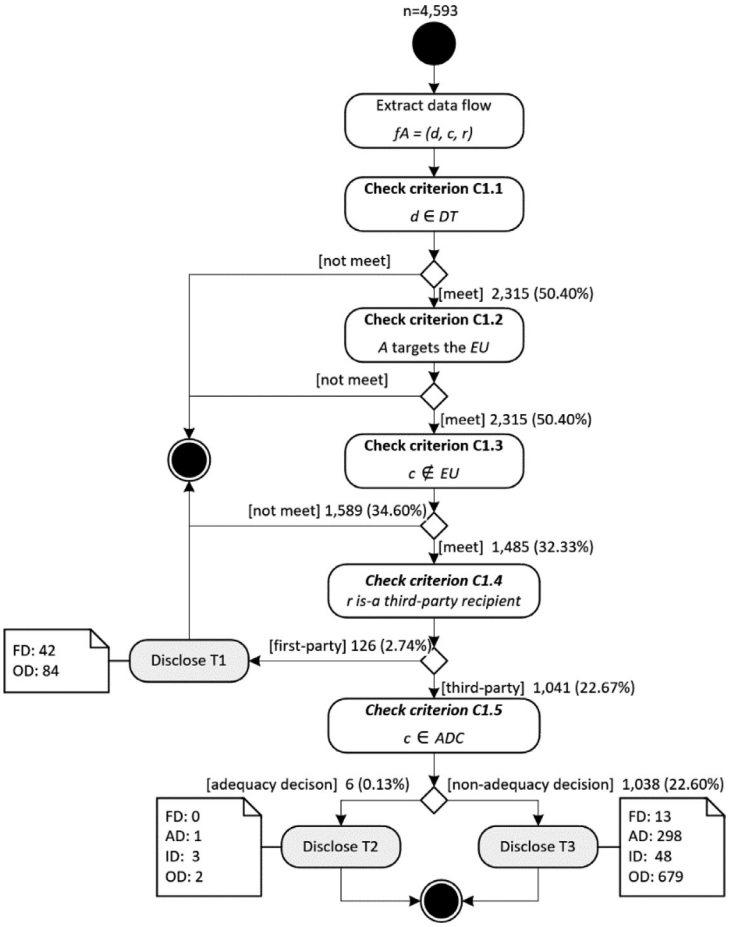


Fig. 11. The number of apps performing cross-border transfers.

Table 2
Distribution of apps per downloads.

# downloads	# apps (%)
10,000,000,000	2 (0.04%)
5000,000,000	1 (0.02%)
1000,000,000	8 (0.17%)
500,000,000	5 (0.11%)
100,000,000	92 (2.00%)
50,000,000	125 (2.72%)
10,000,000	636 (13.85%)
5000,000	455 (9.91%)
1000,000	1100 (23.95%)
500,000	425 (9.25%)
100,000	919 (20.01%)
50,000	249 (5.42%)
10,000	377 (8.21%)
<10,000	199 (4.33%)

5.2. Compliance results

The number of apps that fulfill the criteria leading to the four types of cross-border transfers is detailed in Fig. 11. It was found that *half* (2315) of the apps transferred some type of personal data during the testing period. From them, a subset of 830 apps (18% of the total) transferred personal data solely to other EU countries. As explained in Section 3, these do not imply further requirements in terms of GDPR. The remaining 1485 (32%) apps transferred personal data outside the EU. This subset of apps branches out into three groups that imply different transparency requirements: a small subset of 126 apps transferred personal data to non-EU first-party recipients (Section 5.2.1), a reduced subset of 6 apps transferred personal data to third-party recipients covered by an adequacy decision (Section 5.2.2), and, finally, a substantial subset of 1038 apps transferred personal data to third-party recipients not covered by an adequacy decision (Section 5.2.3). Note that some apps have performed more than one cross-border transfer type. For example, out of 1041 apps that transferred personal data to a non-EU country (C1.3.2 in Fig. 11), a subset of 3 apps transferred personal data to both non-EU third-party recipients and non-EU first-party recipients. Thus, the sum up of both subsets of apps (1038 and 6) exceeds its input in 3.

5.2.1. Transfers to non-EU data controllers

During the testing time, a subset of 126 apps performed cross-border transfers to domains belonging to them but hosted on servers outside the EU.[19] In particular, the United States is the country that hosts the vast majority (88%) of first-party recipient domains. The remaining 12% of these apps host their domains in 12 different countries, including the United Kingdom, Russia, Singapore, South Korea, Brazil, India, China, Hong Kong, Egypt, Australia, Iran, and Japan.

A third (42) of these apps inform accordingly about an EU representative or data controller[20] and therefore have been classified as **full cross-border transfer disclosures (FD)**, while the remaining two-thirds **omitted cross-border transfer disclosures (OD)**.

[19] Details on non-EU data controller cross-border transfers can be found the *T1_results.csv* available at the replication package.

[20] Strictly speaking, it is only a representative, but we have not noticed that some privacy policies refer to it as an EU data controller.

To perform a more in-depth analysis, we relied on the *Issuer Locality* information of each app digital certificate to identify the app provider location and thus distinguish those that appear to be established within the EU from those established outside the EU. To be sure of the reliability of this assumption, we manually checked the consistency between the supposed issuer country extracted from 40 randomly selected apps and the organization's country (if any) disclosed in their respective privacy policies, finding 13 inconsistencies.[21]

As a result, a remarkable 65% of apps providers (83) would be established outside the EU, compared to a small 12% of apps (16) within the EU, while the *Issuer Locality* of the remaining 23% (27) of apps is missing from their digital certificates. Focusing on the 83 non-EU app providers, a substantial 62% (52) of apps omitted the disclosure of an EU representative/data controller in the EU in their privacy policies, thus raising a potential compliance issue. Four key different app providers' attitudes can be stemmed after analyzing in-depth some involved privacy policies.

Several app providers do disclose the intention to perform a transfer outside the EU to their premises in a specific country but appeal to tacit consent as a unique enabling mechanism to perform a cross-border transfer, which is not valid under the GDPR. For example, the *flipboard.app*, a News & Magazine app resorts to the following statement: "*As a California-based company, we store and use personal data outside the EU. By using our websites or submitting your personal data, you consent to such transfer, storing and processing*".

Second, some app providers explicitly state that the underlying service targets a specific country market, other than the EU, although disclaim liability for any cross-border transfers of personal data of those who still choose to use them. For example, the *com.hulu.plus*, a popular video stream app disclaims the following: "*Hulu is headquartered in the U.S. and the Hulu Services are intended for users in the U.S. By viewing any Content or otherwise using the Hulu Services, you consent to the transfer of information to the U.S. to the extent applicable, and the collection, storage, and processing of information under U.S. laws*.".

Some app providers inform about the collection of personal data but omit any applicable regulations at all, so they cannot be expected to designate an EU data representative or controller.

Finally, a minor number of apps (8) still rely on the Privacy Shield Framework as an enabling mechanism to perform a cross-border transfer without further constraints, as it was covered by an EU adequacy decision. Yet, as already mentioned, it was invalidated by the EU Court of Justice on July 16, 2020.

5.2.2. Transfers covered by an adequacy decision

A small set of 6 apps transferred personal data to third-party domains hosted on servers located in countries covered by an EU adequacy decision.[22] In particular, Japan (3 apps), Canada (1), and the United Kingdom (2) were targeted by the apps involved. Since all these countries maintain an adequacy decision any transfer can be made without any further safeguards but, as explained in Section 3, they must disclose the transfer intention, names of targeted countries, and the existence of an adequacy decision itself.

However, no app disclosed all three transparency elements and therefore have been classified as **FD.** One app has been classified as **ambiguous cross-border transfer disclosure (AD),** as despite it disclosed the intention to perform a cross-border transfer, it failed to inform the target countries. Three apps have been classified as **inconsistent cross-border transfer disclosures (ID)**, because despite disclosing the transparency elements required, there is a disagreement between the countries to which the data are transferred, and the countries disclosed in their privacy policies. Finally, two apps omitted disclosure of cross-border transfer practices at all and have been therefore classified as **OD.**

5.2.3. Transfers not covered by an adequacy decision

A large subset of apps (1038) transferred personal data to third-party recipients located in countries that are not covered by an adequacy decision.[23] In particular, we found that the vast majority of these apps (98%) performed cross-border transfers to the United States. Other less popular third-party recipients are located in seven different countries: Singapore (144, 13%), Russia (18, 1.7%), China (16, 1.5%), Hong Kong (3, 0.3%), India (2, 0.2%), Brazil (1, 0.1%), Iran (1, 0.1%). The sum of the percentages exceeds 100% because 168 of the apps performed cross-border transfers to more than one of the mentioned destinations.

A reduced 1.3% (13) of these apps have disclosed the required transparency elements and therefore have been classified as **FD.** Besides, 4.6% (48) of these apps were classified as **ID** as they do disclose the intention to perform a cross-border transfer but there is a disagreement between the countries targeted by transfers and the countries disclosed in the apps' privacy policies. Also, 28.7% (298) of these apps disclose the intention to perform a cross-border transfer but omit one or more of the other three transparency elements and therefore were classified as **AD.** They all fail to inform of the *destination country.* Interestingly, 44 of them performed transfers to the United States and still rely on the Privacy Shield, which was invalidated in 2020, and thus fall into non-compliant ones. Another remarkable aspect is that several apps (64) disclose the implementation of appropriate safeguards (almost all of them through the establishment of *Standard Data Protection Clauses*) but fail to provide data subjects a means to obtain a copy of these safeguards (e.g., an email or download URL). Also, a concern that arises for all types of cross-border transfers is that the privacy policies of several applications use ambivalent statements, such as "*countries around the world*", *"outside the EEA"* or "*any country in which we do business",* to refer to the targeted countries. Finally, a significant 65.4% (679) of apps have been classified as **OD** as neither the transfer intention, the recipient countries, nor the appropriate safeguards were disclosed by their privacy policies.

5.3. Comparison with state-of-the-art approaches

Eskandari et al. (2017) propose PDTLoc, an analysis tool that uses static analysis to detect violations of article 25.1 of the old EU Data Protection Directive (replaced now by the GDPR). Both the Directive and GDPR set similar requirements for cross-border transfers. Yet, PDTLoc did not consider either the recipient type or the privacy policy as a means of disclosing the measures that do enable cross-border transfers. In contrast, our approach leverages the automated privacy policy analysis to extract all transparency elements from stated cross-border transfer practices.

Our approach enables (i) more accurate detection of non-compliant apps and (ii) a broader scope not only considering international transfers but also other types of cross-border transfers (i.e., T1, T2, and T3 in Fig. 1). In the experiment, the subset of 126 apps performing transfers to non-EU data controllers (T1-Section 5.2.1) would have been classified as non-compliant by PDTLoc. However, 42 apps (33%) do inform accordingly about an EU representative or data controller and should therefore be classified

[21] Details can be found in the *Checking_issuer_locality.csv* sheet available at the replication package.

[22] Details on adequacy decision-based cross-border transfers can be found in the *T2_results.csv* available in the replication package.

[23] Details on non-adequacy decision-based cross-border transfers can be found the *T3_results.csv* available at the replication package.

as **full cross-border transfer disclosures (FD),** i.e. compliant. Thus, PDTLoc would have had a 33% false positive rate.

Similarly, the subset of 6 apps performing transfers covered by adequacy decision (T2-Section 5.2.2) would have been directly classified as non-compliant by PDLoc. Although our approach has also classified the 6 apps as non-compliant, it provides useful results by informing stakeholders that the disclosure of the cross-border transfer of one app is ambiguous (as failed to inform target countries), the disclosure of the cross-border transfer of three apps is inconsistent (as there is a disagreement between countries disclosed in the privacy policies and countries transferred) or justly two apps omitted the disclosure of the cross-border transfer practices.

Finally, the subset of 1038 apps performing transfers not covered by an adequacy decision (T3-Secction 5.2.3) would have been directly classified as non-compliant by PDLoc. Yet, 13 apps are compliant as they do disclose all required transparency elements. The rest are non-compliant because the disclosure is ambiguous, inconsistent, or omitted. In this case, PDTLoc would have had a false positive rate of 1.33% and would not have provided useful reports on the type of non-compliance.

6. Discussion

A substantial 48%[24] of analyzed apps that send personal data are potentially non-compliant with the GDPR cross-border transfer requirements. Despite efforts to ensure cross-border transfers through GDPR, the results reveal that there is still a very significant gap between what app providers do in practice and what is intended by GDPR. The results show that 48% of popular apps that send personal data from an EU country are potentially non-compliant with GDPR cross-border transfer requirements. In particular, 33% of those apps that send personal data do not disclose these practices *at all*, while the remaining 15% partially disclose them in an ambiguous and/or inconsistent way.

Explicit consent is a fair enabler of cross-border transfers, but it is being misused. In the absence of an *adequacy decision* or any *appropriate safeguards*, the *explicit consent*[25] could also enable cross-border transfers. Nevertheless, *explicit consent* requires a clear affirmative action of the data subjects, e.g., ticking a box, to be obtained **after** providing **precise details** of the international transfers. As such, explicit consent removes the possibility of using the dark pattern of pre-ticked boxes or tacit consent. To observe the prevalence of disclosing (explicit) consent as an enabler of cross-border transfers, we further analyzed the 359 apps that disclose the intention to perform a cross-border transfer but not the appropriate safeguard implementation. Around 72% (260) of them resort to consent as the only enabler of cross-border transfers. We randomly selected a subset of 50 privacy policies and examined the consent-labeled statement. Interestingly, in all cases, app providers appealed to *tacit consent* stating that the usage of the app by the data subject implies consent to transfer personal data outside the EU. In this line, we observed that most of them (255) performed cross-border transfers during the idle testing stage, i.e., before the user interacts with the app, thus nullifying any attempt to underpin the transfer by *explicit consent*.

Automated means for compliance assessment are key. The need for automated methods and tools to evaluate privacy requirements is essential in an evolving regulatory landscape. In 2016, the EU-US Privacy Shield was introduced as a limited adequacy decision to allow the transfer of personal data to US-based third-party recipients that were certified under the terms of this framework. US-based app providers extensively used this framework, as evidenced in this study, which allowed these providers to legally target EU data subjects and transfer personal data between the EU and the US. However, this framework was invalidated by the Court of Justice of the EU on July 16, 2020 (Publications Office of the European Union, 2020). The automated approach presented in this paper found the still prevalent use of the EU-US Privacy Shield in the privacy policies of apps (Section 5.2.3) despite its invalidity. This shows the relevance of our approach in a landscape that is constantly evolving.

Along the same line, fully automated methods and techniques offer a viable alternative for large-scale assessment of app compliance with different requirements. The open model of the Android mobile ecosystem allows a vast number of apps to be offered globally by developers everywhere. Hence, manual review by control authorities or distribution platform operators is impossible. The mean time for downloading and tagging cross-border transfer practices in a privacy policy was 10s, so with a single instance up to 8640 privacy policies can be tagged in 24h. We strongly believe that our approach can be leveraged for large-scale compliance scrutiny of cross-border transfers.

A supporting tool for developers. Whilst our proposed method mainly aims to support audits by authorities or distribution platforms, such as Google Play Store, it can also support app developers in ensuring the compliance of their apps. It should be recognized that the mobile application ecosystem is a mix of formal requirements established by prevailing regulations such as GDPR, along with informal developer training. App developers can range from hobbyists to experienced professionals in large companies. As found in previous work (Balebako et al., 2014), while large companies may be able to form multidisciplinary teams to enforce legal requirements, small business developers may struggle to understand the privacy and data protection implications of their code. We consider that the necessary multidisciplinary knowledge, including evaluation criteria supported by legal and not only technical interpretations, can be simplified into indicators that can be checked automatically. This paper, as well as other related work (Zimmeck et al., 2017, 2019), proved that (at least part of) such multidisciplinary knowledge can be embedded into GDPR automated assessment pipelines. While these automatic approaches do not act as *infallible judges*, they have the potential to alert developers about possible non-compliance issues.

6.1. Threats to validity

6.1.1. Construct validity

The classification models build upon the Corpus IT-100, which is an annotated dataset of legal requirements laid down in the GDPR. Therefore, there is a risk that such a Corpus does not reflect the construct under study when moving legal requirements to the technical domain. To mitigate this threat, the elaboration of the Corpus IT-100 was undertaken by privacy and data protection experts who comprehensively guided the building of the annotation process, annotation scheme, and the simplified assumptions used during the annotation process of cross-border transfer practices in privacy policies.

6.1.2. Internal validity

If policies in languages other than English were excluded, a bias toward the evaluation of non-EU-based applications could occur. Surprisingly, we observed that the providers of the applications usually include privacy policies in English. In our dataset, only

[24] Note that that some apps have performed more than one type of cross-border transfer. Therefore, an app has been classified as fully compliant only if all individual transfers have been classified as full cross-border transfer disclosures. Detailed results of each app can be found at the replication package.

[25] GDPR Art. 49

20.9% of apps published non-English policies exclusively, and 15% were in Spanish.

Our automated privacy policy analysis approach, like any approach based on statistical learning, exhibits problems of misclassification that should be considered. Thus, as pointed out in the different results in Section 4, although high F-measure values are achieved (from 85.7% to 100%) there is the possibility of a small number of other classifications. All in all, while it certainly does not act as an infallible judge, we highlight the performance of the cross-border transfer intention classifier which does not exhibit any misclassification in a subset of randomly selected privacy policies, demonstrating its potential to alert stakeholders of potential non-compliance issues.

6.1.3. External validity

Since the current implementation of the assessment method is built upon dynamic analysis techniques, it inherits the same limitation faced by them. The use of non-standard encoding mechanisms (Reyes et al., 2018), unusual TLS certificate pinning implementations (Razaghpanah et al., 2017), and sub-optimal coverage of app execution paths (Patel et al., 2018) are some particular open orthogonal challenges to our proposal, which can generate false negatives. Therefore, it cannot ensure completeness and the results of fully compliant apps should not be misleadingly generalized. The fact that we have not observed a cross-border transfer during our testing period does not mean that an app will not do so if its developers, e.g., use customized encoding mechanisms.

All in all, potential false negatives do not put at risk the validity of the results of non-compliant apps, which is remarkably high (48%). The strength of dynamic analysis techniques is that evidence of non-compliant apps stems from real app behavior and does not generate false positives. Therefore, we consider that our proposal, as well as the results, are valuable for app providers, app distribution platforms such as Google Play Store, and supervisory authorities to detect the lower bound of non-compliance issues with GDPR cross-border transfers.

7. Conclusion and future work

In this work, we presented a fully automated method to assess the compliance of mobile apps with the cross-border transfer requirements established by the GDPR. With an F-measure ranging from 85.7% to 100% in identifying the different cross-border transfer transparency elements, our approach can be exploited to extract these privacy practices with a high degree of certainty and at scale.

We applied the automated compliance assessment method to determine the extent to which apps from the Google Play Store comply with the cross-border transfer requirements of the GDPR. After evaluating 4593 apps, the results revealed that there is still a great gap between what app providers do in practice and what is intended by GDPR. Notably, 1115 (48%) apps that sent personal data failed (completely or partially) to comply with the regulations, either because their privacy policies include ambiguous or inconsistent disclosures about cross-border transfers, or they simply omit them.

In a complex and evolving regulatory landscape, automated methods and tools to evaluate privacy requirements are essential to several stakeholders, including supervisory authorities, distribution platforms, and developers. Our current efforts are aimed at extending the analysis of privacy policies disclosed in other languages. Likewise, we aim to extend our method to address other requirements of the GDPR that have not yet been the subject of research efforts, in particular those related to *automated decision-making*.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. Jose M. del Alamo reports financial support was provided by Polytechnic University of Madrid.

CRediT authorship contribution statement

Danny S. Guamán: Conceptualization, Methodology, Data curation, Software, Writing – original draft, Writing – review & editing. **David Rodriguez:** Data curation, Software, Writing – review & editing. **Jose M. del Alamo:** Conceptualization, Methodology, Writing – original draft, Writing – review & editing, Supervision. **Jose Such:** Conceptualization, Writing – review & editing.

Data availability

The link to the data has been added to the manuscript

Acknowledgment

This research has been partially supported by the project TED2021-130455A-I00 funded by MCIN/AEI/10.13039/5011000 11033 and the Europea Union “NextGenerationEU”/PRTR, and by the Escuela Politécnica Nacional in Ecuador.

References

del Alamo, J.M., Guaman, D., Balmori, B., Diez, A., 2021. Privacy assessment in android apps: a systematic mapping study. Electronics 10, 1999. doi:10.3390/ELECTRONICS10161999, (Basel)Page2021;10:1999.

del Alamo, J.M., Guaman, D.S., García, B., Diez, A., 2022. A systematic mapping study on automated analysis of privacy policies. Computing 104, 2053–2076. doi:10.1007/S00607-022-01076-3/FIGURES/5.

Andow, B., Whitaker, J., Enck, W., Reaves, B., Mahmud, S.Y., Singh, K., et al., 2020. Actions speak louder than words: entity-sensitive privacy policy and data flow analysis with PoliCheck. In: Proceedings of the 29th USENIX Security Symposium, pp. 985–1002.

Android Developers. Google play protect 2020. https://developers.google.com/android/play-protect/phacategories?hl=en (accessed March 11, 2023).

Apple Developers. App privacy details on the app store 2020. https://developer.apple.com/app-store/app-privacy-details/(accessed March 11, 2023).

Balebako, R., Marsh, A., Lin, J., Hong, J., Cranor, L.F., 2014. The privacy and security behaviors of smartphone app developers. Internet Society doi:10.14722/usec.2014.23.

Bird S., Loper E. NLTK: the natural language toolkit 2004:214–7.

Choudhary, S.R., Gorla, A., Orso, A., 2016. Automated test input generation for android: are we there yet? In: Proceedings of the - 30th IEEE/ACM International Conference on Automated Software Engineering, pp. 429–440. doi:10.1109/ASE.2015.89 ASE 2015.

Cozar, M., Rodriguez, D., Del Alamo, J.M., Guaman, D., 2022. Reliability of IP geolocation services for assessing the compliance of international data transfers. In: Proceedings of the IEEE European Symposium on Security and Privacy Workshops (EuroS&PW). IEEE, pp. 181–185. doi:10.1109/EuroSPW55150.2022.00024.

Eskandari, M., Kessler, B., Ahmad, M., Oliveira, A.S.de, Crispo, B., 2017. Analyzing remote server locations for personal data transfers in mobile apps. Undefined 2017, 118–131. doi:10.1515/POPETS-2017-0008.

European Commission. Guidelines on transparency under Regulation 2016/679. 2018.

European Commission. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance) 2016.

European Commission. Opinion 02/2013 on apps on smart devices 2013.

European Cyber Security Organization. European Cyber Security Certification: A Meta-Scheme Approach v1.0. 2017. https://ecs-org.eu/?publications=https-ecs-org-eu-documents-publications-5a3112ec2c891-pdf (Accessed: January 5, 2023).

European Union, 2010. Charter of Fundamental Rights of the European Union, 53. European Union, Brussels.

European Union Agency For Network and Information Security. A study on the app development ecosystem and the technical implementation of GDPR 2017. 10.2824/114584.

Fan, M., Yu, L., Chen, S., Zhou, H., Luo, X., Li, S., et al., 2020. An empirical evaluation of GDPR compliance violations in android mhealth apps. In: Proceedings of the - International Symposium on Software Reliability Engineering, ISSRE 2020, pp. 253–264. doi:10.1109/ISSRE5003.2020.00032 -October.

Ferrara, P., Spoto, F., 2018. Static analysis for GDPR compliance. In: Proceedings of the Italian Conference on Cybersecurity, 2058, pp. 1–10.

Guaman, D.S., del Alamo, J.M., Caiza, J.C., 2021. GDPR compliance assessment for cross-border personal data transfers in android apps. IEEE Access 9, 15961–15982. doi:10.1109/ACCESS.2021.3053130.

Gurses, S., 2014. Can you engineer privacy? Commun. ACM 57, 20–23. doi:10.1145/2633029.

Hevner, March, Park, Ram, 2004. Design science in information systems research. MIS Quarterly 28, 75. doi:10.2307/25148625.

Jia, Q., Zhou, L., Li, H., Yang, R., Du, S., Zhu, H., 2019. Who leaks my privacy: towards automatic and association detection with GDPR compliance. Lecture Notes Comput. Sci. 11604, 137–148. doi:10.1007/978-3-030-23597-0_11/TABLES/3, (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)LNCS.

Lin, X., Liu, H., Li, Z., Xiong, G., Gou, G., 2022. Privacy protection of China's top websites: a multi-layer privacy measurement via network behaviours and privacy policies. Comput. Secur. 114, 102606. doi:10.1016/J.COSE.2022.102606.

Lindorfer, M., Neugschwandtner, M., Weichselbaum, L., Fratantonio, Y., Veen, V.v.d., Platzer, C., 2016. ANDRUBIS - 1,000,000 apps later: a view on current android malware behaviors. In: Proceedings of the 3rd International Workshop on Building Analysis Datasets and Gathering Experience Returns for Security, BADGERS 2014, pp. 3–17. doi:10.1109/BADGERS.2014.7.

Mangset P.L. Analysis of mobile application's compliance with the general data protection regulation (GDPR). NTNU (Master's Thesis) 2018.

Moguerza, J.M., Muñoz, A., 2006. Support vector machines with applications. Stat. Sci. 21, 322–336. doi:10.1214/088342306000000493.

Nissenbaum, H., 2004. Privacy as contextual integrity. Washington Law Rev. 79, 101–139.

Patel, G., et al., 2018. On the effectiveness of random testing for android: or how I learned to stop worrying and love the monkey; on the effectiveness of random testing for android: or how i learned to stop worrying and love the monkey. In: Proceedings of the IEEE/ACM 13th International Workshop on Automation of Software Test (AST), 18 doi:10.1145/3194733.3194742.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al., 2012. Scikit-learn: machine learning in python. J. Mach. Learn. Res. 12, 2825–2830. doi:10.48550/arxiv.1201.0490.

Publications Office of the European Union. Judgment of the court of justice of 16 July 2020 2020. https://op.europa.eu/en/publication-detail/-/publication/d17ef5a0-c749-11ea-adf7-01aa75ed71a1/language-en (accessed January 5, 2023).

Razaghpanah, A., Nithyanand, R., Vallina-Rodriguez, N., Sundaresan, S., Allman, M., Kreibich, C., et al., 2018. Apps, trackers, privacy, and regulators: a global study of the mobile tracking ecosystem. Internet Soc. doi:10.14722/ndss.2018.23353.

Razaghpanah, A., Sundaresan, S., Niaki, A.A., Amann, J., Vallina-Rodriguez, N., Gill, P., 2017. Studying TLS usage in Android apps. In: CoNEXT - Proceedings of the 13th International Conference on Emerging Networking EXperiments and Technologies, pp. 350–362. doi:10.1145/3143361.3143400.

Reyes, I., Wijesekera, P., Reardon, J., On, A.E.B., Razaghpanah, A., Vallina-Rodriguez, N., et al., 2018. Won't somebody think of the children?" Examining COPPA compliance at scale. Proc. Privacy Enhancing Technol. 2018, 63–83. doi:10.1515/popets-2018-0021.

Saldana, J., 2015. The Coding Manual For Qualitative Researchers. SAGE Publications Ltd..

UI/Application Exerciser Monkey|Android Developers. n.d. https://developer.android.com/studio/test/other-testing-tools/monkey (accessed October 21, 2022).

Weber, P.A., Zhang, N., Wu, H., 2020. A comparative analysis of personal data protection regulations between the EU and China. Electron. Commerce Res. 20, 565–587. doi:10.1007/s10660-020-09422-3.

Wilson, S., Schaub, F., Dara, A.A., Liu, F., Cherivirala, S., Leon, P.G., et al., 2016. The creation and analysis of a website privacy policy corpus. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. Long Papers, pp. 1330–1340. doi:10.18653/v1/P16-1.

Wilson, S., Schaub, F., Liu, F., Sathyendra, M., Smullen, D., Zimmeck, S., et al., 2018. Analyzing privacy policies at scale: from crowdsourcing to automated annotations. ACM Trans. Web 13. doi:10.1145/3230665.

Zhang, J., Tian, C., Duan, Z., 2021. An efficient approach for taint analysis of android applications. Comput. Secur. 104, 102161. doi:10.1016/J.COSE.2020.102161.

Zimmeck, S., Story, P., Smullen, D., Ravichander, A., Wang, Z., Reidenberg, J., et al., 2019. MAPS: scaling privacy compliance analysis to a million apps. In: Proceedings of the on Privacy Enhancing Technologies, pp. 66–86. doi:10.2478/popets-2019-0037.

Zimmeck S., Wang Z., Zou L., Iyengar R., Liu B., Schaub F., et al. Automated analysis of privacy requirements for mobile apps 2017. 10.14722/ndss.2017.23034.

PhD DANNY S. GUAMÁN is currently an Assistant Professor with the Escuela Politécnica Nacional, Ecuador. His-main research interests include the analysis of data disclosure and the assessment of privacy compliance in information systems.

MSc DAVID RODRIGUEZ is a PhD student and Graduate Teaching Assistant at Universidad Politécnica de Madrid (UPM). His-research interests include issues related to personal data disclosure, privacy, automated data extraction and personal data flows interception from real Android devices.

PhD JOSE M. DEL ALAMO is currently an Associate Professor (with tenure) with the Universidad Politécnica de Madrid (UPM). His-research interests include issues related to personal data management, including personal data disclosure, identity, privacy, and trust management, and considering these aspects to advance the software and systems engineering methodologies applying approaches, such as privacy-by-design and privacy-by-default.

PhD JOSE SUCH is a Professor in the Department of Informatics at King's College London, Director of the King's Cybersecurity center, an EPSRC–NCSC Academic Center of Excellence in Cyber Security Research (ACE-CSR), and Champion of the Security Hub at the Department of Informatics. Previously, he was Reader at King's from 2018 to 2021, Senior Lecturer at King's from 2016 to 2018, and Lecturer (Assistant Professor) at Lancaster University from 2012 to 2016.

Computing (2024) 106:163–184
https://doi.org/10.1007/s00607-023-01209-2

REGULAR PAPER

ROI: a method for identifying organizations receiving personal data

David Rodriguez[1] · Jose M. Del Alamo[1] · Miguel Cozar[1] · Boni García[2]

Received: 19 December 2022 / Accepted: 7 August 2023 / Published online: 29 August 2023

Abstract

Many studies have exposed the massive collection of personal data in the digital ecosystem through, for instance, websites, mobile apps, or smart devices. This fact goes unnoticed by most users, who are also unaware that the collectors are sharing their personal data with many different organizations around the globe. This paper assesses techniques available in the state of the art to identify the organizations receiving this personal data. Based on our findings, we propose Receiver Organization Identifier (ROI), a fully automated method that combines different techniques to achieve a 95.71% precision score in identifying an organization receiving personal data. We demonstrate our method in the wild by evaluating 10,000 Android apps and exposing the organizations that receive users' personal data. We further assess the transparency of these data-sharing practices by analyzing the apps' privacy policies. The results reveal a concerning lack of transparency in almost 78% of apps, suggesting the need for regulators to take action.

Keywords Domain · Identification · Company · Third-party · NER · Personal data · Data controller · Privacy · Android · Apps

Mathematics Subject Classification 68U15 · 68T50 · 68M12 · 68M15 · 68M25 · 68M11 · 68P27 · 68M14

✉ David Rodriguez
david.rtorrado@upm.es

✉ Jose M. Del Alamo
jm.delalamo@upm.es

Miguel Cozar
m.cozar@upm.es

Boni García
boni.garcia@uc3m.es

[1] ETSI Telecomunicación, Universidad Politécnica de Madrid, Madrid, Spain

[2] Universidad Carlos III de Madrid, Campus Leganés, Madrid, Spain

1 Introduction

The widespread adoption of fancy new smart devices, including many different sensors, facilitates the collection of personal data from individuals anywhere and anytime through the websites they visit and the apps they use. The distributed nature of the Internet further facilitates sharing these data with organizations worldwide [1].

Identifying the organizations that receive these personal data is becoming increasingly crucial for different stakeholders. For example, supervisory authorities may leverage this information to conduct investigations on the relationship between the source and destination of some personal data flows to understand a system's compliance with, for instance, legal requirements for international transfers of personal data [2]. Also, privacy and legal researchers can use this information to discover what companies are collecting massive amounts of personal data [3]. Additionally, app and web developers may want to check what organizations they send their users' personal data to, sometimes even without their knowledge [4], to meet transparency requirements set, e.g., by privacy regulations. Even app marketplaces can take advantage of it in their app review processes (e.g. [5]), still in beta phase in June 2023) to help less experienced developers with their app regulatory compliance.

However, identifying the organizations receiving personal data is not an easy task. The app's or website's privacy policies, if present, often fail to include the third parties with which the collector is sharing the personal data [6]. Although a dynamic analysis of the collecting system and its network traffic can reveal the personal data flows [7] and the destination domains [8], identifying the organizations receiving the data may become challenging due to, e.g. WHOIS accuracy and reliability issues [9]. According to Libert et al. [10]: "we find that 36% of domains in our dataset have anonymous whois registration".

We aim to advance the fundamental understanding of the domains receiving personal data flows and the organizations holding them. To this end, we have assessed two techniques available in the state of the art to identify the organization holding a domain, namely WHOIS service consultation and SSL certificate inspection. Our results show the performance of these individual techniques is far from desirable. Thus, we have developed a new technique based on the analyses of privacy policies and combined it into a new method (ROI - Receiver Organization Identifier), showing a high precision level (95.71%) in identifying the organization that receives personal data flows, and significantly outperforming similar methods available in the state of the art.

Moreover, to demonstrate its applicability in the wild, we have applied ROI to discover the companies receiving personal data on a sample of 10,000 Android apps, and to understand whether Android developers provide transparent disclosures of these data-sharing practices. Our results show that two-thirds of them do not make an adequate disclosure.

Our original contributions are:

1. A reliable and precise method to identify organizations holding domains that receive personal data flows, demonstrated in the wild in the Android ecosystem.

2. Two datasets supporting the validation of our method and the individual techniques, together with the assessment results. The first dataset includes 142 privacy policies URLs annotated with the identity of the organization collecting the data. The second one consists of 300 domains and the organizations holding them.
3. An additional dataset of 1112 unique domains receiving personal data from Android apps together with the personal data types received, obtained in our experiment.

The datasets are publicly available for download at https://doi.org/10.17632/3mdyg53c94 [11].

2 Background and related work

2.1 Background

Identifying an organization receiving personal data requires a method capable of matching the receiver domain to the organization holding it. This section analyzes different techniques providing the necessary technical knowledge to comprehend our proposal.

A domain on the Internet is an authority that controls its own resources (e.g., a network or an IP address), and a domain name is a way to address these resources. Domain names are based on a hierarchy where Top Level Domains (TLD) represent the highest level (e.g., .org, .com, or .es) followed by Second Level Domains (SLD) (e.g., mozilla, google, or amazon). SLDs are managed by companies (i.e., domain name registrars) who register the information on authorities holding domain names in a global registry database. An authority can create subdomains to delimit areas or resources under its own domain (e.g., aws.amazon.com or www.amazon.com). A Fully Qualified Domain Name (FQDN), also known as an absolute domain name, is a domain name that specifies its exact location in the domain hierarchy.

WHOIS [12] is the standard protocol for retrieving information about registered domains and their registrants, including the domain holder's identity, contact details, domain expiration date, etc. Nevertheless, several issues [10] have been reported, including inconsistencies and lack of integrity in registrants' identity information.

Previous research has used WHOIS information for different purposes e.g., to extract registration patterns in the com TLD [13], to categorize organizations that own Autonomous Systems on the Internet [14], or to identify domains that redirect to malicious websites [15]. However, Watters et al. [16] pointed out that the basic deficiency in WHOIS data is a lack of consistency and integrity in the registrants' identity data. This was backed by an ICANN report [17] recognizing extended accuracy failures with only 23% of WHOIS records with 'No failure'. Aiming to address these concerns, the ICANN created the Accuracy Reporting System project [18], whose third phase (i.e. the one addressing registrant identity details) is in a to-be-defined status. Recent studies (e.g. [14]) still report that registrars inconsistently collect, release, and update basic information about domain registrants.

An SSL certificate can be another source of information about the authority holding a domain as it digitally binds a cryptographic key, a domain, and, sometimes, the domain holder's details. The cryptographic key allows for setting up secure connections (HTTPS) between the server and any requesting client.

Thus, whenever an HTTPS connection is set, the client can analyze the certificate used to get information on the server domain holder. HTTPS connections have grown over time, reaching 95% of the total connections in November 2022 [19].

SSL certificates are usually issued by a Certificate Authority (CA), which checks the right of the applicant organization to use a specific domain name and may check some other details depending on the certificate type issued. Extended Validation (EV) certificates are issued after the CA conducts a thorough vetting of the applicant and include information on their legal identity. Organization Validated (OV) certificates are issued after the CA conducts some vetting of the applicant and include information about the applicant's Organization Name under the ON field. Domain Validated (DV) certificates are issued with no vetting of the organization's identity, and no information about the applicant is included. Some studies report that DV certificates account for around 70% of all certificates [20].

An alternative technique to identify a domain holder is to search for it in the publicly available privacy policy that should be associated with that domain on the internet. A privacy policy, also known as a privacy notice, is typically presented as a textual document [21] through which an organization informs its users about the operations of their personal data (e.g., collection and transfer) and how it applies data protection principles. In many jurisdictions, e.g., the European Economic Area (EEA), the United Kingdom, or China, the privacy policy must also include the identity and the contact details of the personal information handler, or first-party or data controller in data protection parlance. It is reasonable to assume that the data controller for a given domain is also the authority holding that domain and vice versa.

2.2 Related work

WHOIS and SSL certificates are legit ways of identifying a domain holder. However, given the problems shown in the previous section, we have had to resort to a new method based on identifying the domain's owner through its privacy policy.

A set of activities are mainly needed to achieve this goal, primarily finding and analyzing the policy, and previous works have partially addressed them. For example, PolicyXray [10] tries to find the privacy policy for a specific URL by crawling all possible resources under that domain. Our method improves PolicyXray by considering other means to find the privacy policy associated with a domain, such as keywords (e.g., privacy, legal) search on the domain's home page and through external search engines (i.e., Google). Furthermore, ROI limits the number of requests to the domain to five, thus outperforming PolicyXray, since crawling a whole domain usually requires hundreds of requests that can overload the domain server [22].

Once a privacy policy has been found, the information identifying the data controller needs to be extracted from the text.

Del Alamo et al. [23] have provided an extensive review of the available techniques for the automated analysis of privacy policies and the information extracted from them over symbolic ones for extracting data controller information, and within them. According to this survey, statistical Natural Language Processing (NLP) techniques are favored supervised machine learning algorithms are mostly reported.

Supervised learning algorithms are usually employed to select the policy segments (roughly speaking, a paragraph) where the content of interest is to be found [24]. They need a labeled (annotated) dataset to learn a specific characteristic of the text they will select. Although different authors have proposed techniques for crowdsourcing annotations of privacy policies (e.g., [25]), researchers' annotations supported by legal experts' assistance (e.g., [2]) are easier to collect for small datasets.

The techniques above show a good performance for classification problems i.e., determining the presence/absence of specific information in a privacy policy. For example, Torre et al. [26] followed this approach to determine the presence of a data controller's identity in a privacy policy. Costante et al. [27] applied it to understand whether the policy disclosed the data controller's contact details, e.g., postal address or phone number. Unfortunately, none extracted the controller's identity, just determining whether or not it was disclosed.

However, we aim to find and extract an organization's identity, and Named-Entity Recognition (NER) techniques are usually applied. Closer to our work, Hosseini et al. [28] used NER techniques to identify third-party entities on privacy policies. They trained three NER models with different word embeddings to obtain their results. This work differs from ours as their goal was to recognize all entities of a class (i.e., organization) in a policy. Instead, we aim to get only one output (the data controller identity) from all possible organizations (i.e., first party, third parties) disclosed in the policy text.

Analogously to our work, WebXray [29] also provides information about the holder of a given domain by combining WHOIS information with other information available on the web (e.g., Wikipedia). Several authors [30–32] have leveraged WebXray to identify organizations receiving personal data flows. Therefore, WebXray is the closest approach to compare our results with, which we do in Sect. 3.4.

3 Method

WHOIS service consultation, SSL certificates inspection, and privacy policies analysis are three different techniques to obtain information on an organization receiving personal data. We detail our approach to extracting information from the WHOIS service and privacy policies below, together with their evaluation results. Finally, we propose and evaluate ROI, a new method that combines the techniques showing the best performance.

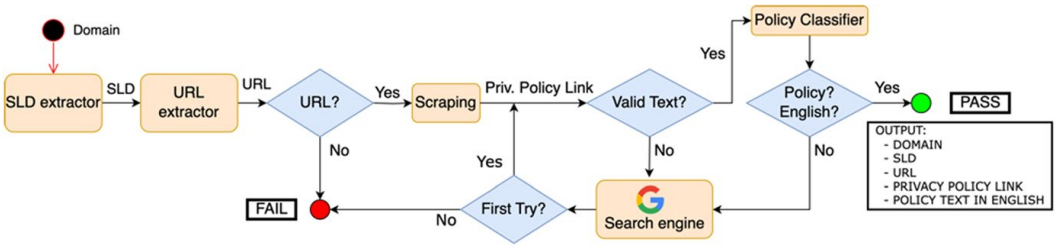


Fig. 1 Privacy policy retrieval and analysis

3.1 WHOIS consultation

We have followed two different approaches to query and parse the WHOIS records. First, leveraging a well-known Python library that queries and parses different WHOIS records. Second, developing our own module focused on extracting the registrant details.

For the first approach, WHOIS domain registration information was retrieved using the python-whois library [33], with over eight million downloads (over 130,000 in October 2022). After an in-depth analysis of the information recovered, we observed incomplete or missing fields that were not correctly parsed, particularly those related to the Registrant Organization identity. This is probably due to the absence of a consistent schema during the domain registration process, as noted by previous research [13]. Thus, we developed our own code to query the WHOIS service using the command line tool and parse the Registrant Organization details. We applied a final filter to discard values hidden for privacy reasons. This filter detects keywords, e.g., "redacted" or "privacy", from a bag of words.

3.2 Privacy policy analyzer

This technique departs from an absolute domain name, finds the privacy policy governing that domain, and analyzes it to extract the data controller identity (Fig. 1).

Finding the privacy policy governing a domain is not straightforward, as we depart from an absolute domain name and must get the website's home page URL to start the search process. To this end, we first obtain the SLD from the domain and then send an HTTP request to it, aiming to be redirected to a valid resource. In case of failure,[1] we leverage search engines (i.e., Google) to find the home page for the given SLD. Once the home page is found, we search for the privacy policy. Again, we have followed two different approaches (1) Scraping the home page with Selenium to find the link to the privacy policy and, in case a valid policy is not found, and (2) searching the policy on the Google search engine.

When a potential privacy policy is found, its text is downloaded and kept for further analysis. Previous work has highlighted [34] that dynamic Javascript code is

[1] Our tests with different methods showed that scraping the target website first and then searching the policy in Google if the scraping didn't work yielded the best results.

sometimes used to display a privacy policy. We relied on Selenium [35] to retrieve the complete text of the privacy policies, which deals with dynamic Javascript code. In our experimental tests, these techniques correctly found 65% of the privacy policies governing the target domain.

Once a potential privacy policy is collected, its language is checked with the langdetect [36] python package, and non-English texts are discarded. Afterward, a supervised Machine Learning (ML) model based on Support Vector Machines (SVM) checks whether the text is indeed a privacy policy.

We applied the SVM approach to determine the optimum separation hyperplane for dividing the analyzed texts into privacy policies or other texts. SVM has empirically proved superior performance in high-dimensional spaces over a range of other ML algorithms, remaining successful even when the number of dimensions exceeds the number of samples [37].

Prior work [24, 38] revealed that SVM outperforms Logistic Regression and Convolutional Neural Networks for categorization of privacy practices. Relying on our previous experience building these kinds of classifiers [39], the hyper-parameters used are the *Modifier-Huber loss function* and *SVM alpha* of 10^{-3}.

We trained the model with 195 manually classified texts, achieving 98.76% precision, 97.56% recall, and 98.15% F1 score[2] when evaluated against 100 unseen English texts.

To identify the data controller in the privacy policy, we first select the paragraphs of the text where it is likely to appear. This selection is based on a bag of words that seeks keywords empirically demonstrated to be closer to the data controller disclosure (e.g., keywords such as "we" and "us" found in the TikTok app privacy notice as shown in Fig. 2). Following previous research in privacy policies analysis [24, 38], our initial approach to identify paragraphs containing controller details was based on a machine learning SVM model trained with 100 manually annotated privacy policies. Nonetheless, privacy policies typically follow a common format and structure. Specifically, the consistent structure of paragraphs where the policy's data controller is declared has led to better results using alternative techniques such as keyword search (i.e., Bag of Words), which was finally implemented.

Named Entity Recognition (NER) techniques are applied to the selected paragraphs to identify the data controller. We have used SpaCy [40] for this, which provides two different trained NER models, one prioritizing efficiency and another favoring accuracy. After testing both, the efficiency model showed poorer results, so the accuracy-based model was implemented. We assessed the performance of the combination of the bag of words and the NER with 142 privacy policies, obtaining the results shown in Table 1.

[2] Accuracy, precision, recall, and F1 score are measures used in Machine Learning to evaluate the performance of categorization algorithms. Accuracy is calculated by dividing the number of correctly classified cases by the total number of instances. Precision is the proportion of true positives among all positive instances, whereas recall is the proportion of true positives among those cases that genuinely belong to the positive class. The F1 score is the harmonic mean of precision and recall, providing a balance between the two measurements.

```
Welcome to TikTok (the "Platform"). The Platform is provided
and controlled by TikTok Inc. ("TikTok", "we" or "us"). We
are committed to protecting and respecting your privacy. This
Privacy Policy covers the experience we provide for users age
13 and over on our Platform.
```

Fig. 2 Privacy policy example (TikTok app) showing the data controller disclosure

Table 1 Privacy policy analyzer metrics

	Accuracy	Precision	Recall	F1-score
Data controller extraction	92.25%	95.45%	94.59%	95.02%

3.3 Individual techniques evaluation

The main goal of this study is to identify the organizations receiving flows of personal data. These organizations can present substantial differences, e.g., company size, location, etc. To make our evaluation as fair as possible, we used a subset of 100 domains randomly chosen from a larger set of 1004 domains that we found receiving personal data from a previous experiment we carried out [2]. For each domain, we manually searched the privacy policy and the data controller disclosed herein.

We used this random dataset to evaluate the performance of WHOIS service consultation, SSL certificates inspection, and privacy policies analysis to identify the controller behind a given domain (Table 2). The output of each technique is either (1) a given value for the domain holder, which can be right (i.e., true positive - TP) or wrong (i.e., false positive—FP), or (2) no value (i.e., false negative—FN) in case the technique cannot determine a specific domain holder. A result cannot be considered as true negative as every domain must have a holder, even if a technique is not able to find it.

The inspection of SSL certificates found 99 certificates out of 100 domains fed, with 30 of them containing the organization name. The missing certificate could not be obtained because this domain uses an HTTP connection. Twenty of the organization names retrieved were correct and ten were wrong; the remaining 69 certificates did not contain the organization name. We did not find any kind of relation between the identity of the CA issuer and the absence of the organization's identity in the

Table 2 Comparative between all the individual techniques

	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
SSL certificate inspection	20.00	66.67	22.23	33.34
WHOIS consultation	30.00	94.87	37.75	54.01
Privacy policy analyzer	56.00	93.34	58.33	71.79

certificates. Only two of those certificates contained IP addresses in addition to the CN. These results translate into a 66.67% precision score, but only 20% identified organizations.

We evaluated the python-whois library as well as our own implementation. Python-whois couldn't find information for ten domains. From the remaining 90 registries, 34 were obtained, 30 were hidden for privacy reasons, and 26 were not correctly parsed. Our own WHOIS-parsing implementation obtained 37 valid owners and 2 incorrect ones. For the remaining 61 domains, 24 did not contain the Registrant Organization field, 2 had an empty value on this field, and 35 registries were hidden for privacy reasons. These results entail the python-whois library performed an 87.18% precision score while our implementation scored 94.87% precision, obtaining more correct results.

Finally, our privacy policy analyzer was evaluated, achieving the best results of all tested techniques. The evaluation is applied to the whole pipeline, including extracting the privacy policy associated with the targeted domain and extracting the data controller name. This pipeline is therefore affected by the performance of each step. Nevertheless, its results outperformed the other techniques, with 56 correct and 4 incorrect outputs, meaning a 93.34% precision score.

3.4 ROI: receiver organization identifier

Given the results achieved by the individual techniques and after a detailed analysis, we combined the privacy policy analyzer with the WHOIS consultation into ROI, a new method to identify an organization receiving personal data. The SSL certificate inspection was discarded due to their low precision score. The python-whois library was also discarded in favor of our implementation.

Interestingly, the combination of the privacy policy analyzer as the first choice and our WHOIS implementation as the second choice outputs the best results, showing even better precision score (95.71%) as the individual techniques while considerably reducing the number of false negative results, achieving 67 true positive results with only 3 false positive results. ROI operating scheme is represented in Fig. 3.

We further analyzed the three false positives. Our NER failed to identify the data controller in two of them. Interestingly, in one of the cases the privacy policy did not mention any data controller at all, which goes against the transparency requirements set by GDPR. As for the third false positive case (unseenreport.com), ROI wrongly attributed this domain to Google as our HTTP request to unseenreport.com was redirected to google.com, raising a red flag due to the redirection to a different

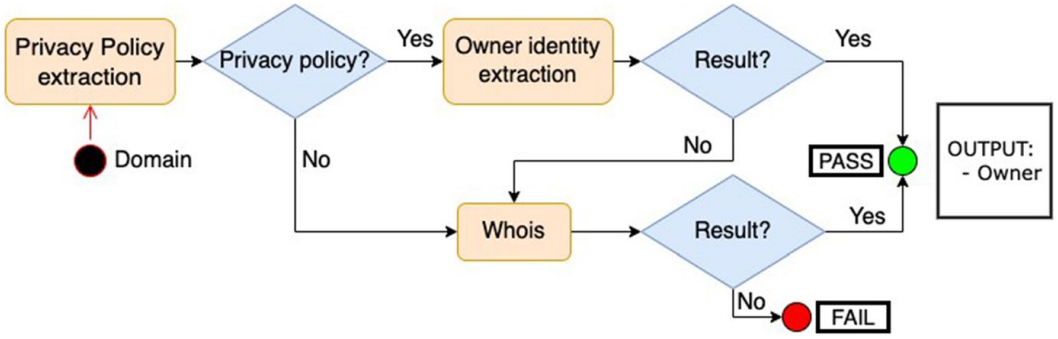


Fig. 3 ROI operation flow diagram

SLD. Unfortunately, we were not able to find the holder organization, even after carrying out an in-depth search of this domain. The domain has been categorized as a malicious website by ANY.RUN [41].

We did a manual inspection on the 30 domains that ROI could not identify. Eleven of these domains did not provide a landing page and are probably only used for back-end purposes.

We did find a landing page for other 13 domains but could not find their privacy policy even while receiving personal data, which may raise compliance issues according to GDPR. Three domains had a non-English website, which is a ROI limitation analyzed in Sect. 5. From the remaining three, one provided the privacy policy through JavaScript expanding elements which were not automatically triggered. These results prove the good performance of ROI, showing that non-identified organizations are usually not providing the information mandated by privacy regulations and thus deserving closer inspection, opening new research lines.

Like ROI, the WebXray tool identifies the organization behind a given domain receiving personal data [29]. Thus, we have compared ROI and WebXray performance against two unseen ground-truth datasets. The first dataset includes 100 URLs (homepages) held by Fortune-500 companies. The second dataset, similar to the one described in Sect. 3.3, consists of a random sample of 100 URLs receiving personal data flows.

Figure 4 illustrates the comparison between the True Positive, False Positive, and False Negative cases (a) and the precision, recall, accuracy, and F1-score metrics (b) from the evaluation of the Fortune-500 dataset. The same comparison is performed on the random sample of personal data receivers (Fig. 5).

There is a noticeable difference in ROI's performance against each dataset. Indeed, ROI achieves slightly better results (97.44% F1-score) against well-known domains than against other less-known receivers (80.24% F1-score). This difference is due to the performance of the WHOIS consultation for the latter since the privacy policy analyzer behaved similarly in both cases.

Receiver organization identifier is a Python-based tool running in a Docker container with all its dependencies installed, so it can be easily deployed in new settings. However, searches are carried out through the Google official paid API and, therefore, the use of ROI requires a Google API token for each new deployment.

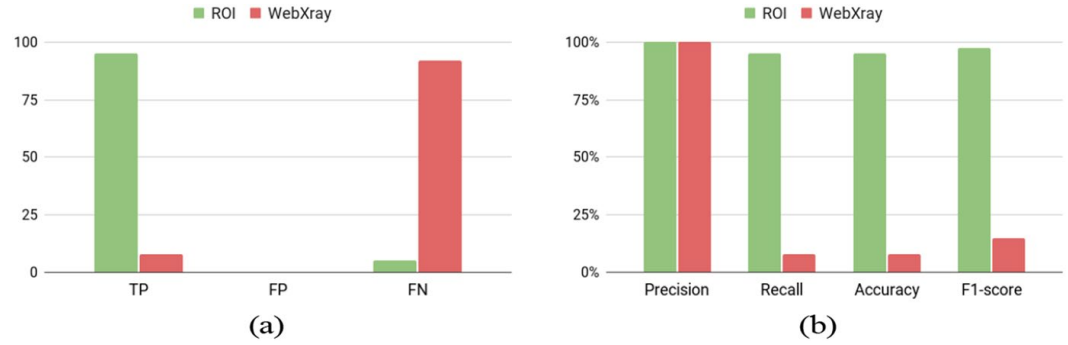


Fig. 4 Comparison between ROI & WebXray results (**a**) and metrics (**b**) on the Fortune-500 dataset

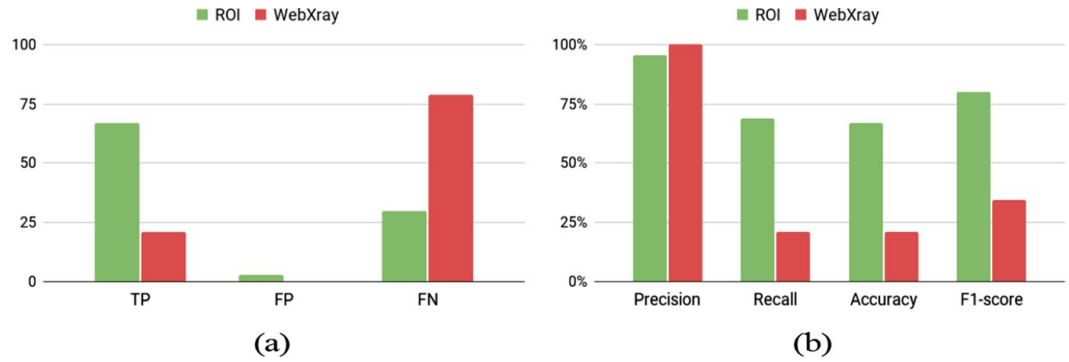


Fig. 5 Comparison between ROI & WebXray results (**a**) and metrics (**b**) on a random sample of personal data receivers

Currently, ROI is able to process privacy policies written in English but not in other languages (only 7.34% of the policies we found in our experiments are not written in English).

All in all, ROI outperforms existing tools serving the same purpose, and its notable scalability and its low number of false negative results support applying it in identifying personal data receivers in the wild.

Receiver organization identifier can easily serve various stakeholders. For instance, researchers in fields such as data protection and privacy can leverage ROI in their research e.g. to uncover companies collecting massive amounts of personal data (as shown in Sect. 4.2). In turn, data protection authorities can utilize ROI to assess mobile applications at scale and discover those sharing data with third parties without declaring it in their privacy policies (as demonstrated in Sect. 4.3). Similarly, it can assist developers in correctly identifying and declaring these recipients in their privacy policies, mitigating substantial fines for non-compliance.

4 Demonstration: android apps evaluation in the wild

In this section, we demonstrate ROI by evaluating 10,000 Android apps from the Google Play Store, analyzing what personal data they send out and the organizations receiving them, and checking whether the recipients have been properly disclosed in the apps' privacy policies. To this end, we describe our experimental environment, and report and analyze the results obtained.

4.1 Experiment setup

We developed a controlled experiment leveraging our previous work [2] on personal data flow interception and analysis in Android apps. This is a pipelined microservices-based platform made up of different modules able to automatically (1) search, download, install, run, and interact with Android apps, and (2) intercept and analyze outgoing network connections.

Specifically, the Download module logs into the Google Play Store simulating a real device and downloads the applications, storing them in the Storage module. The Traffic module is a multi-threaded Python script handling multiple real devices connected at the same time. It gets applications from the Storage module and installs them in each device through the Android Debug Bridge (ADB) [42] connection. After the installation, it runs the apps first in an idle phase (without app stimulation) and then in a dynamic phase (with automated stimulation using Android Monkey [43]). At the same time, a Man-in-the-Middle proxy and an application instrumentation tool (Frida) are used to intercept and decrypt secured connections. The connections' payloads are decoded trying different formats (e.g., Base64, SHA, MD5) and inspected looking for personal data. The results are logged to our centralized logging platform based on ElasticSearch.

Previous research has extensively addressed the detection of personal data leaks in Android apps following two approaches, namely static and dynamic analysis. Static techniques [44] focus on detecting data leakages by analyzing the code without executing it. On the other hand, dynamic techniques require the apps' execution and a further interception of the communications either inside the device (e.g., setting up a virtual private network and analyzing the outgoing traffic [45]). Our setup favors dynamic analysis techniques to capture network packets generated by the real execution of apps rather than static analysis techniques that analyze approximate models that, while ensuring high recall, could generate a high rate of false positives.

Our platform was fed with a list of 10,000 random Google Play Store apps from the top-downloaded category. The apps were collected, downloaded, installed, and executed in September 2022 on five mobile devices Xiaomi Redmi 10, running Android 11 (API 30). Following common practices for dynamic analysis in Android [46], the idle phase was performed for two minutes and the Android Monkey was used to interact with each application for an extra three minutes. Considering the five devices running uninterruptedly and ignoring devices' bugs (which forces us to

manually restart the device affected) it required 6 days 6h and 35 minutes to finish our analysis.

4.2 Recipients' analysis

Our platform managed to execute 7037 apps, identifying 40,493 personal data flows from 3526 apps to 1112 unique domains during the experiment. A vast portion (99.2%) of these data flows correspond to HTTPS connections, which are aligned with the HTTPS encryption level observed on the Web [19]. Interestingly, we found 320 (0.8%) HTTP connections containing personal data, which is an insecure practice. Alarmingly, these HTTP connections included all types of personal data we found except the device's software build number. Therefore, personal data such as the Google advertising identifier or the device location are being sent without adequate protection.

Figure 6 shows the number of apps sending out each personal data type (top), and the number of apps that sent personal data to the top-20 destination SLDs (bottom). Interestingly, most apps sent out the device model name (97.13%), and more than half of apps (61.68%) sent the Google advertising identifier, which is closely aligned with what was observed by previous research [47].

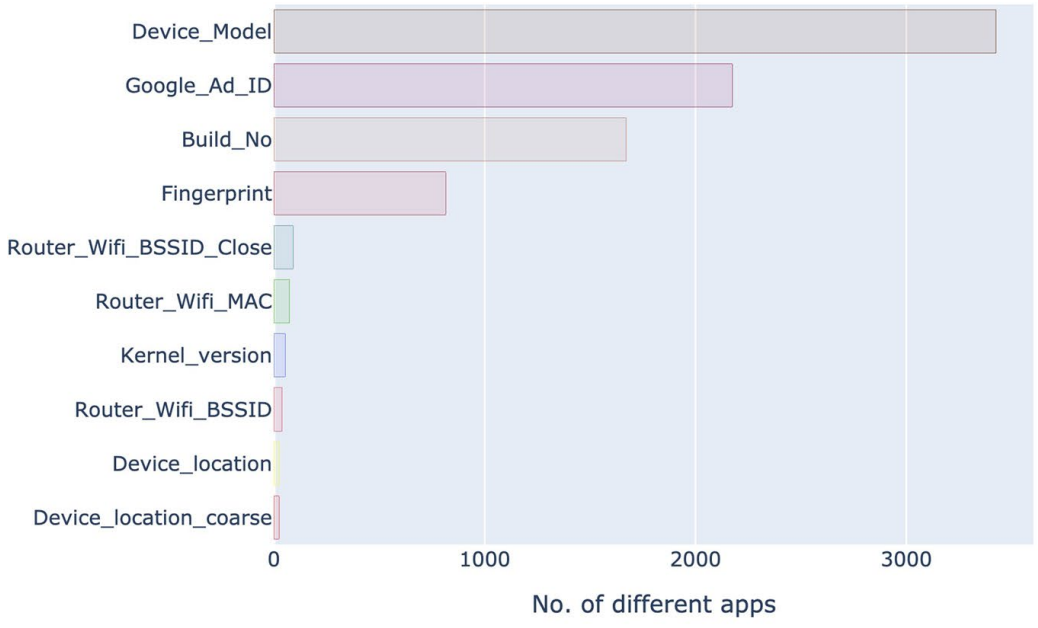
Figure 7 further details the types of personal data that the top-10 domains are receiving. We can see that nine out of ten domains are collecting the Google advertising identifier, commonly used for monetization tracking and personalized advertising. As could be expected, most of the top-20 domains receiving personal data are for analytics, marketing, or monitoring purposes (e.g., firebaselogging-pa.googleapis.com, supersonicads.com, adcolony.net).

We further applied ROI to identify the companies holding the domains receiving the personal data. Overall, we determined them in 82.37% (33,356) of the personal data flows, representing 68.7% (764) of the unique destination domains. Figure 8 shows how many apps sent personal information to the top collectors.

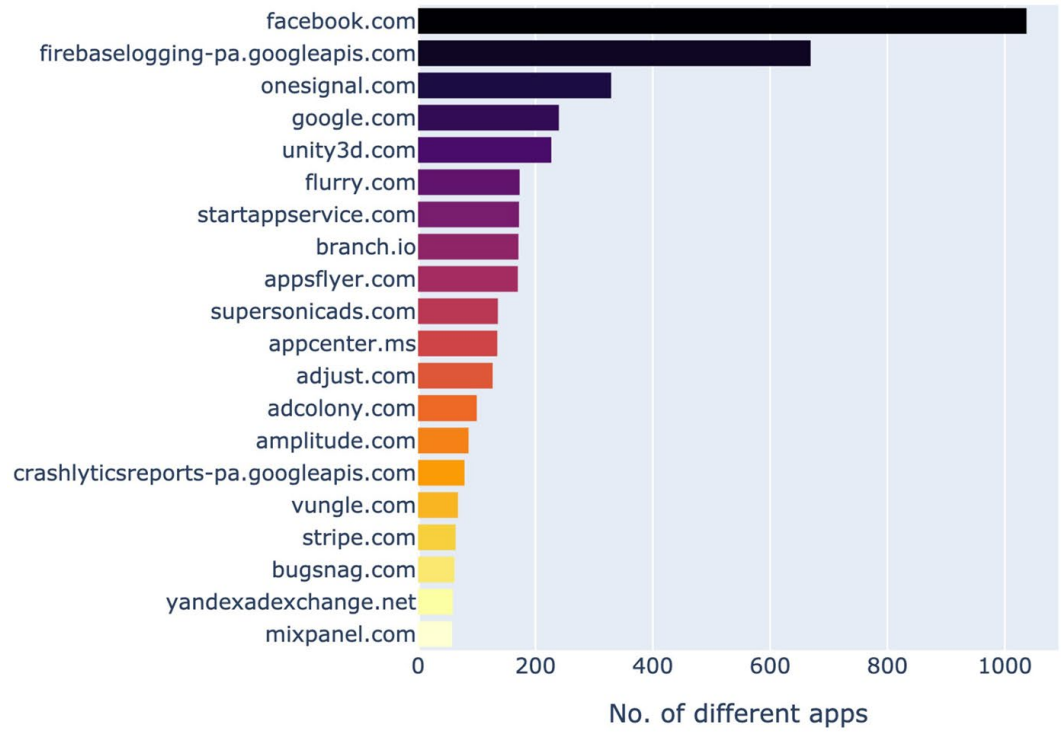
The top-6 companies to which most apps send personal data provide analytics and marketing services. Furthermore, Meta and Google lead this list, receiving data from 1037 (29.41%) and 1006 (28.53%) apps, respectively. Indeed, half of the apps (51.56%) sent personal data to either Meta or Google.

Other companies from this top 10, e.g., Unity or Supersonic Studios Ltd., support games development and publishing, which means the importance of the gaming category in the Google Play store market. On a curious note, Sentry, which provides error and crash monitoring services, also receives users' personal data.

We further analyzed the hierarchy of relationships of the companies we found. The purpose is to make a representative illustration (Fig. 9) of the companies that collect the greatest volume of data, showing the head company as the representative. For example, Microsoft is the parent company of GitHub and LinkedIn. To achieve this, we employed the Crunchbase API [48] to enrich information about an organization, including the hierarchy of relationships between companies, e.g., parent and subsidiaries. Specifically, we searched in Crunchbase for the name of the company we found with ROI and obtained its country, description, and relationships



(a)



(b)

Fig. 6 Personal data sent off the device (**a**) and the popular destinations (**b**)

	Domains									
Type of personal data	1	2	3	4	5	6	7	8	9	10
Device_Model	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Google_Ad_ID	✓		✓	✓	✓	✓	✓	✓	✓	✓
Build_No	✓	✓			✓	✓		✓	✓	✓
Fingerprint		✓								
Router_Wifi_BSSID_Close									✓	
Router_Wifi_MAC									✓	
Kernel_version										
Router_Wifi_BSSID										
Device_location										
Device_location_coarse			✓			✓				

Domains
1. facebook.com
2. firebaselogging-pa.googleapis.com
3. onesignal.com
4. google.com
5. unity3d.com
6. flurry.com
7. startappservice.com
8. branch.io
9. appsflyer.com
10. supersonicads.com

Fig. 7 Type of personal data received by popular domains

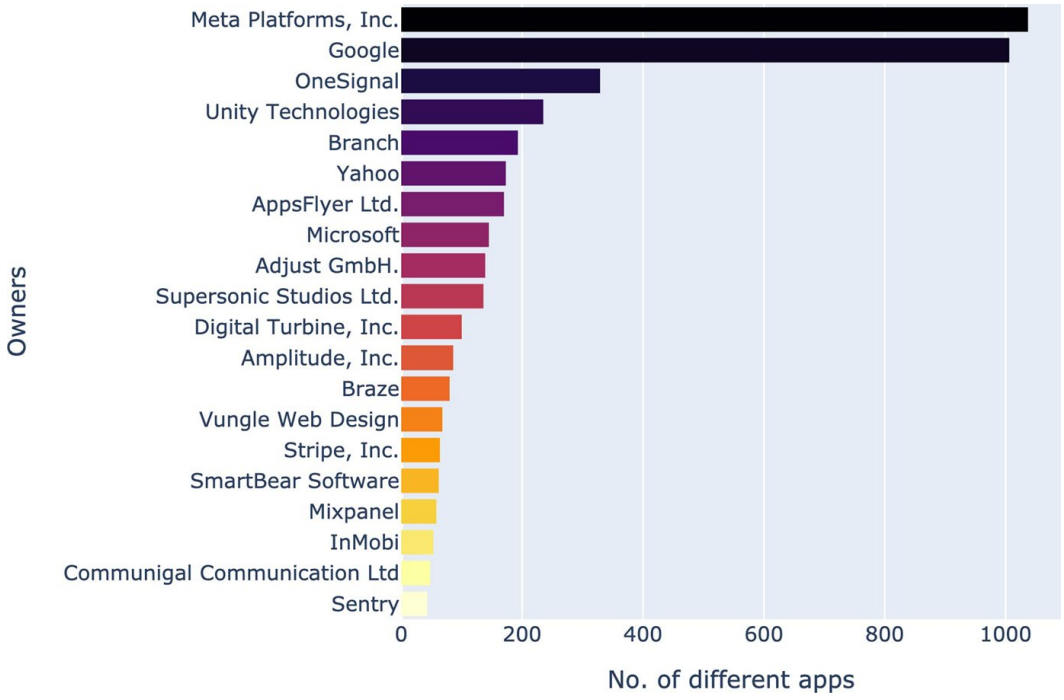


Fig. 8 Companies receiving personal data

with other companies. We repeated the process iteratively until we found the head company and all the companies under its control.

Figure 9 demonstrates that some companies might receive data from several subsidiaries. Thus, the actual amount of personal data collected might be higher than expected, as for Fig. 8. The example of AppLovin is quite representative. While AppLovin provides users with monetization tools, they also have Adjust as a subsidiary for helping developers and MoPub for advertisement serving. The result is a whole ecosystem of companies collecting data that situate the corporation in the top-6 according to our data, above Verizon. Microsoft is another example of a company with several subsidiaries collecting data, i.e., LinkedIn and GitHub.

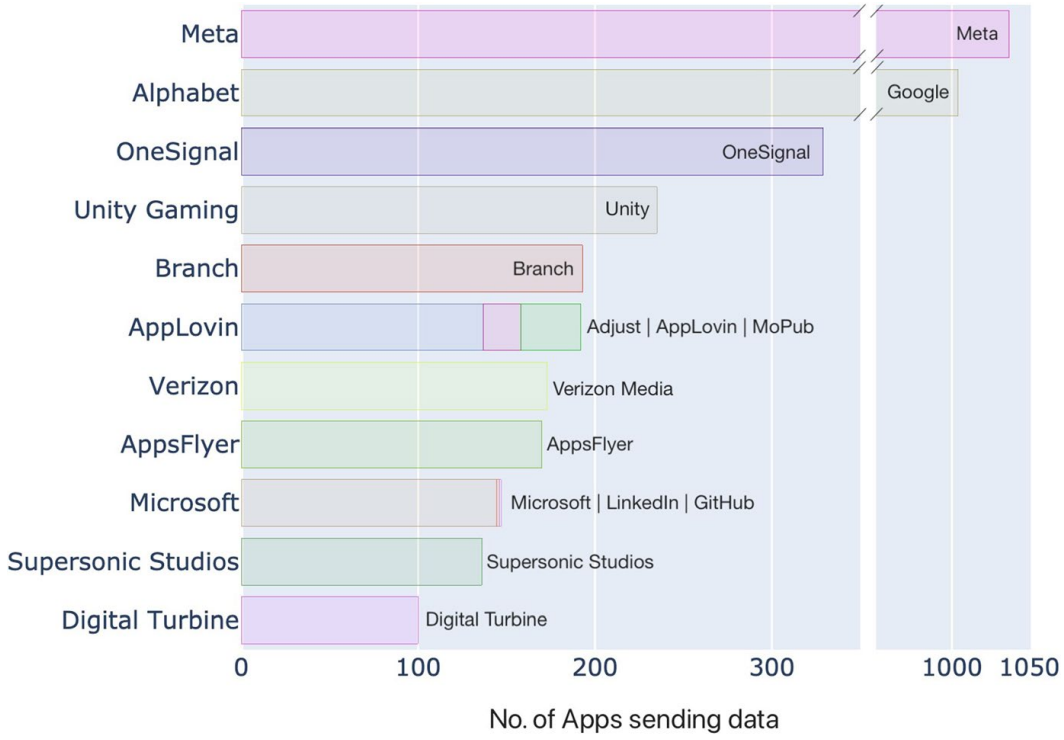


Fig. 9 Top 10 head companies with subsidiaries receiving personal data

Interestingly, the top 8 head companies provide well-known Software Development Kits (SDK) for Android, appearing on the Google Play SDK index [49]. Therefore, it is fair to assume that most personal data receivers are third-party organizations. This is a common practice in the mobile ecosystem and was addressed in previous research [29].

4.3 Recipients disclosure

Receiver organization identifier facilitates transparency and accountability in data protection practices. For example, valuable insight can be gained by cross-referencing the third-parties disclosed in privacy policies with those identified through ROI. When app developers and recipient organizations provide transparent information about the entities involved in data exchange, they demonstrate a commitment to accountability and foster user trust. Through improved transparency, users gain a deeper understanding of how their personal data is handled, empowering them to make informed decisions about apps.

To further demonstrate the ROI potential, we have checked if the recipients identified in the previous section are actually disclosed in the apps' privacy policies. To this end, we retrieved and processed the privacy policies of 2155 applications, extracting the third-parties mentioned therein. We could not process 1371 apps from our initial dataset as we could not find their privacy policy,

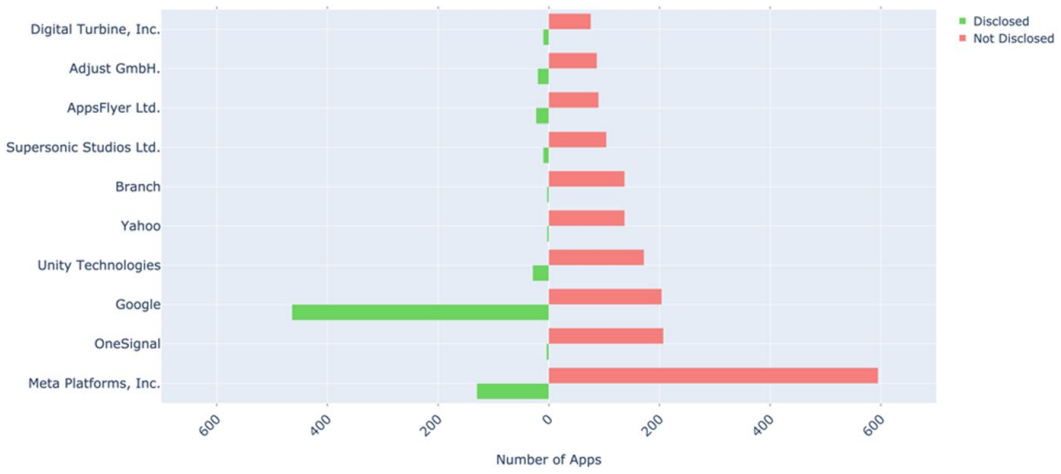


Fig. 10 Third-party disclosure of top organizations in Android apps' privacy policies

which already flags a huge bunch of apps potentially non-compliant with applicable data protection laws such as GDPR.

Android apps generally fail to disclose their data-sharing practices. Only 476 (22%) apps accurately reflect all the entities receiving personal data in their privacy policies, while 1327 (61%) do not disclose any of these entities. The remaining 352 applications fail to declare at least one recipient, providing only partial disclosure. This makes an outstanding 78% of the apps analyzed failing to fully disclose their data-sharing practices as mandated by data protection laws (e.g. GDPR Art. 13 (1)(e)).

The lack of disclosure does not equally affect all personal data recipients. Figure 10 illustrates that Google is the recipient most frequently disclosed by the applications (464 out of 1327 apps properly disclose the transfer of personal data to this company), significantly ahead of the others. Conversely, our observations show that applications very often fail to mention Meta in their privacy policies. A total of 595 applications out of the 828 that send data to Meta (consciously or unconsciously) fail to declare it.

Some specific cases are alarming. For example, "com.wallapop", a leading second-hand selling application in Spain with 10 million downloads on the Google Play Store, fails to disclose recipients like Tapjoy and Google (among others), which are receiving the Google advertising ID and the device's fingerprint, respectively. Fortunately, we have also found full disclosures e.g. the "jigsaw.puzzle.free.games" application, which has 50 million downloads on the Google Play Store, accurately declares data transfers to Meta, Unity, Google, Amazon, and other major tech entities.

Compliance with privacy and data protection laws is critical in today's regulatory world. As we have demonstrated, ROI can substantially aid in identifying personal data recipients, thus supporting compliance assessment processes.

5 Threats to research quality

Receiver organization identifier is a reliable and highly precise method to identify organizations holding domains receiving personal data and can identify the majority of the tested domains. Nevertheless, some limitations have been identified during the development process.

Finding the privacy policy for a given domain is at the core of ROI. How- ever, the disclosure of privacy policies is not standardized; thus, many ad-hoc means are used to present the policy text, e.g., contained in popup elements. Following best practices in the field [25], we have relied on Selenium to address most of these issues as it deals with dynamic JavaScript code.

The information disclosed within a policy text (i.e., the data controller) is not standardized either. Although privacy policies are mandatory in some jurisdictions, e.g., the European Union as for its General Data Protection Regulation (GDPR) [50] Article 13, this information is often missing or wrong (e.g., frequently the app name is used to refer to the data controller, even if the app name is not a legally registered organization). We have addressed this challenge by partially validating the controller extraction method, achieving a nearly 95% F1 score, proving the good performance of this method.

Another limitation comes from the privacy policy text language. For the time being, our NER works with English texts, and it cannot process texts in other languages. Thus, we discarded non-English texts in our analysis, corresponding to only 7.34% of the policies found. To reduce the amount of non-English policy texts, we configured our tools to favor English texts. This was achieved by setting the accept-language parameter in the requests' headers and the lang argument in Selenium's configuration. Nevertheless, we are working on translation methods with NLP techniques that will help us to improve the number of privacy policies analyzed.

Our experiment involved results from 3526 apps. The results with this amount of apps are representative, but outliers may appear when speaking about some data receivers. The results can be extended to consider a more significant number of applications, thus supporting the generalization of the results.

Finally, automated access to web pages might be viewed as unethical if it overloads the website. However, ROI only makes a maximum of five requests per domain instead of crawling them with hundreds of queries. This was pos-sible thanks to the bag of words technique (cf. Sect. 3.2 for details) we applied.

6 Conclusion

This paper has described ROI, a new method that leverages the information available in privacy policies and the WHOIS service to identify organizations receiving personal data flows. ROI achieves a 95.71% precision, greatly outperforming similar methods in the state of the art. We have demonstrated its applicability in the Android context by identifying the companies receiving personal data from

3526 apps in the wild. Unfortunately, we have also shown that a huge portion of these apps fail to properly disclose these organizations in their privacy policies.

ROI brings benefits to various stakeholders. Data protection authorities can leverage it to understand the compliance of personal data collecting systems with privacy and data protection regulations. App developers can gain valuable insights into how their applications adhere to them. Researchers can gain a better understanding of the destinations of massive amounts of personal data.

Our future work points towards contributing to new techniques that support the privacy engineering community in automating the assessment processes of digital systems and services. To this end, we are leveraging ML and NLP techniques to automate the extraction of transparency elements from the privacy policies and check them against the actual behavior observed in the systems under analysis.

Acknowledgements This work has been partially supported by the TED2021-130455A-I00 project funded by MCIN/AEI/https://doi.org/10.13039/501100011033 and by the European Union "NextGenerationEU"/PRTR; and, by the Comunidad de Madrid and Universidad Politécnica de Madrid through the V-PRICIT Research Programme 'Apoyo a la realizaci´on de Proyectos de I+D para j´ovenes investigadores UPM-CAM', under Grant APOYO-JOVENES-QINIM8-72-PKGQ0J. It was possible to identify the relationships between parent and subsidiary companies thanks to Crunchbase, who kindly allowed us free access to its API for this research.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature.

Declarations

Conflict of interest The authors declare through the submission of this document that they have no conflicts of interest. This work was partially supported by the European Union, the Comunidad de Madrid. These are public funds granting that there are no financial or personal interests related to the research. The authors received no financial support or other benefits from any organization or individual with a stake in the research.

References

1. Razaghpanah A, Nithyanand R, Vallina-Rodriguez N, Sundaresan S, Allman M, Kreibich C, Gill P (2018) Apps, trackers, privacy, and regulators a global study of the mobile tracking ecosystem. In: Proceedings of the network and distributed systems security (NDSS) symposium, pp 1–15. https://doi.org/10.14722/ndss.2018.23353
2. Guaman DS, Alamo JMD, Caiza JC (2021) Gdpr compliance assessment for cross-border personal data transfers in android apps. IEEE Access 9:15961–15982. https://doi.org/10.1109/ACCESS.2021.3053130

3. Schindler C, Atas M, Strametz T, Feiner J, Hofer R (2022) Privacy leak identification in third-party android libraries. In: Proceedings of the 2022 7th international conference on mobile and secure services, MobiSec- Serv 2022. https://doi.org/10.1109/MOBISECSERV50855.2022.9727217
4. Balebako R, Marsh A, Lin J, Hong J, Cranor LF (2014) The privacy and security behaviors of smartphone app developers. Workshop on usable security (USEC'14), pp. 1–10.
5. Compliance Intelligence—Checks. https://checks.area120.google.com/. Accessed: 2023-06-08
6. Verderame L, Caputo D, Romdhana A, Merlo A (2020) On the (un)reliability of privacy policies in android apps. In: Proceedings of the international joint conference on neural networks (2020). https://doi.org/10.1109/IJCNN48605.2020.9206660
7. Enck W, Gilbert P, Han S, Tendulkar V, Chun BG, Cox LP, Jung J, McDaniel P, Sheth AN (2014) Taintdroid. ACM Transactions on Computer Systems (TOCS) 32. https://doi.org/10.1145/2619091
8. No Body's Business But Mine: How Menstruation Apps Are Shar- ing Your Data. Privacy international. https://privacyinternational.org/long-read/3196/no-bodys-business-mine-how-menstruations-apps-are-sharing-your-data. Accessed: 2023-06-08
9. Current Issues—ICANN WHOIS. https://whois.icann.org/en/current-issues. Accessed: 2023-06-08
10. Libert T (2018) An automated approach to auditing disclosure of third-party data collection in website privacy policies. In: The Web Conference 2018—Proceedings of the World Wide Web Conference, WWW 2018, pp 207–216. https://doi.org/10.1145/3178876.3186087
11. Rodriguez D, Del Alamo JM, Cozar M, García B (2023) ROI: a method for identifying organizations receiving personal data. Mendeley Data. https://doi.org/10.17632/3mdyg53c94
12. RFC 954—CNAME/WHOIS. https://datatracker.ietf.org/doc/html/rfc954. Accessed: 2023-06-08
13. Liu S, Foster I, Savage S, Voelker GM, Saul LK (2015) Who is .com? learning to parse whois records. In: Proceedings of the ACM SIGCOMM internet measurement conference, IMC 2015-October, pp 369–380. https://doi.org/10.1145/2815675.2815693
14. Ziv M, Izhikevich L, Ruth K, Izhikevich K, Durumeric Z (2021) Asdb: A system for classifying owners of autonomous systems. In: Proceedings of the ACM SIGCOMM internet measurement Conference, IMC, pp 703–719. https://doi.org/10.1145/3487552.3487853
15. Thao TP, Yamada A, Murakami K, Urakawa J, Sawaya Y, Kubota A (2017) Classification of landing and distribution domains using whois' text mining. In: Proceedings of the 16th IEEE conference on trust, security and privacy in computing and communications (IEEE TrustCom-17), pp. 1–8. https://doi.org/10.1109/Trustcom/BigDataSE/ICESS.2017.213
16. Watters PA, Herps A, Layton R, McCombie S (2013) Icann or icant: Is whois an enabler of cybercrime? In: Proceedings—4th cybercrime and trust-worthy computing workshop, CTC 2013, pp 44–49. https://doi.org/10.1109/CTC.2013.13
17. WHOIS Policy Review (2012). https://www.icann.org/en/system/files/files/final-report-11may12-en.pdf. Accessed: 2023-06-08
18. WHOIS Accuracy Reporting System (ARS)—ICANN WHOIS. https://whois.icann.org/en/whoisars. Accessed: 2023-06-08
19. HTTPS encryption on the web—Google Transparency Report. https://transparencyreport.google.com/https/overview. Accessed: 2023-06-08
20. SSL Survey—Netcraft. https://www.netcraft.com/internet-data-mining/ssl-survey/. Accessed: 2023-06-08
21. Victor M, Raúl P (2020) SoK: Three Facets of Privacy Policies. In: Proceedings of the 19th workshop on privacy in the electronic society (WPES'20). Association for Computing Machinery, New York, NY, USA, pp 41–56. https://doi.org/10.1145/3411497.3420216
22. Ahmad SS, Dar MD, Zaffar MF, Vallina-Rodriguez N, Nithyanand R (2020) Apophanies or epiphanies? How crawlers impact our understanding of the web. In: The Web Conference 2020—Proceedings of the World Wide Web Conference, WWW 2020, pp 271–280. https://doi.org/10.1145/3366423.3380113
23. Alamo JMD, Guaman DS, García B, Diez A (2022) A systematic mapping study on automated analysis of privacy policies. Computing 104:2053–2076. https://doi.org/10.1007/s00607-022-01076-3
24. Zimmeck S, Wang Z, Zou L, Iyengar R, Liu B, Schaub F, Wilson S, Sadeh N, Bellovin SM, Reidenberg J, Louis S (2017) Automated analysis of privacy requirements for mobile apps. In: The 24th annual network and distributed system security symposium, NDSS, (2017). https://doi.org/10.14722/ndss.2017.23034
25. Wilson S, Schaub F, Ramanath R, Sadeh N, Liu F, Smith NA, Liu F (2016) Crowdsourcing annotations for websites' privacy policies: can it really work? 25th International World Wide Web Conference. WWW 2016:133–143. https://doi.org/10.1145/2872427.2883035

26. Torre D, Abualhaija S, Sabetzadeh M, Briand L, Baetens K, Goes P, Forastier S (2020) An ai-assisted approach for checking the completeness of privacy policies against gdpr. In: Proceedings of the IEEE international conference on requirements engineering 2020-August, pp 136–146. https://doi.org/10.1109/RE48521.2020.00025
27. Costante E, Sun Y, Petkovic M, Hartog JD (2012) A machine learning solution to assess privacy policy completeness. In: Proceedings of the ACM conference on computer and communications security, pp 91–96. https://doi.org/10.1145/2381966.2381979
28. Hosseini MB (2020) Identifying and classifying third-party entities in natural language privacy policies. In: Proceedings of the 2nd workshop privacy, pp 18–27. https://doi.org/10.18653/v1/2020.privatenlp-1.3.
29. Libert T, Desai A, Patel D (2021) Preserving needles in the haystack: A search engine and multi-jurisdictional forensic documentation system for privacy violations on the web (2021). https://timlibert.me/pdf/Libert_et_al-2021-Forensic_Privacy_on_Web.pdf
30. Binns R, Lyngs U, Kleek MV, Zhao J, Libert T, Shadbolt N (2018) Third party tracking in the mobile ecosystem. In: WebSci 2018—Proceedings of the 10th ACM conference on web science, pp 23–31. https://doi.org/10.1145/3201064.3201089
31. Binns R, Zhao J, Kleek MV, Shadbolt N (2018) Measuring third-party tracker power across web and mobile. ACM Transactions on Internet Technology (TOIT) 18 (2018). https://doi.org/10.1145/3176246
32. Kleek MV, Liccardi I, Binns R, Zhao J, Weitzner DJ, Shadbolt N (2017) Better the devil you know: exposing the data sharing practices of smartphone apps. In: Conference on human factors in computing systems—proceedings 2017-May, pp 5208–5220. https://doi.org/10.1145/3025453.3025556
33. python-whois—PyPI. https://pypi.org/project/python-whois/. Accessed: 2023-06-08
34. Harkous H, Fawaz K, Lebret R, Schaub F, Shin KG, Aberer K (2018) Polisis: automated analysis and presentation of privacy policies using deep learning. In: 27th USENIX Security Symposium. https://doi.org/10.48550/ARXIV.1802.02561
35. García B, Gallego M, Gortázar F, Munoz-Organero M (2020) A survey of the selenium ecosystem. Electronics 9(7):1067. https://doi.org/10.3390/ELECTRONICS9071067
36. langdetect—PyPI. https://pypi.org/project/langdetect/. Accessed: 2023-06-08
37. Moguerza JM, Muñoz A (2006) Support vector machines with applications. Stat Sci 21(3):322–336. https://doi.org/10.1214/088342306000000493
38. Wilson S, Schaub F, Liu F, Sathyendra K M, Smullen D, Zimmeck S, Ramanath R, Story P, Liu F, Sadeh N, Smith N A. (2018) Analyzing privacy policies at scale: From crowdsourcing to automated annotations. ACM Trans Web 13(1). https://doi.org/10.1142/3230665
39. Guamán DS, Rodriguez D, del Alamo JM, Such J (2023) Automated GDPR compliance assessment for cross-border personal data transfers in android applications. Comput Security 130:103262. https://doi.org/10.1016/J.COSE.2023.103262
40. Honnibal M, Montani I, Van Landeghem S, Boyd A (2020) spaCy: industrial-strength natural language processing in python. Zenodo. https://doi.org/10.5281/zenodo.1212303
41. ANY.RUN—Interactive Online Malware Sandbox. https://any.run/. Accessed: 2023–06–08
42. Android Developers. (n.d.). Android Debug Bridge (ADB). https://developer.android.com/tools/adb. Accessed: 2023-06-08
43. Android Developers. (n.d.). UI/Application Exerciser Monkey. https://developer.android.com/studio/test/other-testing-tools/monkey. Accessed: 2023-06-08
44. Laperdrix P, Mehanna N, Durey A, Rudametkin W (2022) The price to play: A privacy analysis of free and paid games in the android ecosystem. In: WWW 2022—Proceedings of the ACM Web Conference 2022, pp 3440–3449 https://doi.org/10.1145/3485447.3512279
45. Razaghpanah A, Vallina-Rodriguez N, Sundaresan S, Kreibich C, Gill P, Allman M, Paxson V (2015) Haystack: A multi-purpose mobile vantage point in user space (2015). https://doi.org/10.48550/arxiv.1510.01419
46. Choudhary SR, Gorla A, Orso A (2015) Automated test input generation for android: Are we there yet? In: 2015 30th IEEE/ACM international conference on automated software engineering (ASE), pp 429–440. https://doi.org/10.1109/ASE.2015.89. IEEE
47. Kollnig K, Shuba A, Binns R, Van Kleek M, Shadbolt N (2022) Are iPhones Really Better for Privacy? A Comparative Study of iOS and Android Apps. In: Proceedings on privacy enhancing technologies (Vol. 2022, Issue 2, pp. 6–24). Privacy enhancing technologies symposium. https://doi.org/10.2478/popets-2022-0033.
48. Crunchbase. https://www.crunchbase.com/home. Accessed: 2023-06-08
49. Google Play SDK Index. https://play.google.com/sdks. Accessed: 2023-06-08

50. European Commission: Regulation (EU) 2016/679 of the European Par- liament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance). European Commission (2016). https://eur-lex.europa.eu/eli/reg/2016/679/oj

2023 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW) | 979-8-3503-2720-5/23/$31.00 ©2023 IEEE | DOI: 10.1109/EuroSPW59978.2023.00022

Comparing Privacy Label Disclosures of Apps Published in both the App Store and Google Play Stores

David Rodriguez*, Akshath Jain†, Jose M. del Alamo*, Norman Sadeh†
*ETSI Telecomunicación, Universidad Politécnica de Madrid
{david.rtorrado, jm.delalamo}@upm.es
†School of Computer Science, Carnegie Mellon University
arjain@andrew.cmu.edu, sadeh@cs.cmu.edu

***Abstract*—Apple and Android introduced privacy labels in 2020 and 2022 respectively as a way of providing consumers with succinct summaries of mobile apps' more salient data practices. A number of apps are published in both stores, offering us the opportunity to compare their privacy label disclosures in the two app stores. This paper compares the data practices privacy labels are intended to capture in each store. It then proceeds to analyze the disclosures of 822 apps published in both app stores, focusing on possible discrepancies. This analysis reveals that privacy label disclosures of what is ostensibly the same mobile app can be quite different. We discuss the different possible reasons behind these differences, including the possibility that these discrepancies might be indicative of potential privacy compliance issues. In particular, focusing on data collection disclosures of five different data types (location, contact info, sensitive info, identifiers, and health & fitness) we find discrepancies between iOS and Google Play privacy label disclosures in 66.5% of the mobile apps we analyze.**

1. Introduction

The internet's and digital technologies' explosive growth in recent years has greatly influenced how people interact, consume media, and transact business. Yet, when personal data is involved, the rising use of digital technology also poses grave concerns to people's privacy. In fact, a KPMG study [14] reports that 86% of Americans claim that data privacy is a recent growing concern for them.

In response to these worries, big tech companies (i.e., Google & Apple) have developed privacy labels, which provide users with information about the data collected and shared by apps and the way the data is protected [13]. These labels are designed with the aim of helping users make informed decisions before installing applications based on their privacy practices.

Apple, one of the leading participants in the mobile operating system industry, has a privacy label section built into the App Store. The iOS privacy labels, which went into effect in December 2020, ask app developers to provide explicit breakdowns into the various types of data they gather, including contact information, location data, and browser history. In a similar manner, Android introduced its own privacy labels into the data safety section on the Google Play Store in April 2022.

Privacy labels in both ecosystems must be declared by app developers, which invite users to rely on the veracity of their statements. Nevertheless, app developers may intentionally or unintentionally be omitting information. Our study attempts to shed light on this issue by presenting a comparison between Android and iOS privacy labels. The contributions of this work are as follows:

- A Mapping between iOS and Android labels defining the practices and data types disclosed that could be directly compared between each platform.
- The design of a method for reliably finding iOS applications on Google Play Store (and vice versa).
- A comprehensive comparison between the mapped privacy labels for 822 identical Android and iOS applications.
- A static analysis of 560 Android applications' source code looking for precise and coarse location collection.

To the best of our knowledge, this is the first research work comparing Android and iOS privacy labels usage by app developers. The proposed mapping along with a reliable method for finding the same applications in both marketplaces will enable new studies to be carried out for the benefit of all mobile device users.

The outline of this document is as follows. Section 2 describes iOS and Android privacy labels, presents the mapping identified between these labels, and documents the related work. In section 3, the method followed for the privacy labels comparison is presented. An analysis at scale is conducted in Section 4, highlighting the differences identified. Those differences are discussed in section 5 with the relevant findings. Potential threats to validity are exposed in Section 6 and the paper's final conclusions are reported in Section 7.

2. Background & Related Work

Privacy labels have emerged as a result of the readability and comprehension problems of privacy policies [5], [25], [26]. Their scope is to encourage developers to disclose their applications' privacy practices following a template that allows a better understanding by users. This section will define the particularities of iOS and Android privacy labels. Then, we propose a mapping between the correspondent practices and data types in both ecosystems. Finally, the closest related work is presented highlighting the differences with our contributions.

DOI 10.1109/EuroSPW59978.2023.00022

2.1. iOS Privacy Labels

Since the addition of Apple's privacy labels in December 2020 [20], developers have been asked to describe four privacy aspects of their apps (see Figure 1, bottom): data item, data type, data purpose, and data practice. The data item is the specific data to be collected by the app, which belongs to a higher data type (i.e., category), for example, the "name" data item belongs to the "contact info" data type. The purpose is the app's reason for accessing these data e.g. analytics or app functionality. Data practice describes to what extent the piece of data will be linked to the user identity i.e Data not Linked to You, Data Linked to You, and Data used to Track You. Data not Linked to You refers to de-identified or anonymized data. Data Linked to You refers to data linked to the user's identity e.g via their account, device, or other details. Data used to Track You refers to further linking the data with new third-party data for advertising purposes, or sharing it with a data broker. It's worth noting that in Apple parlance, data collection implies "*sending the data off-device in a way that allows developers or third-parties to access it for a period longer than what is necessary to service the transmitted request in real time*" [4].

iOS app developers are encouraged to declare the type of data to which each practice alludes, as well as the category in which the data may fall and the purpose for accessing it. Data items, data categories, and purposes are declared in the privacy labels from among those provided by apple [4].

2.2. Android Privacy Labels

Android privacy labels went into effect in April 2022, following a similar overall format for data item and data types, but with some remarkable differences in purposes and practices, when compared to iOS labels. Android practices distinguish between data collection and data sharing. Android data collection refers to the same concept as iOS but, interestingly, according to Android terms, a piece of data does not need to be disclosed as "collected" when it is sent off-device over an encrypted connection.

Data sharing refers to a broader concept than iOS Data Used to Track, where "Sharing" refers to transferring user data to a third party. This practice might not be necessarily disclosed in the labels if the data is previously anonymized. As can be seen in Figure 1 (left), data categories, data types, and purposes follow a similar disclosing format compared to iOS [6].

2.3. Privacy Labels mapping

The mapping between iOS and Android labels is not straightforward. By comparing data practices, data items, and purposes we could observe many-to-many relationships between iOS and Android labels. An example of the intricacies of the relationship between labels is shown in Figure 2, where iOS Developer's Advertising or Marketing purpose is mapped to Android Advertising or marketing and Developer communications purposes, while iOS Third-Party Advertising is also mapped to Android Advertising or marketing.

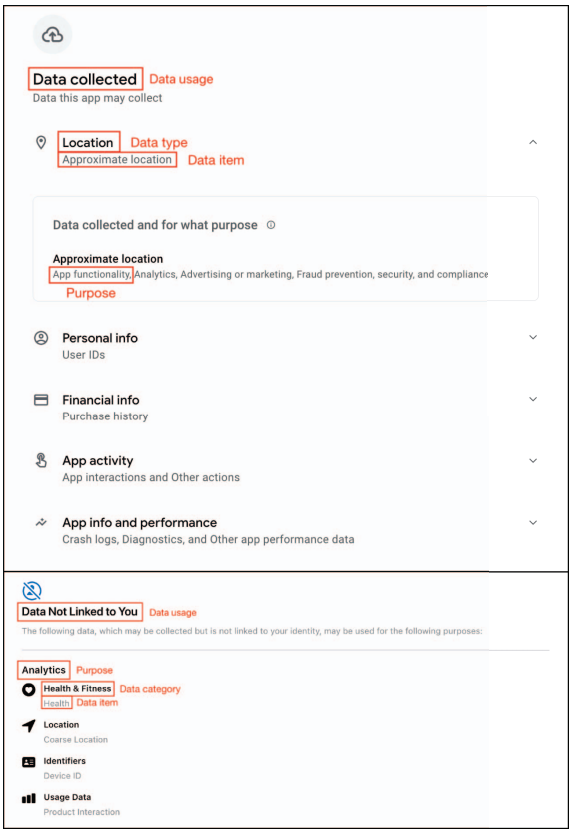


Figure 1. Android (top) and iOS (bottom) privacy labels example

Figure 2. Example of mapping between iOS (left) and Android (right) purpose labels

The many-to-many relationship between the platforms' labels hinders making a straightforward comparison between all iOS and Android data labels. We have therefore limited our comparison to exclusively identical practices and data (i.e., one-to-one relationships). Data Used to Track in iOS alludes to the specific purpose of tracking, which is a subset of possible uses given to data shared in Android. The great difference between these two labels made us limit our comparison to data collection practices and the data items shown in table 1. Purposes were excluded from our analysis.

2.4. Related Work

The notion of privacy labels is an adaptation of nutrition labels introduced by Kelley et al. [7]. Their work aimed to create an information design (i.e., privacy labels) that could improve the comprehensiveness and understanding of privacy policies. In addition, their privacy labels were intended to disclose the collection, use, and sharing of personal data by organizations in an easy-to-understand format. However, the use of privacy labels for

TABLE 1. MAPPING BETWEEN IOS AND ANDROID PRIVACY LABELS

	iOS Label	Android Label
Data practice	Data Linked to You Data Not Linked to You	Data collected
Data items	Sensitive Info	Race and ethnicity Political or religious beliefs Sexual orientation
	Precise Location	Precise location
	Coarse Location	Approximate location
	Name	Name
	Email Address	Email address
	Phone Number	Phone number
	Physical Address	Address
	Other User Contact Info	Other info
	User ID	User IDs
	Device ID	Device or other IDs
	Health	Health info
	Fitness	Fitness info

the mobile ecosystem was lately proposed by Kelley et al. [8].

Since the adoption of privacy labels by Apple, their various uses and advantages are under discussion. Zhang et al. [25] checked the effectiveness of iOS privacy labels by comparing their readability, comprehensibility, salience, and relevance with those of privacy policies. They conducted an in-depth interview study with 24 iPhone users to investigate their experiences, understanding, and perceptions of Apple's privacy labels. The research concluded that "*Apple's privacy labels still do not fully support users' understanding of disclosed application privacy practices*".

Other related works have analyzed the content of iOS privacy labels [11], [13], [18], and assessed their trustworthiness [9], [24]. In particular, Xiao et al. [24] conducted the first comparison between apps' privacy labels and their actual behavior by conducting a dynamic analysis on 5,102 iOS apps, reporting inconsistencies in 3,423 privacy labels. A study of a similar nature was conducted by Koch et al. [9], analyzing 1,687 iOS apps for privacy labels' correctness. During their analysis, they could observe that "*At least 276 [...] apps violate their privacy label by transmitting data without declaration, showing that the privacy labels' correctness was not validated during the app approval process*".

But even fewer studies [3], [19] have addressed Android privacy labels. Closer to our work, Mozilla has conducted a study [19] on the Android top 20 free apps and top 20 paid apps comparing the privacy policy with the privacy labels for each app. The study concludes by finding discrepancies between privacy policies and privacy labels for nearly 80% of the apps reviewed.

Likewise, a reduced number of studies have addressed the comparison between iOS and Android from a privacy perspective [2], [10], [12]. For example, Kollnig et al. [10] used static and dynamic analysis techniques to assess iOS and Android applications identifying personal data leaks. During this comprehensive work, they also analyzed the recipient's identity and location to uncover compliance issues.

The related works described above have either focused on analyzing privacy labels in a single ecosystem or have compared apps' privacy behavior in both domains. To the best of our knowledge, no prior work has addressed a comparison between iOS and Android privacy labels.

3. Method

In this Section, we describe our analysis methodology. We begin by detailing the iOS and Android apps' selection process in Section 3.1. Afterward, in Section 3.2, we provide details on how we collected the iOS and Android privacy labels. Finally, in Section 3.3, we describe the method to perform a static analysis on the apps and check whether the privacy labels match with actual apps' code.

3.1. iOS and Android apps selection

Conducting a comparison between the privacy labels of App Store and Google Play Store applications requires a dataset containing matching applications from both markets. To create a dataset of this kind we pursued the following steps.

iOS apps selection. We scraped the App Store website to collect the name and other details (e.g., privacy policy URL) of the whole list of available applications. From the list, we randomly selected a subset to conduct this study.

Finding matches in Google Play Store. We created an automated method to find the Android app matching each iOS app in our dataset. The method follows a two-phase pipeline: 1) looking for the iOS app name in the Play Store apps' search bar, and 2) comparing both apps' information to determine if they are actually the same app.

The first phase was a straightforward process where we selected the first result (i.e., potentially the most similar according to Play Store). In order to perform the second phase, we first evaluated several approaches on which we based our comparison: application name, developer name, website URL, privacy policy URL, and app's logo. Our evaluation determined that the logo comparison is the most reliable approach to determining if two applications are indeed the same on both platforms.

The logo comparison method downloads both logos and then performs a comparison based on the Structural Similarity Index Measure (SSIM) [23] between them. Our initial validation in a random sample of 30 applications outputted a 1.0 precision score when using an SSIM threshold of 0.9 in the range of [0-1]. However, we found a few false positive cases when validating a larger dataset.

Apps filtering. Since we needed to ensure that the dataset contained only matching applications for the labels comparison, we conducted an additional step to discard potentially incorrectly tagged apps. This filtering consisted in comparing the privacy policies of both apps available on their corresponding platforms.

Before comparing the privacy policies, we required collecting and ensuring they are indeed privacy policies.

We employed Selenium to retrieve those policies loaded through dynamic code. Likewise, it was necessary to discard those URLs leading to non-privacy policies' websites (e.g., landing pages). To do so, we used a machine learning-based classifier that allows us to differentiate between privacy policies and other texts. This classifier is based on the Support Vector Machines (SVM) algorithm and was trained with 195 manually classified texts, achieving 98.76% precision, 97.56% recall, and 98.15% F1 score when evaluated against 100 unseen English texts. After discarding non-privacy policies texts, we compared them by computing the cosine similarity [21] to dismiss the applications that did not perfectly match (i.e., cosine similarity of 1.0) based on the similarity between privacy policies.

3.2. Privacy labels collection and comparison

iOS labels collection. We iterated the process of scraping the privacy labels for each iOS application. The iOS privacy labels are dynamically loaded in a pop-up window after clicking on a "See details" button. This mandatory interaction requires the use of Selenium to trigger the button and load the HTML. Afterward, the BeautifulSoup python library [16] parses the HTML code and we iterate the collection of practices, types of personal data, and purposes.

Android labels collection. Google Play Store does not reuse the same web resource for the privacy labels disclosure as iOS does. Instead, it serves a different resource where the privacy practices of the app are disclosed (i.e., the safety section). This allows us to scrap this info and load it with BeautifulSoup to parse the HTML. Thus, we can collect the practices, types of personal data, and purposes in a shorter period of time than for iOS apps.

Labels comparison. As we explained in Section 2, we have two types of practices in iOS and Android privacy labels. The Android "data collected" practice could be directly compared with the "Data Linked to You" and "Data Not Linked to You" in iOS, while the data shared in Android cannot be compared with the "Data Used to Track You" due to the different meaning explained in Section 2. Therefore, we limit the comparison of privacy labels to data collected usage. Along the process two different topics will be compared: 1) the aggregated personal data types in each ecosystem and 2) the data collected by the same apps (i.e., Android app and iOS app).

3.3. Comparing Android privacy labels to actual app code

The popularity of third-party libraries has grown to the point that it has surpassed the amount of developers' source code [22]. In some cases, developers can be unaware of the whole behavior of these libraries, which could lead them to incorrectly select their app's privacy labels. Sometimes developers simply do not know or understand in detail how their application behaves in terms of privacy. We also aim to look at the code of Android apps and compare our analysis of the code with disclosures provided in privacy labels.

To do so, we relied on the following pipeline to perform static app analysis. First, we use an unofficial API [1] to download the Android applications. Afterward, we decompile and re-build the java code of the apk file with jadx [17], obtaining the Manifest file along with the application's smali and java code.

Once the application's source code is built, we check the permissions in the manifest file to see whether the app is requesting access to personal data. Additionally, we automatically inspect all java files looking for the API calls that access the personal data. We use this information to perform the comparison with the app privacy labels looking for dissimilarities.

4. Evaluation in the wild

This section describes the evaluation conducted to compare Android and iOS privacy labels following the method described in Section 3.

To collect the dataset of apps and privacy labels we started by looking for details on 35k randomly selected apps in the App Store. While scraping the website to get the name and information for each app, we simultaneously searched for the matching Android application. After this process, we found almost 11k matching candidate pairs of iOS and Android apps. We further analyzed each pair by applying the methods described in Section 3.1, successfully identifying 1,423 exact matches.

Out of these 1,423 exact matches, 1,106 of the iOS apps have privacy labels, while only 911 Android apps do. The intersection set yields 822 apps, namely apps that have published labels in both app stores. Our analysis focuses on these 822 apps.

Figure 3 shows the overall number of Android (green) and iOS applications (blue) claiming to collect each data type. As can be seen, there are mismatches in the privacy labels for some data types. For example, precise and approximate location have a difference of 54% in favor of iOS and 36% in favor of Android respectively for each of these data types. This suggests that when collecting location data, iOS apps tend to favor precise location and Android apps favor approximate location.

Another significant mismatch can be seen between user and device identifiers. 43% more iOS apps disclose collecting user identifiers, while 30% more Android apps claim to collect device identifiers. However, the most alarming gap is between the number of apps claiming to collect sensitive information. According to Apple, sensitive information refers to racial or ethnic data, sexual orientation, disability, and religious or philosophical beliefs, among others. Although we can observe a low number of apps reporting the collection of this data type, we would expect to see exactly the same number of apps on both platforms due to its sensitivity. Interestingly too, almost twice as many iOS apps declare to collect user health data, considered by privacy regulations (i.e., GDPR) as sensitive data.

Nevertheless, the most compelling comparison that can be done between Android and iOS in relation to privacy labels is to determine whether the same types of personal data are claimed to be collected for the same application (in both ecosystems). This is shown in Figure 4, where we report the number of Android apps that disclose the collection of a given data type 1) only in Android (green), 2) only in iOS (blue), 3) in both (yellow). The

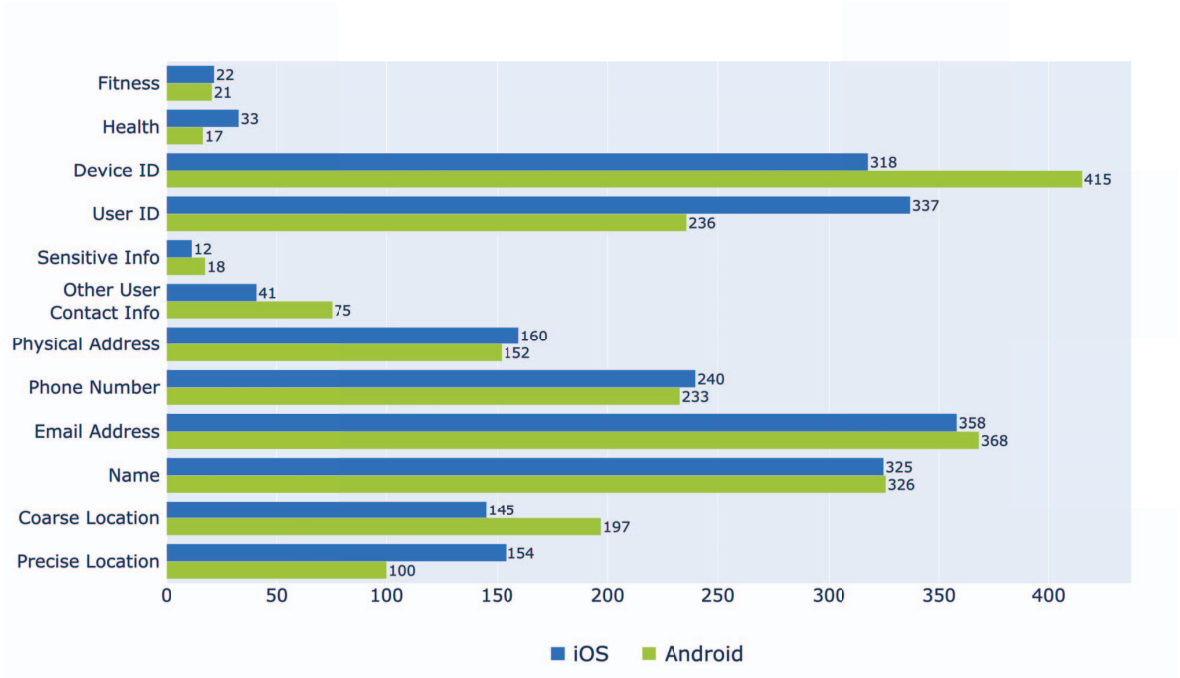


Figure 3. Comparison between the number of Android (green) and iOS (blue) apps declaring the collection of each data type

remaining applications up to the 822 analyzed correspond to those that do not declare the collection of the data in any marketplace. Again, a remarkable gap between Precise location, Coarse location, User ID, and Device ID can be observed Surprisingly, we can observe that only seven applications match reporting the collection of sensitive information, while 16 differ.

The comparison of privacy labels is limited to what developers report their applications to do. However, apps may intentionally or unintentionally be accessing non-reported personal data. To assess their statements, we managed to successfully download and perform static analysis on 560 out of the 822 Android apps.

First, we checked whether applications have access to two types of personal data: coarse location and precise location. Figure 5 shows a comparison between the apps that request access to these data and whether their collection is declared in the privacy labels. As can be seen, 36.6% and 38.2% of the apps that do not declare collecting coarse and precise location respectively, request permissions to access these data.

We further reconstructed the java source code out of the smali code for the 560 applications. As described in Section 3.3, we have looked for the API calls that retrieve the precise and coarse location, observing that for the 54 apps found accessing these two data types, none of them disclose their collection in the privacy labels. Although not conclusive due to static analysis limitations [15] our findings suggest a mismatch between what labels disclose and what apps may actually be doing.

5. Discussion

A **substantial 66.5%** of the 822 applications analyzed **show potential discrepancies between iOS and Android privacy labels**. Moreover, out of the 503 that claim to collect personal data in both marketplaces, **only 16 (3.2%) agree on all the data types mapped** in Section 2. These results suggest notable differences in data practice disclosure in the iOS and Google Play app stores for apps that one would otherwise have expected to have identical or nearly identical data practice disclosures.

There are two major possible explanations for our observations: 1) apps indeed behave differently in each ecosystem and the privacy labels are correct and consistent with that, or 2) the difference suggests apparent inconsistencies of the data practice disclosures of the privacy labels. If it is the first case, we would have applications with exactly the same privacy policy while carrying out different privacy practices. In contrast, the second case is supported by evidence that Android privacy labels show apparent inconsistencies with apps' code. As noted in Section 4, 248 apps (44.3%) request permission to collect coarse or precise location even though none of them are disclosing it in the privacy labels. Moreover, we found 54 of those applications accessing these data in their source code, neither reported on the labels.

Of course, accessing personal data that has not been reported in the data collected section of the privacy labels does not imply non-compliance. As stated in Section 2, in privacy labels parlance "data collected" refers to sending the data off-device, and therefore it might be the case that data is retrieved but never sent out by apps. Nevertheless, it is remarkable to notice that none of the apps found accessing the data did report the collection.

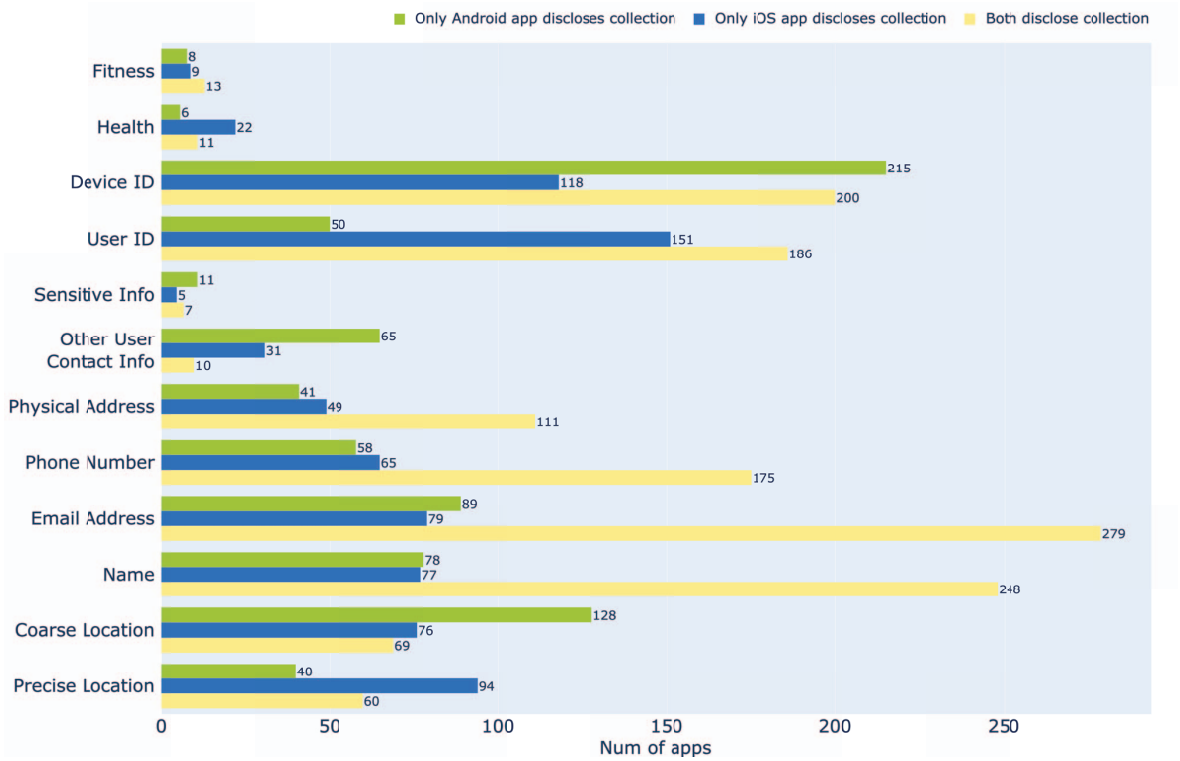


Figure 4. Number of apps disclosing a data item collection in the privacy labels

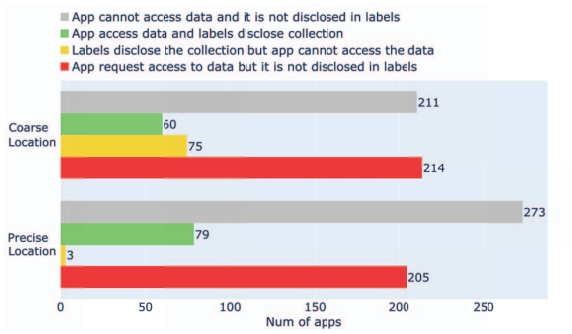


Figure 5. Comparison between Android apps' permissions requested and data collection disclosed in the privacy labels

Differences between the number of applications disclosing collection of precise location and coarse (approximate) location in Android and iOS. In figures 3 and 4 we could see that iOS apps mostly collect precise location while Android does the same with coarse location. Apple considers as precise location the latitude and longitude coordinates with three or more decimal places, equivalent to a dispersion of 110 meters. Google does not define what it considers as precise location but determines that the approximate location is the one capable of locating the user in an area of 3 square kilometers. On the other hand, Apple considers approximate location to be the latitude and longitude coordinates with two decimal places or less, which is equivalent to 1.1 km of dispersion. Thus, the divergences between the definitions of these two data items highlight the disparity between iOS and Android privacy labels, but do not justify the noticeable differences seen in our results.

Apparent inconsistencies found about users' health data being collected in the same applications. We manually inspected apps in two situations: 1) an iOS app that claims to be collecting health data while the same Android app does not; and, 2) the exact opposite situation. Matching with the first situation (1), the Forever GoFit app claims to collect users' health data in iOS but not in Android (with over 50k downloads in Google Play Store). Nevertheless, in the Google Play Store app's description, they state their app "*counts your steps, calories, active time, distance, and record and analyze your sleep and heart rate*". Interestingly, it was last updated during the last month and these functionalities also appear in the Apple Store app's description, along with the same apps' screenshots in both marketplaces. We found the opposite (2) in the VitalFlo Health app, which describes exactly the same functionalities and app screenshots in both marketplaces, while only Android discloses health collection. The main app's purpose is to "*record and track your lung function and symptoms, and automatically sync with your doctor*", where evident health data collection is occurring.

6. Threats to validity

Construct validity. Not all data types can be compared among platforms, only those for which we have found a one-on-one relationship. This may make our comparison not generalizable to other data types. The

number of apps we have analyzed is large and the proposed method for finding iOS apps on Android has a high accuracy, which has been increased by performing an in-depth comparison between privacy policies. This makes negligible the possible error of incorrectly matching one app to another. However, obtaining such a high accuracy along with the fact that not all applications disclose the privacy labels, involved a considerable reduction in the size of the evaluation dataset. The jadx tool decompiles and builds Java files for all Android applications, even if the construction of the source code is not properly achieved. This may lead to an increase in false negative cases of applications retrieving location data type which nevertheless does not introduce false positive results.

External validity. App Store unlike Google Play Store asks developers for a monthly fee to maintain apps on the market. This divergence could lead to greater attention on the app's details provided, and a lower number of unmaintained apps when compared to Google Play Store. This may also be a bias in favor of the iOS applications when comparing the privacy labels.

7. Conclusions

Increasing privacy awareness by users and regulatory pressure by supervisory authorities has led large technology companies (i.e., Google & Apple) to focus their efforts on creating privacy labels for their apps marketplace. In this article, we have compared these labels to check if they are consistent, or if they involve considerable differences for the same applications in the different ecosystems.

Following the proposed method, we collected and compared the privacy labels of 822 apps. Through the comparison, we observed that only 3.2% of the apps disclosing the collection of data coincide in both ecosystems, while a remarkable 44.3% of the analyzed apps are requesting permissions to retrieve data they have not disclosed in their labels. The divergences between iOS and Android privacy labels for applications with the same privacy policy confirm the existence of apparent privacy inconsistencies. We hope that these findings serve as a call to action for regulators. In future work, we aim to conduct a dynamic analysis of the apps to further analyze the actual apps‘ behavior and compare it with the privacy labels.

8. Acknowledgements

This research has been partially supported by the project TED2021-130455A-I00 funded by MCIN/AEI/10.13039/501100011033 and the Europea Union "NextGenerationEU"/PRTR. Grants from the National Science Foundation Secure and Trustworthy Computing program (CNS-1801316, CNS-1914486) and an unrestricted Privacy Faculty Award from Google helped fund this study.

References

[1] Rehmat Alam. Command line google play apk downloader. github repository. https://github.com/rehmatworks/gplaydl, 2021. Accessed: 2023-03-29.

[2] Ying Chen, Heng Xu, Yilu Zhou, Sencun Zhu, and George Washington. Is this app safe for children?: a comparison study of maturity ratings on android and ios applications. pages 201–212. Association for Computing Machinery (ACM), 5 2013.

[3] Lorrie Faith Cranor. Mobile-app privacy nutrition labels missing key ingredients for success. *Communications of the ACM*, 65:26–28, 11 2022.

[4] Apple Developer. App privacy details. https://developer.apple.com/app-store/app-privacy-details/. Accessed: 2023-03-29.

[5] Jack Gardner, Yuanyuan Feng, Kayla Reiman, Zhi Lin, Akshath Jain, and Norman Sadeh. Helping mobile application developers create accurate privacy labels. *Proceedings - 7th IEEE European Symposium on Security and Privacy Workshops, Euro S and PW 2022*, pages 212–230, 2022.

[6] Play Console Help. Provide information for google play's data safety section. https://support.google.com/googleplay/android-developer/answer/10787469?hl=en, 2022. Accessed: 2023-03-29.

[7] Patrick Gage Kelley, Joanna Bresee, Lorrie Faith Cranor, and Robert W. Reeder. A "nutrition label" for privacy. In *Proceedings of the 5th Symposium on Usable Privacy and Security*, SOUPS '09, New York, NY, USA, 2009. Association for Computing Machinery.

[8] Patrick Gage Kelley, Lorrie Faith Cranor, and Norman Sadeh. Privacy as part of the app decision-making process. *Conference on Human Factors in Computing Systems - Proceedings*, pages 3393–3402, 2013.

[9] Simon Koch, Malte Wessels, Benjamin Altpeter, Madita Olvermann, Martin Johns, and : Datenanfragen. Keeping privacy labels honest. In *Privacy Enhancing Technologies Symposium*, pages 486–506, 2022.

[10] Konrad Kollnig, Anastasia Shuba, Reuben Binns, Max Van Kleek, and Nigel Shadbolt. Are iphones really better for privacy? comparative study of ios and android apps. *Proceedings on Privacy Enhancing Technologies*, 2022:6–24, 9 2021.

[11] Konrad Kollnig, Anastasia Shuba, Max Van Kleek, Reuben Binns, and Nigel Shadbolt. Goodbye tracking? impact of ios app tracking transparency and privacy labels. *ACM International Conference Proceeding Series*, 22:508–520, 6 2022.

[12] Lydia Kraus, I. Wechsung, and S. Möller. A comparison of privacy and security knowledge and privacy concern as influencing factors for mobile protection behavior. In *Workshop on Privacy Personas and Segmentation (PPS)*, 2014.

[13] Yucheng Li, Deyuan Chen, Tianshi Li, Yuvraj Agarwal, Lorrie Faith Cranor, and Jason I. Hong. Understanding ios privacy nutrition labels: An exploratory large-scale analysis of app store data. *Conference on Human Factors in Computing Systems - Proceedings*, 4 2022.

[14] Orson Lucas, Martin Sokalski, and Rob Fisher. Corporate data responsibility: Bridging the consumer trust gap. https://advisory.kpmg.us/articles/2021/bridging-the-trust-chasm.html?utm_source=vanity&%3Butm_medium=referral&%3Bmid=m-00005652&%3Butm_campaign=c-00107353&%3Bcid=c-00107353. Accessed: 2023-03-29.

[15] Dimitri Prestat, Naouel Moha, and Roger Villemaire. An empirical study of android behavioural code smells detection. *Empirical Software Engineering*, 27:1–34, 12 2022.

[16] PyPI. beautifulsoup4. https://pypi.org/project/beautifulsoup4/. Accessed: 2023-03-29.

[17] GitHub repository. Dex to java decompiler. https://github.com/skylot/jadx. Accessed: 2023-03-29.

[18] Gian Luca Scoccia, Marco Autili, Giovanni Stilo, and Paola Inverardi. An empirical study of privacy labels on the apple ios mobile app store. *Proceedings - 9th IEEE/ACM International Conference on Mobile Software Engineering and Systems, MOBILESoft 2022*, pages 114–124, 2022.

[19] Anne Stopper and Jen Caltrider. See no evil: Loopholes in google's data safety labels keep companies in the clear and consumers in the dark. mozilla foundation. https://foundation.mozilla.org/en/campaigns/googles-data-safety-labels/, 2021. Accessed: 2023-03-29.

[20] Apple Support. About privacy information on the app store and the choices you have to control your data. https://support.apple.com/en-us/HT211970, 2021. Accessed: 2023-03-29.

[21] Vikas Thada and Vivek Jaglan. Comparison of jaccard, dice, cosine similarity coefficient to find best fitness value for web retrieved documents using genetic algorithm. *International Journal of Innovations in Engineering and Technology*, 2, 2013.

[22] Haoyu Wang, Yao Guo, Ziang Ma, and Xiangqun Chen. Wukong: A scalable and accurate two-phase approach to android app clone detection. pages 71–82. Association for Computing Machinery, Inc, 7 2015.

[23] Zhou Wang, Alan Conrad Bovik, Hamid Rahim Sheikh, and Eero P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13:600–612, 4 2004.

[24] Yue Xiao, Zhengyi Li, Yue Qin, Xiaolong Bai, Jiale Guan, Xiaojing Liao, and Luyi Xing. Lalaine: Measuring and characterizing non-compliance of apple privacy labels at scale. *arXiv*, June 2022.

[25] Shikun Zhang, Yuanyuan Feng, Yaxing Yao, Lorrie Faith Cranor, and Norman Sadeh. How usable are ios app privacy labels? In *Privacy Enhancing Technologies Symposium*, pages 204–228, 2022.

[26] Shikun Zhang and Norman Sadeh. Do privacy labels answer users' privacy questions? In *Network and Distributed System Security Symposium*, 2023.

IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 17 November 2023, accepted 14 December 2023, date of publication 3 January 2024,
date of current version 11 January 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3349425

RESEARCH ARTICLE

Sharing is Not Always Caring: Delving Into Personal Data Transfer Compliance in Android Apps

DAVID RODRIGUEZ[1], JOSE M. DEL ALAMO[1], CELIA FERNÁNDEZ-ALLER[2], AND NORMAN SADEH[3], (Member, IEEE)
[1]ETSI Telecomunicación, Universidad Politécnica de Madrid, 28040 Madrid, Spain
[2]ETSI Sistemas Informáticos, Universidad Politécnica de Madrid, 28040 Madrid, Spain
[3]School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213, USA

Corresponding author: Jose M. Del Alamo (jm.delalamo@upm.es)

This work was supported in part by the Ministerio de Ciencia e Innovación (MCIN)/Agencia Estatal de Investigación (AEI)/10.13039/501100011033 under Project TED2021-130455A-I00, and in part by European Union "NextGenerationEU"/Plan de Recuperación, Transformación y Resiliencia (PRTR). The work of Jose M. Del Alamo was supported by Spanish "Ministerio de Universidades" through "Movilidad" Sub-Program of "Programa Estatal para Desarrollar, Atraer y Retener Talento," within "Plan Estatal de Investigación Científica, Técnica y de Innovación 2021–2023."

ABSTRACT In an era marked by ubiquitous reliance on mobile applications for nearly every need, the opacity of apps' behavior poses significant threats to their users' privacy. Although major data protection regulations require apps to disclose their data practices transparently, previous studies have pointed out difficulties in doing so. To further delve into this issue, this article describes an automated method to capture data-sharing practices in Android apps and assess their proper disclosure according to the EU General Data Protection Regulation. We applied the method to 9,000 random Android apps, unveiling an uncomfortable reality: over 80% of Android applications that transfer personal data off device potentially fail to meet GDPR transparency requirements. We further investigate the role of third-party libraries, shedding light on the source of this problem and pointing towards measures to address it.

INDEX TERMS Android, compliance assessment, data protection, data transfer, dynamic analysis, GDPR, large language model, personal data, privacy policy, third-party.

I. INTRODUCTION

Data privacy, often identified as data protection, has become a hot topic, gaining increasing attention over recent years. This surge is primarily fueled by growing user concerns, prompting the formulation and update of data protection laws. Among such regulations, the General Data Protection Regulation (GDPR) [1] stands out as the European mandate for data protection with a global impact worldwide [2]. In fact, this regulation has been a model for drafting similar legal provisions in other countries [3].

The GDPR is founded on seven key principles [4]: lawfulness, fairness, and transparency; purpose limitation; data minimisation; accuracy; storage limitation; integrity and confidentiality (security); and accountability. Among them, our study particularly emphasizes the principle of transparency.

The transparency principle mandates that any information and communication relating to personal data processing be presented "*in a concise, transparent, intelligible, and easily accessible form, using clear and plain language*" (GDPR, Art. 5(1)(a)). Furthermore, Article 13 specifies the information that the data controller (i.e., the entity determining the purposes and the means of processing of personal data) must provide to the data subject (i.e., the individual whose data is being processed) when collecting their personal data. In particular, Article 13(1)(e) asserts that the data controller shall specify the "*recipients or categories of recipients of the personal data, if any*". In GDPR terms (Article 4 (9)), a recipient is "*a natural or legal person, public authority, agency or another body, to which the personal data are disclosed, whether a third party or not.*". Third

The associate editor coordinating the review of this manuscript and approving it for publication was Alessandro Pozzebon.

parties are defined in the Article 4 (10) as "*a natural or legal person, public authority, agency or body other than the data subject, controller, processor and persons who, under the direct authority of the controller or processor, are authorised to process personal data*". Therefore, Article 13(1)(e) requires disclosing the identity or categories of any recipient of personal data other than the data controller.

Personal data recipients are particularly abounding in the mobile app ecosystem. Modern software development paradigms call for the integration of online services, exposed through Application Programming Interfaces (APIs), to streamline app development and monetize them. These services offer convenient functionalities and features the app provider does not need to create from scratch e.g., app analytics, identity services, or ads serving, offering significant time-saving and revenue opportunities. To this end, they usually carry out personal data transfers to different recipients. Oftentimes these services are wrapped as code libraries (also known as third-party libraries or Software Development Kits - SDKs), which are packed in the app and delivered jointly with the app's own code into the user device. The popularity of some SDKs has escalated to the point where these libraries now contribute more code to the app than the app developers' own code [5]. Our research specifically focuses on these libraries and their personal data transfers, investigating how their involvement in data processing often remains obscured or inadequately disclosed to the users, thus potentially breaching GDPR transparency requirements.

While SDKs are capable of collecting and transmitting users' personal data to external recipients, it is important to clarify that the responsibility for GDPR compliance is in the app's data controller. When an app is granted permission to access certain resources, all integrated libraries, including these SDKs, automatically inherit these permissions. However, they do not independently determine their use. Consequently, it is crucial for the apps' data controllers to have a comprehensive understanding of the functionalities of these SDKs. They must ensure that their use of these SDKs aligns with GDPR requirements, particularly concerning transparency and lawful processing of personal data. This is essential to mitigate privacy risks for users and to uphold the app's compliance with GDPR mandates.

Official guidelines on transparency under GDPR state that "*the actual (named) recipients of the personal data, or the categories of recipients, must be provided. In accordance with the principle of transparency and fairness, controllers must provide information on the recipients that is most meaningful for data subjects. In practice, this will generally be the named recipients, so that data subjects know exactly who has their personal data. If controllers opt to provide the categories of recipients (either because their identity may regularly change, or because the list would be overwhelmingly long), the information should be as specific as possible by indicating the type of recipient (i.e. by reference to the activities it carries out), the industry, sector and sub-sector and the location of the recipients*" [6]. The same idea is supported by other guidelines provided by the European Commission [7], by the Information Commissioner's Office (ICO) in the UK [8], and by the Court of Justice of the European Union [9]. These official guidelines are summarized in Table 1, where both possible mechanisms —actual name or categories— are presented.

However, disclosing the personal data recipients' categories with the level of detail required by the aforementioned guidelines — including specifics like activity, industry, sector, sub-sector, and location — is notably more challenging and less precise than simply identifying recipients by name. Consequently, the disclosure of such detailed information would be rare. In fact, we manually reviewed 100 privacy policies from applications identified as potentially non-compliant (i.e., transferring personal data to unnamed recipients) during our experiments. Notably, none of them detailed the activity, sector, sub-sector, or location of the unnamed recipients. This omission suggests non-adherence to regulatory guidelines, potentially leading to GDPR non-compliance. In response to these insights, our work is specifically designed to focus on the identification and verification of explicitly named recipients.

Our study aims to shed light on the transparency of personal data transfers in the Android ecosystem and the role of code libraries in this process. To this end, we describe an automated method to capture data-sharing practices in Android apps, assess their proper disclosure according to the GDPR transparency requirements, and understand the source of discrepancies.

To demonstrate the applicability of the method in the wild, we have applied it to 9,000 random Android apps from the Google Play Store, unveiling an uncomfortable reality: over 80% of Android applications that transfer personal data off device potentially fail to meet GDPR transparency requirements. Our findings further suggest that libraries seem to be at the core of the nondisclosure issues.

Aiming to improve adherence to data protection principles in the Android ecosystem, this study describes a fully automated approach that serves multiple stakeholders. Regulators may leverage this method for preliminary, large-scale examinations of privacy practices, subsequently narrowing their focus to apps exhibiting potential non-compliance for detailed investigation. App providers, often unaware of the activities of the SDKs integrated into their applications, can utilize this system to gain insight into SDK behaviors, ensuring accurate disclosures in their privacy policies. Moreover, this research is a component of the autoGDPR initiative (http://autogdpr.org), which aspires to establish a public portal presenting app analyses, thus empowering users with knowledge of privacy implications associated with app usage.

The remaining of the article is organized as follows. Section II presents the related works and contrasts them with our study. Section III introduces and explains the

TABLE 1. Transparency requirements of personal data recipients according to official guidelines.

Required transparency element	Description	Specific requirements
Named Recipients	Naming the recipients refers to disclosing the actual names of the recipients of personal data.	-
Categories of Recipients	If providing named recipients is impossible, controllers may choose to disclose categories of recipients.	• **Type of Recipient**: Description based on the activities carried out by the recipient. • **Industry**: The industry to which the recipient belongs. • **Sector and Sub-sector**: Detailed sector and sub-sector classification of the recipient. • **Location**: Geographical location of the recipient.

method we have developed and its components along with their validation. The method's application on a set of 9,000 Android applications is detailed in Section IV, where the results obtained are also presented. Section V discusses these findings and Section VI concludes the paper.

II. RELATED WORK

Our study identifies personal data recipients in Android apps and assesses whether they are transparently disclosed, further analyzing the extent to which code libraries are implicated in potential compliance issues. This section details those previous works that have touched upon these topics, and how our work compares to them.

Researchers have leveraged static, dynamic, or hybrid techniques to spot personal data transfers in mobile apps [10]. For example, Ferrara and Spoto [11] relied on static code analysis to flag potential personal data leaks in the apps' source code. Jia et al. [12] leveraged dynamic techniques to detect personal data disclosures in network packets. Jia's work could be seen as complementary to ours as we also leverage dynamic analysis techniques to identify personal data transfers, yet we further focus on assessing a transparent disclosure of these data transfers according to GDPR requirements.

Once a personal data transfer is detected the recipient needs to be identified so as to understand if it is disclosed in the app privacy policy. There have been numerous prior works focusing on identifying personal data recipients in both the web and mobile ecosystems [13], [14], [15]. Often referred to as trackers because they specialize in advertising and marketing, these kinds of organizations are the focus of most studies [16], [17]. In a previous work [18], we elaborated a method to reveal the identity of the recipient of a personal data transfer. In addition to the contribution of that work, this paper has the goal of checking whether the apps' disclosures meet the transparency requirements set forth by the GDPR.

The behavior of the code libraries that apps integrate has been the focus of previous research too. Despite most related works concentrating on malware detection [5], [19], [20], [21], some previous studies have attempted to identify code libraries and their personal data leaks. Again, the library identification can be accomplished through static [21], [22], [23], dynamic [24] or hybrid [25], [26] analysis approaches. While these previous works have examined the data transfer behaviors of libraries, demonstrating that some of them pose a significant privacy risk, they have not flagged them as the source of potential compliance issues, as we do in this paper. For the identification of data transfers carried out by libraries, we have leveraged state-of-the-art dynamic analysis techniques that combine the interception of connections in the network with the analysis of stack traces captured during the app execution.

Apps' privacy practices, including the declaration of personal data recipients, have been typically disclosed through privacy policies written in natural language [27]. Machine learning techniques have been widely used to extract information from them. Generally, classifiers are trained with annotated policies [10], which demand a significant time for their coding. Nevertheless, the recent rapid, widespread adoption of chatbots, like ChatGPT based on Large Language Models that do not require specific training, emerges as a promising alternative [28], [29], [30]. Specifically, ChatGPT has demonstrated remarkable performance in processing legal information [31], making it an alternative tool for extracting practices and general information from privacy policies. Our work leverages these state-of-the-art techniques for extracting information from privacy policies, achieving high-performance levels.

Privacy labels, based on the concept of nutrition labels [32], have been recently introduced to disclose privacy practices in mobile apps [33], [34]. Apple introduced privacy labels in the iOS App Store in 2020, compelling app providers to disclose their data practices through a structured schema. A year later, Google introduced the same privacy label concept in its Play Store via its Data Safety Section. Recent studies have shown that privacy labels often contain mistakes [35], [36] and discrepancies with the privacy policies [37], either overstating or understating the apps' privacy practices. In this work, we also analyze the apps' privacy labels to understand if and to what extent they disclose the personal data transfers.

A few previous works have dealt with the assessment of data transfers compliance with GDPR requirements [38], [39], [40]. Razaghpanah et al. [38] analyzed the landscape of tracking services in the mobile ecosystem further discussing GDPR and ePrivacy (the EU legislation on privacy in communications) impact. However, their analysis did not go

deeper into compliance assessment, as they acknowledged "*Our methodology is limited [. . .] and is therefore unable to identify [Advertising and Tracking Services] that are (or, will be) in violation of these regulations*".

Andow et al. [40] and Tan and Song [39] proposed a flow-to-policy check method, which similarly to our work analyzes personal data transfers and then compares the destination entities with those disclosed in the privacy policy. Unlike our proposal, they perform static code analysis to ascertain personal data transfers yet, as stated [39], "*in general, static analysis can obtain more comprehensive data flows, while dynamic analysis can ensure the realness of the detected data flows*". For this reason, and as our method aims to check compliance with GDPR requirements, we have used dynamic analysis to prioritize soundness over completeness.

Furthermore, in our opinion, both papers present limitations to 1) identify named recipients in the privacy policy and, 2) determine the app's data controller. Indeed, both works identify recipients from the destination domain name of the data transfer (e.g., *Adjust* from *http://app.adjust.com*), which precludes detecting organizations that do not resemble their domain name (e.g., *https://teleport.soom.la/* belongs to *ironSource Ltd.*). Andow et al. also follow this strategy for data controller identification by parsing the app package name (e.g., com.mycompany.app). In turn, Tan et al. carry out a manual identification of data controllers, thus obviously limiting the method's scalability. Our method, in contrast, can overcome both, not only accurately identifying the data controller and recipients but also considering whether they are different legal entities, as detailed in the next section.

Therefore, to the best of our knowledge, this is the first work proposing a scalable assessment of the GDPR transparency principle regarding personal data transfers to recipients in Android apps.

III. METHOD

This section details the components of our method to identify personal data recipients in Android apps and assess their transparent disclosure in the apps' privacy policies. Initially, we introduce the method from a high-level perspective, including its context, followed by a thorough explanation of each component's functioning and the validations supporting their performance.

A. OVERVIEW

The method is integrated into a platform that the authors have developed in previous work [41]. This platform automatically downloads applications from Google Play Store including their metadata, privacy policy and privacy labels, installs and runs the apps on real devices with simulated user interactions and events, and intercepts their network communications. To intercept the network connections, this platform utilizes a Man in the Middle (MitM) proxy that enables capturing both HTTP and HTTPS connections. Eventually, the platform searches for already known personal data — extracted from the real devices used — in the body and URL of the intercepted connections, and finally, logs the results (e.g., personal data being transferred, local port used to set up the connection, destination domain, etc.).

Our method receives as input an app's privacy policy and a domain name where the app sends personal information (recipient domain), as observed by the platform. We further modified the platform described above to log the traces leading to a connection setup in the apps, which are also used as inputs to our analysis. The output consists of a disclosure issue flag: positive if the method detects that the recipient has not been adequately disclosed in the privacy policy, or negative otherwise. In the former case, the output also includes the recipient undisclosed and the library initiating the data transfer.

Figure 1 shows the method's main modules, which are responsible for 1) Identifying the personal data recipient (i.e., Recipient Analyzer), 2) assessing their proper disclosure in the app's privacy policy (i.e., Disclosure Checker), and 3) identifying the library triggering the data transfer (i.e., Library Analyzer).

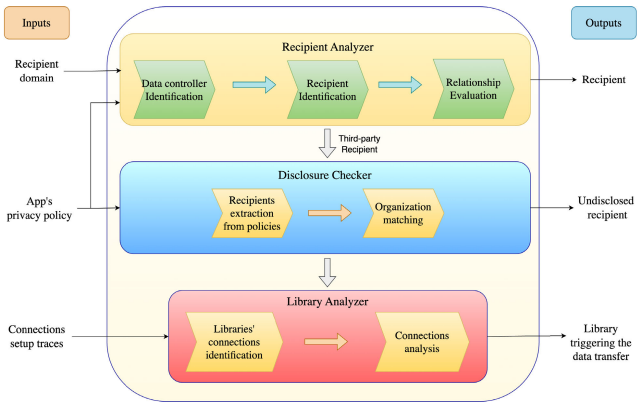


FIGURE 1. Method overview.

The Recipient Analyzer allows determining if the app's data controller and the personal data recipient are actually the same legal entity. The Disclosure Checker evaluates whether the app's privacy policy transparently declares the transfer of personal data to the recipient. The Library Analyzer identifies the library triggering the transfer of personal data to the recipient. Thus, it helps us understand if potential non-compliance issues are related to the use of these libraries.

B. RECIPIENT ANALYZER

This pivotal component aids in determining whether a specific personal data transfer targets the data controller or a recipient. It receives the app's privacy policy and the transfer destination domain, and performs three separate steps: 1) Apps' data controller identification, 2) Recipient identification, and 3) Data controller-to-recipient relationship determination.

1) DATA CONTROLLER IDENTIFICATION

The identification of the app's data controller can be achieved by processing the app's privacy policy text. This is the

most reliable way to identify the organization responsible for the specific app under analysis since the GDPR regulation requires to disclose the data controller. For this task, we utilized the ChatGPT API, specifically the *gpt-3.5-turbo* model, using the *gpt-3.5-turbo-16k* model for lengthy policies.

This step automates a prompt inquiring about the data controller, where the privacy policy is also included in the prompt. According to our tests, the optimal prompt requests the data controller's name for the policy, or NONE if it is unknown (not disclosed). The used prompt is shown in Listing 1.

```
Which company do you think this privacy policy belongs to? Please ONLY name it. Answer NONE if you do not know.
PRIVACY POLICY BEGINS HERE:

# Privacy policy content #
```

LISTING 1. Prompt inquiring about data controller in a privacy policy.

This prompt is optimized to facilitate the model comprehension of key information. Capitalized words are used to indicate where the attention should be focused on, as it has been empirically observed to improve the results. The prompt also specifies where the privacy policy begins, thereby avoiding information confusion.

We validated this step on a set of 50 random applications. We manually checked their policies discarding eight non-English texts. Table 2 details the Data Controller Identification performance metrics.

TABLE 2. Performance metrics for the data controller identification and recipient identification steps.

	Data Controller Identification (chatGPT)	Recipient Identification (ROI)
Precision	97.14%	95.71%
Accuracy	88.09%	67.00%
Recall	89.47%	69.07%
F1-score	93.15%	80.24%

2) RECIPIENT IDENTIFICATION

When intercepting personal data transfers, our platform obtains the transfer contents and the destination domain. However, to determine if the transfer is made to a recipient as for GDPR terms, we first need to know the organization owning the destination domain. We leveraged our previous work on a "*Receiver Organization Identifier*" (ROI) tool [18] to this end.

Briefly, this tool uses Selenium and a web search engine to find the privacy policy governing a web domain. It then checks if the text found is indeed a privacy policy, and extracts the data controller's identity using the SpaCy library [42] for the entity recognition process. This method partially meets the goal of the *Data Controller Identification* described above, although it also includes a thorough search process to find the privacy policy governing the target domain. Besides, it applies SpaCy for the data controller identification instead of ChatGPT, as the latter was unavailable at the time of ROI development. The precision of ROI in identifying the organization is 95.71% (Table 2), close to the metric obtained using ChatGPT, although clearly favoring the latter.

3) RELATIONSHIP EVALUATION

Determining the role of an organization as a recipient of personal data transfers involves discerning if the organization receiving the personal data is other than the data controller. Previous works [40], [41] compared the target domain name with the mobile application's name to that end. We significantly improve this process by using their comparison only as a first step. If no (apparent) similarities are found between both fields, ChatGPT is used once again to determine if both the data controller and recipient point at the same entity, with the prompt in Listing 2.

```
Are #company1 and #company2 the same company? Please answer only with YES or NO. They would be the same company if they refer to the same legal entity. They cannot be considered the same company only because they belong to the same corporate network.
```

LISTING 2. Prompt inquiring if both companies are actually the same entity.

ChatGPT serves the specific purpose of dealing with different entities where string matching could lead to false negatives (e.g., renamed companies or corporate designations). This component outputs a boolean answer, allowing us to categorize the recipient as either the data controller or a recipient. We utilize the Disclosure Checker component to verify if the recipients are adequately disclosed in the privacy policies according to the GDPR guidelines.

4) PIPELINE VALIDATION

The Recipient Analyzer pipeline was validated with a dataset of apps and recipients observed in personal data transfers from previous experiments we carried out. Our dataset included 50 destination domains coded as data controllers and 50 destination domains coded as recipients. The codification of a domain as data controller or recipient for an app required manual checks via Crunchbase [43]. The method's performance is detailed in Table 3.

C. DISCLOSURE CHECKER

This component assesses whether the recipients of personal data transfers are transparently disclosed in an app's privacy

TABLE 3. Performance metrics for the recipient analyzer and the disclosure checker components.

Performance metric	Recipient Analyzer	Disclosure Checker
Accuracy	88.00%	95.00%
Precision	84.91%	100.00%
Recall	91.84%	86.36%
F1-score	88.24%	92.68%

```
The following text is a privacy policy of an
Android application.

# Privacy policy content #

End of the privacy policy.

Which third-party companies, service providers or
partners are described in the privacy policy?
```

LISTING 3. Prompt inquiring about the recipients described in the privacy policy.

policy. To this end, we leveraged ChatGPT again to extract from apps' privacy policies the recipients that may receive personal data transfers. The ChatGPT prompt used for this is shown in Listing 3.

We process the output set to remove unnecessary business designations (e.g., Corp or Inc). Afterward, we check whether the organizations observed to receive personal data are on this set to determine if the transfers are explicitly disclosed as per GDPR requirements.

Like the previous pipeline, this one underwent a validation using an untapped dataset consisting of 60 randomly chosen privacy policies where the recipients disclosed were coded by one of the authors. The performance in identifying the recipients disclosed in the privacy policy is detailed in Table 3. The remarkable precision highlights the method's capability to correctly identify when a privacy policy transparently discloses the identity of a personal data recipient.

D. LIBRARY ANALYZER

This component aims to discern whether a library initiated a personal data transfer to a recipient. This will allow us to match transparency disclosure issues in privacy policies to libraries, thus shedding light on potential sources of compliance issues.

This method leverages Frida [44], a dynamic analysis tool used to instrument mobile applications and injects new code into a running process. The app behavior can thus be traced and modified at runtime. We capitalized on Frida to capture network connections established by an app where a socket is set up through the Android (standard) platform API. Thus, when a connection is set up its metadata is logged, including the destination domain, the local port used in the device, and the call stack trace. In turn, our MitMProxy also intercepts the device connections in the network and logs the associated metadata (destination domain, local port), reads their content (even if encrypted), and searches for personal data transfers. Thus, we can easily associate the connections identified with the MitMProxy with those logged by Frida by matching their local (origin) ports. Figure 2 depicts this process.

The benefit of matching connections captured with both techniques is that we enrich our knowledge on the data transfer with information on the source code that established these connections, as for the stack traces obtained with Frida. As a result, we are able to understand if a library is the source of issues in the data transfer disclosures.

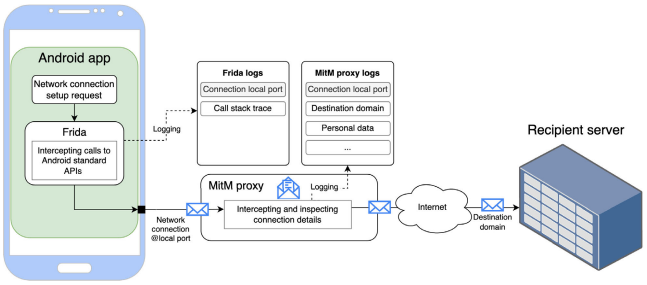


FIGURE 2. Data transfers interception with Frida and a MitM proxy.

1) UNDERSTANDING STACK TRACES

To identify libraries in traces, we first need to understand how a trace is structured. Figure 3 shows an example of a trace, where a connection being setup through the Socket.connect method of the Android standard API is at the top of the trace, as this was one of the methods we monitored. Conversely, the method that triggered the connection setup appears at the bottom of the stack trace, in this case, the *Thread.run* method of the Android standard API. In between, two main pieces of code are found, as for the package names of the classes involved: *com.android.okhttp* belonging to the OKHttp library [45], widely used in Android for handling HTTP and HTTP2 client-side communications; and, *com.my.tracker* belonging to the myTracker library [46], an analytics library.

Upon analyzing the various cases we have encountered in the traces we logged, we observed that oftentimes library methods are the only ones involved in the connection setup leading to a data transfer. That is, the Android standard API methods – i.e. android/dalvik/java/javax namespaces, the JetPack library ones – androidx namespace, or any non-library code was not invoked. In these cases, we consider that it wasn't the app who deliberately initiated the connection since it involved only the library code (despite being legally responsible at all moment for their app's behavior).

2) LIBRARIES IDENTIFICATION

The search for libraries in the stack traces requires knowing their package names, e.g., com.my.tracker in Figure 3. For this purpose, we departed from the Google Play SDK Index [47], which lists the 129 most popular commercial libraries on Google Play.

We leveraged the Maven Central Repository [48] to expand this initial set as it contains a large number of commonly used libraries in the Android ecosystem. To this end, we started from 55,444 stack traces we logged with Frida and, after eliminating duplicates we were left with 18,011 unique package names. Among them, we identified 672 first-party package names (i.e. presumably developed by the data controller) and 2,385 third-party ones (developed by other entities) already found in the Google Play SDK Index. Therefore, our unassigned sample was reduced to 14,954 package names. Among these, numerous package

```
java.lang.Exception
    at java.net.Socket.connect(Native Method)
    at com.android.okhttp.internal.Platform.connectSocket(Platform.java:182)
    at com.android.okhttp.internal.io.RealConnection.connectSocket(RealConnection.java:145)
    at com.android.okhttp.internal.io.RealConnection.connect(RealConnection.java:116)
    at com.android.okhttp.internal.http.StreamAllocation.findConnection(StreamAllocation.java:186)
    at com.android.okhttp.internal.http.StreamAllocation.findHealthyConnection(StreamAllocation.java:128)
    at com.android.okhttp.internal.http.StreamAllocation.newStream(StreamAllocation.java:97)
    at com.android.okhttp.internal.http.HttpEngine.connect(HttpEngine.java:302)
    at com.android.okhttp.internal.http.HttpEngine.sendRequest(HttpEngine.java:245)
    at com.android.okhttp.internal.huc.HttpURLConnectionImpl.execute(HttpURLConnectionImpl.java:465)
    at com.android.okhttp.internal.huc.HttpURLConnectionImpl.getResponse(HttpURLConnectionImpl.java:411)
    at com.android.okhttp.internal.huc.HttpURLConnectionImpl.getResponseCode(HttpURLConnectionImpl.java:542)
    at com.android.okhttp.internal.huc.DelegatingHttpsURLConnection.getResponseCode(DelegatingHttpsURLConnection.java:106)
    at com.android.okhttp.internal.huc.HttpsURLConnectionImpl.getResponseCode(HttpsURLConnectionImpl.java:30)
    at com.my.tracker.obfuscated.r.a(Unknown Source:50)
    at com.my.tracker.obfuscated.x.a(Unknown Source:15)
    at com.my.tracker.obfuscated.x.a(Unknown Source:0)
    at android.activity.c.run(:7)
    at java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor. java:1167)
    at java.util.concurrent.ThreadPoolExecutor$worker.run(ThreadPoolExecutor.java:641)
    at java.lang.Thread.run(Thread.java:923)
```

FIGURE 3. Example of stack trace invoking the Socket.connect method.

names had been deliberately obfuscated, so their filtering left us with 6,626 package names. They were searched in the Maven Central Repository, resulting in 152 new libraries being identified. This takes us from the 129 libraries initially reported in the Google Play SDK Index to 281 (eliminating duplicates), translating to a dataset increase of 117.83%.

We applied the library identification step on a set of 20,125 traces coming from old experiments. Interestingly, we observed that 58% of the personal data transfers to recipients are initiated by libraries identified in our dataset, giving an idea of the prevalent role of these libraries in personal data transfers to recipients.

IV. EVALUATION

In this section, we apply the proposed method to the analysis of the personal data transfers to recipients and their disclosure through privacy policies on a randomly selected set of Android applications. In a bid for a more profound understanding, we compare these data transfers with those made by libraries. Finally, we further inspect the apps' privacy labels to discern if this novel disclosure mechanism of disseminating privacy practices proves more reliable.

A. EXPERIMENT DESIGN

We started by randomly selecting 9,000 apps from the Google Play Store. Figure 4 shows the distribution of these apps per category, number of downloads, and average users' ratings. We leveraged our platform to download the apps and their privacy policies and labels, install and run the apps, capture the connections they made during execution, and analyze their contents for detecting personal data transfers. We used five Redmi10 mobile devices running Android 30 for the apps' execution.

The execution of the applications resulted in 202,088 successfully intercepted connections, where we observed 23,840 connections transferring personal data off the device in 4,335 applications (48.17%). The quantity and type of personal data transferred can be observed in Table 4.

Surprisingly, unsecured HTTP connections (3.25%) are still being established by a few apps, some even carrying personal data such as the Google Ad ID or even the precise location, posing a severe privacy risk for their users.

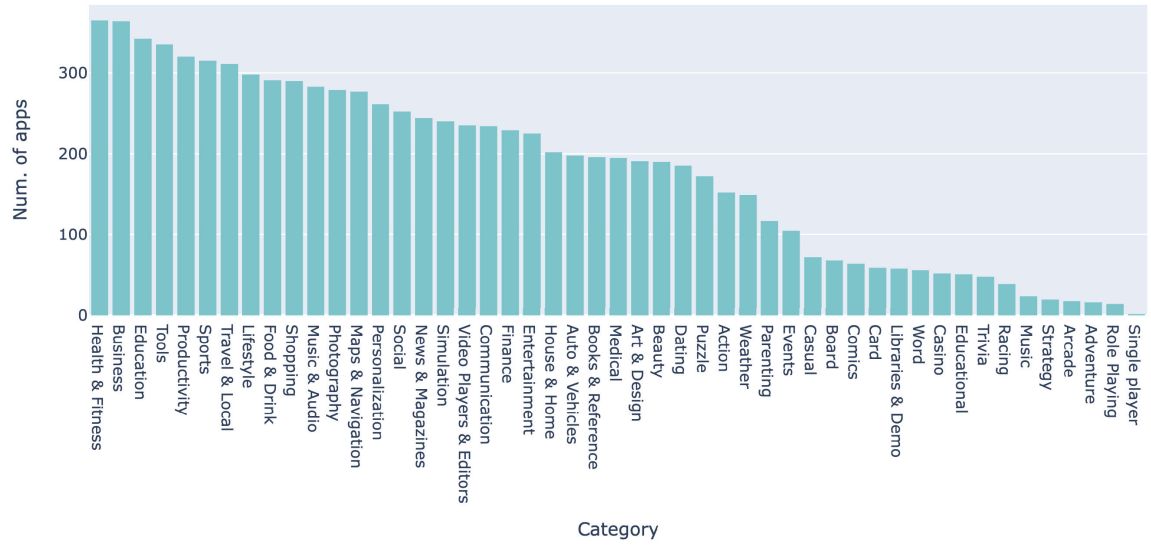
TABLE 4. The type and count of personal data flows captured.

Data type category	Data type description	No. connections (%)
Device_Model	Device model name	22,004 (92.3%)
Google_Ad_ID	Google unique device identifier for advertising purposes	10,846 (45.5%)
Build_No	Build number of Android software version	5,822 (24.4%)
Fingerprint	Unique device identifier based on multiple hardware and software details	1,540 (6.5%)
Router_Wifi_BSSID_Close	Name of near Wifi routers	164 (0.7%)
Router_Wifi_MAC	MAC address of the router connected to the device	136 (0.6%)
Device_location_coarse	Approximate device location	112 (0.5%)
Device_location	Precise device location	106 (0.4%)
Router_Wifi_BSSID	Name of the router connected to the device	82 (0.3%)
Kernel_Version	Android Kernel version	71 (0.3%)

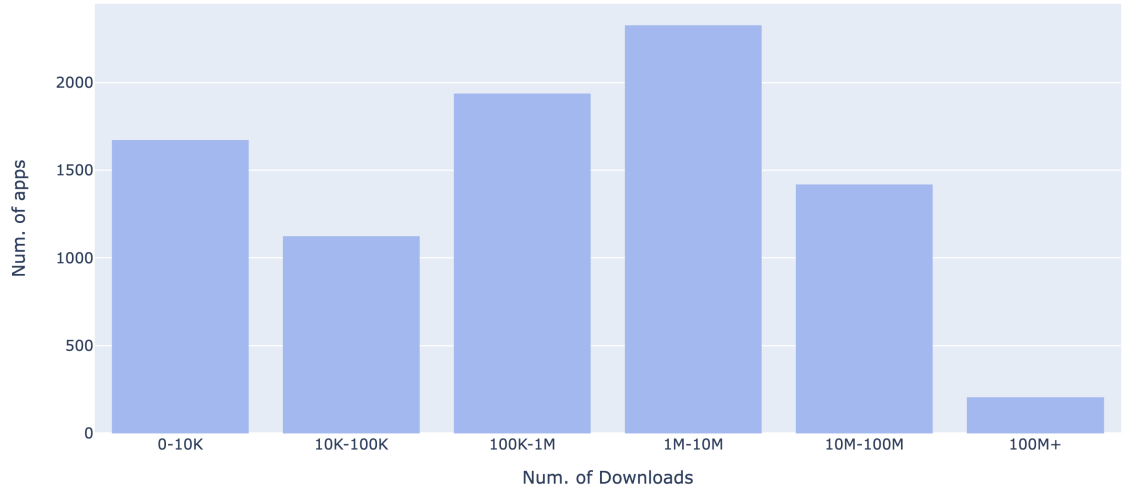
In the following section, we will analyze the recipients of these personal data transfers and the level of transparency we observe in the apps' privacy disclosures.

B. APPS TRANSPARENCY ASSESSMENT

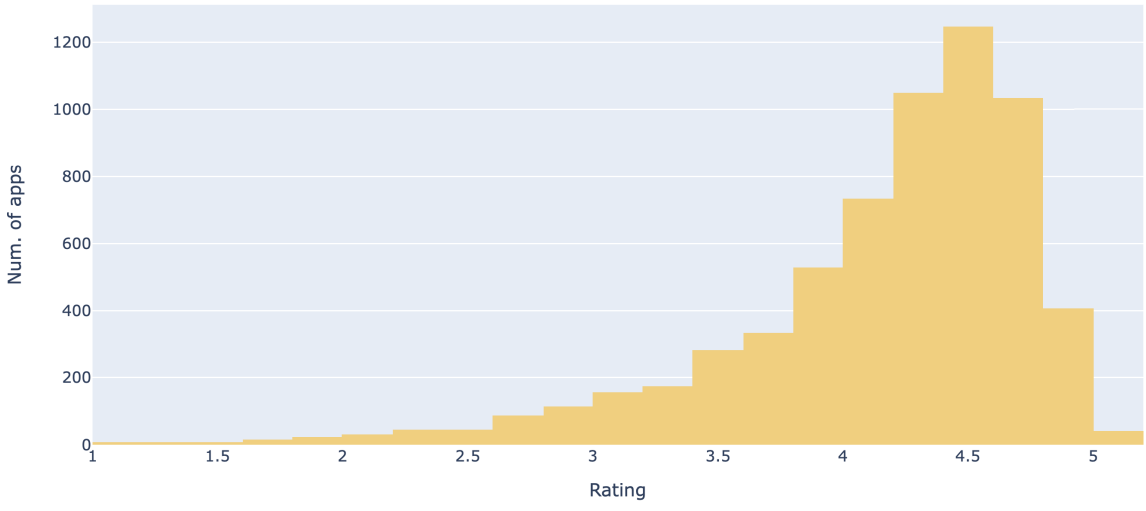
Building on the method detailed in section III, we have been able to identify the recipients of personal data and verify if the apps' privacy policies disclose them adequately.



(a)



(b)



(c)

FIGURE 4. **Dataset distribution based on a) categories, b) number of downloads, and c) rating.**

We successfully pinpointed the recipient's identity in 17,340 of the connections carrying personal data (corresponding to 3,621 apps). This resulted in 206 distinct recipients, with Google (24.86% connections), Meta (17.93%), and Unity (7.66%) being the most prominent recipients of personal data transfers. Notably, the top-10 identified recipients received the 71.89% of personal data transfers, which is a clear indication of the data collection concentration among very few participants in the Android ecosystem.

We employed the *Data Controller Identification* component to identify each app's data controller by looking into the app's privacy policy. Throughout the process, we identified the data controllers of 1,536 apps responsible for 8,232 data transfers. A significant 25.94% of the supposed app's privacy policies were not actually privacy policies (e.g., they were landing pages instead) or were in languages other than English, leading us to discard them. Interestingly, we also observed that a considerable 22% of apps in our dataset are not transparently declaring or identifying the data controller in their privacy policies, which is a requirement according to GDPR since they are processing personal data (as observed in the connections we intercepted). This points to potential compliance issues by these apps.

Upon identifying the apps' data controllers and the recipients of the data transfers, the *Relationship Evaluation* component determined if these data transfers ended up in the data controller or a different recipient. We were able to determine this relationship in 8,220 (1,536 apps) of the 8,232 connections, noting that 95.4% of the personal data transfers were made to other recipients (1,510 apps). The top recipients were again Google (23.90%), Meta (16.85%), and Unity (9.55%).

We then employed the *Disclosure Checker* component to assess if the apps' privacy policy transparently discloses the data transfers to the identified recipients. This component revealed that 1,225 (81.12%) of the 1,510 apps where the privacy policy could be evaluated failed to disclose the personal data recipients according to the GDPR transparency requirements. Figure 5 shows the 20 recipients that most often are not disclosed by the apps' privacy policies, meaning that these apps' users (in our dataset there are apps with more than one billion downloads) are not informed about who their personal data are transferred to.

These results depict a worrying situation as they show that personal data transfers, while abundant and concentrated in a few organizations, are seldom disclosed to data subjects in the apps' privacy policies. Next, we delve into this issue to understand if the use of libraries may be related to these problems.

C. IDENTIFYING LIBRARIES' TRANSFERS

We further analyzed the source of the connections transferring personal data where the data controller could be identified. This led us to identify the code triggering 6,596 data transfers to other recipients.

Table 5 shows the top 10 recipients whose libraries transfer personal data off the device. It's clear that Google's libraries are the most frequently used, with Google Mobile Services (*com.google.android.gms*) occupying the first place and Firebase Services (*com.google.firebase*) taking the fourth.

TABLE 5. Top-10 recipients whose libraries transfer personal data off the device.

Package name	Library provider (Recipient)	% of connections
com.google.android.gms	Google LLC	25.92%
com.facebook	Meta Platforms, Inc.	9.67%
com.unity3d	Unity Technologies	8.00%
com.google.firebase	Google LLC	5.62%
com.flurry	Flurry	4.52%
com.safedk	AppLovin	3.40%
com.ironsource	IronSource	2.06%
com.adjust.sdk	Adjust	1.76%
com.mbridge.msdk	Mintegral	1.64%
com.inmobi	InMobi	1.56%

We then cross-checked this data with the undisclosed data transfers, to understand where the source of the issues is. We found that the libraries are establishing a worrying 73.68% of the undisclosed data transfers. Again, Google's libraries take first place with 23.57% of undisclosed transfers, followed by Unity's (14.52%) and Meta's (13.57%) in second and third places, respectively. These libraries send personal data like device model, Google advertising ID, and software build number to the different recipients.

D. DATA SHARING DISCLOSURES IN PRIVACY LABELS

We have further analyzed the apps' privacy labels to understand if this new form of privacy disclosure reflects the data-sharing practices of apps. It should be noted that privacy labels present severe limitations in meeting GDPR transparency requirements. As advanced by Novovic [49] the privacy labels "*cannot convey the mandatory obligations required by the GDPR*", being only a complement to other disclosure means such as privacy policies. Indeed, privacy labels allow disclosing the type of data shared with a recipient and the purpose, but their design does not allow disclosing the recipient identity. However, their analysis is still useful to let us know about the apps' awareness of the undergoing data sharing.

To this end, we departed from the 1,510 applications where personal data transfers to recipients were observed and found privacy labels for 1,266 (83.38%) of them. Although privacy labels are mandatory for apps updated since July 20, 2022, we observed that 11 of the 238 (4.62%) apps that do not include privacy labels have been updated after the deadline, which is a potential compliance issue with the Google Play Store policy.

After comparing the data transfers to recipients with the data sharing disclosures in the apps' privacy labels,

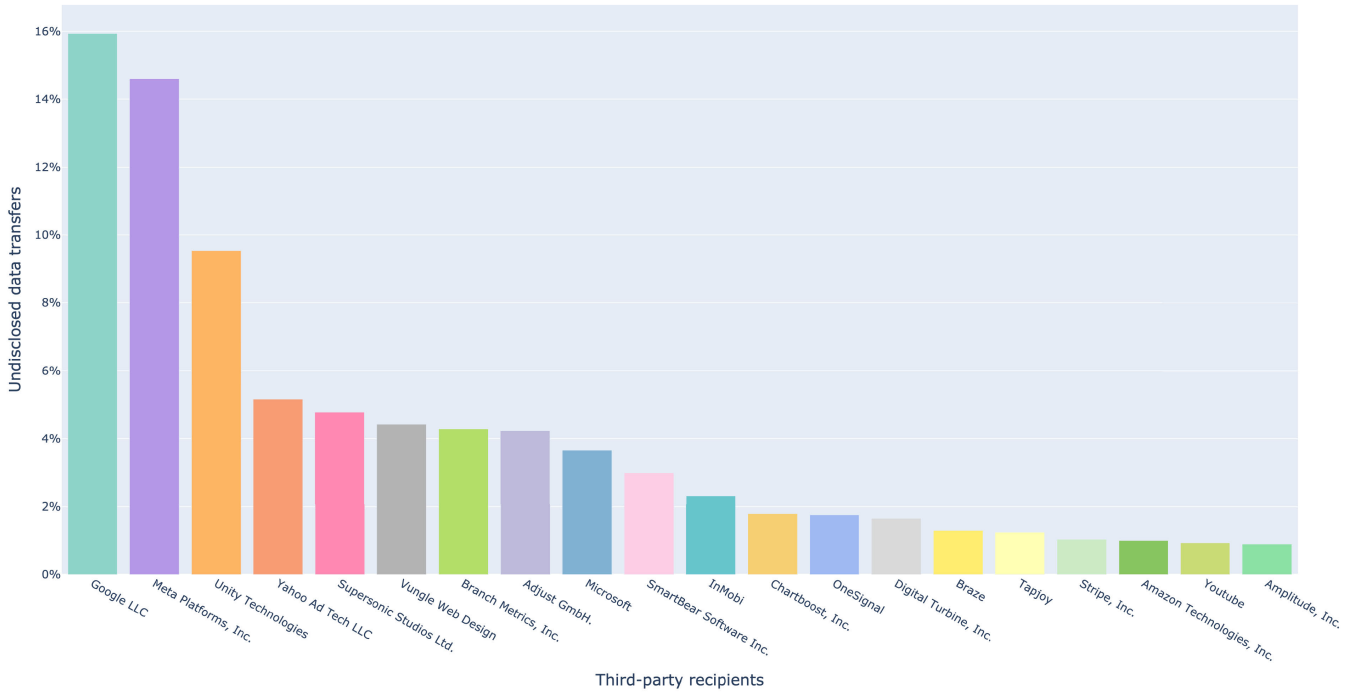


FIGURE 5. Top-20 recipients not being transparently disclosed in the apps' privacy policies. The Y-axis shows the rate of undisclosed data transfers for each recipient over the total amount of undisclosed data transfers.

we observed that 420 (33.17%) apps sent personal data to recipients without disclosure in the labels, which is a sign of these apps' unawareness of their data sharing practices. Notably, all of them were observed sending the *Device or other IDs* without declaring them, except one of them that was found sending the user's *Precise location*.

Interestingly, after analyzing the libraries responsible for the connections undisclosed in the privacy labels, significant differences were observed. The top-five libraries, in descending order, consist of *com.unity3d*, *com.facebook*, *com.safedk*, *com.flurry*, and *com.google*.

A comparative analysis of these 420 applications that fail to disclose the data transfer through their labels paints a dire picture, as 317 of them (75.48%) present severe issues: They neither adhere to their labels disclosures nor do they correctly declare their data-sharing practices in their privacy policies. This presents a dual risk for users, who are left with no apparent way of discerning that their personal data is being dispatched to other organizations. This revelation underscores the magnitude and urgency of the transparency challenges that must be addressed.

On the other hand, these results also indicate that more than half (668 apps, 54.48%) of the apps do acknowledge sharing data in their privacy labels, though they do not disclose the recipient identity as the labels do not allow for it. This finding suggests that most apps are aware of the information being shared, yet they still struggle to disclose it in their privacy policies properly. One possible explanation yields in the features provided by development tools and the Play Store itself, which leverage the information available in the libraries' manifest files to warn developers about the permissions requested by the libraries they integrate, thus calling their attention to disclose them. Unfortunately, these tools do not yet support the automated extraction of finer details such as the entities with which the libraries would be sharing information, which might be used to warn data controllers to disclose the data-sharing practices of their applications transparently.

V. DISCUSSION

The Popularity of an App is Not an Indicator of Greater Transparency: We analyzed if popular applications, according to their number of downloads, better disclose their data-sharing practices. To that end, we computed the rate of disclosing apps over the number of apps in each group in our dataset (Figure 4b). The results did not reveal any particular group of applications as more transparent than the others based on this feature. Nevertheless, it's undeniable that the reach of these apps varies depending on their download counts, posing the most popular apps a greater risk due to their broader user base.

For example, the application with the highest number of downloads in our dataset, com.lenovo.anyshare.gps, fails to disclose the recipients of the personal data it shares. This application, with over a billion downloads on the Google Play Store, has been observed to send the Google advertising ID to Meta Platforms, Inc., Adjust GmbH, and AppsFlyer. However, the app's privacy policy remains ambiguous, indicating that they may share data "*With advertisers and marketing partners in order to display advertisements on our App and support our business, to show how many users of the App have clicked or viewed an advertisement,*

and third-party measurement companies for the purposes of measurement, analytics, engagement technologies and optimization of our Services''. According to the transparency guidelines [50] provided by the European Commission, these privacy policy statements fail due to leaving room for different interpretations and using ambivalent terms (e.g., engagement technologies and optimization of our Services).

Another intriguing case is the *flipboard.app*, which has accumulated over 500 million downloads on the Google Play Store. This application has been spotted transmitting personal data to Adjust GmbH., Microsoft, InMobi, and Upcraft. The Google advertising ID is sent to Adjust and InMobi via the libraries of *com.adjust.sdk* and *com.inmobi*, respectively. The app's privacy policy indicates that personal data can be sent to third parties, including advertising partners. However, none of these organizations are explicitly mentioned, nor activity, sector, sub-sector or location of these recipients are disclosed along with the categories.

Apps' Providers Lack Proper Support to Disclose Their Data-Sharing Practices Accurately: Libraries are at the core of 82% of undisclosed data transfers. Thus, we have checked the websites and the Google Index SDK page for the libraries shown in Table 5 to understand if they properly disclose their data practices. Websites for mainstream libraries generally provide extensive documentation on their privacy practices, including the information they collect and share and, even in some cases, information on how to complete the app's privacy labels (e.g., details on the Unity3D library for Android can be found at https://docs.unity.com/ads/en-us/manual/ImplementingDataPrivacy). On the other hand, the information available at the Google Index SDK is scarce, as it only provides information on the permissions requested by each library (basically, the information available in the library's manifest file), lacking a link to a privacy policy, details on the personal data recipients' identity, or the data collected/shared by the library. As a result, while the information on the libraries' privacy practices is available, it remains scattered, which reduces the ability of data controllers to integrate and disclose it properly, resulting in more than 80% of apps failing to disclose their personal data recipients.

In contrast, our findings suggest that most apps' data controllers are aware of the information being shared, as two-thirds of them managed to report that on their apps' privacy labels. One possible explanation for this different behavior yields in the features provided by development tools and the Google Play Store itself, which leverage the information available in the libraries' manifest files to warn developers about the permissions requested by the libraries they integrate, thus calling their attention to disclose them. Unfortunately, these tools do not yet support the automated extraction of finer details such as the organizations with which the libraries would be sharing information, which might be used to warn developers to disclose the data-sharing practices of their applications properly.

Large Language Models are a Useful Tool for Privacy Policy Analysis: To date, extracting practices from privacy policies has largely leaned on Artificial Intelligence and Natural Language Processing, often necessitating annotated datasets to train Machine Learning models. Crafting these annotations demands both legal and technical expertise. Large Language Models, being trained on vast and varied data, offer coherent responses to prompts, eliminating the need for specific annotation or retraining for each policy feature extraction. Remarkably, ChatGPT comprehends intricate queries, synthesizing information from its extensive training data. For example, in this study, we have noted how it can tailor its responses based on interrelated companies within a group. This represents a substantial time and effort saver, bypassing the technical hurdles of creating Named Entity Recognition components, cross-referencing information from business databases, and the requisites of Natural Language Processing to generate a comparable output.

However, ChatGPT's training posed a hefty computational and temporal demand, constraining its data to 2021 and resulting in significant outdatedness. This proves problematic when discerning ever-evolving business affiliations, potentially leading to inaccurate outputs. This Large Language Model has advanced its ability to understand and retain information from lengthy prompts, now accommodating up to 16k tokens. Larger prompts also come with higher economic (via paid API) and computational (in terms of time) costs, which could affect the quality of results. Newer GPT models exhibit improvements, but their increased model size still restricts their usage and prompt length (128k tokens), with significantly higher prices. Yet, our experimental tests reveal a heightened attention span and a clearer distinction between the query's essence and the actual privacy policy passed as input. Notably, ChatGPT operates non-deterministically by default, meaning the same input might yield varied outputs over time. To address this problem, we are testing temperature and seed parameters to achieve deterministic outputs consistently. Our preliminary tests show that GPT-4 offers more consistent results over time, substantially minimizing these constraints. Using GPT-3.5 solidifies Large Language Models as a robust alternative to ''traditional'' privacy policy extraction techniques. Intriguingly, this is just the tip of the iceberg; with each iteration, its performance only seems to soar, promising even more refined outputs in future versions.

A. THREATS TO VALIDITY

1) CONSTRUCT VALIDITY

The construct validity of our approach is primarily influenced by our decision to focus on disclosures that name recipients, as opposed to those that merely categorize them. This decision stems from our preliminary manual inspection of privacy policies, where we observed a consistent lack of adherence to the detailed disclosure requirements set

forth in the GDPR guidelines. Specifically, none of the 100 policies we examined sufficiently detailed recipient categories (activity, sector, sub-sector, or location) in alignment with the official guidelines. Consequently, our automated method was calibrated to scrutinize disclosures that explicitly name recipients, a practice more in tune with the GDPR's transparency principle. While this approach enhances the relevance and specificity of our analysis, it introduces a limitation: the potential oversight of category-based disclosures that may, albeit infrequently, conform to GDPR standards. This exclusion could lead to an under-representation of compliant practices in our findings. Nevertheless, our decision to focus on named disclosures is justified by the higher likelihood of these practices aligning with GDPR transparency requirements, thereby reinforcing the construct validity of our study in capturing a critical aspect of GDPR compliance.

2) REPRODUCIBILITY AND REPEATABILITY

The ChatGPT API has demonstrated exemplary performance, positioning itself through this article as a robust alternative to conventional machine learning methods reliant on annotated privacy policies. The widespread appeal of this tool has culminated in substantial demand, leading OpenAI to impose request limitations and resulting in occasional server-side errors. The inherent non-deterministic nature of ChatGPT introduces variability in its outputs, potentially challenging result reproducibility. However, our observations indicate that newer versions of the GPT models (i.e., GPT-4 Turbo) appear to mitigate this output variability by providing a seed parameter, further affirming its suitability for extracting practices from privacy policies.

3) INTERNAL VALIDITY

Our proposed methods are statistical in nature, which can introduce the risks of false positives and false negatives when evaluating compliance with the GDPR transparency requirements. This poses a challenge to the accuracy of our claims. To address this, we've meticulously curated annotated datasets, allowing for the computation of validation metrics for each described component. Additionally, we validated each method using distinct data sets, sidestepping potential biases. During the crafting of prompts, we prioritized those leading to superior precision metrics, thereby reducing false positives and bolstering the reliability of our findings.

Our pursuit of libraries in the apps' execution traces has illuminated their significant role in personal data transfers, and their core role in the detected issues. Our search is bound by our libraries' white list, which we subsequently scout within the traces. Despite rigorous efforts to expand this list, we acknowledge that some libraries might escape our radar, leading to false negatives in our outcomes. Yet, utilizing a list ensures that identified entities are genuinely libraries. Furthermore, our approach relies on identifying connections made via the standard Android API. While most network connections are established through it, some others might be set up e.g., through native code, leading us to miss these data-sharing practices in Frida (although still intercepting them in the MitM proxy). These potential data transfers might yield false negative cases, yet they do not threaten the validity of our results in setting a lower threshold for undisclosed data-sharing practices.

4) EXTERNAL VALIDITY

For this study, we employed dynamic analysis tools to detect data transfers to third parties from the apps, including the Exerciser Monkey tool [51], which injects pseudo-random inputs into the apps, for the apps' stimulation during the dynamic analysis. As noted in related studies [52], the code coverage of such tools can be restricted, suggesting that many connections might not be triggered, leading to a potentially skewed representation of data transfers. However, our approach emphasizes soundness over completeness, ensuring that our observations and conclusions are truthful, albeit possibly incomplete. Thus, the actual situation might be even more concerning than our findings suggest, but never less severe than we exposed. Moreover, we relied on ChatGPT to check if two legal entities point at the same company and thus discern between the data controller and other recipients. This can be affected by the timeframe of the data used to train ChatGPT, as this can rapidly change in the business world. However, new GPT models like GPT-4 Turbo — Which is a modifiable parameter of our components —, have updated information up to April 2023, minimizing this problem.

VI. CONCLUSION

This paper has described a method to assess whether Android apps meet GDPR transparency requirements when transferring personal data. We applied it to 9,000 applications on the Google Play Store, yielding alarming results. An overwhelming 81.12% of applications fail to transparently declare the recipients of personal data in their privacy policies. This poses a significant risk to user privacy but also to app owners, who may face substantial financial penalties if do not meet GDPR transparency requirements. Furthermore, it also raises the question of the legal basis supporting the data collection by third parties, which in turn may challenge the lawfulness of these extended practices as already decided by the European Data Protection Board in a recent decision [53].

Our future work points to supporting developers to understand better whether they meet transparency requirements or gain awareness of the causes. Upon discovering that libraries are involved in almost three-quarters of the cases of noncompliance, we will investigate whether this may be due to a lack of transparency on the part of the responsible of these libraries, as previous work has shown [54]. We also aim to support data protection authorities in better spotting concerning issues that, due to their scale and impact, deserve their attention. All in all, our findings underscore the urgent need for more comprehensive and transparent data practices in app development and distribution markets.

AUTHOR CONTRIBUTIONS

David Rodriguez: Methodology, Software, Validation, Formal analysis, Investigation, Data Curation, Writing - Original Draft. **Jose M. Del Alamo:** Conceptualization, Investigation, Resources, Writing - Original Draft, Writing - Review & Editing, Supervision, Project administration, Funding acquisition. **Celia Fernández-Aller:** Conceptualization (legal assessment regarding privacy and GDPR requirements), Writing - Review & Editing. **Norman Sadeh:** Conceptualization, Writing - Review & Editing.

REFERENCES

[1] EUR-Lex. (2016). *EUR-Lex—32016R0679-EN*. [Online]. Available: https://eur-lex.europa.eu/eli/reg/2016/679/oj

[2] M. Goddard, "The EU general data protection regulation (GDPR): European regulation that has a global impact," *Int. J. Market Res.*, vol. 59, no. 6, pp. 703–705, Nov. 2017, doi: 10.2501/ijmr-2017-050.

[3] *The EU's Data Strategy from a Multifaceted Perspective. Views from Southern Europe*, PromethEUs, Amherst, NY, USA, Jun. 2023, pp. 1–76. [Online]. Available: https://www.i-com.it/wp-content/uploads/2023/06/PromethEUs_DataStrategy_Joint-Publication-Final-1.pdf

[4] C. J. Hoofnagle, B. van der Sloot, and F. Z. Borgesius, "The European union general data protection regulation: What it is and what it means," *Inf. Commun. Technol. Law*, vol. 28, no. 1, pp. 65–98, Jan. 2019, doi: 10.1080/13600834.2019.1573501.

[5] H. Wang, Y. Guo, Z. Ma, and X. Chen, "WuKong: A scalable and accurate two-phase approach to Android app clone detection," in *Proc. Int. Symp. Softw. Testing Anal.*, 2015, pp. 71–82, doi: 10.1145/2771783.2771795.

[6] European Commission. (2018). *Guidelines on Transparency Under Regulation 2016/679 (wp260rev.01)*. Accessed: Nov. 15, 2023. [Online]. Available: https://ec.europa.eu/newsroom/article29/items/622227

[7] European Data Protection Board. (2023). *Guidelines 01/2022 on Data Subject Rights—Right of Access*. Accessed: Nov. 15, 2023. [Online]. Available: https://edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-012022-data-subject-rights-right-access_en

[8] Information Commissioner's Office. (2020). *Data Protection Act 2018. Enforcement Powers of the Information Commissioner*. Penalty Notice. TikTok Information Technologies U.K. Limited. [Online]. Available: https://ico.org.uk/media/4025182/tiktok-mpn.pdf

[9] Court of Justice of the European Union. (2023). *Judgment of the Court (First Chamber) of 12 January 2023. RW v Österreichische Post AG—Case C-154/21*. Accessed: Nov. 15, 2023. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A62021CJ0154

[10] J. M. Del Alamo, D. S. Guaman, B. García, and A. Diez, "A systematic mapping study on automated analysis of privacy policies," *Computing*, vol. 104, no. 9, pp. 2053–2076, Sep. 2022, doi: 10.1007/s00607-022-01076-3.

[11] P. Ferrara and F. Spoto, "Static analysis for GDPR compliance," in *Proc. Italian Conf. Cybersecur.*, vol. 2058, 2018, pp. 1–10.

[12] Q. Jia, L. Zhou, H. Li, R. Yang, S. Du, and H. Zhu, "Who leaks my privacy: Towards automatic and association detection with GDPR compliance," in *Wireless Algorithms, Systems, and Applications*. Cham, Switzerland: Springer, Jun. 2019, pp. 137–148, doi: 10.1007/978-3-030-23597-0_11.

[13] Z. Yu, S. Macbeth, K. Modi, and J. M. Pujol, "Tracking the trackers," in *Proc. 25th Int. Conf. World Wide Web*, 2016, pp. 121–132.

[14] T. Libert, "Exposing the invisible web: An analysis of third-party http requests on 1 million websites," *Int. J. Commun.*, vol. 9, pp. 3544–3561, Jan. 2015. [Online]. Available: https://ijoc.org/index.php/ijoc/article/view/3646

[15] S. Han, J. Jung, and D. Wetherall, "A study of third-party tracking by mobile apps in the wild," Dept. Comput. Sci. Eng., Univ. Washington, Seattle, WA, USA, Tech. Rep. UW-CSE-12-03, Jan. 2012.

[16] R. Binns, J. Zhao, M. V. Kleek, and N. Shadbolt, "Measuring third-party tracker power across web and mobile," *ACM Trans. Internet Technol.*, vol. 18, no. 4, pp. 1–22, Nov. 2018, doi: 10.1145/3176246.

[17] N. Vallina-Rodriguez, S. Sundaresan, A. Razaghpanah, R. Nithyanand, M. Allman, C. Kreibich, and P. Gill, "Tracking the trackers: Towards understanding the mobile advertising and tracking ecosystem," 2016, *arXiv:1609.07190*.

[18] D. Rodriguez, J. M. Del Alamo, M. Cozar, and B. García, "ROI: A method for identifying organizations receiving personal data," *Computing*, vol. 105, no. 12, Aug. 2023, doi: 10.1007/s00607-023-01209-2.

[19] W. Hu, D. Octeau, P. D. McDaniel, and P. Liu, "Duet: Library integrity verification for Android applications," in *Proc. ACM Conf. Secur. Privacy Wireless Mobile Netw.*, Jul. 2014, pp. 141–152, doi: 10.1145/2627393.2627404.

[20] L. Li, T. F. Bissyandé, J. Klein, and Y. Le Traon, "An investigation into the use of common libraries in Android apps," in *Proc. IEEE 23rd Int. Conf. Softw. Anal., Evol., Reengineering (SANER)*, vol. 1, Mar. 2016, pp. 403–414, doi: 10.1109/SANER.2016.52.

[21] M. Backes, S. Bugiel, and E. Derr, "Reliable third-party library detection in Android and its security applications," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur. (CCS)*. New York, NY, USA: Association for Computing Machinery, Oct. 2016, pp. 356–367, doi: 10.1145/2976749.2978333.

[22] Z. Ma, H. Wang, Y. Guo, and X. Chen, "LibRadar: Fast and accurate detection of third-party libraries in Android apps," in *Proc. IEEE/ACM 38th Int. Conf. Softw. Eng. Companion (ICSE-C)*. New York, NY, USA: Association for Computing Machinery, May 2016, pp. 653–656.

[23] J. Zhang, A. R. Beresford, and S. A. Kollmann, "LibID: Reliable identification of obfuscated third-party Android libraries," in *Proc. 28th ACM SIGSOFT Int. Symp. Softw. Test. Anal.* New York, NY, USA: Association for Computing Machinery, Jul. 2019, pp. 55–65, doi: 10.1145/3293882.3330563.

[24] Y. He, X. Yang, B. Hu, and W. Wang, "Dynamic privacy leakage analysis of Android third-party libraries," *J. Inf. Secur. Appl.*, vol. 46, pp. 259–270, Jun. 2019, doi: 10.1016/j.jisa.2019.03.014.

[25] C. Schindler, M. Atas, T. Strametz, J. Feiner, and R. Hofer, "Privacy leak identification in third-party Android libraries," in *Proc. 7th Int. Conf. Mobile Secure Services (MobiSecServ)*, Feb. 2022, pp. 1–6, doi: 10.1109/MobiSecServ50855.2022.9727217.

[26] H. Cheng, G. Hu, J. Liu, Z. Kang, C. Pan, and Z. Zhang, "Detecting third-party libraries for privacy leakage in packed Android applications," in *Proc. China Autom. Congr. (CAC)*, Nov. 2022, pp. 5053–5058, doi: 10.1109/CAC57257.2022.10054907.

[27] V. Morel and R. Pardo, "SoK: Three facets of privacy policies," in *Proc. 19th Workshop Privacy Electron. Soc.* New York, NY, USA: Association for Computing Machinery, Nov. 2020, pp. 41–56, doi: 10.1145/3411497.3420216.

[28] M. V. Reiss, "Testing the reliability of ChatGPT for text annotation and classification: A cautionary remark," 2023, *arXiv:2304.11085*.

[29] P. Törnberg, "ChatGPT-4 outperforms experts and crowd workers in annotating political Twitter messages with zero-shot learning," 2023, *arXiv:2304.06588*.

[30] F. Gilardi, M. Alizadeh, and M. Kubli, "ChatGPT outperforms crowd-workers for text-annotation tasks," 2023, *arXiv:2303.15056*.

[31] J. H. Choi, K. E. Hickman, A. Monahan, and D. B. Schwarcz, "ChatGPT goes to law school," *J. Legal Educ.*, vol. 71, no. 3, p. 387, Jan. 2023, doi: 10.2139/ssrn.4335905.

[32] P. G. Kelley, L. F. Cranor, and N. Sadeh, "Privacy as part of the app decision-making process," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.* New York, NY, USA: Association for Computing Machinery, Apr. 2013, pp. 3393–3402, doi: 10.1145/2470654.2466466.

[33] Apple Developer. *App Privacy Details—App Store*. Accessed: Nov. 15, 2023. [Online]. Available: https://developer.apple.com/app-store/app-privacy-details/

[34] Google Play Store Support. (Mar. 2023). *Google Play's Data Safety Section*. [Online]. Available: https://support.google.com/googleplay/android-developer/answer/10787469?hl=en

[35] R. Khandelwal, A. Nayak, P. Chung, and K. Fawaz, "Unpacking privacy labels: A measurement and developer perspective on Google's data safety section," 2023, *arXiv:2306.08111*.

[36] D. Rodriguez, A. Jain, J. M. D. Alamo, and N. Sadeh, "Comparing privacy label disclosures of apps published in both the app store and Google Play Stores," in *Proc. IEEE Eur. Symp. Secur. Privacy Workshops (EuroS&PW)*, Jul. 2023, pp. 150–157, doi: 10.1109/EuroSPW59978.2023.00022.

[37] A. Jain, D. Rodriguez, J. M. D. Alamo, and N. Sadeh, "ATLAS: Automatically detecting discrepancies between privacy policies and privacy labels," in *Proc. IEEE Eur. Symp. Secur. Privacy Workshops (EuroS&PW)*, Jul. 2023, pp. 94–107, doi: 10.1109/eurospw59978.2023.00016.

[38] H. Zheng, M. Xue, H. Lu, S. Hao, H. Zhu, X. Liang, and K. Ross, "Smoke screener or straight shooter: Detecting elite Sybil attacks in user-review social networks," in *Proc. Netw. Distrib. Syst. Secur. Symp. (NDSS)*, 2018, doi: 10.14722/ndss.2018.23009.

[39] Z. Tan and W. Song, "PTPDroid: Detecting violated user privacy disclosures to third-parties of Android apps," in *Proc. IEEE/ACM 45th Int. Conf. Softw. Eng. (ICSE)*, May 2023, pp. 473–485, doi: 10.1109/ICSE48619.2023.00050.
[40] B. Andow, S. Y. Mahmud, J. Whitaker, W. Enck, B. Reaves, K. Singh, and S. Egelman, "Actions speak louder than words: Entity-sensitive privacy policy and data flow analysis with POLICHECK," in *Proc. 29th USENIX Secur. Symp. (USENIX Security)*, 2020, pp. 985–1002.
[41] D. S. Guamán, D. Rodriguez, J. M. del Alamo, and J. Such, "Automated GDPR compliance assessment for cross-border personal data transfers in Android applications," *Comput. Secur.*, vol. 130, Jul. 2023, Art. no. 103262. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167404823001724
[42] I. Montani et al., Aug. 2023, "Explosion/spaCy: V3.6.1: Support for Pydantic v2, find-function CLI and more," *Zenodo*, doi: 10.5281/zenodo.8225292.
[43] Crunchbase. (2023). *Crunchbase*. [Online]. Available: https://www.crunchbase.com/
[44] Frida. (2023). *A World-Class Dynamic Instrumentation Framework*. [Online]. Available: https://frida.re/
[45] I. Square. (2019). *Okhttp*. [Online]. Available: https://square.github.io/okhttp/
[46] (Feb. 2023). *MyTracker Android SDK*. [Online]. Available: https://github.com/myTrackerSDK/mytracker-android
[47] Google Play Store. *Google Play SDK Index*. [Online]. Available: https://play.google.com/sdks/?hl=en-419
[48] (2023). *Maven Repository*. [Online]. Available: https://mvnrepository.com/
[49] M. Novović, "Privacy nutrition labels, app store and the GDPR: Unintended consequences?" *J. Data Protection Privacy*, vol. 5, no. 3, pp. 267–280, 2022.
[50] E. Commission. (Apr. 2018). *Article29—Transparency Guidelines*. [Online]. Available: https://ec.europa.eu/newsroom/article29/items/622227/en
[51] Google Android Developers. (Apr. 2023). *UI/Application Exerciser Monkey*. [Online]. Available: https://developer.android.com/studio/test/other-testing-tools/monkey
[52] P. Patel, G. Srinivasan, S. Rahaman, and I. Neamtiu, "On the effectiveness of random testing for Android: Or how I learned to stop worrying and love the monkey," in *Proc. IEEE/ACM 13th Int. Workshop Autom. Softw. Test (AST)*. New York, NY, USA: Association for Computing Machinery, May 2018, pp. 34–37.
[53] European Data Protection Board. (Dec. 2022). *Binding Decision 4/2022 on the Dispute Submitted by the Irish SA on Meta Platforms Ireland Limited and its Instagram Service (Art. 65 GDPR)*. [Online]. Available: https://edpb.europa.eu/system/files/2023-01/edpb_binding_decision_202204_ie_sa_meta_instagramservice_redacted_en.pdf
[54] K. Kollnig, R. Binns, P. Dewitte, M. Van Kleek, G. Wang, D. Omeiza, H. Webb, and N. Shadbolt, "A *Fait Accompli*? An empirical study into the absence of consent to third-party tracking in Android apps," in *Proc. 17th USENIX Conf. Usable Privacy Secur. (SOUPS)*. Berkeley, CA, USA: USENIX Association, 2021, pp. 181–195.

DAVID RODRIGUEZ received the B.S. degree in telecommunications engineering and the M.S. degree in cybersecurity from Universidad Politécnica de Madrid (UPM), Spain, in 2021 and 2022, respectively, where he is currently pursuing the Ph.D. degree.

From 2021 to 2023, he was a Research Assistant with the Telematic Systems Department. Since 2023, he has been a Graduate Teaching Assistant with the Telematic Systems Department, UPM. He holds an intelectual property registration and several papers in conferences and journals. His research interests include the development of static and dynamic automated tools for auditing mobile devices and the creation and integration of machine learning methods and large language models to assess legal compliance regulations in the mobile ecosystems.

Mr. Rodriguez has been awarded the Best Presentation Paper and the Second Best Presentation Paper in the International Workshop on Privacy Engineering, in 2022 and 2023, respectively.

JOSE M. DEL ALAMO received the Ph.D. degree in data privacy and protection from the Department of Telematic Systems Engineering, Universidad Politécnica de Madrid.

He has imparted various cybersecurity courses across degrees and master's programs and has extensively supervised student research. He is currently an Associate Professor (with tenure) with the Department of Telematic Systems Engineering, Universidad Politécnica de Madrid. He is the Principal Investigator of notable projects, such as Horizon Europe SUNRISE and Spanish Government funded autoGDPR, focusing on topics, such as critical infrastructure resilience and data protection assessment. He holds multiple patents in the field and has a significant presence in academic publications, and chairing the IEEE International Workshop on Privacy Engineering.

Dr. Alamo contributions have garnered several awards, such as the Young Scholar Award at the 8th International Conference on Computers, Privacy, and Data Protection. He received multiple honors, including two Research Sexenniums.

CELIA FERNÁNDEZ-ALLER received the Ph.D. degree in law and technology.

She is currently a Senior Lecturer with Universidad Politécnica de Madrid, where she manages courses on the legal and ethical aspects of computer science. She was previously a Lecturer with Universidad Centroamericana de El Salvador (UCA) and had an internship with Comillas Pontifical University. She is an Active Member with the GIOS Research Group and was appointed by the Spanish Government to draft a digital rights charter. She has been honored as a Visiting Professor at Bristol University and serves on several advisory boards. She has contributed to conferences, such as ECAI and ECSA, and has published extensively in journals, such as DOXA and *IEEE Technology and Society Magazine*. She holds two patents in EDUCERE Project and has issued various legal reports for companies and public institutions. Her research interests include transdisciplinary studies on human rights, data protection, and emerging technologies, such as AI and robotics.

NORMAN SADEH (Member, IEEE) is a Professor in the School of Computer Science at Carnegie Mellon University (CMU). He co-founded and co-directs CMU's Privacy Engineering Program, and also co-founded and for ten years co-directed CMU's PhD Program in Societal Computing. Norman served as lead principal investigator on two of the largest domestic research projects in privacy, the Usable Privacy Policy Project (https://usableprivacy.org) and the Personalized Privacy Assistant Project (https://privacyassistant.org). He was also founding CEO and, until its acquisition by Proofpoint, chairman and chief scientist of Wombat Security Technologies, a company that defined the multi-billion dollar user-oriented cybersecurity market. Technologies Norman developed with colleagues at CMU and Wombat are used to protect tens of millions of users around the world against cybersecurity attacks such as phishing. Dr. Sadeh is well known for his pioneering work on AI-based privacy enhancing technologies, including the development of privacy assistants, the development of automated privacy compliance tools but also work on modeling people's privacy expectations and preferences and on privacy and security nudging. This work has been credited with influencing the development of privacy-enhancing solutions at companies such as Apple, Google and Facebook, and results of his research have also informed activities at regulatory agencies, including the Federal Trade Commission and the California Office of the Attorney General. In the late nineties Norman served as Chief Scientist of the EUR 550 million European Union's e-Commerce initiative, which included all pan-European research in cybersecurity and privacy as well as contributions to several major European public policy initiatives.

Computing (2024) 106:3879–3903
https://doi.org/10.1007/s00607-024-01331-9

REGULAR PAPER

Large language models: a new approach for privacy policy analysis at scale

David Rodriguez[1] · Ian Yang[2] · Jose M. Del Alamo[1] · Norman Sadeh[2]

Received: 17 January 2024 / Accepted: 24 July 2024 / Published online: 22 August 2024

Abstract

The number and dynamic nature of web sites and mobile applications present regulators and app store operators with significant challenges when it comes to enforcing compliance with applicable privacy and data protection laws. Over the past several years, people have turned to Natural Language Processing (NLP) techniques to automate privacy compliance analysis (e.g., comparing statements in privacy policies with analysis of the code and behavior of mobile apps) and to answer people's privacy questions. Traditionally, these NLP techniques have relied on labor-intensive and potentially error-prone manual annotation processes to build the corpora necessary to train them. This article explores and evaluates the use of Large Language Models (LLMs) as an alternative for effectively and efficiently identifying and categorizing a variety of data practice disclosures found in the text of privacy policies. Specifically, we report on the performance of ChatGPT and Llama 2, two particularly popular LLM-based tools. This includes engineering prompts and evaluating different configurations of these LLM techniques. Evaluation of the resulting techniques on well-known corpora of privacy policy annotations yields an F1 score exceeding 93%. This score is higher than scores reported earlier in the literature on these benchmarks. This performance is obtained at minimal marginal cost (excluding the cost required to train the foundational models themselves). These results, which are consistent with those reported in other domains, suggest that LLMs offer a particularly promising approach to automated privacy policy analysis at scale.

Keywords Large language models · Natural language processing · Privacy policies · Data protection · Privacy · Feature extraction

Mathematics Subject Classification 68M11 · 68M14 · 68M15 · 68M25 · 68P27 · 68T50 · 68U15

Extended author information available on the last page of the article

1 Introduction

The digital era has led to an unprecedented expansion of web and mobile applications and a myriad of online services. This growth is a testament to technological advancement and the increasing reliance of businesses and organizations on digital platforms for various operations. A central aspect of this digital proliferation is the extensive use of technologies for personal data collection, primarily driven by the objective of enhancing targeted marketing strategies. The ability to collect, analyze, and utilize user data has become a cornerstone of modern commerce, offering businesses invaluable insights into consumer behavior and preferences.

However, the increasing collection and utilization of personal data has raised significant privacy concerns. Users' privacy is at risk as their data becomes valuable in the digital marketplace. This concern has led to the emergence of regulatory bodies and the formulation of data protection legislation aimed at safeguarding user privacy. These legislations, such as the General Data Protection Regulation (GDPR) in the European Union or the California Consumer Privacy Act (CCPA) in California, impose stringent requirements on how organizations should handle personal data.

Ensuring compliance with these legislations, however, poses a formidable challenge. The proliferation of online services, compounded by globalization, makes it impractical, if not impossible, for regulators to manually assess each service's adherence to applicable privacy laws. This situation is further exacerbated by the dynamic nature of online services, where data processing practices and the privacy policies disclosing them are subject to frequent changes [1]. In response to these challenges, automated methods have been proposed for analyzing the text privacy policies at scale [2]. These techniques can in turn be used to develop a variety of useful tools [3, 4], some to help individual consumers (e.g., privacy question answering assistants, or browser extensions to help users take advantage of privacy choices buried deep in the text of privacy policies), some to help developer, companies, app stores and regulatory agencies (e.g., tools to help automatically identify potential compliance issues).

The automated analysis of privacy policies has leveraged Natural language processing (NLP) techniques [2]. Symbolic and statistical state-of-the-art NLP techniques are proposed to address this task, although each has drawbacks. Symbolic approaches rely on pre-defined rules, leading to lower performance when compared to statistical approaches due to the lack of adaptability to differences present in legal texts. Thus, state-of-the-art research has predominantly relied on statistical approaches such as machine learning (ML) techniques, and particularly supervised learning models, to train and evaluate models identifying privacy practice disclosures such as personal data collection or sharing [5]. These models require the use of manually annotated datasets [6–8], which are often expensive, time-consuming to create, and prone to errors [9]. Furthermore, building and training those models require advanced technical expertise, contributing to a higher barrier to entry. As a result, their practical application is mainly suitable for large-scale projects where the benefits can outweigh these significant costs.

On the other hand, modern Generative Artificial Intelligence (GenAI), particularly Large language models (LLMs), represents a significant advancement in the NLP domain, being able to understand and generate human-like text, making it particularly well-suited for parsing and analyzing the complex language present in privacy policies without needing annotated datasets. In this context, this paper proposes the application of LLMs for the effective and efficient extraction of privacy practices from privacy policies. In particular, we focus on ChatGPT, which relies on Generative Pre-trained transformer (GPT) models.

Our study identifies the optimal configuration of ChatGPT prompts, parameters, and models, integrating advanced techniques such as few-shot learning. Additionally, we conduct a comparative analysis of our proposed ChatGPT configuration with Llama 2 and other state-of-the-art techniques. Our findings reveal that our proposal competes with and even outperforms these traditional methods. Moreover, we discuss its advantages regarding lower upfront costs, reduced processing times, and greater ease of use.

Thus, we propose LLM-based solutions and our specific ChatGPT configuration as a viable replacement for traditional NLP techniques in the task of automated privacy policy processing. Our research contributes to the ongoing discourse on the application of GenAI in legal and regulatory contexts [10–12], suggesting a paradigm shift towards more efficient, cost-effective, and accessible tools for privacy policy analysis.

2 Background

The development of new privacy policy analysis methods leveraging statistical NLP approaches frequently requires labeled corpora for training and validation. In the domain of privacy policy analysis, several datasets manually annotated by legal experts have been employed to build supervised learning methods.

2.1 MAPP dataset

The MAPP corpus [13] plays a crucial role in our study of privacy policy analysis. It comprises 64 privacy policies from Google Play Store apps, segmented into paragraphs. Each segment is meticulously annotated by legal experts to indicate whether it discloses the collection or sharing of various types of personal data. For example, a typical annotation might label a paragraph as disclosing the collection of geographical location data. This granularity enables precise training and validation of NLP models designed to identify specific data handling practices. Notably, the MAPP corpus is one of the few multilingual datasets available, including annotations in both English and German, which broadens its applicability and utility in cross-lingual legal studies.

2.2 OPP-115 dataset

The OPP-115 dataset [14] is among the most utilized corpora in privacy policy research and one of the oldest in the field. It follows a similar annotation model to MAPP but encompasses a broader scope with 115 annotated privacy policies. Each policy is annotated for nearly identical practices and data types, as seen in MAPP, which allows for comparative analysis and benchmarking across studies. An example annotation in OPP-115 might involve identifying clauses related to third-party data sharing.

2.3 APP-350 dataset

As the largest dataset in this domain, APP-350 [3] includes 350 privacy policies, annotated with even finer details regarding data collection and sharing practices than the previous datasets. This dataset serves as a comprehensive tool for training models to detect and interpret complex legal language regarding data privacy. For instance, the annotations distinguish whether data collection or sharing is conducted by the first party or a third party. Such distinctions may be particularly valuable for assessing compliance with data protection regulations.

2.4 IT-100 dataset

The IT100 dataset [15] focuses on international data transfer disclosures in privacy policies, containing annotations for 100 different policies. This dataset highlights specific statements that indicate cross-border data transfer, such as clauses pertaining to the European Union's General Data Protection Regulation (GDPR). An example from this dataset would be annotations identifying the mechanisms used for data transfer, like Standard Contractual Clauses.

3 Related work

Privacy policies are documents written in plain text that outline how organizations handle personal data. However, the complexity and length of these documents often make them challenging to understand and process [16]. This has spurred interest in automated methods for analyzing privacy policies [2], which fall into two major categories, namely, symbolic and statistical NLP.

Symbolic NLP approaches [17–19] are relevant but come with inherent limitations when processing new texts: these techniques model language through grammar rules and lexicons, thus requiring extensive manual effort to create and code these rules. This process is both time-consuming and hard to scale, especially when dealing with intricate aspects of privacy policies. Symbolic NLP is effective in morphological and lexical analysis, such as identifying privacy practices through keyword analysis. It also handles more complex tasks like syntactic and

semantic analysis, using tools like the Stanford dependency parser [20]. PolicyLint [21] is a state-of-the-art tool based on this symbolic NLP approach that employs ontologies to detect contradictions in privacy policy statements about personal data collection and sharing. Its ability to identify negative sentences-a challenging task for statistical NLP techniques–highlights its potential for specific privacy policy analysis tasks. However, it faces challenges with unanticipated variations, including typos or infrequent cases, thus limiting its applicability to new cases.

Statistical NLP approaches, on the other hand, leverage machine learning techniques for language processing: supervised, unsupervised, and Artificial neural networks (ANN)-based techniques. Supervised methods are the predominant technique usually employed for automated privacy policy analysis, with geometric algorithms like Support vector machine (SVM) [14, 22–24] and Logistic Regression (LR) [25, 26] being the most prevalent. Unsupervised techniques, although less common, utilize models like Hidden Markov models (HMM) [14, 27] and Latent Dirichlet Allocation (LDA) [28] for clustering practices during policy analysis. ANN-based techniques are also in use for this task, including Convolutional Neural Networks (CNNs) [8, 29], Recurrent neural networks (RNN) [30], and Google's BERT [6, 31], sometimes showing superior performance than supervised learning methods [6, 31].

Expanding upon symbolic and statistical NLP methods, LLMs can generate coherent text based on a given input, such as GPTs [32] and Llama 2 [33]. Building on those LLMs, ChatGPT and Llama 2-Chat are chatbots trained to provide meaningful answers to pieces of text inputs (i.e., prompts) and with adjustable performance through parameters like "temperature" that influence the results' variability [34], and response times. The ability to provide relevant answers is achieved through a combination of unsupervised and supervised learning techniques underpinned by neural networks trained on extensive datasets. Additionally, the relevance and format of the responses are typically enhanced through prompt augmentation [35], which involves modifying the given input prompt to improve the output performance or to steer the output in a specific direction. Notably, LLMs' proficiency in processing lengthy input texts is boosted by the attention mechanisms inherent in transformer architectures [36]. A recent study conducted by Qin et al. [37] has analyzed to what extent LLMs like ChatGPT can perform various tasks -reasoning, language inference, Q&A, dialogue, summarization, entity recognition, and sentiment analysis—using 20 well-established NLP datasets to benchmark their performance, showing high reasoning capabilities.

Integrating LLMs into the automated analysis of privacy policies and legal texts [38] represents a significant evolution in assessing compliance with data protection regulations. Tang et al. [12] have explored their application in this context, highlighting its potential to surpass traditional methods in extracting and classifying general, coarse-grained privacy practices within legal texts. Our research extends this exploration by thoroughly analyzing LLMs' ability to identify more detailed data practices in privacy policies. We provide insights into the optimal model configuration for this task and further demonstrate LLMs' generalization capabilities, particularly ChatGPT's, to identify varied privacy practices. Our findings reveal that ChatGPT, leveraging few-shot learning, outperforms traditional symbolic and statistical

NLP methods in key areas, including classification performance, time efficiency, and cost-effectiveness.

4 Experimental design

GPT models have an intrinsically complex behavior dependent on the prompt design, configuration parameters, and model selection. We rely on the Design Science Research (DSR) methodology [39] to propose a ChatGPT framework for privacy policy analysis and evaluate its effectiveness. DSR guides the design of new artifacts through an iterative and systematic process. Specifically, we followed an iterative split testing process [40] to assess the performance of the prompt, parameter, or model selection changes within each iteration. Finally, we check our proposed configuration performance against two unseen sets of policies, conduct a set of comparative analyses with state-of-the-art solutions, and demonstrate its generalization capabilities. Through this systematic process, we propose a well-performing and generalizable configuration of ChatGPT as a novel and effective approach for the automated analysis of privacy policies.

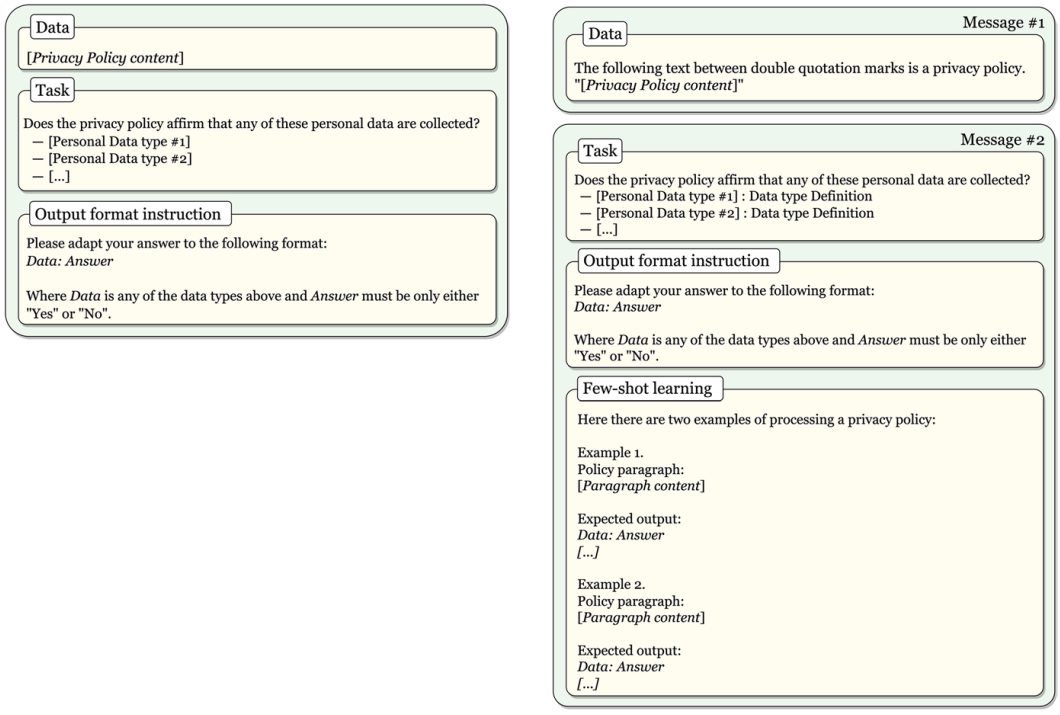
4.1 Ground truth

Determining the optimal ChatGPT configuration that offers the best performance requires using a ground truth dataset to validate and quantify results. We relied on the MAPP dataset [13] for this task, retaining an experimental set on which to apply changes and measure their impacts and a control set to validate the final configuration's overall performance. Unlike traditional NLP techniques, using a ground truth dataset is only required while designing the configuration framework. Afterward, we can generalize it to identify other privacy practices without generating new annotated datasets or validating new methods, as demonstrated in Sect. 5.

The MAPP dataset is inherently unbalanced, presenting a challenge for objective analysis. To address this, we employed stratified sampling to create balanced experimental and control subsets, ensuring that each subset was representative of the overall dataset's diversity in data types collected. Thus, we stratified sampling to generate the experimental (33 policies) and control (31 policies) subsets. The distribution of data types in these subsets is presented in Table 1. With a standard deviation between both datasets of 2.44 for all data categories, with categories annotated in almost all policies (e.g., IP address and device IDs, in 95%) and others in only one of them (e.g., Political, religious, or philosophical belief). This is contextualized against the backdrop of the mean annotation counts per policy, which are 8.13 and 8.18 for the experimental and control sets, respectively. The observed standard deviation, in relation to the means, suggests a moderate degree of variance in annotation frequency per policy across the datasets. This degree of variability is within acceptable limits for the intended analytical scope, affirming a balanced and representative data stratification for the empirical analysis.

Table 1 Distribution of personal data types between 'experimental' and 'control' sets following stratified sampling

Personal data type	Experimental set	Control set
Computer information	27	22
Contact information	28	29
Cookies and tracking elements	29	25
Demographic data	21	22
Financial	23	17
Generic personal information	28	31
Health, genetic, or biometric data	7	7
IP address and device IDs	31	30
Location	23	23
Other	27	26
Personal identifier	9	7
Political, religious, or philosophical belief	1	1
Social media data	14	11
Unspecified	30	27
User online activities	31	30



(a) Baseline prompt design. (b) Final prompt design.

Fig. 1 Baseline and final ChatGPT prompt designs for the identification of privacy practices in privacy policies

4.2 Prompt design

We departed from the prompt design depicted in Fig. 1a. This baseline prompt is structured in *Data*, *Task* and *Output* format instruction segments. *Data* is the privacy policy for identifying practices. *Task* is the actual practice in which identification in the policy is requested. *Output* format instruction provides the guidelines to obtain responses that can be processed automatically.

We applied this baseline prompt to the specific task of identifying types of personal data purportedly collected or shared as per the privacy policy. This required providing the privacy policy in the Data segment, asking about each data type in the Task segment, and steering the formatting of responses, including "*Data: Answer*" in the Output format instruction segment. In our experiment, this baseline prompt achieved the following metrics: 0.79 accuracy, 0.92 recall, 0.78 precision and 0.84 F1 score.

Crafting the optimum prompt design requires a split testing process to reveal the effects of various changes in the prompt. The only ChatGPT parameter adjusted during this phase is the temperature value, set to zero to provide more deterministic responses [41], and using the GPT-4 Turbo model to take advantage of its speed and input prompt size capabilities up to 128k tokens. While numerous tests were conducted, this section will focus solely on those that involved a significant change in performance metrics. A detailed summary of these metrics, derived from each test, is encapsulated in Table 2. In this testing sequence, each technique that demonstrated a performance improvement was systematically integrated into the subsequent tests. Thus, each new test was benchmarked against the last updated configuration.

Specifying data boundaries. Incorporating the phrase "The following text is a privacy policy" improved the metrics, specifically a +1.47% increase in accuracy (from 0.793 to 0.804), +0.4% in recall, +1.36% in precision, and +0.92% in F1 score. This improvement is attributed to enhancing the model's ability to discern the limits of the privacy policy text. A minor adjustment involving the indication that the privacy policy text is enclosed in double quotes led to an additional +0.8% rise in recall, precision, and F1 score, and a +1.16% increase in accuracy.

Data placement. We evaluated the impact of the placement of the privacy policy within the prompt-either at the beginning or the end-on the performance metrics. Positioning the privacy policy at the end, contrary to the beginning, slightly diminished the overall metrics, including a decrease in accuracy by −1.15%, recall by −1.98%, precision by −0.07%, and F1 score by −0.97%.

Augmenting task description. The initial prompt version primarily focused on enumerating the types of data to be identified. However, given the inherent complexity of data categorization–a challenge even for human annotators as substantiated in related literature [42]—the prompt was augmented to include the internal definitions used for manual annotations in the MAPP dataset. While this expansion resulted in a lengthier prompt, it significantly enhanced all metrics except for recall, with an increase of +4.58% in accuracy, a decrease of −0.79% in recall, +5.96% in precision, and +2.72% in F1 score.

Message splitting. We have tested splitting the prompt into two different messages, passed to ChatGPT one after the other. Specifically, we separated the privacy

Table 2 Metrics obtained with each technique tested on ChatGPT and Llama 2. The tests have been executed sequentially and are incremental: when a technique exhibits superior performance (highlighted in bold within the table), it is incorporated into the subsequent test. The sole exception to this is the Data Segmentation test, which demonstrated a minimal improvement in the F1 score at the expense of significantly increased processing time and a decrease in precision; hence, this technique has not been ultimately integrated. GPT-4 Turbo and Llama 2 70B-Chat models are used unless otherwise stated, as they showed the best performance

Prompt technique	Accuracy	Precision	Recall	F1 Score
ChatGPT				
Baseline prompt	0.793	0.785	0.922	**0.848**
Specify Data boundaries	0.804	0.796	0.926	**0.856**
Specify Data boundaries (double quotes)	0.814	0.803	0.933	**0.863**
Data placement (Bottom)	0.804	0.802	0.915	0.855
Augmenting Task description	0.851	0.850	0.926	**0.887**
Message splitting	0.855	0.864	0.915	**0.888**
Data pruning	0.848	0.847	0.926	0.885
Data segmentation	0.850	0.837	0.948	0.889
Task segmentation	0.804	0.919	0.756	0.829
Data & Task segmentation	0.571	0.871	0.374	0.523
One-shot prompting	0.862	0.889	0.893	**0.891**
Two-shot prompting	0.874	0.886	0.919	**0.902**
Three-shot prompting	0.858	0.872	0.907	0.889
Llama 2				
Baseline prompt	0.846	0.880	0.873	**0.877**
Data segmentation	0.625	0.625	1.000	0.770
Task segmentation	0.613	0.870	0.490	0.627
Data & Task segmentation	0.846	0.847	0.921	**0.882**
Two-shot prompting	0.623	0.623	1.000	0.768

policy (Data segment) from the remainder prompt. The results show a better overall understanding and comprehension of the privacy policy, reflected in a +1.56% increase in precision, +0.56% increase in accuracy and +0.22% increase in F1 score, at the cost of a −1.2% decrease in recall.

Data pruning. This technique eliminates the paragraphs of the policy that do not have information regarding collecting or sharing personal data. We crafted a specific prompt for this task. The results show the overall policy metrics have remained practically the same.

Segmentation. We also assessed the role of input processing in the results with three different configurations: (1) Data segmentation, i.e., analyzing each individual paragraph at a time; (2) Task segmentation, i.e., asking only for one specific practice (e.g., a given data type collection) at a time, and (3) Data and Task segmentation, i.e., asking for one specific practice in one specific paragraph. Data segmentation did not show significant improvement. However, Task segmentation had a surprising effect on the result: accuracy decreased by −5.48%, recall decreased by −18.4%,

precision increased by +8.06%, and overall leading to a decrease in the F1 score of −6.46%. This suggests that asking for each practice individually may lead to a loss of a broader contextual understanding of the model, negatively impacting its overall performance. Finally, Data and Task segmentation showed the worst results, dropping recall by −59.6% and decreasing F1 score by −40.97%, reinforcing the importance of context for ChatGPT when analyzing privacy policies.

It's worth noting that while Data segmentation showed a similar performance to keep the whole policy, it increases the cost of the queries since they are computed according to the prompt and response size and not specifically by the number of requests. Furthermore, it also increases the processing time for each policy, as more requests (as many as policy paragraphs) must be processed. Thus, we have discarded this option in favor of processing the whole policy.

Few-shot prompting. Few-shot prompting [43] refers to providing a set of examples (shots) with the prompt to guide the model. We have tested this technique, including in the prompt one, two, and three examples-randomly chosen-of paragraph annotations. The best result was obtained with two-shot examples, showing a significant improvement in the metrics (+3.31% accuracy, +0.0% recall, +4.29% precision, +2.18% F1 score).

Final prompt design. Fig. 1b presents the prompt configuration that our tests have consistently found to be most effective in identifying privacy practices. In this optimized prompt structure, the Data segment is introduced in an initial message, followed by the Task segment in a subsequent message. This Task segment incorporates definitions of the targeted practice-in this instance, data types-along with the same Output format instruction used in the Baseline prompt and a Few-shot learning component. The Few-shot learning part includes two illustrative examples (Two-shot) of processing paragraphs from privacy policies and the corresponding expected outputs.

4.3 Parameter tuning

ChatGPT offers a number of parameters that can be configured to modify its responses [41], namely temperature, top p, and system inputs. For testing these parameters, we used the final prompt design presented in Fig. 1b.

Temperature. Temperature is a hyperparameter that allows controlling the randomness and creativity of the text generated by a GenAI. If the temperature is low, the model will probably produce the most "correct" text, but with little variation. Conversely, a higher temperature value shows greater variation (i.e., creativity). Lower temperature values are preferable for the development of a deterministic framework. We automated queries to ChatGPT to measure this feature, using the same prompt for all 33 privacy policies in the experimental dataset and performing the requests on 3 different days and at 5 different times of the day (from 9 am to 9 pm). This amounts to 495 different requests and responses, and we observed 52 discrepancies (i.e., different responses compared to the typical answer), which means 89.5% consistency in ChatGPT responses. This-although far from absolute

determinism—highly increases the 59.6% percentage of determinism achieved with the default temperature value of 1.0.

The GPT-4 Turbo model introduced a new feature called seed, specifically for obtaining consistent responses over time with the same prompt. Even though determinism is its declared purpose, we observed that using a seed value and default temperature provided only 84.65% of similar responses. Nonetheless, the combination of 0 temperature and seed shows 90.51%, being the most reliable combination of these two parameters.

In our evaluation of the ChatGPT's performance across different temperature settings, we found that higher temperature values inversely impact the consistency of the metrics, with deterministic responses being optimal. This tendency is notable as the quality of the outputs deteriorates with increasing temperature. Concurrently, a manual inspection of the responses revealed a propensity for incomplete data type coverage. Specifically, responses frequently reported only the initial data type queried. This issue not only aggravates the decline in the F1 score but also results in a significant proportion of the data types-nearly half-remaining unaddressed in the responses. Such findings underscore the importance of temperature configuration in ensuring both the accuracy and completeness of the information extracted by ChatGPT.

Top p. Top p, or "*nucleus sampling*", consists of selecting the next token from the "nucleus" or subset of the vocabulary that constitutes the cumulative probability mass of the top p most probable tokens. For example, setting $top\_p = 0.1$ means only tokens comprising the top 10% probability mass are considered [41]. We have observed little performance variability when testing different values of $top\_p =$ [0, 1] while keeping the default temperature value ($T = 1$). Furthermore, in its official documentation, OpenAI recommends modifying the temperature value or the top p parameter, but not both simultaneously. Thus, we chose the default value of top p for our implementation and set the temperature to zero. These settings allow us to obtain more reproducible results.

System inputs. Using the OpenAI API, messages can be assigned to different roles (i.e., user, assistant, or system), where the system instruction can give high-level instructions for the conversation. We tested two system instructions: (1) "*You are a helpful assistant with extensive knowledge in data protection and privacy engineering*." and (2) "*You are a helpful assistant with extensive knowledge in data protection and law*", which specifically indicate areas of knowledge that are important for our task analyzing privacy policies. Neither of the two system instructions improved the results obtained but rather worsened them.

4.4 Fine-tuning

OpenAI facilitates model customization through fine-tuning, which involves retraining a model on a specific dataset to enhance its performance. This approach is beneficial for augmenting response consistency and can enable the use of shorter prompts while still achieving the desired format. During our experimentation phase,

Table 3 Comparison between the baseline and the fine-tuned GPT−3.5 model analyzing privacy policies by chunks and GPT-4 Turbo model analyzing privacy policies as a whole

Metrics	Chunked privacy policy processing		Whole privacy policy processing
	GPT-3.5-turbo-0613 (fine-tuned)	GPT-3.5-turbo-0613	GPT-4 Turbo
Accuracy	0.867	0.677	0.916
Precision	0.803	0.519	0.898
Recall	0.803	0.944	0.963
F1 score	0.803	0.670	0.935

fine-tuning was available only for the *gpt-3.5-turbo-0613* model. Thus, we tested the effect of model fine-tuning using this ChatGPT version.

The *gpt-3.5-turbo-0613* model sets a maximum prompt size of 4096 tokens. Thus, we segmented the policies into smaller subsets (chunks), each conforming to the condition that the combined length of the policy text (T) and prompt (P) did not exceed 4096 tokens ($T + P < 4096$). This process led to the creation of a training set comprising 73 chunks, aligned with the manual annotations from the MAPP corpus and subjected to a default training configuration of three epochs as determined by OpenAI based on dataset size.

This fine-tuned model demonstrated superior performance when compared to the baseline (not fine-tuned) model: accuracy increased from 0.677 to 0.867, precision increased from 0.519 to 0.803, and the F1 score increased from 0.670 to 0.803. Still, it could not beat the GPT-4 Turbo model (which does not require chunking the policies thanks to the 128K tokens limit), probably due to its ability to retain context (see Table 3).

4.5 Validation

We first applied our proposed configuration framework to the MAPP ground-truth control set, comprising 31 privacy policies, utilizing the prompt, parameters, and model based on our findings in prior sections. The prompt employed is the one described in Sect. 4.2. The selection of parameter values was based on determinism consideration: *temperature* = 0, a fixed seed, and *top_p* = 1 (the default setting). Finally, the GPT-4 Turbo model is employed for its performance, speed, and significantly higher input token limit balance. This configuration yields an accuracy of 0.916, a recall of 0.976, a precision of 0.898, and an F1 score of 0.935 on the control set of the MAPP corpus.

We further validated our prompt design and model configuration against a larger ground truth, i.e., OPP-115 dataset [14], renowned for its fine-grained manual annotations of privacy practices. This validation yielded consistent results: 0.904 accuracy, 0.912 recall, 0.949 precision, and 0.930 F1 score, indicating that our proposal exhibits robust performance even when applied to a larger and more varied set of privacy policies.

5 Demonstration

This section aims to demonstrate why LLMs, specifically ChatGPT, can be considered a competent technique for privacy policy analysis at scale. First, we compare ChatGPT to its closest GenAI rival, Llama 2, in terms of extracting the same privacy practices across identical test sets. We then compare ChatGPT with state-of-the-art statistical and symbolic NLP approaches to evaluate its performance and verify whether it can rival or even replace them. Finally, we analyze our proposal's generalization capabilities for identifying other privacy practices, namely the declaration of international transfers in privacy policies.

5.1 Comparison with Llama 2

Llama 2 [33] is a family of open-source LLMs released by Meta that competes with ChatGPT in the GenAI space. Specifically, Meta has released versions with 7, 13, and 70 billion parameters, each with a fine-tuned "Chat" version optimized for dialogue. For a more direct comparison with the ChatGPT models, we focus on the chat variant of each of the Llama 2 models.

We downloaded the 7B directly from Meta via their GitHub repository and ran it using four NVIDIA GeForce RTX 2080 Ti GPUs. Due to GPU constraints, we used the Python library from Together.AI to run the 70B model [44]. We initially tried to run the 13B model in our local environment, but as we achieved poor performance, we also used the Together.AI installation.

Prompt. We departed from the final prompt design shown in Fig. 1b and followed another split testing process to identify the best-performing Llama 2 prompt design. All the Llama 2-Chat models have a 4,096 token limit, which forced us to segment the privacy policies (i.e., the Data part) to ensure that our prompts are under the maximum token limit. Additionally, we removed the few-shot learning part from the prompt, as this yielded worse performance in our experiments with Llama 2. We observed that this technique resulted in outputs that did not conform to the requested format and additionally caused overfitting to the provided examples. Finally, we tried segmenting the Task in the prompt by asking for each data practice at a time, improving the results. Table 2 summarizes the different tests and the resulting performance.

Parameters. Just as for the ChatGPT models, we parameter-tuned across the temperature and the top p values. Similar to observations with ChatGPT, our experiments suggest setting a temperature value of 0 and a default top p of 1.0 as the best-performing configuration.

We carried out our experiments with the three Llama 2 versions. As the Llama 2 70B-Chat model consistently showed better results, we used this version to assess its performance against the MAPP control set (31 annotated policies) (Table 4). Llama 2 demonstrates comparable but slightly lower performance in identifying privacy practices in this dataset compared to our ChatGPT-4 proposal.

Table 4 Metrics comparison between the best-performing Llama 2 70B-Chat (Llama 2) and ChatGPT-4 Turbo configurations

		Accuracy	Precision	Recall	F1 score
MAPP	Llama 2	0.846	0.847	0.921	0.882
	GPT-4 Turbo	0.916	0.898	0.976	0.935
OPP-115	Llama 2	0.749	0.700	0.814	0.753
	GPT-4 Turbo	0.904	0.949	0.912	0.930

We further evaluated the performance of the Llama 2-70B configuration against the OPP-115 dataset. The results (Table 4) show that Llama 2 obtains worse performance against this new dataset, suggesting that, unlike ChatGPT, this Llama 2 configuration does not generalize well to different datasets.

5.2 Comparison with state-of-the-art techniques

We propose LLMs and, specifically, ChatGPT as a new technique for automating privacy policy information processing and extraction. To confirm it as such, we compare its performance in extracting fine-grained practices from policies with state-of-the-art statistical and symbolic approaches.

5.2.1 Statistical approaches

In this study, we conducted a comparative analysis of our configuration framework proficiency in identifying fine-grained privacy practices against statistical classifiers based on Support Vector Classifiers (SVC)–a subtype of SVM-, which were trained and validated using the APP-350 corpus [3]. To ensure a rigorous comparison, the same policy dataset was employed to evaluate the performance of both methods.

The primary objective was to assess ChatGPT's ability to accurately identify particular types of personal data collection as stated in privacy policies. For this purpose, we selected 10 distinct data types, with an emphasis on higher specificity (for instance, choosing "Contact email address" over the broader "Contact information"), spanning various categories such as contact data, identifiers, and social login data.

Table 5 delineates the comparative performance of ChatGPT against the pre-trained SVC classifiers for identifying each specified data type. The results indicate a comparable level of performance across most data types. However, a notable exception was observed with the SIM identifier, where ChatGPT's performance was significantly lower despite achieving 100% precision. A detailed manual review of the annotations for this data type in the original policies revealed a common annotation issue: Human annotators wrongly coded this data type. Specifically, the annotators coded "device serial number" under the "SIM serial number" category. However, the former is issued by the device manufacturer, while the latter is provided by the mobile carrier. This discrepancy likely contributed to the lower F1 score for ChatGPT in identifying the SIM identifier.

Table 5 ChatGPT vs. traditional machine learning classifiers' performance for identifying first-party data collection per data type

Data type	ChatGPT			SVC classifiers		
	Precision (%)	Recall (%)	F1 score (%)	Precision (%)	Recall (%)	F1 score (%)
Contact Email Address	95	95	95	97	94	96
Contact Phone Number	100	85	92	94	94	94
Identifier Cookie	92	97	95	95	100	98
Identifier IMEI	83	88	86	94	94	94
Identifier Device ID	74	89	81	96	87	91
Identifier MAC	94	84	89	88	79	83
Identifier Mobile Carrier	79	90	84	100	57	73
Identifier SIM Serial	100	13	**22**	73	100	84
Location WiFi	70	58	64	48	92	63
Social login	77	65	71	83	81	82

Excluding the analysis of the SIM serial identifier, which was identified as an anomaly, the comparative evaluation yielded an average F1 score of 84.1% for ChatGPT, as opposed to 86% achieved by the SVC-based classifiers for the selected data types. This outcome illustrates that while traditional SVC-based classifiers are recognized for their reliability and accuracy, ChatGPT presents a comparable level of performance. ChatGPT offers the added advantage of significantly simpler usability, making it a viable alternative for similar tasks in data practice identification.

5.2.2 Symbolic approaches

PolicyLint [21], a tool designed to analyze privacy policies, employs a symbolic approach based on ontologies to detect contradictions in statements regarding personal data collection and sharing. This tool identifies negative sentences, which are often challenging for conventional machine learning techniques. The public repository of PolicyLint's code [45], as referenced, was utilized to process the privacy policies in our control set, facilitating a comparative analysis with our proposal.

PolicyLint operates by identifying sentence structures characterized by [actor] [action] [data_object] [entity]. Here, "actor" signifies a first or third party involved in data handling, "action" denotes the nature of data interaction (positive or negative, such as collection or non-collection), "data_object" pertains to the type of data in question, and "entity" refers to the recipient of the data (for instance, advertisers).

Given that the data_objects in PolicyLint do not align format-wise with those in our MAPP corpus, a manual matching process was undertaken by two authors to correlate PolicyLint's classifications with the data types in our corpus. This matching was independently conducted, followed by an agreement phase for resolving

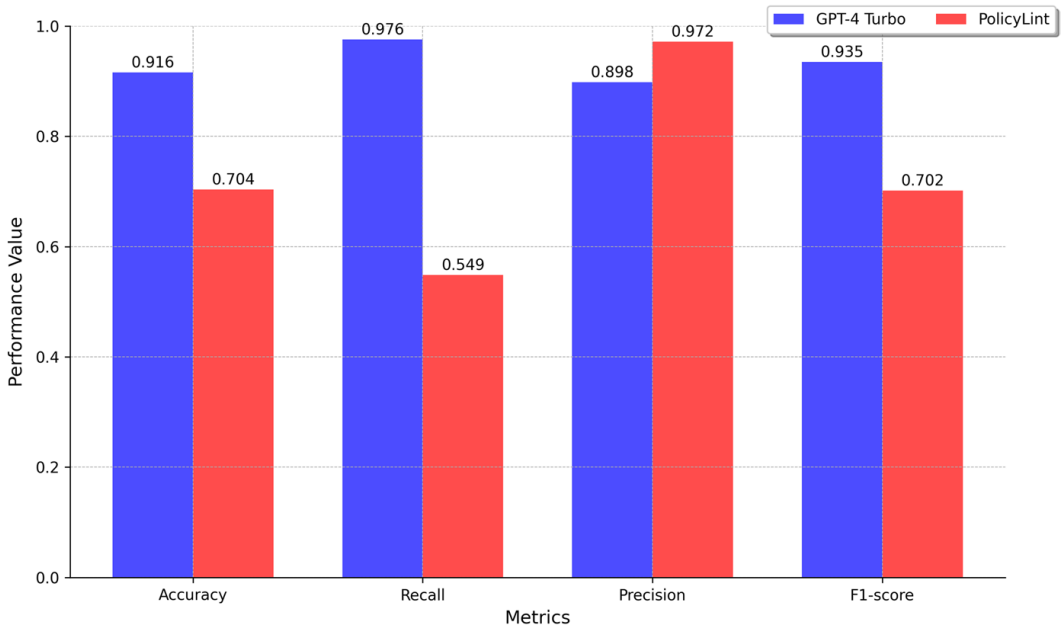


Fig. 2 Metrics comparison between the ChatGPT-based method and PolicyLint

discrepancies. The matching criteria were aligned with the definitions provided in the MAPP corpus. Subsequently, the comparative performance metrics of both methods were analyzed and presented in Fig. 2.

Our analysis revealed that PolicyLint exhibits high precision, surpassing the metrics its authors reported. This discrepancy might stem from our methodology, where we assess whether a data type is identified at least once in a policy, instead of PolicyLint's validation across all relevant statements. However, PolicyLint's approach overlooks negative cases, leading to a lower recall. Overall, our evaluation indicates that our ChatGPT configuration framework significantly outperforms PolicyLint's F1 score, highlighting its efficacy in extracting and analyzing data practices from privacy policies.

5.3 Generalization capabilities

The proficiency of ChatGPT in extracting data collection and sharing practices from privacy policies has been notably demonstrated in our study. In this section, we extend the evaluation to assess ChatGPT's generalization capabilities in identifying a broader range of practices within privacy policies. This extension is grounded in our prior research [46], which focused on analyzing privacy policies to find disclosures related to international data transfers. This previous study produced a dataset (IT100) comprising 100 privacy policies where privacy practices related to international data transfers were manually annotated by legal experts [15]. A Support Vector Machines (SVM)-based classifier was trained to identify these specific practices.

In Fig. 3, we present the comparative analysis of the performance metrics between our configuration proposal of ChatGPT and the SVM-based classifier, utilizing the IT100 dataset for evaluation. ChatGPT was configured as per the

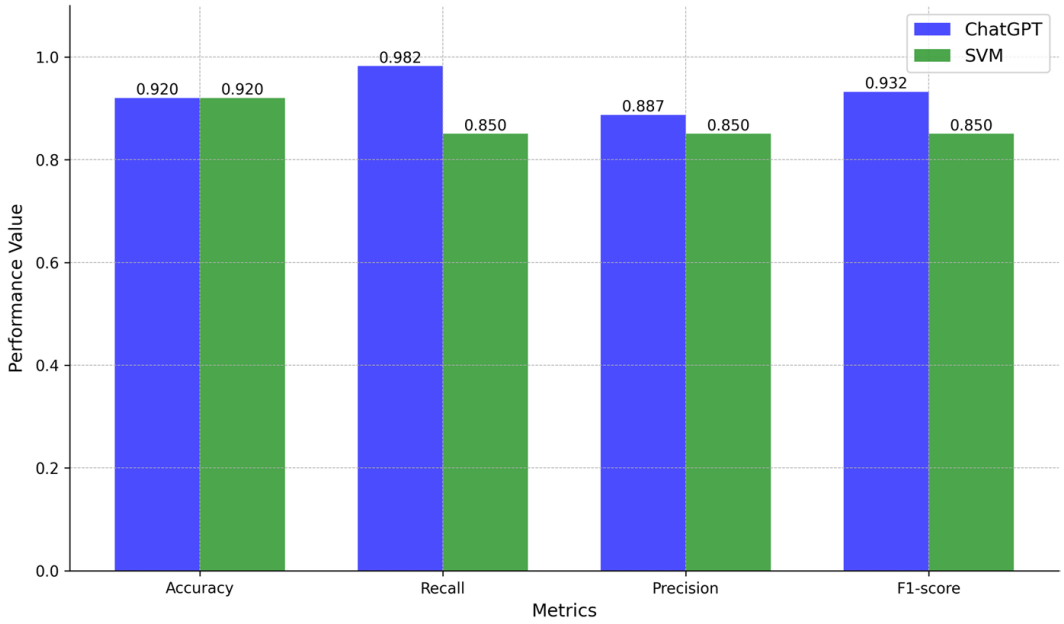


Fig. 3 Metrics comparison between the ChatGPT and SVM machine learning classifier performance identifying international data transfer practices

parameters delineated in Sect. 4, which included an instantiation of our enhanced prompt in Sect. 4.2, and settings like temperature=0, top_p=1, and an absence of system_input. The results displayed by ChatGPT were significantly superior in most metrics, reinforcing its efficacy in extracting information about diverse practices from privacy policies.

6 Discussion

ChatGPT demonstrates a more balanced and adaptable performance in privacy policy analysis compared to traditional symbolic and statistical methods, overcoming the limitations of manual annotations and varying data across different corpora. Symbolic methods are characterized by their rigidity, which is reflected in their performance metrics. High precision in symbolic methods indicates their well-defined patterns and rules are closely aligned with specific instances in the data. However, this precision comes at the cost of completeness, as evidenced by their lower recall. In contrast, ChatGPT demonstrates a more balanced performance, achieving a notably higher F1 score than PolicyLint. This suggests that ChatGPT—and our proposed configuration, while less rigid in its approach, captures the breadth of privacy practices within policies more effectively.

When comparing ChatGPT with statistical methods such as SVM, we find that these traditional classifiers perform similarly in identifying certain data types. However, ChatGPT excels particularly in recognizing practices like international data transfers, which are complex and multifaceted. This superior performance is notable, given that statistical methods often depend on extensive manual annotations, which can introduce errors. As seen in the Identifier SIM Serial case in Table 5,

such annotation errors can significantly impact classifier performance. Wagner et al. [47] supports this observation, indicating that the average agreement among human annotators for attribute values is considerably lower than for top-level categories. This discrepancy highlights the challenges in achieving consensus among annotators and the advantage of ChatGPT's approach, which is not constrained by the limitations of manual annotations.

Furthermore, our analysis of different corpora, specifically the MAPP and OPP-115 datasets, sheds light on the variance in annotations across datasets. The performance disparities observed for *Social media data* and *Personal identifier data* between these two corpora suggest that the annotations for these data types likely vary, underscoring the issues associated with training classifiers on manually annotated data [48]. This reinforces the need for approaches like ChatGPT that rely less on such annotations, offering a more adaptable and potentially more accurate solution for privacy policy analysis.

Economic considerations play a significant role in the choice of the technique to process privacy policies. Manual annotators in the United States are reported to earn approximately $8.5 per hour, while rates in lower-income countries range between $3–$4 per hour [49]. However, the annotation of privacy policies demands legal expertise for accurately identifying data protection practices, entailing a higher pay rate, assumed here at a minimum of $10 per hour. For the MAPP corpus, three experts annotated each policy, averaging 1 h and 52 min each [13]. Multiple annotations of the same content by different experts ensure reliable and high-quality data where inter-annotation agreement can be measured. Previous research [46] demonstrated that training classifiers with 100 policies can be sufficient, which raises costs by up to $5,601.

Setting aside the technical expertise required for classifier development, the cost-effectiveness of traditional classifiers becomes behooveful with GPT-4 Turbo at approximately 81,500 privacy policies and with GPT−3.5 Turbo at around 825,000 policies (Fig. 4). This cost difference suggests that depending on specific application needs and constraints, GPT−3.5 Turbo, with an F1 score of 87.2% measured on the MAPP corpus control set, might be a pragmatic choice compared to GPT-4 Turbo, which achieved an F1 score of 93.5%[1] in our evaluation. Furthermore, Llama 2 models, specifically 7B and 70B, may be considered in terms of cost discussion. Both models were publicly released for free use, but our hardware limitations imposed by the latter forced us to use Together.AI API for that version. The current API cost for the Llama 2-70B model is 10% lower than the GPT−3.5 Turbo model while showing an even higher performance −88.2% F1 score-making it even more convenient in terms of cost by performance. Llama 2-7B has significantly lower computational requirements, leading to no other cost but computation and achieving an 80.1% F1 score. Thus, GPT-4 Turbo offers the best performance of the LLMs compared, but at the highest cost. Whereas if the computing capabilities are sufficient to run it locally, Llama 2-70B offers good performance at a low cost.

[1] This difference is not only due to the model performance but also because the few-shot prompting technique that can be applied to the GPT-4 Turbo model thanks to its increased token limit.

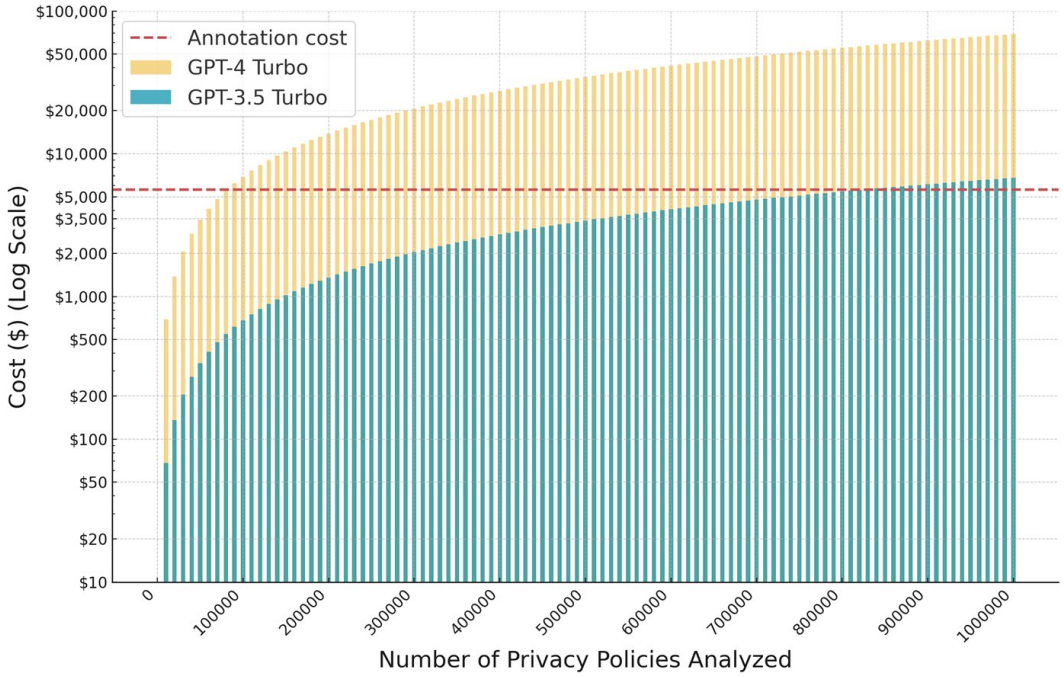


Fig. 4 Cost Comparison of analyzing privacy policies with ChatGPT API

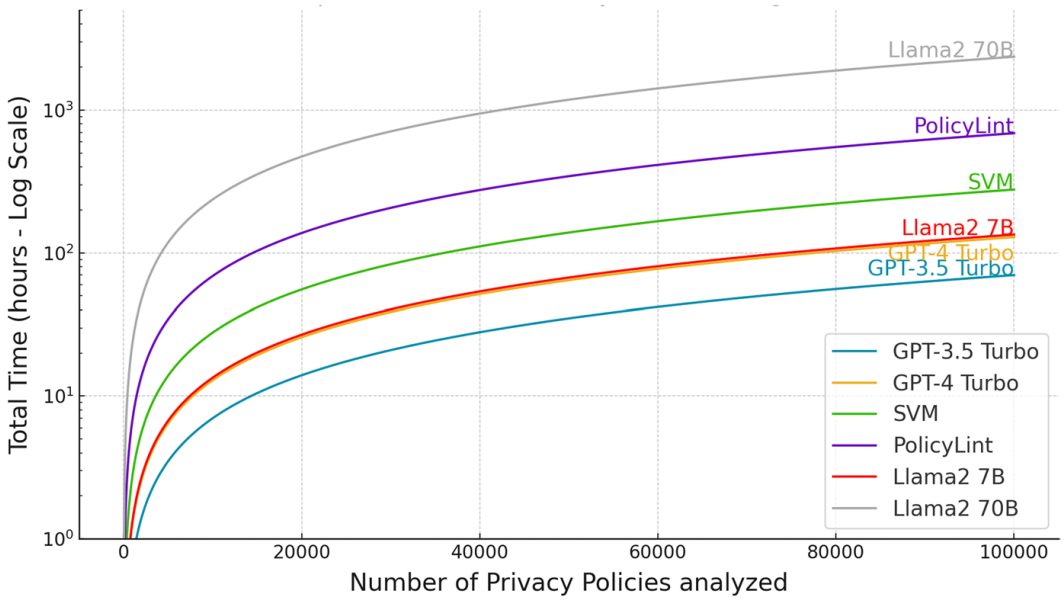


Fig. 5 Time comparison between ChatGPT models and SVM to process privacy policies

In assessing the processing capabilities of ChatGPT models, our analysis indicates a marked efficiency advantage over traditional machine learning and symbolic AI techniques. Acknowledging the operational constraints imposed by OpenAI on these models, specifically regarding token throughput per minute is critical. GPT-4 Turbo is limited to 300,000 tokens per minute, while GPT−3.5 Turbo can process up to 1,000,000 tokens within the same timeframe.

With an average of 6652 tokens required in average to fully process a privacy policy, GPT-4 Turbo can analyze up to 45 policies per minute, in contrast to the 150 policies per minute capability of GPT−3.5 Turbo. Figure 5 depicts this variance in processing capacity, with GPT-4 Turbo necessitating slightly more time for large-scale privacy policy analyses when compared to GPT−3.5 Turbo. Furthermore, the SVM-based classifier takes approximately double the time of the slower GPT model to process an equivalent number of policies. In stark contrast, PolicyLint, while being at the forefront of privacy policy analysis symbolic-based tools, demands up to six times the processing time of GPT−3.5 Turbo for comparable tasks. The two versions of Llama 2 show remarkably different processing times. The Llama 2-7B, locally analyzing each policy at once (truncating policies when length limit required), shows a similar processing time compared to GPT-4 Turbo, while Llama 2-70B through Together.AI API (analyzing policies by chunks), shows the slowest performance of all techniques.

These findings underscore the superior speed of LLM models and highlight the need to balance performance with processing time, especially when scaling to analyze vast numbers of privacy policies. Thus, organizations may find the trade-off between the slightly lower speed of GPT-4 Turbo and its enhanced accuracy acceptable, particularly in scenarios where quality of analysis is paramount. Conversely, for applications where time efficiency is a priority, GPT−3.5 Turbo presents a compelling option, offering rapid analysis with a modest compromise in performance metrics.

For the GPT models, parallel processing can be employed to concurrently analyze up to 150 and 45 policies per minute for GPT−3.5 and GPT-4, respectively, adhering to the stipulated token rate limits. To scale up concurrent processing capabilities with ChatGPT, users may opt for multiple paid accounts, which entails additional costs due to the subscription requirements for accessing the API via ChatGPT Plus. Another avenue is to request OpenAI for elevated rate limits, a request that hinges on the company's approval. Anticipation of expanded rate limits by OpenAI in the future could potentially democratize access to more extensive parallel processing for all users, thereby broadening -even more- the scope of large-scale privacy policy analysis.

The rapid progression of generative AI technology is evident in the quick succession of ChatGPT models introduced. Within the span of mere months, we have witnessed the release of successive ChatGPT iterations, namely GPT−3.5 Turbo, GPT-4, and GPT-4 Turbo. Alongside the expected speed and cost efficiency enhancements, a notable shift has been observed in model determinism. For instance, the determinism observed in ChatGPT−3.5 (99.19%) significantly exceeds that of GPT-4 Turbo, suggesting a potential trade-off between response variability and model robustness.

This rapid succession has introduced variations in the models' performance, particularly regarding prompt responsiveness and temperature settings. Current outputs from most recent models align more closely with expectations even at increased temperature settings, evidencing an enhanced capacity of ChatGPT to interpret prompts with fewer instructions and diminishing the necessity for techniques such as prompt augmentation.

The execution speeds of the Turbo models are noteworthy, achieving significant throughput without compromising performance for the task at hand. Moreover, the cost efficiencies introduced with these models -threefold less for GPT-4 Turbo and tenfold less for GPT−3.5 Turbo-consolidate ChatGPT's position as a vying competitor to state-of-the-art tools for large-scale studies.

We have observed that the token limit per minute has substantially increased-up to 30 times for GPT-4 and nearly 10 times for GPT−3.5 Turbo. This escalation, coupled with the models' improved response times, results in more expedient processing of privacy policies, as evidenced in Fig. 5. Regarding F1 score performance, the new GPT-4 Turbo model remains consistent with its predecessors, albeit with notable variations: a 1.36% increase in the F1 score for the MAPP corpus and a similar decrease for the OPP-115. The intricacies of these models make it challenging to pinpoint the exact causes of these variations, but it is remarkable that the optimization inherent in the Turbo models has not detrimentally impacted performance for this specific task.

7 Conclusion

Throughout this article, we have substantiated the applicability of LLMs in analyzing and extracting privacy practices from privacy policies. Specifically, ChatGPT has proven to be as effective as traditional NLP techniques, offering significant advantages in terms of cost, runtime, and ease of development. This article also reported on prompts and configurations of LLMs, that were found to yield particularly high performance in identifying and categorizing a variety of data practice disclosures in the text of privacy policies. Results reported in this paper are not limited to the tuning of parameters but also include comparisons of zero-shot and few-shot learning approaches.

Overall our results are consistent with those reported in other domains and suggest that LLM techniques can be configured to outperform more traditional NLP approaches to analyzing the text of privacy policies. Future work will focus on integrating LLM-based analysis into automated systems for privacy compliance. We expect that this approach will yield more efficient and accurate tools for consumers, enterprises and regulators.

Acknowledgements This work has been partially supported by the TED2021-130455A-I00 project funded by MCIN/AEI/10.13039/501,100,011,033 and by the European Union "NextGenerationEU"/PRTR. Jose M. del Alamo has received a grant from the Spanish "Ministerio de Universidades" through the "Movilidad" sub-programme of the "Programa Estatal para Desarrollar, Atraer y Retener Talento", within the "Plan Estatal de Investigación Científica, Técnica y de Innovación 2021–2023". This research has also been partially supported by the National Science Foundation under its Security and Trustworthy Computing Program (grant CNS-1914486).

Author contributions David Rodriguez: Methodology, Software, validation, formal analysis, investigation, data curation, writing-original manuscript. Jose M. Del Alamo: Conceptualization, investigation, resources, writing-original manuscript, writing-review and editing, supervision, project administration, funding acquisition. Ian Yang: Software, validation, formal analysis, data curation. Norman Sadeh: Conceptualization and methodology.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature.

Data availability No datasets were generated or analysed during the current study.

Declarations

Conflict of interest The authors declare no competing interests.

References

1. Srinath M, Matheson L, Venkit PN, Zanfir-Fortuna G, Schaub F, Giles CL, Wilson S (2023) Privacy now or never: Large-scale extraction and analysis of dates in privacy policy text. In: Proceedings of the ACM Symposium on Document Engineering 2023. https://doi.org/10.1145/3573128.3609342. ACM
2. Del Alamo JM, Guaman DS, García B et al (2022) A systematic mapping study on automated analysis of privacy policies. Computing 104:2053–2076. https://doi.org/10.1007/s00607-022-01076-3
3. Zimmeck S, Story P, Smullen D, Ravichander A, Wang Z, Reidenberg JR, Russell NC, Sadeh N (2019) Maps: scaling privacy compliance analysis to a million apps. Proc Priv Enhanc Tech 2019:66
4. Bannihatti Kumar V, Iyengar R, Nisal N, Feng Y, Habib H, Story P, Cherivirala S, Hagan M, Cranor L, Wilson S, Schaub F,Sadeh N, (2020) Finding a choice in a haystack: automatic extraction of opt-out statements from privacy policy text. In: Proceedings of the web conference 2020, pp. 1943-1954. https://doi.org/10.1145/3366423.3380262
5. Zimmeck S, Wang Z, Zou L, Iyengar R, Liu B, Schaub F, Wilson S, Sadeh N, Bellovin SM, Reidenberg J (2017) Automated analysis of privacy requirements for mobile apps. In: 24th Annual Network and Distributed System Security Symposium, NDSS 2017
6. Wilson S, Schaub F, Liu F, Sathyendra KM, Smullen D, Zimmeck S, Ramanath R, Story P, Liu F, Sadeh N et al (2018) Analyzing privacy policies at scale: From crowdsourcing to automated annotations. ACM Trans Web (TWEB) 13(1):1–29
7. Bui D, Shin KG, Choi J-M, Shin J (2021) Automated extraction and presentation of data practices in privacy policies. Proc Priv Enhanc Technol 2021(2):88–110
8. Harkous H, Fawaz K, Lebret R, Schaub F, Shin KG, Aberer K (2018) Polisis: Automated analysis and presentation of privacy policies using deep learning. In: 27th USENIX Security Symposium (USENIX Security 18), pp 531–548
9. Klie J-C, Webber B, Gurevych I (2023) Annotation error detection: analyzing the past and present for a more coherent future. Comput Linguist 49(1):157–198. https://doi.org/10.1162/coli_a_00464
10. Choi JH, Hickman KE, Monahan A, Schwarcz D (2022) Chatgpt goes to law school. J Legal Educat 71:387. https://doi.org/10.2139/ssrn.4335905
11. Tan J, Westermann H, Benyekhlef K (2023) Chatgpt as an artificial lawyer? Artificial Intelligence for Access to Justice (AI4AJ 2023)
12. Tang C, Liu Z, Ma C, Wu Z, Li Y, Liu W, Zhu D, Li Q, Li X, Liu T, Fan L (2023) PolicyGPT: automated analysis of privacy policies with large language models. Preprint at https://arxiv.org/abs/2309.10238
13. Arora S, Hosseini H, Utz C, Bannihatti VK, Dhellemmes T, Ravichander A, Story P, Mangat J, Chen R, Degeling M, Norton T, Hupperich T, Wilson S, Sadeh N (2022) A tale of two regulatory regimes: Creation and analysis of a bilingual privacy policy corpus. In: Proceedings of the

thirteenth language resources and evaluation conference, pp 5460–5472. https://aclanthology.org/2022.lrec-1.585
14. Wilson S, Schaub F, Dara A, Liu F, Cherivirala S, Leon PG, Andersen MS, Zimmeck S, Sathyendra K, Russell NC, Norton TB, Hovy E, Reidenberg JR, Sadeh N (2016) The creation and analysis of a website privacy policy corpus. In: Proceedings of the 54th annual meeting of the association for computational linguistics (Volume 1: Long Papers), pp 1330–1340. Association for Computational Linguistics, Berlin, Germany. https://doi.org/10.18653/v1/P16-1126
15. PrivApp: IT100-Corpus. Accessed: January 10, 2024 (2024). https://github.com/PrivApp/IT100-Corpus
16. Reidenberg JR, Breaux T, Cranor LF, French B, Grannis A, Graves JT, Liu F, McDonald A, Norton TB, Ramanath R, Russell NC, Sadeh N, Schaub F (2015) Disagreeable privacy policies: mismatches between meaning and users' understanding. Berkeley Technol Law J 30:39–88
17. Oltramari A, Piraviperumal D, Schaub F, Wilson S, Cherivirala S, Norton TB, Russell NC, Story P, Reidenberg J, Sadeh N (2018) Privonto: a semantic framework for the analysis of privacy policies. Semant Web 9(2):185–203. https://doi.org/10.3233/SW-170283
18. Evans MC, Bhatia J, Wadkar S, Breaux TD (2017) An evaluation of constituency-based hyponymy extraction from privacy policies. In: 2017 IEEE 25th International Requirements Engineering Conference, pp 312–321. https://doi.org/10.1109/RE.2017.87
19. Hosseini MB, Wadkar S, Breaux TD, Niu J (2016) Lexical similarity of information type hypernyms, meronyms and synonyms in privacy policies. In: 2016 AAAI Fall Symposium Series
20. Chen D, Manning CD (2014) A fast and accurate dependency parser using neural networks. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp 740–750
21. Andow B, Mahmud SY, Wang W, Whitaker J, Enck W, Reaves B, Singh K, Xie T (2019) Policylint: investigating internal privacy policy contradictions on google play. In: 28th USENIX security symposium (USENIX Security 19), pp 585–602. USENIX Association. https://www.usenix.org/conference/usenixsecurity19/presentation/andow
22. Guntamukkala N, Dara R, Grewal G (2015) A machine-learning based approach for measuring the completeness of online privacy policies. In: 2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA), pp 289–294. IEEE
23. Silva AR, Caramujo J, Monfared S, Calado P, Breaux T (2016) Improving the specification and analysis of privacy policies. ICEIS 2016:336
24. Story P, Zimmeck S, Ravichander A, Smullen D, Wang Z, Reidenberg J, Russell NC, Sadeh N (2019) Natural language processing for mobile app privacy compliance. In: AAAI Spring symposium on privacy-enhancing artificial intelligence and language technologies, vol 2, pp 24–32
25. Sathyendra KM, Schaub F, Wilson S, Sadeh N (2016) Automatic extraction of opt-out choices from privacy policies. In: 2016 AAAI Fall Symposium Series
26. Sathyendra KM, Wilson S, Schaub F, Zimmeck S, Sadeh N (2017) Identifying the provision of choices in privacy policy text. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pp 2774–2779
27. Liu F, Ramanath R, Sadeh N, Smith NA (2014) A step towards usable privacy policy: Automatic alignment of privacy statements. In: Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers, pp 884–894
28. Massey AK, Eisenstein J, Antón AI, Swire PP (2013) Automated text mining for requirements analysis of policy documents. In: 2013 21st IEEE International Requirements Engineering Conference (RE), pp 4–13. IEEE
29. Keymanesh M, Elsner M, Sarthasarathy S (2020) Toward domain-guided controllable summarization of privacy policies. In: NLLP@ KDD, pp 18–24
30. Liu F, Fella NL, Liao K (2016) Modeling language vagueness in privacy policies using deep neural networks. In: 2016 AAAI Fall Symposium Series
31. Ravichander A, Black AW, Wilson S, Norton T, Sadeh N (2019) Question answering for privacy policies: Combining computational and legal perspectives. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), pp 4949–4959. Association for Computational Linguistics, Hong Kong, China. https://doi.org/10.18653/v1/D19-1500
32. Radford A, Narasimhan K, Salimans T, Sutskever I et al. (2018) Improving language understanding by generative pre-training

33. Touvron H et al. (2023) Llama 2: Open foundation and fine-tuned chat models. Preprint at https://arxiv.org/abs/2307.09288
34. Ghanadian H, Nejadgholi I, Al Osman H (2023) ChatGPT for suicide risk assessment on social media: quantitative evaluation of model performance, potentials and limitations. In: Barnes J, De Clercq O, Klinger R (eds) Proceedings of the 13th Workshop on Computational Approaches to Subjectivity, Sentiment, & Social Media Analysis, pp 172–183. Association for Computational Linguistics, Toronto, Canada. https://doi.org/10.18653/v1/2023.wassa-1.16
35. Shum K, Diao S, Zhang T (2023) Automatic prompt augmentation and selection with chain-of-thought from labeled data. In: Bouamor H, Pino J, Bali K (eds) Findings of the Association for Computational Linguistics: EMNLP 2023, pp 12113–12139. Association for Computational Linguistics, Singapore. https://doi.org/10.18653/v1/2023.findings-emnlp.811
36. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I (2017) Attention is all you need. Advances in neural information processing systems 30
37. Qin C, Zhang A, Zhang Z, Chen J, Yasunaga M, Yang D (2023) Is ChatGPT a General-Purpose Natural Language Processing Task Solver? Preprint at https://arxiv.org/abs/2302.06476
38. Savelka J, Ashley KD (2023) The unreasonable effectiveness of large language models in zero-shot semantic annotation of legal texts. Front Artif Intell 6:1279794
39. Brocke J, Hevner A, Maedche A (2020) Introduction to Design Science Research, pp 1–13. Springer. https://doi.org/10.1007/978-3-030-46781-4_1
40. Kohavi R, Longbotham R (2015) Online controlled experiments and a/b tests. Encycl Mach Learning data Min. https://doi.org/10.1007/978-1-4899-7502-7_891-2
41. OpenAI: Chat API Reference. Accessed: January 10, 2024 (2024). https://platform.openai.com/docs/api-reference/chat/create
42. Fredriksson T, Mattos DI, Bosch J, Olsson HH (2020) Data labeling: An empirical investigation into industrial challenges and mitigation strategies. In: Morisio M, Torchiano M, Jedlitschka A (eds) Product-Focused Softw Process Improv. Springer, Cham, pp 202–216
43. Brown T, Mann B, Ryder N, Subbiah M, Kaplan JD, Dhariwal P, Neelakantan A, Shyam P, Sastry G, Askell A et al (2020) Language models are few-shot learners. Adv Neural Inform Process Syst 33:1877–1901
44. Together AI: Inference Python Documentation. Accessed: January 10, 2024 (2024). https://docs.together.ai/docs/inference-python
45. Andow B (2024) PrivacyPolicyAnalysis. Accessed: January 10, 2024. https://github.com/benandow/PrivacyPolicyAnalysis
46. Guamán DS, Rodriguez D, del Alamo JM, Such J (2023) Automated GDPR compliance assessment for cross-border personal data transfers in android applications. Comput Secur 132:103262. https://doi.org/10.1016/j.cose.2023.103262
47. Wagner I (2023) Privacy policies across the ages: content of privacy policies 1996–2021. ACM Trans Privacy Secur 26(3):1–32. https://doi.org/10.1145/3590152
48. Yan Y, Rosales R, Fung G et al (2014) Learning from multiple annotators with varying expertise. Mach Learning 95:291–327. https://doi.org/10.1007/s10994-013-5412-1
49. Allen Institute for AI: Crowdsourcing, Pricing, Ethics, and Best Practices. https://blog.allenai.org/crowdsourcing-pricing-ethics-and-best-practices-8487fd5c9872. Accessed: January 10, 2024 (2024)

Authors and Affiliations

David Rodriguez[1] · Ian Yang[2] · Jose M. Del Alamo[1] · Norman Sadeh[2]

✉ Jose M. Del Alamo
jm.delalamo@upm.es

David Rodriguez
david.rtorrado@upm.es

Ian Yang
iany@andrew.cmu.edu

Norman Sadeh
sadeh@cs.cmu.edu

[1] ETSI Telecomunicación, Universidad Politécnica de Madrid, Madrid, Spain

[2] School of Computer Science, Carnegie Mellon University, Forbes Ave, Pittsburgh, PA 15213, USA

Data Retention Disclosures in the Google Play Store: Opacity Remains the Norm

1st David Rodríguez
ETSI Telecomunicación,
Universidad Politécnica de Madrid
Madrid, Spain
david.rtorrado@upm.es

2nd Celia Fernández-Aller
ETSI Sistemas Informáticos,
Universidad Politécnica de Madrid
Madrid, Spain
mariacelia.fernandez@upm.es

3rd Jose M. Del Alamo
ETSI Telecomunicación,
Universidad Politécnica de Madrid
Madrid, Spain
jm.delalamo@upm.es

4th Norman Sadeh
School of Computer Science,
Carnegie Mellon University
Forbes Ave, Pittsburgh, United States
sadeh@cs.cmu.edu

2024 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW) | 979-8-3503-6729-4/24/$31.00 ©2024 IEEE | DOI: 10.1109/EUROSPW61312.2024.00009

***Abstract*—Privacy policies serve as the primary channel through which users are informed about the handling of their personal data, as required by regulations such as the General Data Protection Regulation (GDPR). This paper presents an evaluation of Android applications' privacy policies, focusing on how they articulate and disclose data retention periods. In this paper, we introduce a systematic approach that leverages Large Language Models to evaluate GDPR compliance regarding data retention disclosure across a diverse sample of 2,235 apps, demonstrating the applicability of the method at scale. Our approach reports a 0.904 F1 score, validated with a ground truth dataset manually annotated by legal experts and publicly released. Results show that over half of the examined policies are potentially non-compliant, with a significant subset indicating indefinite data retention and a high ratio of overlapping retention periods on the same privacy policy. This lack of compliance implies that those policies either fail to specify a retention period or provide unclear criteria for determining how long user data is kept. Thus, our study highlights the critical need to improve the clarity and enforcement of privacy policy practices, laying the groundwork for more transparent data governance.**

1. Introduction

In recent years, the technological landscape has undergone a dramatic transformation, underscored by the widespread adoption of smartphones and the explosion of mobile applications. These advancements, fueled by devices bristling with sensors, have catalyzed the vast collection and transmission of users' personal data across networks, leading to an era of unprecedented data accumulation. As the volume of stored data swells, so too do the privacy risks associated with its retention, ranging from data breaches to the misuse of aged information that could infringe upon individual privacy. The General Data Protection Regulation (GDPR) [1] addresses changing privacy needs by establishing clear limits on the collection and use of personal data.

In the context of rapid technological progress and increasing privacy concerns, this study examines the transparency of data retention disclosures in Android app privacy policies to ensure they meet GDPR standards for informing users about data handling and retention. Utilizing Large Language Models (LLMs), we analyzed 2,235 apps to assess GDPR compliance in their privacy policies. Our goal is to expose current practices, pinpoint transparency gaps, and highlight the importance of clear data retention disclosures.

2. Background and Related Work

GDPR emphasizes transparency as one of its seven key principles [6] upon which it is based, ensuring that individuals are informed. Among the specific mandates of this regulation is the requirement for clear disclosure of data retention periods in privacy policies, as detailed in Article 13(2)(b). It compels data controllers to transparently communicate either the specific duration for data storage or the criteria determining such periods, thus protecting user rights and facilitating informed decisions regarding their use of mobile applications.

Privacy policies serve as the primary conduit through which data controllers disclose their data handling practices to users, including data retention. Their inherent legal and technical jargon, combined with vague or overly broad statements, often obfuscates the true nature of data practices, leaving users ill-informed [9]. The difficulty in interpreting these policies can affect user comprehension and pose challenges for regulators and stakeholders in assessing compliance.

Given the challenges in understanding privacy policies, the pursuit of automation in privacy policy analysis has emerged as a promising solution, allowing evaluations at scale that provide a global perspective on personal data usage and retention practices. This technological approach can also enable Data Protection Agencies (DPAs) to save resources and enhance compliance monitoring efforts. Previous studies have demonstrated the applicability of automation in assessing legal compliance with respect to international transfers [5]. They used traditional Machine Learning methods based on pre-trained classifiers and observed a non-compliance ratio of over half of the

DOI 10.1109/EuroSPW61312.2024.00009

apps. Thus, Machine Learning-based classifiers have been proven as an effective approach for privacy policy processing, specifically for the evaluation of compliance with other legal practices, such as the collection of personal data [14]. Recent studies have showcased the effectiveness of LLMs in analyzing and processing privacy policies [12]. Additionally, LLMs have been utilized to evaluate compliance with specific aspects such as personal data sharing, revealing that over 80% of apps did not meet the GDPR required standards [10].

Notably, related work [13] reveals that data retention, despite being one of the less frequently mentioned practices in privacy policies, occupies a substantial portion of the policy text when addressed. Another study [8] indicated that user preferences lean towards less intrusive data practices, including shorter data retention periods, underscoring the importance of our investigation into privacy policy transparency regarding data retention. To the best of our knowledge, our study is the first to systematically explore the disclosure of data retention periods in privacy policies based on an automated method and to demonstrate its effectiveness on a large-scale sample of privacy policies.

3. Method

This section presents our method, which leverages ChatGPT and its GPT-4 model [7] to assess the extent to which privacy policies comply with Article 13(2)(b) of the GDPR, regarding the disclosure of personal data retention periods. The method encompasses the creation of a ground-truth dataset, formulating the ChatGPT-based approach that identifies data retention periods, and validating it against the ground truth. The dataset used for this analysis, containing both the ground truth and validation results, has been publicly released [11] to contribute to the broader research community's efforts in scrutinizing privacy policy compliance with GDPR.

3.1. Ground truth

The foundation of our analysis rests on the OPP-115 dataset [13], widely recognized for its comprehensive annotations on privacy practices. This dataset is valuable for its explicit categorization of data retention practices, a focal point of our study. To tailor this dataset to our specific GDPR compliance assessment, two authors, including a senior data protection lawyer, manually reviewed each privacy policy segment that contained data retention statements as reported by OPP-115 annotators. This review process led to annotating six distinct categories that reflect the variance in GDPR compliance and transparency regarding data retention statements, as delineated in Table 1. These categories, labeled C0 through C5, establish the benchmarks for our evaluation of privacy policies, distinguishing between those that meet GDPR transparency requirements (C1-C5) and those that do not (C0).

Categories C1 and C2 represent policies with explicit disclosures of data retention periods, whether finite or indefinite. In contrast, categories C3, C4, and C5 encompass policies that define the retention period based on specific criteria or conditions rather than stating an explicit length of time. Although GDPR compliance for C5 necessitates additional scrutiny of its data retention purposes [4] (defined in Article 5(1)(b)), this study deems C5 valid without such analysis, acknowledging it goes beyond our current scope.

The dataset will be made publicly available upon the acceptance of this paper.

3.2. Method description

Our methodology employs the GPT-4 model, enhanced with the few-shot learning technique [3] and incorporates the case definitions outlined in Table 1 within the prompt. Additionally, we prompt the model whether the privacy policy expressly states no data retention, to exclude it from subsequent analyses. This approach enables a comprehensive classification of privacy policies according to various levels of transparency regarding data retention periods.

We have adjusted our approach to allow ChatGPT to identify and report multiple cases in which the privacy policy could be classified. This allows us to conduct a comprehensive study, identifying as many different data retention periods as the privacy policy declares. We encouraged the model to output its results in a Python list, aiding in the automated parsing of its outputs, a crucial feature for applying our method on a large scale.

3.3. Validation

Our method achieved outstanding results in evaluating GDPR compliance with data retention period disclosures. We measured an accuracy of 0.939, precision of 0.846, recall of 0.971, and an F1 score of 0.904. These metrics highlight the effectiveness of our methodology in accurately assessing privacy policies' compliance with GDPR's transparency mandates. Notably, the high recall rate indicates the method's proficiency in capturing instances of potential non-compliance. This serendipitous result reinforces the robustness of our approach, ensuring that few non-compliant policies are overlooked.

4. Evaluation in the wild

This section details the experiment conducted on a set of 2,235 Android applications. The proposed method has been utilized to assess these applications' compliance with the transparency requirement of personal data retention periods mandated in the GDPR.

4.1. Experiment design and app selection

Our primary objective is to ascertain whether applications communicate their data retention periods transparently. This requires prior analysis of two conditions to determine that the application is retaining data and must declare the retention period(s): 1) the applicability of GDPR Article 13 to the application, inferred if the app processes personal data of EU subjects, and 2) the absence of a declaration in the privacy policy regarding non-retention of user data. The latter was rigorously evaluated

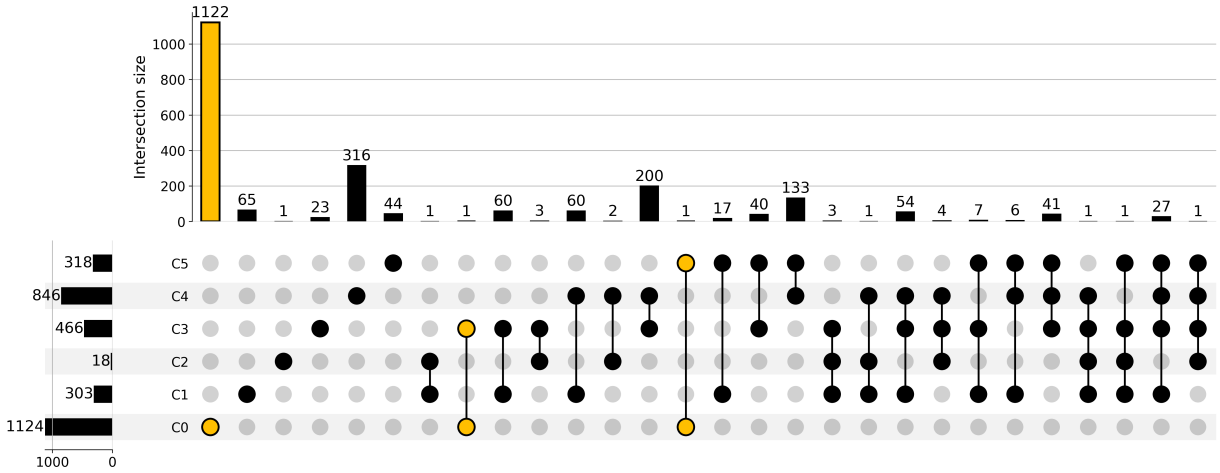


Figure 1: Number of privacy policies annotated according to each case (C0-C5). Dots represent those policies annotated with a single case (first six vertical bars) or multiple cases (from the seventh vertical bar onwards). Horizontal bars on the left side show the occurrence frequency of each case. The yellow color highlights potentially non-compliant policies. *Note: The two policies overlapping C0 and other cases should be considered wrongly annotated by ChatGPT.*

TABLE 1: Cases where privacy policies are categorized according to transparency and personal data retention periods.

Case	Description
C0	No data retention period is indicated in the privacy policy.
C1	A specific data retention period is indicated (e.g., days, weeks, months...).
C2	Indicate that the data will be stored indefinitely.
C3	A criterion is determined during which a defined period during which the data will be stored can be understood (e.g., as long as the user has an active account).
C4	It is indicated that personal data will be stored for an unspecified period, for fraud prevention, legal, or security reasons.
C5	It is indicated that personal data will be stored for an unspecified period, for purposes other than fraud prevention, legal, or security.

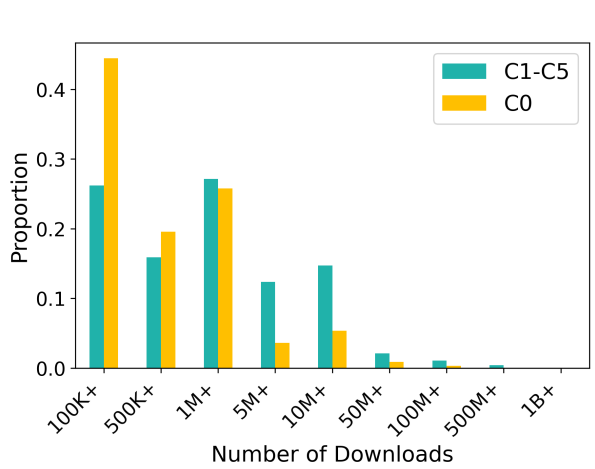
through the method delineated in Section 3.2. To fulfill the first condition, we employed our dynamic app behavior evaluation platform.

This platform employs a network of interconnected Docker containers designed to streamline the analysis process. Applications are systematically downloaded along with their respective privacy policies and stored for further analysis. Subsequently, these applications are installed on Xiaomi Redmi 10 devices to simulate real-user interaction, generating authentic network traffic. This traffic is captured via MiTM proxy and scrutinized to identify personal data transmissions. This allows us to determine if the applications meet criterion 1) outlined above and, therefore, must adhere to Art. 13(2)(b).

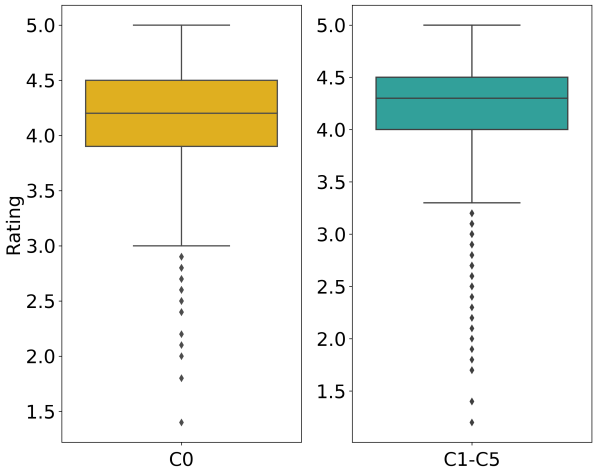
The selection process began with a dataset comprising over 4 million applications from the AndroZoo dataset [2]. To ensure diverse representation, we categorized these applications into three tiers based on their download counts as proxies for their popularity: from 100K to less than 1M for moderately popular apps, from 1M to less than 10M for highly popular apps, and 10M upwards for the most popular apps. From these tiers, we aimed to select a balanced sample of 10,000 applications, maintaining proportionality across the categories reflective of the original distribution, with 65.26% of the sample from the first tier, 27.40% from the second, and 7.33% from the third. These applications were subsequently downloaded and analyzed using a dynamic analysis platform. We successfully downloaded, installed, and executed 5,759 applications, intercepting their network connections. Of these, 3,478 applications, accounting for 60.39% of the subset, engaged in the transmission of personal data, and 2,307 had their privacy policies available in English, making them eligible for further examination under our study's criteria.

4.2. Results

In the analysis of 2,307 privacy policies, 72 applications were excluded due to a) 71 declaring no data storage and b) one providing no response, resulting in 2,235 policies for detailed evaluation. Our analysis divided these policies into two primary categories: those potentially non-compliant with GDPR (C0) and those likely compliant (C1-C5). Notably, a slight majority, 50.20%, were categorized as potentially non-compliant (C0), underscoring a significant challenge in meeting GDPR's transparency criteria for data retention periods. Figure 1 further highlights that data retention disclosures within privacy policies are unevenly distributed, with a predominant number of policies not specifying retention periods, and a considerable segment employing various criteria for determining data retention, indicative of the nuanced and often complex nature of privacy policy statements. This complexity and overlapping of criteria disclosed mirrors



(a) Bar chart comparison between the app distribution of each set based on the number of downloads.



(b) Box plot comparison between the app distribution of each set based on the rating.

Figure 2: Comparison between the potential compliant (C1-C5) and non-compliant (C0) sets based on (a) the number of downloads and (b) the rating.

findings in the literature regarding the extension of such disclosures [13].

Our examination further explored the relationship between app popularity and GDPR compliance status. The analysis revealed a discernible pattern: applications categorized as potentially non-compliant (C0) generally exhibited lower download counts and marginally reduced user ratings compared to their counterparts in compliance categories C1-C5. Although a t-test indicated no significant differences in user ratings between the two groups ($p > 0.05$), a χ^2 (chi-squared) test on download counts yielded a statistically significant difference ($p < 0.05$). This observation strongly indicates that those responsible for the most popular applications are more likely to focus on legal and regulatory compliance.

Conclusion

This extensive analysis of Android applications has revealed a landscape where adherence to GDPR-mandated transparency in data retention disclosures is notably deficient, with over half of the evaluated policies potentially failing to meet the requirements. While 20.85% of policies explicitly allow for indefinite data retention, this study's focus on transparency acknowledges such declarations as compliant in that specific context, yet it underscores the broader privacy concerns they introduce. Moreover, the prevalence of policies detailing multiple data retention periods—34.18% of those analyzed—exemplifies the intricate nature of these documents, posing significant understanding difficulties for the average user.

Our findings advocate for enhanced regulatory oversight to ensure that privacy policies are compliant, clear, and accessible to users. In forthcoming research, we intend to broaden the scope of our analysis to encompass further aspects of transparency, examining how effectively users are informed about their rights concerning data erasure, modification, or access.

Acknowledgements

This work has been partially supported by the TED2021-130455A-I00 project funded by MCIN/AEI/10.13039/501100011033 and by the European Union "NextGenerationEU"/PRTR. This research has also been partially supported by the National Science Foundation under its Security and Trustworthy Computing Program (grant CNS-1914486).

References

[1] EUR-Lex—32016R0679 - EN, 2016. [Online]. Available: https://eur-lex.europa.eu/eli/reg/2016/679/oj.

[2] Kevin Allix, Tegawendé F. Bissyandé, Jacques Klein, and Yves Le Traon. Androzoo: Collecting millions of android apps for the research community. In *Proceedings of the 13th International Conference on Mining Software Repositories*, MSR '16, pages 468–471, New York, NY, USA, 2016. ACM.

[3] Tom B. Brown, Benjamin Mann, Nick Ryder, et al. Language models are few-shot learners. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS'20, Red Hook, NY, USA, 2020. Curran Associates Inc.

[4] Bart van der Sloot Chris Jay Hoofnagle and Frederik Zuiderveen Borgesius. The european union general data protection regulation: what it is and what it means*. *Information & Communications Technology Law*, 28(1):65–98, 2019.

[5] Danny S. Guamán, David Rodriguez, Jose M. del Alamo, and Jose Such. Automated gdpr compliance assessment for cross-border personal data transfers in android applications. *Computers Security*, 130:103262, 7 2023.

[6] Chris Jay Hoofnagle, Bart van der Sloot, and Frederik Zuiderveen Borgesius. The european union general data protection regulation: What it is and what it means. *Information and Communications Technology Law*, 28(1):65–98, Jan 2019.

[7] OpenAI, Josh Achiam, Steven Adler, Sandhini Agarwal, et al. Gpt-4 technical report. *arXiv*, March 2023. Accessed: 2024-03-27.

[8] Dimitris Potoglou, Fay Dunkerley, Sunil Patil, and Neil Robinson. Public preferences for internet surveillance, data retention and privacy enhancing services: Evidence from a pan-european study. *Computers in Human Behavior*, 75:811–825, 2017.

[9] Joel R. Reidenberg, Travis Breaux, Lorrie Faith Cranor, Brian French, Amanda Grannis, James T. Graves, Fei Liu, Aleecia McDonald, Thomas B. Norton, Rohan Ramanath, N. Cameron Russell, Norman Sadeh, and Florian Schaub. Disagreeable privacy policies: Mismatches between meaning and users' understanding. *Berkeley Technology Law Journal*, 30(1):39–88, 2015.

[10] David Rodriguez, Jose M. Del Alamo, Celia Fernández-Aller, and Norman Sadeh. Sharing is not always caring: Delving into personal data transfer compliance in android apps. *IEEE Access*, 12:5256–5269, 2024.

[11] David Rodríguez, Celia Fernández, Jose M del Alamo, and Norman Sadeh. Data retention period disclosures in privacy policies, 2024. Publisher: Mendeley Data, Version: V1, doi: 10.17632/c4x958pzpm.1.

[12] Chenhao Tang, Zhengliang Liu, Chong Ma, Zihao Wu, Yiwei Li, Wei Liu, Dajiang Zhu, Quanzheng Li, Xiang Li, Tianming Liu, and Lei Fan. Policygpt: Automated analysis of privacy policies with large language models, 2023.

[13] Shomir Wilson, Florian Schaub, Aswarth Abhilash Dara, Frederick Liu, Sushain Cherivirala, Pedro Giovanni Leon, Mads Schaarup Andersen, Sebastian Zimmeck, Kanthashree Mysore Sathyendra, N. Cameron Russell, Thomas B. Norton, Eduard Hovy, Joel Reidenberg, and Norman Sadeh. The creation and analysis of a website privacy policy corpus. *54th Annual Meeting of the Association for Computational Linguistics, ACL 2016 - Long Papers*, 3:1330–1340, 2016.

[14] Sebastian Zimmeck, Peter Story, Daniel Smullen, Abhilasha Ravichander, Ziqi Wang, Joel Reidenberg, N Cameron Russell, and Norman Sadeh. Maps: Scaling privacy compliance analysis to a million apps. *Proceedings on Privacy Enhancing Technologies*, 2019:66–86.

Privacy Settings of Third-Party Libraries in Android Apps: A Study of Facebook SDKs

David Rodriguez
ETSI Telecomunicación, Universidad Politécnica de Madrid
Madrid, Spain
david.rtorrado@upm.es

Joseph A. Calandrino
Washington, D.C., USA
jcalandr@alumni.princeton.edu

Jose M. Del Alamo
ETSI Telecomunicación, Universidad Politécnica de Madrid
Madrid, Spain
jm.delalamo@upm.es

Norman Sadeh
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA
sadeh@cs.cmu.edu

Abstract

Previous studies have demonstrated that privacy issues in mobile apps often stem from the integration of third-party libraries (TPLs). To shed light on factors that contribute to these issues, we investigate the privacy-related configuration choices available to and made by Android app developers who incorporate the Facebook Android SDK and Facebook Audience Network SDK in their apps. We compile these Facebook SDKs' privacy-related settings and their defaults. Employing a multi-method approach that integrates static and dynamic analysis, we analyze more than 6,000 popular apps to determine whether the apps incorporate Facebook SDKs and, if so, whether and how developers modify settings. Finally, we assess how these settings align with the privacy practices that developers disclose in the apps' privacy labels and policies.

We observe widespread inconsistencies between practices and disclosures in popular apps. These inconsistencies often stem from privacy settings, including a substantial number of cases in which apps retain default settings over alternatives that offer greater privacy. We observe fewer possible compliance issues in potentially child-directed apps, but issues persist even in these apps. We discuss remediation strategies that SDK and TPL providers could employ to help developers, particularly developers with fewer resources who rely heavily on SDKs. Our recommendations include aligning default privacy settings with data minimization principles and other conservative practices and making privacy-related SDK information both easier to find and harder to miss.

Keywords

Third-party libraries, software development kits, privacy settings, Facebook SDK, Android applications, dynamic analysis, default settings, compliance analysis, privacy labels, privacy policies

1 Introduction

Mobile applications are integral to modern life, from how we communicate with others and entertain ourselves to how we manage our health and finances. Today, mobile app developers rely heavily on software development kits (SDKs). SDKs comprise a collection of software tools and programs, and they routinely incorporate third-party libraries (TPLs) into apps. SDKs help developers produce sophisticated apps rapidly and efficiently. They can do anything from assisting developers in building, deploying, and managing apps to facilitating targeted advertising, social media logins, and much more. Despite the benefits of SDKs, their use in mobile apps has raised concerns regarding privacy.

SDK providers often offer developers free or subsidized services in exchange for collecting and utilizing user data across all apps that use their SDKs [63]. Given their market share and business practices, the data available to the parties behind some widely used SDKs may be significant [63]. The integration of TPLs into Android apps—whether via SDKs or not—introduces a well-documented array of privacy concerns [1, 37, 38, 59, 66, 68, 82]. These libraries can access user data ranging from location to personal communications, potentially without explicit user consent [2] or even developer awareness [11]. Beyond jeopardizing user privacy, this access places app developers at risk of failing to comply with their privacy promises and laws like GDPR or CCPA, which may create obligations from data minimization to consumer controls on personal information [31, 51].

While privacy concerns regarding TPLs may be well known, a gap exists in understanding the impact of TPL privacy settings and their defaults on the privacy of apps. Even if a TPL offers developers extensive privacy-related settings options, prior work suggests that defaults tend to favor functionality over privacy, encouraging practices that may be unnecessary or opaque [14]. Furthermore, developers may be reluctant to change default settings for TPLs [59]. If developers fail to choose appropriate privacy-related TPL settings, apps may not adhere to the developers' privacy promises and obligations, and users may lack knowledge of and control over apps' privacy practices. Addressing these issues requires a deeper investigation into how developers interact with privacy-related TPL settings, ideally contextualized by developers' privacy commitments and obligations.

We consider the privacy-related settings and defaults provided by Facebook (Meta) SDKs for Android, specifically the Facebook Android SDK [19] and Facebook Audience Network SDK [18]. These SDKs offer valuable case studies due to their widespread use and configurable privacy settings, which can significantly influence app

Proceedings on Privacy Enhancing Technologies YYYY(X), 1–15

https://doi.org/XXXXXXX.XXXXXXX

Table 1: Analyzed Facebook SDKs and their modules.

SDK name	Module (SDK) name	Maven artifact ID[a]	Description
Facebook Android SDK	Facebook Core SDK	`facebook-core`	Provides analytics and functionality for other SDK modules.
	Facebook Login SDK	`facebook-login`	Allows users to authenticate using Facebook credentials.
	Facebook Share SDK	`facebook-share`	Enables sharing from an app on Facebook.
	Facebook Messenger SDK	`facebook-messenger`	Integrates Facebook Messenger functionality.
	Facebook App Links SDK	`facebook-applinks`	Supports links into other apps (Android deep links).
	Facebook Marketing SDK	`facebook-marketing`	Facilitates integration of Facebook marketing capabilities.
Audience Network SDK	-	`audience-network-sdk`	Enables advertising and audience monetization.

[a] Maven is a project management system that works with a public repository to integrate Android SDKs. We omit groupId prefix (`com.facebook.android`) from artifact IDs.

privacy practices. Our study focuses on whether and how developers modify these settings. Understanding how developers interact with privacy settings, particularly defaults, is crucial to addressing privacy concerns that TLPs and SDKs raise. Our research employs a combination of static and dynamic analysis, with both approaches providing complementary insights and robust validation of developers' privacy settings choices. We also examine how developers' choices align with apps' stated privacy practices, which can not only uncover discrepancies but also suggest failures of developers to appreciate available choices and implications.

Our analysis reveals that a large proportion of apps retain default SDK privacy-related settings over privacy-enhancing alternative settings. We also identified potential compliance issues in privacy labels and policies when compared with actual app behavior, and many of the concerns are associated with privacy-related SDK settings and their defaults. These findings offer valuable insights into privacy-relevant choices that app developers make when utilizing SDKs, with potential implications for developers, providers of TPLs/SDKs, and policymakers.

Contributions. This study examines the privacy settings available to and made by app developers integrating the Facebook Android SDK and Facebook Audience Network SDK. We employ static and dynamic analysis on more than 6,000 Android apps. Key contributions include:

(1) *Compilation and Detailed Examination of SDK Privacy Settings*: We document and explain the privacy-related settings available in both Facebook SDKs along with their defaults, highlighting their privacy impact.
(2) *Static Analysis*: This analysis assesses Facebook SDK integration in apps, and it compiles and offers insights into certain developer choices regarding privacy-related settings.
(3) *Dynamic Analysis*: We develop new methods that validate and expand on our static analysis. These methods utilize runtime details to confirm SDK integration, determine SDK version, and analyze or confirm privacy-related settings choices. Our findings suggest that static analysis alone may not offer a complete picture.
(4) *Compliance Analysis*: We examine apps' privacy labels and policies, identifying potential discrepancies between declared and apparent privacy practices stemming from SDK use. This analysis highlights possible compliance issues and offers hints of underlying causes.

The remainder of this paper is organized as follows. Section 2 introduces the two Facebook SDKs and their privacy-related settings. Section 3 reviews related work. Section 4 describes our research approach. Section 5 presents findings from our analysis of more than 6,000 apps, discussing SDK integration, privacy-related settings modifications, and compliance issues observed. Section 6 contextualizes our findings with existing developer studies and explores mitigation strategies, and Section 7 addresses the study's limitations. Section 8 concludes and outlines future research directions.

2 Facebook SDKs

We focus on two Facebook SDKs for Android, the Facebook Android SDK and the Audience Network SDK, that are among the most popular social and ad network SDKs in the Android ecosystem [7]. Both SDKs bundle TPLs with apps and provide configurable privacy-related settings that developers can modify through an app's Android Manifest file, app code, or the Meta Developers Platform [57], a centralized developer hub that includes features for managing apps and configuring Facebook SDKs. As Section 4 discusses, our analysis can infer modifications that developers make to these settings via these three methods.

2.1 Facebook Android SDK

The Facebook Android SDK (or "Facebook SDK for Android") offers an extensive range of functionality, including user authentication via Facebook login and content sharing on the platform. This SDK employs a modular architecture that enables developers to integrate specific features independently [28]. All modules rely on essential functionality from the required Facebook Core SDK (or simply Core SDK), discussed below, but developers can otherwise incorporate each module into apps as desired [28, 32]. Table 1 provides an overview of the SDK and its modules.

The Facebook Core SDK module manages privacy-related settings for all modules, providing an array of such settings. These settings have been available in their current form since version 4.34.0 (June 2018) [64]. We compiled privacy-related settings and confirmed the default value for each via both Meta's official documentation [24] and manual examination of the Facebook Core SDK's source code [32]. We provide details on each setting below (and in Table 2). Although Meta's documentation mentions possibilities like delaying data collection "*to obtain user consent or fulfill legal obligations*" [20], the default for each privacy-related setting

Table 2: Privacy-related settings available in the analyzed Facebook SDKs.

SDK	Setting	Definition	Default	Available Since Version (Date)
Facebook Android SDK	AutoLogAppEvents	Enables collection of events and other data for user interaction and engagement tracking.	Enabled	4.34.0 (June 2018)
	AutoInit	Controls automatic initialization of the Facebook Android SDK upon app launch.	Enabled	4.34.0 (June 2018)
	AdvertiserIDCollection	Allows collection of Advertising Identifier (AdID) for personalized advertising.	Enabled	4.34.0 (June 2018)
	LimitEventAndDataUsage	Restricts logged events from being used for purposes other than analytics or conversions.	Disabled	4.34.0 (June 2018)
Audience Network SDK	DataProcessingOptions	Allows constraints on Facebook's use and sharing of user data.	Disabled	5.5.0 (August 2019)
	MixedAudience	Adjusts data collection and use to assist compliance with children's privacy laws.	Disabled	5.6.0 (October 2019)

notably is the option that immediately facilitates or minimizes restrictions on data collection, use, and sharing.

AutoLogAppEvents. Determines whether the SDK collects and logs user interactions, events, and other data, such as app downloads, in-app purchases, ad interactions, email address, name, phone number, physical address details (city, state/province, zip/postal code, and country), gender, and date of birth. Collection and logging is enabled by default. Developers can change this setting via the app's Manifest, code instructions, or the Meta Developers Platform [57].[1]

AutoInit. Controls the automatic initialization of the SDK upon app launch. This setting is enabled by default, which allows the SDK to start functioning immediately and collect data for analytics without additional initialization code. According to Meta's documentation, developers can disable this setting via the app's Manifest. Although the SDK can be re-enabled through code instructions, it is unclear whether this method also allows for disabling the SDK.

AdvertiserIDCollection. Governs the collection of Android's Advertising Identifier (also known as AAID, GAID, or AdID; we use AdID alone for consistency). This identifier is generally used by SDKs to track user activity across apps and deliver personalized advertising [74]. The collection of AdID is enabled by default but can be disabled through the Manifest or via code instructions.

LimitEventAndDataUsage. Can restrict whether logged user interactions and events sent to Facebook are used for purposes beyond analytics and conversions. These restrictions are disabled by default, but when enabled, the data collected will not be used for targeted advertising or detailed marketing profiling [26]. This setting is stored on the device and persists across app launches. Unlike other settings, this one is not detailed in Meta's official documentation, but we identified it in the SDK code. This setting can be changed via code instructions.

[1]The automatic logging of events can also be managed through the Events Manager [58], a tool within Meta's Business Suite [56] that allows for monitoring, analyzing, and managing events tracked by the SDK.

2.2 Facebook Audience Network SDK

Facebook's Audience Network SDK [21] is for advertising and monetization, providing tools for integrating Facebook ads into apps. We compiled privacy-related settings for this SDK from the official documentation's section on best practices [22, 23].

DataProcessingOptions. Allows developers to specify how Facebook should handle user data. This setting is also known as Limited Data Use (LDU). By default, data usage is not limited, but developers can change this setting to restrict data processing based on user location or Meta's geolocation. These restrictions support compliance with U.S. state privacy regulations [22] by limiting data use for personalized ads, sharing with third parties, and data retention duration. This setting can be modified only through code instructions.

MixedAudience. Helps developers manage apps used by both children and adults, facilitating compliance with the Children's Online Privacy Protection Act (COPPA) in the U.S. [23]. This setting is disabled by default, but when enabled, the SDK adjusts data collection practices to limit personal data collection from children and restrict data use for targeted advertising. The setting can be modified only via code instructions or the Meta Developers Platform.

3 Related Work

The incorporation of SDKs and TPLs into mobile apps can create substantial privacy and compliance risks. Previous research has identified risk stemming from the inheritance framework of Android permissions: all libraries in an app can use the permissions granted to the app [68, 70, 77]. SDKs that enable app monetization through advertising frequently collect personal data to deliver ads based on targeted audiences. Such data collection may lack transparency [27, 84] and could violate privacy regulations such as GDPR, CPPA, and COPPA, especially in apps targeting children [1].

Additionally, developers face challenges understanding the implications of SDK integration in their apps, which can lead to unintentional data leaks and further exacerbate risk [1, 84, 85]. The widespread adoption of SDKs and TPLs creates risks that many apps transmit personal data without proper user consent [13, 38].

Malicious Android libraries may even target SDKs and TPLs from other vendors in the same app to extract user data, since third-party components of apps are not isolated from each other [75].

To assess the real-world privacy impact of TPLs and SDKs in the mobile ecosystem, detecting their presence in apps and analyzing their behavior can be useful. TPL detection techniques and tools can be categorized based on their operational principles. Detection techniques based on package structure analyze the organizational hierarchy of code packages, utilizing predefined patterns to identify the presence of TPLs [55]. Class-dependency analysis evaluates the interdependencies among classes to detect modular components indicative of TPLs [53]. Control flow graph (CFG) and opcode analysis techniques trace control flow and low-level code structure, looking for library-specific patterns that facilitate identification of TPLs [78]. Signature-based detection utilizes a database of known TPL signatures, matching code segments to signatures to detect libraries [79]. Heuristic and machine learning tools leverage algorithms to recognize code patterns and adapt based on data [16].

Zhan et al. [80] reviewed tools available for identifying TPLs in Android apps. Their empirical study compared these tools based on various characteristics, including effectiveness, accuracy of version identification, resilience to code obfuscation, and ease of use. In this study, LibScout [10] outperformed other tools in terms of effectiveness and accurate TPL detection, albeit with higher processing times. Consequently, we leverage LibScout in our study to help identify TPLs.

Various prior approaches evaluate the behavior of Android apps. These approaches can generally be categorized as static and dynamic analysis techniques. Static analysis involves examining app code and other material without executing apps. FlowDroid [8], IccTA [52], and Amandroid [76] are widely used static analysis tools. Dynamic taint analysis and network traffic analysis are common dynamic analysis approaches. Dynamic taint analysis involves tracking the flow of data through an app during execution to identify app behavior and data uses [44, 69]. Network traffic analysis typically employs traffic interception tools such as Mitmproxy to obtain HTTP(S) communication, decrypt as necessary, and identify (potentially concerning) transmitted data [36, 47, 50, 66, 83].

Some prior work has sought to determine whether TPLs are responsible for particular app behavior and data practices. Hao et al. [37] dynamically instrumented APIs and inspected their call stack to assess whether a TPL is causing data leaks. Other related work relies on the same conceptual approach and uses different tools like Frida to inspect stack traces for API calls of interest [39, 66, 68]. One component of our analysis applies a similar approach to analyze the integration of SDKs in apps, including details of the SDKs' privacy-related settings.

Existing literature has linked SDK configurations with privacy leaks [27] and urged developer caution when using SDKs [81]. Surveys and interviews with developers reveal a reluctance to modify default settings for advertising SDKs [59]. Default settings can facilitate extensive data collection, potentially undermining user consent and putting privacy compliance at risk [12].

Closer to our work, Kollnig et al. [50] explored changes to TPLs' default privacy settings by Android and iOS developers. Their static analysis focused only on the inspection of the apps' Manifest files. They conclude that modifications are infrequent, risking violation of the GDPR's data minimization principle. Our research extends this inquiry with a deep examination of two prominent SDKs' settings, defaults, and options alongside developer choices and app privacy disclosures. We complement static analysis with dynamic analysis. This approach allows us to monitor for changes made outside of an app's Manifest file, including changes at runtime.

We also consider the possibility that SDK settings result in mismatches between app behavior and an app's privacy policy or label. Prior studies have documented potential discrepancies (and associated risks) between these privacy disclosures and app practices, considering diverse issues including repercussions of SDK integration [12, 30, 50, 66, 84]. These discrepancies may undermine transparency-oriented marketplace policies, user trust, and compliance with formal privacy policies [46].

4 Research Method

We combined state-of-the-art static, dynamic, and compliance analysis techniques into an analysis platform that sheds light on developer choices regarding the Facebook Android SDK and Facebook Audience Network SDK's privacy-related settings. Our static and dynamic analyses are complementary, with dynamic analysis validating and expanding on observations from static analysis. We analyzed app code, execution behavior, communications, and metadata. Our compliance analysis compares findings from the static and dynamic analyses with developers' representations in privacy labels and privacy policies. To facilitate analysis of a large volume of apps, we segmented the analysis tasks (as well as the downloading and storing of apps) into Docker modules and used RabbitMQ asynchronous queries to coordinate the modules (see Figure 1).

This analysis provides a detailed view of developer practices surrounding the Facebook SDKs' privacy-related settings, identifies risks, and yields insights into how developers approach the settings. The following sections describe our analysis platform.

4.1 Download and Storage

Our download module uses an unofficial Google Play Store API [33] to fetch apps (APKs) and metadata, such as download statistics and privacy policy URLs. Multiple "workers" operate simultaneously and independently. Each simulates a real device connection through an individual Google account. This multi-worker approach allows us to parallelize the downloading of terabytes of app data, achieving a peak download speed of six apps per minute (approximately 360MB per minute) using three workers. The module also uses Selenium to download privacy policies and extracts privacy labels from apps.

Our storage module acts as a centralized API server, storing and efficiently serving APKs, privacy policies, and privacy labels to other components of the analysis platform. This centralized storage works well, meeting the high-throughput demands of our platform.

4.2 Static Analysis

Our static analysis approach is designed to determine the presence of the Facebook SDKs in an Android app and collect evidence of how developers configure certain privacy-related settings. Illustrated in Figure 1, the static analysis pipeline is structured into two principal phases: SDK presence identification and settings analysis. This

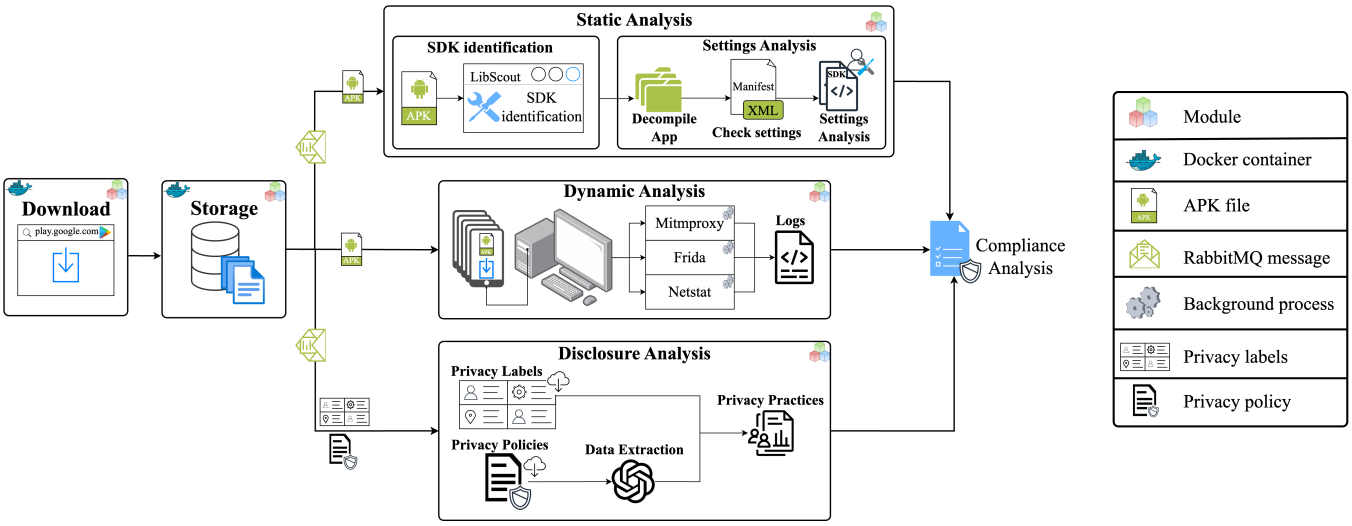


Figure 1: Architecture of our analysis platform.

component's modular design facilitates parallel execution across multiple machines.

SDK Identification. To identify SDKs in apps, our static analysis pipeline relies on LibScout [10]. LibScout uses TPL profiles to recognize SDKs in Android apps and is robust to code obfuscation like identifier renaming and API hiding [80]. LibScout constructs detailed profiles based on the class hierarchy and method signatures extracted from an SDK's compiled .jar or .aar files.

LibScout comes with default profile data for the general Facebook Android SDK, but developers may choose to include or exclude modules of this SDK (see Section 2.1). Therefore, we seek to identify the required Facebook Core SDK submodule, and we also must identify the Audience Network SDK. To address this, we developed scripts to crawl relevant SDKs available in the Maven repository and automate LibScout profile generation.[2]

To handle the raw unstructured logs that LibScout outputs, we wrote a parsing script that uses predefined text-processing rules. Our module logs each relevant SDK identified.

Settings Analysis. If the prior phase suggests an app contains a relevant SDK, the analysis focuses on evidence of privacy-related SDK settings in the app's Manifest file. We use Apktool [6] to extract and reconstruct the Manifest file from an app's compiled APK without data loss. Given the Manifest file, we analyze the XML elements and attributes to locate any assignment of relevant settings that developers can modify via this file: Facebook Android SDK's `AutoLogAppEvents`, `AutoInit`, and `AdvertiserIDCollection` settings. This analysis reveals modifications as well as instances where default settings are explicitly or implicitly unchanged.

[2]While we considered a similar approach to identify versions of SDKs, we were concerned about reliability given the sometimes-small code differences between versions. The scripts to crawl TPLs and generate LibScout profiles are available at https://github.com/DavidRodriguezTorrado/PrivacySDKSettingsAnalyzer

4.3 Dynamic Analysis

Our dynamic analysis validates, extends, and occasionally offers a different perspective from static analysis. We examine apps' runtime behavior and monitor traffic. The analysis allows us to verify the integration of Facebook SDKs and learn versions. It also reveals adjustments to privacy-related Facebook SDK settings via methods beyond an app's Manifest file alone, including code and Meta Developers Platform. In addition, we capture app communication and assess whether transmissions stem from a Facebook SDK. The combination of static and dynamic analysis provides a nuanced view of developer configuration choices and actual SDK usage.

Our dynamic analysis module utilizes five Redmi 10 devices running Android API 30 (Android 11) physically located and running in Spain. This allows parallelization and mitigates potential bottlenecks. Devices are equipped with active Frida servers for app instrumentation.

Installation and Execution. After the download module acquires an app, a RabbitMQ message initiates dynamic analysis. The app is installed on a Redmi 10 device, and background applications are halted. The app undergoes a 120-second idle phase with no user interaction followed by a 180-second interactive phase with pseudo-random events triggered via Android Monkey. Following analysis of an app, we restore the device configuration to its initial state.

SDK and Version Identification. To identify Facebook SDK integration and version, we manually inspected the code of both Facebook SDKs and identified methods (primarily getters) that reveal the SDK version and configuration values. Frida, a dynamic instrumentation toolkit, allows real-time interaction with and manipulation of processes running in user space. Frida enables monitoring and interception of all methods and calls during execution. Through injected JavaScript, we dynamically triggered the SDK-identifying methods, allowing us to capture the actual SDK version. This complements LibScout's SDK identification—ensuring the reliability of both identification methods—and reveals SDK version.

Settings Analysis. We use Frida to monitor privacy-related Facebook SDK settings.[3] Our use of Frida to infer SDK behavior and configuration changes at runtime extends our static analysis of Manifest files. All settings in Table 2 have setter methods, allowing developers to modify the settings at runtime. All of the Facebook Android SDK settings also provide getter methods for retrieving current values.

After confirming the presence of Facebook SDKs, we continuously check for SDK initialization every second to avoid triggering it prematurely by accessing the privacy settings. Once initialization is detected, we query the getters every five seconds to capture initial values and track any subsequent changes. During Frida execution, we also intercept the setters, recording both previous and new values to track configuration changes.

Network Traffic Monitoring. During both the idle and interactive execution phases, we monitor network traffic using Mitmproxy and Frida. Mitmproxy intercepts HTTP traffic from the app and decrypts encrypted (HTTPS) traffic. We use Netstat to identify open ports on the mobile device, ensuring that connections are exclusively made by the app under analysis. Our Frida scripts manipulate certificate validation to bypass certificate pinning and ensure Mitmproxy can decrypt HTTPS traffic. Frida also helps us trace the source of network communications. Our scripts intercept calls to networking methods (e.g., sockets) and log contextual information, such as the specific port used and the stack trace. This allows us to identify the code triggering the communication, including whether the code is in a third-party library. Leveraging work by Rodriguez et al. [66], we cross-reference this data with Mitmproxy logs to associate communication content with the source app and libraries. This approach enables us to determine what data is being transmitted by the applications and which SDKs are responsible for transmission.

4.4 Disclosure and Compliance Analysis

We compare evidence of apps' practices from our static and dynamic analysis against developers' declared practices in apps' privacy labels and privacy policies. This comparison can expose potential mismatches stemming from privacy-relevant Facebook SDK settings and can hint at underlying causes.

Privacy Labels. In 2022, Google implemented privacy labels [48, 49] as an accessible format for users to learn of apps' privacy practices [40]. Before the adoption of privacy labels by major app stores, privacy policies were the primary method for informing users about these practices.

Meta offers guidance regarding practices that developers should disclose on privacy labels for apps that integrate the Facebook SDKs [29]. That guidance depends on the specific SDK and the custom events configured within the SDKs. For example, developers can modify automatic event logging (via `AutoLogAppEvents`) to limit collection of user data (see Section 2.1). Both the Facebook Android SDK and the Audience Network SDK collect device identifiers; however, it is not explicitly stated how adjusting SDK settings (e.g., `AdvertisingIDCollection`) should impact privacy label disclosures.

We checked if AdID collection is declared in privacy labels and if AdID is collected and transmitted by apps. If an app integrates the Facebook Core SDK and the `AdvertisingIDCollection` setting is either unmodified or enabled, the privacy label should declare the collection of AdID. Additionally, we analyzed network traffic to verify whether the AdID is transmitted by the Facebook SDKs and cross-referenced these findings with the corresponding privacy labels to assess the alignment between actual practices and disclosures.

Privacy Policies. Privacy policies have traditionally been the primary means of informing users of Android apps' privacy practices. Google mandates that all apps have a privacy policy, which must be accessible from both the Play Store listing and the app itself [42]. We seek to assess whether app behavior aligns with practices declared in privacy policies and to identify discrepancies stemming from integrated Facebook SDKs. We also note apparent inconsistencies between privacy labels and privacy policies.

Because privacy policies are written in natural language, extraction of relevant information is challenging. To address this, we leverage an existing LLM-based privacy policy analysis tool [67]. This tool leverages ChatGPT with a carefully designed prompt that integrates advanced prompting techniques, iterative refinements, and context retention strategies to detect privacy practices. The method demonstrates high accuracy in identifying statements related to the collection of the AdID identifier, achieving an F1-score between 0.984 and 1.0 across two datasets of privacy policies annotated by legal experts. This allows for an automated approach that facilitates analysis of privacy policy disclosures at the large scale of our study and enables focused manual verification of more critical findings.

5 Results

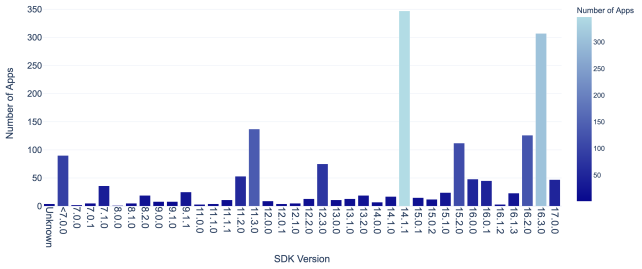
To evaluate whether and how developers modify privacy-related settings of Facebook SDKs in Android apps, we compiled a set of popular apps from AndroZoo [3], a regularly updated repository containing metadata for over four-million apps. We focused on apps with metadata collected in 2023 or later to ensure the timeliness of our analysis and the availability of apps in the Google Play Store.

We analyzed apps in order of popularity, ultimately downloading 8,848 apps to send through our analysis pipeline (see Figure 1). All analyzed apps had more than one-million downloads. During static or dynamic analysis, some apps experienced issues, ranging from LibScout processing errors to app installation and communication problems. Installation issues may stem from factors like the rooted device environment used for dynamic analysis, and execution errors when using tools like Mitmproxy or Frida also affected app analysis. We ultimately successfully processed 6,203 apps. To ensure a snapshot of practices at roughly a single point of time, we downloaded and analyzed all apps in April and May 2024.

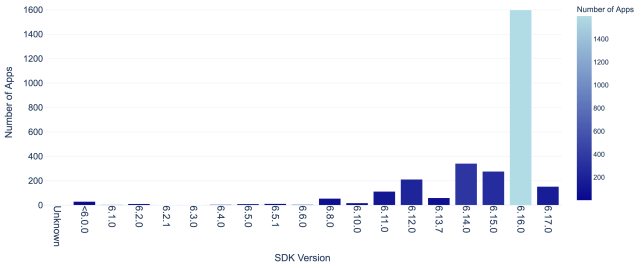
5.1 Facebook SDKs Integration

We used both static and dynamic analysis to identify Facebook SDKs in apps. This choice allowed for cross-verification. LibScout and our Frida-based method agreed on the presence or absence of

[3]The Frida script used to detect Facebook SDK integration and monitor privacy-related settings is available at https://github.com/DavidRodriguezTorrado/PrivacySDKSettingsAnalyzer.



(a) Distribution of Facebook Core SDK versions. Versions prior to 7.0.0 (May 2020) are aggregated.



(b) Distribution of Audience Network SDK versions. Versions prior to 6.0.0 (September 2020) are aggregated.

Figure 2: Distribution of Facebook SDK versions.

the Facebook Core SDK and Audience Network SDK in 97.45% of the apps analyzed. As a precaution, we excluded 158 apps where the methods disagreed, leaving 6,045 apps.

The Facebook Core SDK, which manages privacy settings for the Facebook Android SDK and its components, was integrated into 1,693 apps (28.00%). The Audience Network SDK was found in 2,897 apps (47.92%). Additionally, 1,345 apps (22.25%) integrate both the Facebook Core SDK and the Audience Network SDK, while 3,245 apps (53.68%) integrate at least one of these SDKs.

We relied on our Frida-based version identification technique for version identification. While this method faced challenges due to potential changes in the location or accessibility of version-indicating methods, those challenges affected only a small number of apps (four for the Facebook Core SDK and one for the Audience Network SDK).

Figure 2 illustrates the distribution of versions of the Facebook Core SDK and the Audience Network SDK within our sample. The Facebook Core SDK exhibits a broader version distribution, with many developers retaining earlier versions. We did not observe any clear relationship between the Facebook Core SDK version and app popularity in our data. Conversely, the Audience Network SDK trends towards more recent versions being widely adopted. Our most observed version (6.16.0) matches the most deployed version according to the Google Play SDK Index [61].

Based on Google Play Store app categories, Facebook SDKs are particularly popular in gaming apps in our dataset. For instance, the Facebook Audience Network SDK is integrated into 85.7% of 'Arcade' apps, 85.5% of 'Word' apps, and 82.7% of 'Puzzle' apps. Similarly, the Facebook Core SDK is present in 83.3% of 'Casino' apps, 67.7% of 'Strategy' apps, and 66.7% of 'Role Playing' apps. Beyond gaming, categories like 'Music' (38.3% Core SDK, 72.3% Audience Network SDK) and 'Shopping' (35.5% Core SDK) also have significant integration rates. Conversely, Facebook SDK integration rates are lower for 'Medical' apps (15.4% Core SDK, 7.7% Audience Network SDK) and 'Educational' apps (7.4% for both SDKs).

5.2 Privacy-Related Settings Configuration

Our analysis indicates that some developers actively modify Facebook SDK privacy-related settings, which may suggest some level of awareness of these options. However, our observations also reveal instances where settings are configured in ways that do not result in meaningful changes, which could reflect complexities or challenges in navigating the available options and understanding their implications.

AutoLogAppEvents (Facebook Android SDK). A substantial portion of the apps (1,409, 83.23%) that integrate the Facebook Core SDK explicitly set a value for the automatic event logging option in the Manifest file. However, nearly two-thirds of these apps (939) explicitly assigned the default value (enabled) to the setting, and only one-third of the apps explicitly disabled this feature. The remaining 284 apps (16.77%) did not set any value, thus retaining the default (enabled).

In turn, our dynamic analysis revealed that 1,226 apps (72.42%) had automatic event logging enabled at runtime. In five apps, discrepancies between static and dynamic analyses were observed, likely due to adjustments made via the Meta Developers Platform, which allows for SDK configuration.[4]

Furthermore, we detected scarce runtime changes to this configuration attribute via code instructions (i.e., setter methods). We noted 32 apps enabling the setting when it was already enabled, three apps disabling it when it was already disabled, and one app disabling it when it was previously enabled. The repeated overriding of values without effecting changes could imply that developers face difficulties or ambiguities in understanding the impact of these configurations at runtime, as noted in prior studies on SDK documentation and configuration challenges [71, 72]. Section 6 further explores these issues.

AutoInit (Facebook Android SDK). The documented method to prevent the automatic initialization of the Facebook Android SDK is to disable it in the Manifest file [5], but this setting can also be altered in the app code. Our analysis revealed that 186 apps (10.99%) employed the Manifest approach to disable auto-initialization. Although this setting is enabled by default, 66 apps explicitly enabled it via the Manifest.

During dynamic analysis, we used Frida to check the SDK initialization status at one-second intervals. We found that in 194 apps, the SDK did not initialize at app startup. This suggests that eight

[4]Note that a direct comparison between the numbers obtained from static and dynamic analyses is not feasible due to differences in the coverage of each method. While static analysis includes all analyzed apps, dynamic analysis is limited to apps that can be executed successfully during the analysis process. Consequently, some apps included in the static analysis could not be dynamically analyzed, resulting in a mismatch in the datasets.

Table 3: Summary of analyzed Facebook SDKs' privacy settings. The third column is the percentage of apps that did not explicitly assign a value to the setting (default or otherwise) via the Manifest or in code. The fourth column is the percentage of apps that enabled a privacy-enhanced configuration, determined by cross-referencing changes observed in the Manifest, getters, and setters. For reasons we discuss in Section 5.2, we exclude `MixedAudience`.

SDK	Setting	No value explicitly assigned (%)	Privacy-enhanced configuration (%)
Facebook Core	AutoLogAppEvents	24.75%	17.90%
Facebook Core	AutoInit	84.64%	11.46%
Facebook Core	AdvertiserIDCollection	31.25%	6.79%
Facebook Core	LimitEventAndDataUsage	100%	0%
Audience Network	DataProcessingOptions (LDU)	99.76%	0.14%

apps employed alternative methods to delay initialization beyond the Manifest settings. Additionally, we observed three instances where the `AutoInit` configuration was changed from `False` to `True` at runtime, thereby initiating the SDK. According to the official documentation, this setting is intended to allow developers to obtain user consent before SDK initialization. However, we did not assess whether these apps implemented such consent mechanisms. Conversely, one app stopped the SDK by changing its configuration from `True` to `False` at runtime, and 38 set but did not change the previous (`True`) value.

AdvertiserIDCollection (Facebook Android SDK). A large number of apps (1,082, 63.91%) explicitly chose a value for this setting, but most (958) enabled it, which is the default configuration. Conversely, only 124 apps disabled this setting explicitly through the Manifest file.

Our dynamic analysis revealed that 1,406 apps (83.05%) had this setting enabled at runtime, while only 115 apps (6.79%) had it disabled. The remaining apps could not be executed due to errors encountered during the dynamic analysis execution. Notably, five apps appeared to modify this setting at runtime but did not alter the actual state; three remained enabled, and two remained disabled. Only two apps effectively changed the state of this setting: one enabling and one disabling it.

The discrepancy in the number of apps with this setting disabled between the static and dynamic analyses is primarily due to the fact that some apps could not be successfully executed during the dynamic analysis. However, in two specific apps, the values observed in the static and dynamic analyses did not align. This difference is likely because Frida may not have captured configuration changes that occurred very early in the app's execution. However, our multi-method approach, combining Manifest inspection with real-time monitoring via getters, mitigates this issue. By retrieving the actual runtime value of settings through getters at regular intervals, we ensure that even early programmatic changes are eventually captured, providing a comprehensive view of the app's true configuration.

LimitEventAndDataUsage (Facebook Android SDK). This setting can be modified only via code using a provided setter method. This option is disabled by default (thus not limiting usage of collected data). We did not observe any apps modifying this setting, potentially because Meta's official documentation for developers does not discuss it. Consequently, when retrieving the value of this attribute using its corresponding getter, we observed that all apps had it disabled (set to `False`).

DataProcessingOptions (Audience Network SDK). As detailed in Section 2.2, this setting allows developers to modify data processing to comply with U.S. state privacy regulations by adjusting the Limited Data Usage (LDU) option.

We discovered seven apps that explicitly disabled LDU mode (maintaining the default setting), and only four apps enabled it. Given that 2,897 apps integrate this SDK, the configuration rate for this privacy-preserving option is notably low at 0.14%. It is important to note, however, that this setting is designed specifically for compliance with U.S. regulations, and our experiment was conducted on apps available in Spain. The versions of these apps in other regions, particularly the U.S., could differ in this respect, potentially affecting the observed results.

MixedAudience (Audience Network SDK). We did not observe any apps setting this option. We did observe changes in previous tests we ran when developing our infrastructure. This option can be configured through the Meta Developer Portal and has no getter for us to monitor. Therefore, the lack of evidence does not establish that developers are not assigning a value, but only that they are not doing so via the Manifest or in code.

Summary. Our analysis shows varying levels of developer modification of Facebook SDK privacy settings. While some settings, such as `AutoLogAppEvents`, had their default values overridden in 17.90% of apps, `LimitEventAndDataUsage` was not modified in any app. As shown in Table 3, many apps retained the default configurations, with 88.54% leaving `AutoInit` unchanged. The percentage of apps opting for privacy-enhanced configurations remains low across most settings, such as `AdvertiserIDCollection`, where only 6.79% of apps disabled the default data collection setting.

5.3 Data Transfers

General Traffic Analysis. During the dynamic analysis phase, we successfully intercepted 80,449 unique connections from 4,959 apps, with 3,589 of these apps transmitting a range of user data. Our traffic inspection revealed several key trends in data transmission practices. Device model and AdID were the most frequently transferred types of personal data, with 29,784 and 17,332 transfers, respectively, suggesting a focus on advertising and device-specific

optimizations. Conversely, other data such as email addresses had significantly lower transfer frequencies.

Location data transfer, though less frequent, was notable in apps requiring location-based services. Specifically, coarse device location data appeared in 372 transfers, while precise location data was present in 264 transfers. WiFi-related data, including router identifiers such as BSSID and MAC address, was also documented, indicating that some apps collect detailed network connection information, potentially for geolocation services and network optimization.

Our analysis revealed large disparities in the percentage of apps transmitting user data across different categories. For instance, only 16.66% (54) of the apps in the Educational category transmitted user data. In contrast, 90.32% (93) of the apps in the Shopping category sent user data.

Utilizing the IPInfo service [15, 45], our dynamic analysis geolocated the IP addresses of the servers to which data was transmitted. Most of the traffic, which originated in Spain, was directed to servers located within the same country. The United States, Russia, and Singapore emerged as the second, fourth, and fifth most frequent destinations, respectively, underscoring significant cross-border data flows, particularly to non-EU countries. Data was transmitted to servers in a total of 40 countries, including geographically distant nations such as Oman, Malaysia, South Africa, Taiwan, Japan, and South Korea.

Facebook SDK Traffic Analysis. Cross-referencing Mitmproxy and Frida logs enabled us to extract stack traces from 86.46% of the connections that contained known user data, helping us to pinpoint the responsible libraries. Among these, Facebook SDKs were identified as one of the top sources of off-device personal data transmission, with 917 connections containing personal data across 518 apps, ranking second after Google's libraries and surpassing Unity3d.

Among all data types transferred by Facebook's SDKs, AdID was the most prevalent (54.03%), followed by device model (45.89%) and the WiFi router's BSSID (0.08%), as illustrated in Figure 3. Both the Facebook Core SDK and Audience Network SDK transmitted AdID and device model data, but BSSID transmission was observed exclusively in apps integrating the Facebook Core SDK. While we cannot determine exactly how Facebook uses this data, it could be employed for purposes such as profiling users, delivering personalized ads, and measuring ad performance.

Facebook's SDKs appear to transmit fewer types of data compared to other top SDKs. This difference could be due to our limitations in detecting certain data types within network connections (e.g., navigation and shopping history, which the Audience Network SDK's official documentation suggests are collected). Despite these limitations, the data that is transmitted is consistent with a business model that heavily relies on user data for advertising and analytics.

Geographic analysis of data transfers shows concentration of data flows within the EU, potentially due to GDPR regulations. The majority of the data sent by the Facebook SDKs is directed to servers located in Spain (99.58%)—the location of our test devices—and a smaller fraction goes to Portugal (0.42%). The fact that most connections remained within the country of origin suggests a localized approach to data handling.

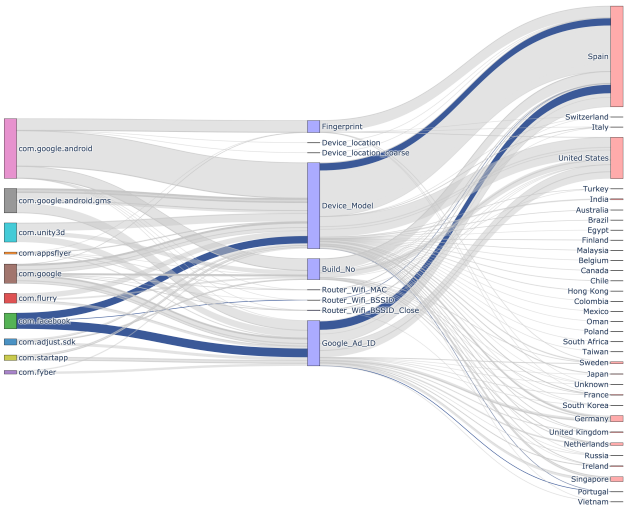


Figure 3: Personal data flows from the top ten third-party libraries to various countries. Facebook SDKs' transfers are highlighted in blue.

5.4 Disclosure Analysis

This section focuses on analyzing compliance by examining the handling of the AdID in apps that integrate the Facebook Core SDK, which regulates the automatic collection of this data through the `AdvertiserIDCollection` privacy setting. Our approach correlates three elements: the `AdvertiserIDCollection` setting in the SDK, the actual transmission of the AdID over the network observed through dynamic analysis, and the disclosures related to device identifiers found in privacy labels and policies. This analysis allows us to identify potential discrepancies and compliance issues specifically related to the collection and transmission of the AdID.

Privacy Label Analysis. We assess the congruence between the declared behaviors in privacy labels and apps' actual data collection practices regarding the AdID. The Google Play Store mandates that apps' privacy labels declare off-device data transmission, including by SDKs: "*This includes user data transmitted off device from your app by libraries and/or SDKs used in your app, irrespective of whether data is transmitted to you or a third-party server*" [40]. Facebook instructs developers to disclose the collection of "*Device or other IDs*" when integrating their SDKs [29].

For apps integrating the Facebook Core SDK, we collected developer choices for the `AdvertiserIDCollection` setting in the app's Manifest file. We also examined evidence of the setting's value from our dynamic analysis. Given those details, we scrutinized privacy labels to identify discrepancies regarding AdID collection practices and to draw insights.

For analytical clarity, we define three distinct sets within our study: L, S, and D, corresponding to labels in Google Play Store, Static analysis, and Dynamic analysis, respectively:

- L comprises privacy labels l_a for each app a, where $l_a = 1$ signifies a label indicating AdID collection, and $l_a = 0$ otherwise. Thus, $L(a)$ represents the label of app a.

- S is the set of apps assessed via static analysis, with a function $M : S \rightarrow \{-1, 0, 1\}$ mapping each app s based on the Manifest's AdID setting: enabled (1), unchanged (0), or disabled (-1). Hence, $M(a)$ represents the AdID setting value of app a in the Manifest.
- D represents apps evaluated through dynamic analysis, with a function $C : D \rightarrow \{0, 1\}$ determining the operational status of AdID collection: enabled (1) or disabled (0). Therefore, $C(a)$ represents the runtime value of AdID collection setting for app a.

Each app a belongs to the set $A = L \cap S \cap D$, as comprising apps that successfully underwent static and dynamic analysis and where the status of the `AdvertiserIDCollection` setting for the Facebook Core SDK was ascertainable.

We define $f_{\text{UCI}}(L(a), M(a), C(a))$ and $f_{\text{ACI}}(L(a), M(a), C(a))$, related to the compliance status of each app a, where $f = 0$ denotes potential non-compliance. For brevity, we omit the explicit reference to each app a in subsequent expressions, using the simplified notation $f_{\text{UCI}}(L, M, C)$ and $f_{\text{ACI}}(L, M, C)$:

(1) **Function $f_{\text{UCI}}(L, M, C)$**: This function flags instances of potentially unaware non-compliance ($f_{\text{UCI}}(L, M, C) = 0$), where a developer leaves the default AdID collection setting, which allows collection, despite not declaring this collection in the privacy label. It is defined as:

$$f_{\text{UCI}}(L, M, C) = \begin{cases} 0 & \text{if } L = 0 \wedge M = 0 \wedge C = 1 \\ 1 & \text{otherwise} \end{cases}$$

This condition may indicate a developer's lack of awareness of the Facebook Android SDK's practices and settings.

(2) **Function $f_{\text{ACI}}(L, M, C)$**: This function identifies cases of non-compliance ($f_{\text{ACI}}(L, M, C) = 0$) where we have evidence of a developer choice to enable AdID collection—explicitly choosing a setting option in the Manifest file—without disclosure in the privacy label. It is defined as:

$$f_{\text{ACI}}(L, M, C) = \begin{cases} 0 & \text{if } L = 0 \wedge M = 1 \wedge C = 1 \\ 1 & \text{otherwise} \end{cases}$$

The proportions of apps exhibiting potentially unaware and aware non-compliance are calculated as:

(1) **Portion of Potentially Unaware Compliance Issues (P_{UCI})**: The fraction of apps where the developer did not explicitly change the default `AdvertiserIDCollection` setting in the Manifest that present a potential compliance issue. $N_{\text{AdID default}}$ represents the total number of apps where we observed no evidence that the setting was changed in the Manifest file or at runtime.

$$P_{\text{UCI}} = \frac{\sum_{a \in A} f_{\text{UCI}}(L, M, C)}{N_{\text{AdID default}}} \tag{1}$$

Among apps that did not alter the default data collection configuration (518 in total), 155 failed to report that AdID was collected. This accounts for 29.25% of the apps in this group (P_{UCI}), suggesting that developers are potentially not compliant, and possibly unaware of it.

(2) **Portion of Potentially Aware Compliance Issues (P_{ACI})**: The fraction of apps where developers explicitly assigned `AdvertiserIDCollection` to `True` in the Manifest that present a potential compliance issue. $N_{\text{AdID explicitly enabled}}$ refers to the total number of apps where the developers explicitly enabled collection.

$$P_{\text{ACI}} = \frac{\sum_{a \in A} f_{\text{ACI}}(L, M, C)}{N_{\text{AdID explicitly enabled}}} \tag{2}$$

For 865 apps, we observed that developers enabled collection of AdID explicitly by setting `AdvertiserIDCollection` to `True` in the app's Manifest, and this matches the setting we observed in our dynamic analysis of the app. Of these apps, 240 did not disclose this collection in their privacy labels. This equates to $P_{\text{ACI}} = 27.75\%$.

These figures highlight a substantial potential compliance issue: 399 out of 1,388 apps (28.75%) have `AdvertiserIDCollection` enabled—whether explicitly set or left as the default—but fail to disclose this in their privacy labels. When considering all 1,693 apps that integrate the Facebook Core SDK, this non-disclosure rate accounts for 23.57% of all apps. Furthermore, 9.16% of all Core SDK-integrating apps and 38.85% of potentially non-compliant apps have compliance issues that may stem from default settings.

Privacy Policy Analysis. We extended our compliance verification to privacy policies, employing the method described in Section 4.4. This method evaluates whether privacy policies disclose the collection of both device IDs and IP addresses together, as these two data types were annotated together in the ground truth dataset used for validation. In contrast, privacy labels treat device IDs and IP addresses as distinct categories, making a direct comparison between privacy policies and labels challenging. Despite these differences in how data types are categorized, our methodology effectively identifies discrepancies in compliance. Specifically, our analysis examined instances where privacy policies did not declare the collection of identifiers or IP addresses while app settings explicitly enabled AdID collection via the Manifest. Our findings reveal that 30 of the 865 apps (3.47%) that enabled the `AdvertiserIDCollection` in the Manifest did not declare this practice in their privacy policies. Moreover, through our network traffic analysis, we observed that half of these apps (15) transmitted the AdID over the network.

Additionally, we investigated the apps where developers did not attempt to configure the `AdvertiserIDCollection` setting. Out of 518 such apps, we successfully analyzed the privacy policies of 374 apps, finding that 28 apps (7.49%) failed to declare AdID collection. Across all 1,388 apps with AdID collection enabled—either by default or explicitly—we analyzed 1,037 privacy policies and identified 58 apps (5.59%) that did not disclose AdID collection. The remaining privacy policies could not be retrieved or were inaccessible via the URLs provided in the Google Play Store.

Labels and Policies Comparison. Due to the differing data types covered under privacy labels and policies, direct comparison is challenging. However, inconsistencies are clear when privacy labels declare the collection of device and other IDs which are absent from the corresponding privacy policies. Our examination identified 445 privacy policies that explicitly stated no collection of device IDs or IP addresses, yet 182 corresponding privacy labels indicated otherwise. Out of the 445 apps whose privacy policies stated no

collection of such data, we observed 120 (26.97%) apps sending the AdID over the network. These discrepancies affirm the findings of previous studies, underscoring persistent misalignments between declared privacy labels, policies [46], and app behavior.

5.5 Children's Apps Analysis

Apps that may be used by children represent a particularly sensitive subset of the app ecosystem. Frameworks such as the Pan European Game Information (PEGI) [60] system in Europe and the Entertainment Software Rating Board (ESRB) [9] in the United States provide widely adopted age-appropriateness ratings for games and apps, which are used by major platforms like the Google Play Store [35].

For apps to be listed in Google Play's family category, the "Google Play Families Policy" [41] sets comprehensive guidelines on app content and data practices, including restrictions on transmitting sensitive personal information. The policy also mandates only advertising SDKs certified under the Families Self-Certified Ads SDK Program [34]. The Audience Network SDK is not certified at the time of writing.

We consider apps that meet all of the following three criteria: 1) have a "PEGI 3" rating (least stringent), 2) are "Teacher Approved" on Google Play, and 3) have committed to the "Play Families Policy." Our evaluation of 73 apps that met these criteria did not uncover issues related to the integration of Facebook SDKs: we did not observe any evaluated apps that integrated the Audience Network SDK, transmitted the AdID or device locations, or had observed inaccuracies in privacy labels. However, we identified four apps that integrated the Facebook Core SDK. Of these, two had the settings `AutoLogAppEventsEnabled` and `AdvertiserIDCollection` enabled, one had these settings disabled, and for one app, this information could not be retrieved through dynamic analysis. While the integration of the Facebook Core SDK itself may not violate the Play Families Policy, enabling these settings could conflict with the data minimization principles mandated by COPPA and GDPR.

We conducted a broader analysis on 779 apps that declared adherence to the Play Families Policy, without considering additional criteria such as PEGI ratings or the Teacher Approved badge. This analysis revealed several potential compliance issues: three apps were found to transmit the AdID, possibly violating policy restrictions on transmitting sensitive data for children or users of unknown age. Of these, one app did not disclose this data collection in its privacy label, and the other two failed to declare it in their privacy policies. Additionally, six apps integrated the Audience Network SDK, which may not be permitted under the Play Families policy. Seventeen apps integrated the Facebook Core SDK, with 11 of these having the `AutoLogAppEventsEnabled` setting enabled and 10 having the `AdvertiserIDCollection` setting enabled.

Previous research has consistently highlighted significant COPPA compliance concerns for potentially child-directed apps. Studies found that over half of the analyzed Android apps targeting children potentially violated COPPA due to data collection practices and the lack of consent mechanisms [50, 65]. Compliance challenges were also observed within Google's family categories, including apps in the "Designed for Families" program, a precursor to the current Play Families Policy where similar privacy violations were identified [84]. A contributor to these violations was the integration of third-party SDKs, some of which were explicitly prohibited in child-directed apps due to their data handling practices [65]. Furthermore, non-compliance was often linked to developers' insufficient knowledge and misconfiguration of privacy settings within SDKs [4, 50], issues that align with the findings of our analysis.

Our findings indicate reduced integration of Facebook SDKs in both the more restrictive subset of 73 apps and the broader set of 779 apps. This suggests potentially improved privacy practices for apps matching the criteria for these groups, particularly regarding the use of third-party SDKs. However, it remains speculative whether this reduction can be attributed solely to policies such as the Play Families Policy. Other factors, including heightened scrutiny, regulatory pressures, and evolving industry standards, likely contribute to these results. Furthermore, significant possible gaps persist in compliance with COPPA and GDPR, particularly in the configuration of privacy settings.

6 Discussion

This section draws on qualitative studies of developers to explore their perspectives on third-party SDK and library integration, highlighting developer motivations, configuration practices, awareness of privacy implications, and challenges faced in managing privacy compliance. Our analysis primarily focuses on the privacy practices of mobile apps and integrated third-party libraries. An understanding of developers' perspectives and challenges provides crucial context for interpreting our findings. We discuss these perspectives and challenges before turning to mitigation approaches.

6.1 Developers' Perspectives and Challenges

SDK Selection. In an ecosystem dominated by free applications, developers often view advertising providers as a "necessary evil" essential for sustaining their business [25]. Research suggests that developers' choice of ad networks, typically integrated into apps via SDKs, is influenced by recommendations from colleagues, information obtained from forums, and trust in large organizations such as Google's AdMob [59]. Additionally, some developers explicitly report relying on major organizations under the assumption that such entities inherently comply with legal standards, further reducing their concerns about privacy risks [4]. These findings align with our observations, as nearly half (47.92%) of the top-downloaded apps integrate Facebook's Audience Network SDK, reflecting a strong preference for established ad networks.

Challenges in SDK Integration. Despite their reliance on these SDKs, developers frequently encounter significant challenges during the integration process. A key frustration stems from the fragmented and inconsistent nature of privacy-related documentation provided by SDK vendors. This documentation is often written in dense legal language, scattered across multiple sources, or presented with inconsistent terminology and formatting, making it difficult for developers to find and correctly implement essential privacy configurations [43, 72]. Our examination of Meta's documentation highlighted this challenge: privacy-related guidance was often scattered and not centralized. Critical information, such as privacy label disclosures for Facebook SDKs in Android apps, was frequently buried in sources like blog posts [29]. This decentralization creates

difficulties for developers in understanding and correctly configuring SDKs, particularly for managing privacy settings and accurately disclosing privacy practices. The *AutoLogAppEvents* setting, which affects whether apps automatically send data to Meta, illustrates this issue: we had difficulty determining the exact types of data transmitted, and uncertainty further complicates developers' efforts to provide clear and accurate privacy disclosures.

Privacy Settings in SDKs. Many developers are unaware of the privacy settings provided by third-party SDKs, leading to incorrect configurations and potential compliance issues [4]. Compounding these challenges, ad networks frequently configure their SDKs with privacy-unfriendly default settings that maximize data collection and targeted advertising [59], as supported by our examination of Facebook's SDKs. These defaults, often specified in documentation and sample code, may implicitly encourage developers, particularly those lacking in-depth privacy knowledge, to adopt configurations that expand data collection, such as personalized ads or broad data-sharing permissions [72]. Developers have expressed a reluctance to modify default settings, which could contribute to practices that may conflict with users' privacy expectations [59], especially given issues surrounding documentation clarity and accessibility.

Privacy Compliance Challenges. The complexity and opacity of third-party SDKs and libraries pose further challenges in terms of privacy compliance. Developers frequently report difficulties in understanding the full extent of data collection practices by these libraries, as their behavior is often unpredictable or insufficiently documented [11, 25, 71]. This lack of transparency complicates developers' efforts to manage privacy settings effectively and comply with legal frameworks such as GDPR, COPPA, and CCPA [4]. The knowledge gap regarding the operation of these SDKs also impacts developers' ability to disclose app behaviors accurately in privacy policies, potentially leading to non-compliance. Smaller development teams or independent developers, in particular, often lack the technical expertise or resources to manage privacy compliance effectively, relying instead on external services, legal templates, or app store guidance [4, 62]. Additionally, the delegation of privacy responsibilities to legal or specialized teams, rather than integrating privacy considerations throughout the development process, further exacerbates compliance challenges [43].

The challenges developers face in complying with privacy requirements when integrating third-party SDKs and libraries are widespread and multifaceted. Studies by Li et al. [54] and Tahaei et al. [73] show that developers perceive privacy compliance as burdensome, offering minimal personal benefit. This leads many developers to adopt a reactive approach, responding primarily to external pressures, such as operating system updates or app store policies, rather than proactively integrating privacy considerations from the outset [54]. This reactive mindset is further complicated by the need to balance functionality with stringent privacy requirements, a struggle that frequently arises when developers draft or update privacy policies [73].

Developers' Approaches to Privacy Management. A significant contributor to this reactive stance is the lack of robust tools and reliable support systems for implementing privacy-preserving measures. Developers often rely on fragmented and informal resources, which exacerbates inconsistencies in compliance efforts [43, 72]. Horstmann et al. [43] emphasize the absence of standardized procedures for verifying privacy implementations, drawing parallels to the more structured guidance found in the security domain. The call for practical, accessible guidelines is echoed by Ekambaranathan et al. [25] and Balebako et al. [11], who argue that current data protection frameworks and SDK documentation are insufficiently clear, leaving developers without the necessary support to make informed decisions.

Moreover, developers' limited engagement in technical testing of SDKs further complicates privacy compliance. Alomar et al. [4] highlight that only a small fraction of developers actively test data collection practices to ensure they align with legal standards, revealing a critical gap in proactive privacy management. This gap reflects a fragmented accountability structure, with developers potentially caught between legal obligations and inadequate tools or guidance. Collectively, these findings underscore a fragmented accountability structure and a significant need for enhanced documentation, comprehensive tools, and better-integrated support systems to empower developers.

6.2 Mitigation Approaches

The prior section discusses research that suggests developer reluctance to change settings, challenges in doing so, and broader compliance difficulties. In combination with that research, our results offer additional evidence that SDK privacy-related settings and their defaults may be contributing to real-world privacy issues. While developers are responsible for their apps, interventions by third-party SDK providers and others may reduce the likelihood of these issues.

One straightforward mitigation strategy is for SDK providers to take a privacy-by-design approach and choose more cautious default privacy-related settings. The trade-offs of this choice depend on the potential privacy harms, benefits of the different settings options to various parties, and developer appreciation of and willingness to change privacy-related settings. While the appropriate choice may depend on the circumstances, we note that arguments that developers can easily switch to more conservative privacy settings might also suggest that developers could easily switch from more conservative defaults to alternatives.

Given the numerous challenges identified in integrating and managing third-party SDKs, addressing gaps in documentation and support mechanisms could also offer meaningful benefits. SDK providers should strive not only to make information related to the privacy and data protection aspects of SDK integration easy to find but also to make critical information hard to miss. Evaluation of the efficacy of SDK provider documentation, guidance, and other support mechanisms with respect to privacy-related settings could suggest further areas for improvement.

Marketplaces also could assist. Beyond implementing privacy compliance mechanisms and checks that address SDK settings risks, marketplaces could mandate that providers consolidate SDK privacy information. The Google Play SDK Index [61] aggregates information about popular SDKs, including versions and required permissions. The index offers a link to each SDK's privacy details, but these links are sometimes missing or outdated. With privacy

manifests [17] in the App Store, Apple offers a more direct solution by mandating that SDKs and apps include detailed files outlining their privacy practices. Unlike more fragile links, Apple proposes integrating privacy manifests directly into the SDK's metadata. This approach provides a centralized and structured format, making it easier for developers to understand what data is collected and shared. The approach also increases the potential for automated compliance assessment.

While Apple's manifests in particular could potentially make privacy information more readily accessible, they do not guarantee full transparency or accuracy. These manifests rely on the SDK providers to self-report data practices. Additionally, neither Apple nor Google's approach fully prevents developers from overlooking critical privacy information, highlighting a need for enhanced warnings or guidance to ensure developers recognize and disclose necessary privacy details in their labels and policies. Future work could empirically evaluate the effectiveness of Apple and Google's approaches in fostering compliance.

7 Limitations

Construct Validity. Our analysis leverages the AndroZoo dataset, which aggregates a comprehensive collection of apps from various sources and is consistently updated over time. AndroZoo collects APK files primarily from the Google Play Store and other third-party marketplaces using automated crawlers. These APKs are selected based on their availability at the time of crawling, without specific criteria for functionality or popularity. This non-selective approach ensures a broad range of apps are archived, but it may not perfectly reflect the current distribution in the Google Play Store. To ensure relevance, we focused on apps with high download numbers, enabling us to analyze privacy practices in apps with significant user bases. While this approach might not capture the full diversity of the Play Store, it allows us to derive meaningful insights from widely used apps.

Internal Validity. Our study confronts challenges to internal validity mainly due to potential obfuscation in the apps analyzed, which could obscure SDK behaviors and impact result accuracy. To mitigate this, we utilized LibScout, known for its resilience to obfuscation. Additionally, our methods based on Frida have been validated against LibScout, demonstrating comparable performance in detecting Facebook SDKs. This consistency increases confidence in our analytical approach and the reliability of our findings.

Our Frida-based methods aim to retrieve values representing privacy-related settings, which can be altered by various means. We conducted a rigorous manual review of the Facebook Core SDK source code to ensure that the variable's value that we retrieve reflects any alterations via the Manifest, code, or the Meta Developers Platform. Although we cannot explicitly inspect changes through the Meta Developers Platform, we can infer them from misalignment between static and dynamic analyses. This gives us greater confidence in these results over simply observing settings in the Manifest file.

Code injection with Frida during dynamic analysis is initiated in spawn mode, ensuring that the app is launched with Frida attached, which helps minimize the possibility of unobserved behavior. While rare instances exist where Frida may not capture a setter at runtime, we mitigate this by cross-referencing multiple sources of information: the initial Manifest configuration, the attribute's value at startup retrieved via getters as available, and continuous retrieval of these values every five seconds during the app's execution. This multi-faceted approach allows us to maintain a comprehensive understanding of the app's privacy settings and minimizes the likelihood of undetected configuration changes.

External Validity. Our study focuses on two popular Facebook SDKs in relatively popular Android apps. Our findings may not generalize to other SDKs, apps, and mobile platforms. Nevertheless, the study yields insights into developer practices when integrating two of the most popular SDKs into apps with many downloads.

8 Conclusions

We conducted a detailed examination of privacy-related settings in Android apps that use two popular Facebook SDKs. Our dynamic analysis, which evaluates SDKs' actual runtime configuration values, exposed discrepancies between settings choices declared statically in a Manifest file and settings values in practice. Relying solely on Manifest analysis would lead to inaccurate estimates of apps modifying privacy-related settings, as key configurations like `AutoLogAppEvents` and `AdvertiserIDCollection` can also be changed through code. Furthermore, the Audience Network SDK's settings and the `LimitEventAndDataUsage` setting cannot be modified via the Manifest, meaning they would go entirely undetected without additional analysis. By accessing getters, we determined actual runtime values, and by capturing setters, we identified changes made during app execution—insights that Manifest analysis alone would miss.

Our findings indicate that developers often fail to accurately reflect SDK-related practices in privacy disclosures—potentially driven in part by default settings—leading to discrepancies between declared and actual practices. One option to address this is for SDK providers to choose more conservative default SDK settings choices. SDK providers should also ensure that documentation, guidance, and tools effectively help developers configure privacy-related settings appropriately. App marketplaces can aggregate SDK privacy details and guidelines, ensuring easier access to accurate information. Additionally, marketplaces could implement privacy compliance mechanisms and checks that address SDK settings risks.

Future work could extend this study by including a broader selection of Android apps and exploring additional popular SDKs. The current study focuses on two Facebook SDKs, and analyzing other SDKs would provide insights not only into the privacy practices and behaviors of the SDKs but also into how developers configure and manage the privacy settings of these SDKs. Expanding to a more diverse set of apps would allow us to explore how different types of apps integrate and utilize SDKs, offering a more comprehensive view of app developer privacy management practices.

Acknowledgments

The work of Jose M. Del Alamo was partially supported by the CEDAR project, funded by the Horizon Europe research program (2021-2027) under grant agreement no. 101135577, and the work of David Rodriguez was partially supported by the PRESECREL project, funded by the Plan Estatal de Investigación Científica y

Técnica y de Innovación 2017-2020 (Ministerio de Ciencia e Investigación (Spain) - MCIN/AEI/10.13039/501100011033) under Grant agreement PID2021124502OB-C43. This research has also been partially supported by the National Science Foundation under its Secure and Trustworthy Computing (SaTC) Program (grant CNS-1914486). The authors would like to thank U.S. Federal Trade Commission staff for feedback regarding this research.

References

[1] A. Akash, S. Chithra, P. Vasuki, T. Shanmughapriya, and N. M. MG. 2022. Towards Privacy for Android Mobile Applications. In *2022 International Conference on Futuristic Technologies (INCOFT)*. IEEE, 1–8.

[2] Ar Akash, S Chithra, P Vasuki, T Shanmughapriya, and Nivas Muthu MG. 2022. Towards Privacy for Android Mobile Applications. In *2022 International Conference on Futuristic Technologies (INCOFT)*. IEEE, 1–8.

[3] Kevin Allix, Tegawendé F. Bissyandé, Jacques Klein, and Yves Le Traon. 2016. AndroZoo: Collecting Millions of Android Apps for the Research Community. In *Proceedings of the 13th International Conference on Mining Software Repositories* (Austin, Texas) *(MSR '16)*. ACM, New York, NY, USA, 468–471. https://doi.org/10.1145/2901739.2903508

[4] Noura Alomar and Serge Egelman. 2022. Developers Say the Darnedest Things: Privacy Compliance Processes Followed by Developers of Child-Directed Apps. , 250–273 pages. https://doi.org/10.56553/popets-2022-0108

[5] Android Developers. 2024. App manifest overview. https://developer.android.com/guide/topics/manifest/manifest-intro. Accessed: 03 October 2024.

[6] Apktool. n.d.. Apktool Official Website. Retrieved May 31, 2024 from https://apktool.org/

[7] AppBrain. 2024. The list of top Ad networks for Android. https://www.appbrain.com/stats/libraries/ad-networks

[8] S. Arzt, S. Rasthofer, C. Fritz, E. Bodden, A. Bartel, J. Klein, and P. McDaniel. 2014. Flowdroid: Precise Context, Flow, Field, Object-Sensitive and Lifecycle-Aware Taint Analysis for Android Apps. *ACM SIGPLAN Notices* 49, 6 (2014), 259–269.

[9] Entertainment Software Association. n.d.. Entertainment Software Rating Board. Retrieved May 31, 2024 from https://www.esrb.org/

[10] M. Backes, S. Bugiel, and E. Derr. 2016. Reliable Third-Party Library Detection in Android and Its Security Applications. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*. 356–367.

[11] Rebecca Balebako, Abigail Marsh, Jialiu Lin, Jason Hong, and Lorrie Faith Cranor. 2014. The Privacy and Security Behaviors of Smartphone App Developers. In *Workshop on Usable Security*. Citeseer, Internet Society, 1–10. https://doi.org/10.14722/usec.2014.23006

[12] Y. Chen, M. Zha, N. Zhang, D. Xu, Q. Zhao, X. Feng, K. Yuan, F. Suya, Y. Tian, K. Chen, X. Wang, and W. Zou. 2019. Demystifying Hidden Privacy Settings in Mobile Apps. In *2019 IEEE Symposium on Security and Privacy (SP)*. 570–586. https://doi.org/10.1109/SP.2019.00054

[13] H. Cheng, G. Hu, J. Liu, Z. Kang, C. Pan, and Z. Zhang. 2022. Detecting Third-Party Libraries for Privacy Leakage in Packed Android Applications. In *2022 China Automation Congress (CAC)*. IEEE, 5053–5058.

[14] Hichang Cho, Sungjong Roh, and Byungho Park. 2019. Of promoting networking and protecting privacy: effects of defaults and regulatory focus on social media users' preference settings. *Computers in Human Behavior* 101 (2019), 1–13.

[15] Miguel Cozar, David Rodriguez, Jose M. Del Alamo, and Danny Guaman. 2022. Reliability of IP Geolocation Services for Assessing the Compliance of International Data Transfers. In *2022 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*. 181–185. https://doi.org/10.1109/EuroSPW55150.2022.00024

[16] H. Cui, G. Meng, Y. Li, Y. Li, Y. Zhang, J. Sun, D. Zhu, and W. Wang. 2022. LibHunter: An Unsupervised Approach for Third-Party Library Detection without Prior Knowledge. In *2022 IEEE Symposium on Computers and Communications (ISCC)*. 1–7.

[17] Apple Developer Documentation. n.d.. Adding Privacy Manifests. Retrieved May 31, 2024 from https://developer.apple.com/documentation/bundleresources/privacy_manifest_files/adding_a_privacy_manifest_to_your_app_or_third-party_sdk

[18] Meta Developer Documentation. n.d.. Audience Network SDK for Android. Retrieved May 30, 2024 from https://developers.facebook.com/docs/audience-network/setting-up/platform-setup/android/add-sdk

[19] Meta Developer Documentation. n.d.. Facebook SDK for Android. Retrieved May 30, 2024 from https://developers.facebook.com/docs/android/

[20] Meta Developer Documentation. n.d.. Meta App Events. Retrieved Retrieved May 30, 2024 from https://developers.facebook.com/docs/app-events/getting-started-app-events-android/

[21] Meta Developer Documentation. n.d.. Meta Audience Network for Android. Retrieved Retrieved May 30, 2024 from https://developers.facebook.com/docs/audience-network/setting-up/platform-setup/android/add-sdk

[22] Meta Developer Documentation. n.d.. Meta Data Processing Options for US Users. Retrieved Retrieved May 30, 2024 from https://developers.facebook.com/docs/audience-network/optimization/best-practices/data-processing-options

[23] Meta Developer Documentation. n.d.. Mixed Audience & COPPA. Retrieved Retrieved May 30, 2024 from https://developers.facebook.com/docs/audience-network/optimization/best-practices/coppa

[24] Meta Developer Documentation. n.d.. Official Documentation. Retrieved Retrieved May 30, 2024 from https://developers.facebook.com/docs/

[25] Anirudh Ekambaranathan, Jun Zhao, and Max Van Kleek. 2021. "Money makes the world go around": Identifying Barriers to Better Privacy in Children's Apps From Developers' Perspectives. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 46, 15 pages. https://doi.org/10.1145/3411764.3445599

[26] Inc. Facebook. 2024. Facebook SDK for Android - Class FacebookSdk. https://developers.facebook.com/docs/reference/android/current/class/FacebookSdk/ Accessed: 2024-09-22.

[27] Á. Feal, J. Gamba, J. Tapiador, P. Wijesekera, J. Reardon, S. Egelman, and N. Vallina-Rodriguez. 2021. Don't Accept Candy from Strangers: An Analysis of Third-Party Mobile SDKs. In *Data Protection and Privacy: Data Protection and Artificial Intelligence*. Vol. 13. 1.

[28] Meta for Developers Blog. 2017. Optimizing and Improving the Android SDK. Retrieved Retrieved May 30, 2024 from https://developers.facebook.com/blog/post/2017/09/26/android-sdk-optimization/

[29] Meta for Developers Blog. n.d.. Resources for Completing App Store Data Practice Questionnaires for Apps That Include the Facebook or Audience Network SDK. Retrieved May 31, 2024 from https://developers.facebook.com/blog/post/2022/07/18/resources-for-completing-app-store-data-practice-questionnaires-apps-facebook-or-audience-network-sdk/

[30] Jack Gardner, Yuanyuan Feng, Kayla Reiman, Zhi Lin, Akshath Jain, and Norman Sadeh. 2022. Helping Mobile Application Developers Create Accurate Privacy Labels. In *2022 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*. 212–230. https://doi.org/10.1109/EuroSPW55150.2022.00028

[31] GDPR 2016. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation). https://eur-lex.europa.eu/eli/reg/2016/679/oj Accessed: 2024-09-22.

[32] GitHub. n.d.. Facebook Android SDK. Retrieved Retrieved May 30, 2024 from https://github.com/facebook/facebook-android-sdk

[33] GitHub. n.d.. Google Play Unofficial Python API. Retrieved May 31, 2024 from https://github.com/marty0678/googleplay-api/

[34] Google Play Console Help. 2024. Participate in the Families Self-Certified Ads SDK Program. https://support.google.com/googleplay/android-developer/answer/12955712 Accessed: 2024-09-22.

[35] Google Play Help. 2024. Apps and Games Content Ratings on Google Play. https://support.google.com/googleplay/answer/6209544 Accessed: 2024-09-22.

[36] D. S. Guamán, D. Rodriguez, J. M. del Alamo, and J. Such. 2023. Automated GDPR Compliance Assessment for Cross-Border Personal Data Transfers in Android Applications. *Computers & Security* 130 (2023), 103262.

[37] X. Hao, D. Ma, and H. Liang. 2022. Detection and Privacy Leakage Analysis of Third-Party Libraries in Android Apps. In *International Conference on Security and Privacy in Communication Systems*. Cham: Springer Nature Switzerland, 569–587.

[38] Y. He, X. Yang, B. Hu, and W. Wang. 2019. Dynamic Privacy Leakage Analysis of Android Third-Party Libraries. *Journal of Information Security and Applications* 46 (2019), 259–270. https://doi.org/10.1016/j.jisa.2019.03.014

[39] Y. He, X. Yang, B. Hu, and W. Wang. 2019. Dynamic Privacy Leakage Analysis of Android Third-Party Libraries. *Journal of Information Security and Applications* 46 (2019), 259–270. https://doi.org/10.1016/j.jisa.2019.03.014

[40] Google Play Console Help. n.d.. Data Safety Section. Retrieved May 31, 2024 from https://support.google.com/googleplay/android-developer/answer/10787469?hl=en

[41] Google Play Console Help. n.d.. Google Play Families Policies. Retrieved May 31, 2024 from https://support.google.com/googleplay/android-developer/answer/9893335?hl=en

[42] Google Play Console Help. n.d.. User Data. Retrieved May 31, 2024 from https://support.google.com/googleplay/android-developer/answer/10144311?visit_id=638525050145726100-2464963387&rd=1

[43] Stefan Albert Horstmann, Samuel Domiks, Marco Gutfleisch, Mindy Tran, Yasemin Acar, Veelasha Moonsamy, and Alena Naiakshina. 2024. "Those things are written by lawyers, and programmers are reading that." Mapping the Communication Gap Between Software Developers and Privacy Experts. , 151–170 pages. https://doi.org/10.56553/popets-2024-0010

[44] H. Inayoshi, S. Kakei, and S. Saito. 2022. Plug and Analyze: Usable Dynamic Taint Tracker for Android Apps. In *2022 IEEE 22nd International Working Conference on Source Code Analysis and Manipulation (SCAM)*. 24–34.

[45] IPinfo. n.d.. IPinfo Official Website. Retrieved May 31, 2024 from https://ipinfo.io/

[46] Akshath Jain, David Rodriguez, Jose M. Del Alamo, and Norman Sadeh. 2023. ATLAS: Automatically Detecting Discrepancies Between Privacy Policies and Privacy Labels. In *2023 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*. 94–107. https://doi.org/10.1109/EuroSPW59978.2023.00016

[47] Q. Jia, L. Zhou, H. Li, R. Yang, S. Du, and H. Zhu. 2019. Who Leaks My Privacy: Towards Automatic and Association Detection with GDPR Compliance. In *Wireless Algorithms, Systems, and Applications: 14th International Conference, WASA 2019, Honolulu, HI, USA, June 24–26, 2019, Proceedings*, Vol. 14. Springer International Publishing, 137–148.

[48] Patrick Gage Kelley, Joanna Bresee, Lorrie Faith Cranor, and Robert W. Reeder. 2009. A "nutrition label" for privacy. In *Proceedings of the 5th Symposium on Usable Privacy and Security* (Mountain View, California, USA) *(SOUPS '09)*. Association for Computing Machinery, New York, NY, USA, Article 4, 12 pages. https://doi.org/10.1145/1572532.1572538

[49] Patrick Gage Kelley, Lorrie Faith Cranor, and Norman Sadeh. 2013. Privacy as part of the app decision-making process. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Paris, France) *(CHI '13)*. Association for Computing Machinery, New York, NY, USA, 3393–3402. https://doi.org/10.1145/2470654.2466466

[50] Konrad Kollnig, Anastasia Shuba, Reuben Binns, Max Van Kleek, and Nigel Shadbolt. 2022. Are iPhones Really Better for Privacy? A Comparative Study of iOS and Android Apps. *Proceedings on Privacy Enhancing Technologies* 2022, 2 (2022), 6–24. https://doi.org/10.2478/POPETS-2022-0033

[51] California State Legislature. 2018. California Consumer Privacy Act of 2018. https://leginfo.legislature.ca.gov/faces/codes_displayText.xhtml?lawCode=CIV&division=3.&title=1.81.5.&part=4.&chapter=55.&article= Accessed: 2024-09-22.

[52] L. Li, A. Bartel, T. F. Bissyandé, J. Klein, Y. Le Traon, S. Arzt, and P. McDaniel. 2015. Iccta: Detecting Inter-Component Privacy Leaks in Android Apps. In *2015 IEEE/ACM 37th IEEE International Conference on Software Engineering*, Vol. 1. 280–291.

[53] M. Li, P. Wang, W. Wang, S. Wang, D. Wu, J. Liu, R. Xue, and W. Huo. 2020. Large-Scale Third-Party Library Detection in Android Markets. *IEEE Transactions on Software Engineering* 46, 9 (2020), 981–1003.

[54] Tianshi Li, Elizabeth Louie, Laura Dabbish, and Jason I. Hong. 2021. How Developers Talk About Personal Data and What It Means for User Privacy: A Case Study of a Developer Forum on Reddit. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW3, Article 220 (jan 2021), 28 pages. https://doi.org/10.1145/3432919

[55] Z. Ma, H. Wang, Y. Guo, and X. Chen. 2016. LibRadar: Fast and Accurate Detection of Third-Party Libraries in Android Apps. In *2016 IEEE/ACM 38th International Conference on Software Engineering Companion (ICSE-C)*. 653–656.

[56] Inc. Meta Platforms. n.d.. Meta Business Suite. Retrieved Retrieved Oct 2, 2024 from https://business.facebook.com/

[57] Inc. Meta Platforms. n.d.. Meta Developers Platform. Retrieved Retrieved May 30, 2024 from https://developers.facebook.com/apps

[58] Inc. Meta Platforms. n.d.. Meta Events Manager. Retrieved Retrieved Oct 2, 2024 from https://www.facebook.com/events_manager2/

[59] Abraham H. Mhaidli, Yixin Zou, and Florian Schaub. 2019. "We Can't Live Without Them!" App Developers' Adoption of Ad Networks and Their Considerations of Consumer Risks. In *Fifteenth Symposium on Usable Privacy and Security (SOUPS 2019)*. USENIX Association, Santa Clara, CA, 225–244. https://www.usenix.org/conference/soups2019/presentation/mhaidli

[60] PEGI. n.d.. Pan European Game Information. Retrieved May 31, 2024 from https://pegi.info/

[61] Google Play. n.d.. Google Play SDK Index. Retrieved May 31, 2024 from https://play.google.com/sdks

[62] Maxwell Prybylo, Sara Haghighi, Sai Teja Peddinti, and Sepideh Ghanavati. 2024. Evaluating Privacy Perceptions, Experience, and Behavior of Software Development Teams. In *Twentieth Symposium on Usable Privacy and Security (SOUPS 2024)*. USENIX Association, Philadelphia, PA, 101–120. https://www.usenix.org/conference/soups2024/presentation/prybylo

[63] Jennifer Pybus and Mark Coté. 2024. Super SDKs: Tracking personal data and platform monopolies in the mobile. *Big Data & Society* 11, 1 (2024), 20539517241231270.

[64] Maven Repository. n.d.. Facebook Core SDK. Retrieved Retrieved May 30, 2024 from https://mvnrepository.com/artifact/com.facebook.android/facebook-core

[65] Irwin Reyes, Primal Wijesekera, Joel Reardon, Amit Elazari Bar On, Abbas Razaghpanah, Narseo Vallina-Rodriguez, and Serge Egelman. 2018. "Won't somebody think of the children?" Examining COPPA compliance at scale. *Proceedings on Privacy Enhancing Technologies* 2018, 3 (April 2018), 63–83. https://doi.org/10.1515/popets-2018-0021

[66] D. Rodriguez, J. M. Del Alamo, C. Fernández-Aller, and N. Sadeh. 2024. Sharing is Not Always Caring: Delving Into Personal Data Transfer Compliance in Android Apps. *IEEE Access* 12 (2024), 5256–5269. https://doi.org/10.1109/ACCESS.2024.3349425

[67] David Rodriguez, Ian Yang, Jose M. Del Alamo, and Norman Sadeh. 2024. Large language models: a new approach for privacy policy analysis at scale. *Computing* 106 (Aug 2024), 3879–3903. https://doi.org/10.1007/s00607-024-01331-9

[68] C. Schindler, M. Atas, T. Strametz, J. Feiner, and R. Hofer. 2022. Privacy Leak Identification in Third-Party Android Libraries. In *2022 Seventh International Conference on Mobile and Secure Services (MobiSecServ)*. IEEE, 1–6. https://doi.org/10.1109/MobiSecServ50855.2022.9727217

[69] J. Schütte, A. Kuechler, and D. Titze. 2017. Practical Application-Level Dynamic Taint Analysis of Android Apps. In *2017 IEEE Trustcom/BigDataSE/ICESS*. 17–24.

[70] J. Seo, D. Kim, D. Cho, I. Shin, and T. Kim. 2016. FLEXDROID: Enforcing In-App Privilege Separation in Android. In *Proceedings of the Network and Distributed System Security Symposium (NDSS)*. Internet Society, San Diego, CA, USA, 1–15. https://doi.org/10.14722/ndss.2016.23485

[71] Mohammad Tahaei, Ruba Abu-Salma, and Awais Rashid. 2023. Stuck in the Permissions With You: Developer & End-User Perspectives on App Permissions & Their Privacy Ramifications. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 168, 24 pages. https://doi.org/10.1145/3544548.3581060

[72] Mohammad Tahaei and Kami Vaniea. 2021. "Developers Are Responsible": What Ad Networks Tell Developers About Privacy. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI EA '21)*. Association for Computing Machinery, New York, NY, USA, Article 253, 11 pages. https://doi.org/10.1145/3411763.3451805

[73] Mohammad Tahaei, Kami Vaniea, and Naomi Saphra. 2020. Understanding Privacy-Related Questions on Stack Overflow. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3313831.3376768

[74] Eeva Terkki, Ashwin Rao, and Sasu Tarkoma. 2016. Investigating User Profiling and Privacy Leaks in Mobile Ad Networks. *Tiny Trans. Comput. Sci.* 4 (2016). https://api.semanticscholar.org/CorpusID:41901100

[75] J. Wang, Y. Xiao, X. Wang, Y. Nan, L. Xing, X. Liao, and Y. Zhang. 2021. Understanding Malicious Cross-Library Data Harvesting on Android. In *30th USENIX Security Symposium (USENIX Security 21)*. 4133–4150.

[76] F. Wei, S. Roy, X. Ou, and Robby. 2018. Amandroid: A Precise and General Inter-Component Data Flow Analysis Framework for Security Vetting of Android Apps. *ACM Transactions on Privacy and Security (TOPS)* 21, 3 (2018), 1–32.

[77] J. Zhan, Q. Zhou, X. Gu, Y. Wang, and Y. Niu. 2017. Splitting Third-Party Libraries' Privileges from Android Apps. In *Information Security and Privacy: 22nd Australasian Conference, ACISP 2017, Auckland, New Zealand, July 3–5, 2017, Proceedings, Part II*. Springer International Publishing, 80–94.

[78] X. Zhan, L. Fan, S. Chen, F. Wu, T. Liu, X. Luo, and Y. Liu. 2021. ATVHunter: Reliable Version Detection of Third-Party Libraries for Vulnerability Identification in Android Applications. In *2021 IEEE/ACM 43rd International Conference on Software Engineering (ICSE)*. 1695–1707.

[79] X. Zhan, L. Fan, T. Liu, S. Chen, L. Li, H. Wang, Y. Xu, X. Luo, and Y. Liu. 2020. Automated Third-Party Library Detection for Android Applications: Are We There Yet?. In *2020 35th IEEE/ACM International Conference on Automated Software Engineering (ASE)*. 919–930.

[80] X. Zhan, T. Liu, Y. Liu, Y. Liu, L. Li, H. Wang, and X. Luo. 2021. A Systematic Assessment on Android Third-Party Library Detection Tools. *IEEE Transactions on Software Engineering* 48, 11 (2021), 4249–4273.

[81] X. Zhang, X. Wang, R. Slavin, T. Breaux, and J. Niu. 2020. How Does Misconfiguration of Analytic Services Compromise Mobile Privacy?. In *Proceedings of the ACM/IEEE 42nd International Conference on Software Engineering*. 1572–1583.

[82] Kaifa Zhao, Xian Zhan, Le Yu, Shiyao Zhou, Hao Zhou, Xiapu Luo, Haoyu Wang, and Yepang Liu. 2023. Demystifying Privacy Policy of Third-Party Libraries in Mobile Apps. In *2023 IEEE/ACM 45th International Conference on Software Engineering (ICSE)*. 1583–1595. https://doi.org/10.1109/ICSE48619.2023.00137

[83] Y. Zhou. 2021. An Automated Pipeline for Privacy Leak Analysis of Android Applications. In *2021 36th IEEE/ACM International Conference on Automated Software Engineering (ASE)*. 1048–1050. https://doi.org/10.1109/ASE51524.2021.9678875

[84] Sebastian Zimmeck, Peter Story, Daniel Smullen, Abhilasha Ravichander, Ziqi Wang, Joel Reidenberg, N. Cameron Russell, and Norman Sadeh. 2019. MAPS: Scaling Privacy Compliance Analysis to a Million Apps. *Proceedings on Privacy Enhancing Technologies* 2019, 3 (2019), 66–86. https://doi.org/10.2478/popets-2019-0037

[85] S. Zimmeck, S. Wang, L. Zou, R. Iyengar, B. Liu, F. Schaub, S. Wilson, N. Sadeh, S. Bellovin, and J. Reidenberg. 2017. Automated Analysis of Privacy Requirements for Mobile Apps. In *24th Network & Distributed System Security Symposium (NDSS 2017)*. 286–296. https://doi.org/10.14722/ndss.2017.23034

Chapter 4

Discussion

In this section, the findings from the studies and publications included in this thesis are brought together and examined, tying them back to the overarching objective: to develop automated methods and artifacts for evaluating the regulatory compliance of mobile applications with the General Data Protection Regulation (GDPR). This objective is pursued through three specific areas: analyzing the behavior of mobile applications, assessing the practices disclosed in privacy policies and labels, and automating regulatory compliance evaluation.

The publications presented in this thesis address key aspects in these areas, offering complementary insights that collectively advance an integrated solution. Ranging from identifying noncompliance patterns in international data transfers to analyzing privacy settings in third-party libraries, each article directly contributes to one or more of the specific objectives. These contributions reinforce the thematic coherence of the thesis and underscore its relevance within regulatory and technical contexts.

The following section describes the platform used to assess a large volume of mobile applications at scale. It then details the methods and artifacts developed to evaluate specific aspects of GDPR Article 13. Following that, the results produced by this platform and its artifacts are presented and interpreted, with a focus on their connection to the research's specific objectives. The section then outlines the academic output stemming from this thesis, along with future lines of research and potential sources of funding. Finally, it offers overall conclusions and future perspectives, highlighting both the individual and collective contributions of the publications, as well as the potential impact of these findings on furthering progress in this field.

4.1 Platform Description

This section describes the modular platform, illustrated in Figure 4.1, which has enabled the production of the results presented in this thesis. Designed as a scalable and flexible infrastructure, the platform is based on Docker containers and employs RabbitMQ as a queue manager to synchronize the various modules. This architecture facilitates the integration of new functionalities, as well as the maintenance and improvement of existing components,

ultimately fostering the development of a more comprehensive solution for addressing the thesis objectives.

The core modules of the platform and its initial design predate this thesis. They were originally conceived and developed as part of undergraduate research projects and a doctoral thesis (Guamán Loachamín, 2022) within the CLIIP project, which focused on privacy engineering. At the onset of this work, the platform was migrated from a Google Cloud-based environment to local servers, reducing operational costs while preserving its functionality and scalability.

Additionally, throughout the development of this thesis, all modules have been maintained and improved. The download and search modules required constant updates to adapt to changes made by Google in the Play Store backend. Moreover, a method for collecting privacy labels, introduced by the marketplace in 2022, was integrated into the download module. The download and storage modules also had to be adapted to support Bundle Apps, a new application format that emerged due to the increasing size of APK files, leading to their segmentation into multiple APK files. The traffic module and the international transfers module were developed as part of the doctoral thesis previously mentioned.

During this thesis, the traffic module was rebuilt to support modern devices such as the Xiaomi Redmi10 and to automatically detect multiple connected devices, enabling parallel analysis. Additional features include support for installing Bundle Apps, halting background processes, identifying connections specific to the target application, and incorporating a new IP geolocation service. The International Transfers module was used along with the updated traffic module to run a large-scale analysis, which led to the publication of the first paper. This thesis contributed to the module with an automated approach that retrieves the traffic module's logs and compares them with the International transfers module logs to perform an automated evaluation of compliance.

The other modules and scripts mentioned throughout this thesis were conceptualized and developed from scratch during the course of this research.

4.1.1 Core Modules

Below is a description of the core modules in the platform, all of which are employed in studies where regulatory compliance is evaluated.

Search Module

The Search module automates the discovery of applications in the Google Play Store through the use of an API (marty0678, 2023). This component enables searches based on customized parameters such as categories, download counts, or keywords. Throughout this thesis, the module played an essential role in generating randomized sets of applications that formed the basis of numerous experiments, ensuring both diversity and representativeness in the analyzed samples.

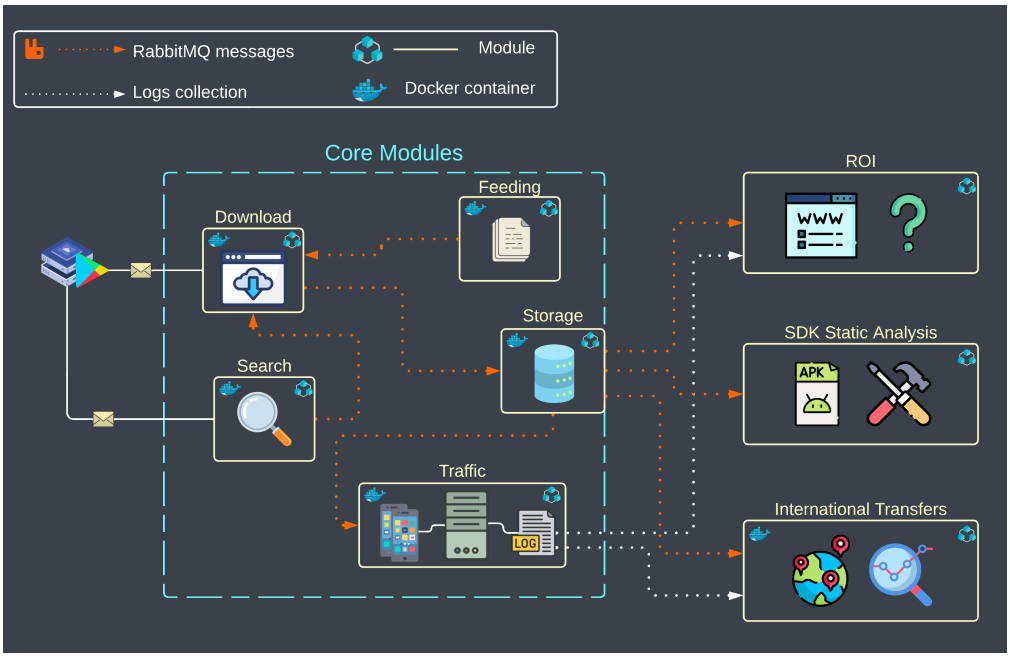


Figure 4.1: Overview of the updated Modular Platform Architecture, maintained and partially developed during the course of this thesis.

Download Module

The Download module manages the acquisition of applications identified by the Search module. Relying on the same API, it simulates connections from real Android devices to obtain the exact versions of APK files later analyzed on physical devices by the Traffic module. In addition, it utilizes Selenium to locate and automatically download the privacy policies associated with each app.

Feeding Module

The Feeding module processes predefined lists of applications for direct download by the Download module. This functionality proved essential in studies requiring the analysis of specific applications or targeted evaluations, streamlining workflows tailored to the needs of each experiment.

Storage Module

The Storage module manages the preservation and retrieval of all data gathered by the Download module, including the APK files themselves. These files are stored on a central server and organized into hierarchical structures for easy retrieval. This module is indispensable for ensuring that the data are accessible to other platform modules—particularly the Traffic module, which needs direct access to the APKs and associated privacy policies during its analysis. Its capacity to handle large volumes of data and to facilitate subsequent usage was necessary for the large-scale studies carried out as part of this thesis.

Traffic Module

The Traffic module, depicted in Figure 4.2, serves as a central component of the platform by capturing and analyzing network traffic from mobile applications running on physical devices. Its implementation and capabilities have been decisive in documenting and evaluating the privacy-related behavior of applications, providing evidence for comparison with declared practices.

The module's workflow begins with a preparation phase in which the physical device is configured, necessary tools (e.g., Mitmproxy and Frida) are verified, and the target APK is retrieved from the Storage module and installed on the device. Permissions required by the application are granted, and background apps are closed to ensure that the analysis focuses exclusively on the target app.

Network interception then proceeds through static and dynamic phases. Static interception collects connections initiated when the application is idle, capturing any background updates or periodic data transfers. Dynamic interception observes the network traffic generated by simulated user interactions, reproduced by the Android Monkey tool. In both phases, Mitmproxy operates as an HTTP proxy, recording all connections and analyzing them to identify IP addresses, domains, data recipients, and any transmitted personal data. These connections are correlated with the specific application under analysis to exclude network activity from other sources.

At the end of the process, the application is uninstalled, and the device's initial configuration is restored. All results—including traffic logs and supplementary analyses—are stored for further examination, creating a clean and reproducible environment for each new test.

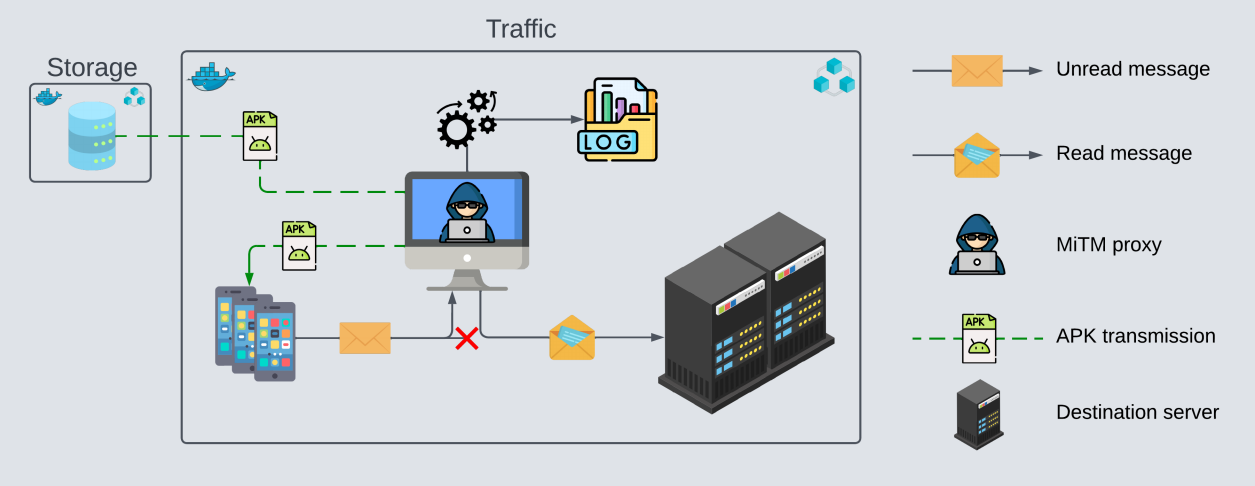


Figure 4.2: Diagram of the Traffic Module Architecture.

4.1.2 Other Modules

In addition to the core modules that form the foundation of the platform, several specialized components were developed to address specific tasks and support advanced analyses. These

complementary modules were designed to extend the platform's capabilities and provide functionalities described in the studies published during this thesis, thereby enabling the automated evaluation of certain GDPR requirements.

A noteworthy module is **ROI (Receiver Organization Identifier)** which automates the identification of organizations that receive personal data. This module integrates multiple sources of information—such as privacy policies, WHOIS queries, and SSL certificate analysis—to accurately determine the entities responsible for the web domains identified by the Traffic module as destinations for personal data transfers. ROI has been pivotal in assessing how applications comply with the GDPR's transparency requirements, particularly those pertaining to the disclosure of third-party data recipients.

Another key component is the **International Transfers module**, aimed at identifying practices related to sending data to countries outside the European Economic Area (EEA). This module examines privacy policies to detect explicit mentions of international data transfers. Although this module was designed and created during a prior doctoral thesis, it was improved to include a method that collects the Traffic module logs and compares them to the international transfers mentioned in privacy policies. Insights derived from this comparison enable the evaluation of an application's compliance with regulatory requirements in this domain and have supported one of the main articles included in this thesis.

Additional complementary modules include the **Static SDK Analysis module**, developed to investigate which SDKs are integrated into applications and how developers configure their privacy settings, based on an inspection of the Manifest file. This module has made it possible to evaluate the correlation between SDK usage and privacy configurations in relation to noncompliance issues, providing a comprehensive view of the low rate at which developers configure SDK privacy settings.

Moreover, various support scripts and tools have been created for specific tasks, including the automated correlation of privacy policy statements with the app behavior observed by the platform. Some of these tools, now integrated into other modules, have enhanced the automation of regulatory evaluations within the mobile ecosystem.

Finally, the platform's capacity to handle large volumes of data has been validated through experiments involving up to 10,000 applications. These experiments not only underscore the platform's resilience but also confirm its scalability for large-scale regulatory analyses.

4.2 Description of Main Contributions

4.2.1 Automated GDPR Compliance Assessment for Cross-Border Personal Data Transfers in Android Applications

Article Objective

The aim of this article is to develop and present a fully automated method for assessing the compliance of Android applications with the GDPR requirements to carry out international transfers of personal data. This work addresses the need for tools that systematically analyze

the alignment of app practices with regulatory standards, identifying any discrepancies that could compromise the transparency and legality of data processing.

Background

The GDPR's evaluable requirements for mobile applications—especially those that can be assessed externally without direct cooperation from the data controller—primarily focus on transparency, a fundamental principle on which Article 13 is based. This article mandates that data controllers must clearly inform users about their data collection and processing practices. This research specifically focuses on the evaluation of compliance with Article 13(1)(f), which states:

"[*the controller shall provide the data subject with:*] *where applicable, the fact that the controller intends to transfer personal data to a third country or international organization and the existence or absence of an adequacy decision by the Commission, or in the case of transfers referred to in Article 46 or 47, or the second subparagraph of Article 49(1), reference to the appropriate or suitable safeguards and the means by which to obtain a copy of them or where they have been made available.*"

Ensuring compliance with this requirement is particularly challenging in the mobile app ecosystem, where international data transfers occur frequently due to the extensive reliance on third-party services, cloud storage providers, and analytics tools. Many apps transfer user data to entities located outside the European Economic Area (EEA), which may be subject to different data protection laws. The GDPR provides mechanisms to ensure that such transfers are lawful, including adequacy decisions, Standard Contractual Clauses (SCCs), and Binding Corporate Rules (BCRs), which should be mentioned in privacy policies in order to comply. This study aims to translate Article 13(1)(f) requirements into programmable compliance rules to assess compliance.

Technical Description of the Developed Artifact

The method integrates several activities that combine the analysis of privacy policies, observation of real app behavior, and the evaluation of regulatory compliance. First, a legal formalization process was conducted in collaboration with data protection experts, transforming the GDPR's transparency obligations for international data transfers into verifiable rules. These rules were adapted to various scenarios, including data transfers to countries with or without adequacy decisions or data collection by non-European controllers. They served as the basis for developing automated methods of verification. Natural language processing techniques were employed to process app privacy policies, using supervised binary classifiers designed to detect key elements such as destination countries, safeguards in place, and instructions for obtaining copies of the measures adopted. These classifiers demonstrated robust performance in the validation phase, achieving F-scores ranging from 85% to 100%.

Second, the apps' technical behavior was evaluated through dynamic analysis that captured network traffic generated on real devices. This relied on the Traffic module, which intercepted and decrypted personal data using a MiTM proxy and Frida. The data collected were subsequently correlated with destination domains using services like IPinfo and webXray to

identify the ultimate recipients of the information.

Finally, a dedicated component integrated and evaluated the evidence generated by the various platform modules. This component consolidated data obtained from the dynamic analysis—reflecting actual app behavior—with the practices concerning international data transfers extracted from privacy policies via the supervised classifiers. It also incorporated technical information from the static analysis of app metadata, including details of the signing certificate's issuer. By correlating these findings, the component determined the applicable assessment scenario for each app based on the rules previously defined in accordance with GDPR requirements. The module then verified whether each app complied with these legal obligations, taking into account the relevant use cases specified in the regulation.

Main Results

The proposed method was applied to a set of 4,593 popular applications downloaded from the Google Play Store in Spain, yielding significant findings. Approximately 32% of the evaluated apps engaged in international transfers of personal data, and among these, nearly 50% exhibited some level of noncompliance with GDPR rules governing international data transfers. The apps were classified into four categories of compliance—complete disclosure, ambiguous disclosure, inconsistent disclosure, and omission of statements. For instance, fewer than 1% of the apps offered information fully aligned with GDPR requirements, whereas over 1,000 apps displayed substantial omissions in their policies.

Another outcome of this research was the creation of the IT-100 corpus, a manually annotated dataset of privacy policy segments that has been made available to the scientific community as an open resource. This corpus was instrumental in training and validating the binary classifiers that examined declared international data transfer practices within the policies, serving as a relevant contribution of this study.

Connection to Thesis Objectives

This article makes a direct contribution to the thesis's overarching and specific objectives by integrating advanced dynamic analysis tools and automated privacy policy processing to assess the GDPR compliance of mobile applications. It achieves the main thesis goal by introducing an automated method that combines the technical analysis of app behavior with the evaluation of practices declared in privacy policies, aligning both with the regulation's transparency and legality requirements.

From the standpoint of the specific objectives, the article addresses several key aims. With regard to technical behavior analysis, the use of MITM proxy and Frida enabled the identification of personal data transfer patterns, including recipients and third parties involved, thus clarifying how data are shared and managed. As for the evaluation of declared practices, the development of supervised classifiers to automatically process privacy policies and detect elements such as destination countries and implemented safeguards met another of the thesis's specific objectives.

Moreover, the formalization of GDPR requirements into verifiable rules was essential for structuring the automated assessment. This element aligns directly with the goal of translating

regulations into programmatically testable criteria. The empirical findings, which highlight a high prevalence of noncompliance and omissions, reinforce the objective of documenting patterns and trends within the mobile ecosystem.

Impact

The article's impact is reflected through various academic metrics and contributions to the field of data privacy. According to Scopus, the work has achieved a Field-Weighted Citation Impact (FWCI) of 4.13, indicating that it has been cited at more than four times the average rate for publications of the same field and type. This places it in the 95th percentile for citation impact, underscoring its considerable influence within the scientific community. The FWCI measure, which accounts for context-specific citation rates, further confirms the article's widespread recognition and above-average impact.

In addition, the article has generated substantial interest in research and outreach platforms, as evidenced by PlumX Metrics. It has amassed 36 captures—mainly readers in academic reference managers—suggesting that it is frequently consulted by researchers and professionals in related areas. There is also a documented media mention, indicating that the findings have resonated beyond purely academic spheres and may potentially influence public or regulatory discussions. Notably, the article was highlighted in *Information Technology Daily* on June 29, 2023, emphasizing its relevance to understanding international data transfers in the current regulatory landscape.

A particularly noteworthy aspect of the article's impact is its citation in a public policy proposal: the document "IICA sociales fundamentales: Una propuesta de reforma CUA," published by the Regional Government of Murcia in March 2024. This reference reflects the direct applicability of the research findings in designing regulatory policies, illustrating how the results can inform and guide normative decisions beyond the academic domain.

Furthermore, the research emerged from an interdisciplinary collaboration among researchers from the Escuela Politécnica Nacional (EPN) in Quito, Ecuador, and King's College London (KCL) in the United Kingdom. This partnership combined technical and legal expertise to tackle the challenges associated with regulatory compliance in mobile applications. Initiated within the scope of this work, the collaboration not only enriched the quality of the research but also laid the groundwork for subsequent joint projects. The academic relationship has led to other relevant scientific contributions, detailed later in this document, highlighting the importance of international alliances in producing high-impact findings.

4.2.2 ROI: A Method for Identifying Organizations Receiving Personal Data

Article Objective

This article introduces ROI (Receiver Organization Identifier), an automated method for identifying the organizations that receive personal data in the context of Android applications. Although ROI is applied here to uncover the recipients of personal data, its broader utility lies in revealing the entity responsible for a given web domain. This work seeks to significantly

enhance current domain-responsibility identification techniques by integrating multiple data sources—such as privacy policies, WHOIS queries, and SSL certificate analysis—achieving over 95% precision. Additionally, it demonstrates the method's applicability in a large-scale study of 10,000 applications, underscoring the lack of transparency in data-sharing practices among a substantial subset of these apps.

Background

The transparency principle of the GDPR ensures that users are informed about who is responsible for processing their data and with whom it is shared. Article 13(1)(a) requires data controllers to disclose their identity and contact details, while Article 13(1)(e) mandates the disclosure of data recipients or categories of recipients. These obligations are particularly relevant in the mobile app ecosystem, where personal data is frequently collected and shared with multiple external entities.

Despite these regulatory requirements, transparency regarding both data controllers and recipients remains highly inconsistent. Privacy policies often omit explicit references to recipients, instead using vague language such as "we may share your data with third-party partners." In addition, apps frequently integrate third-party services (e.g., analytics, advertising SDKs, cloud storage) that collect personal data independently, raising concerns about undisclosed data flows.

This study introduces ROI, an automated method that identifies the entity responsible for a given web domain by integrating multiple data sources. Thus, this method can be used to assess compliance with Article 13(1)(a) when used to process the app's privacy policy or used to assess Article 13(1)(e) when utilized to process the recipient domain's privacy policy.

Technical Description of the Developed Artifact

ROI combines various analytical techniques to accurately identify the organizations that receive personal data. First, a module was developed to parse WHOIS records, substantially improving upon existing tools. This module filters out irrelevant information and pinpoints the entities responsible for a domain with greater precision. In parallel, ROI features a privacy policy analyzer that integrates natural language processing (NLP) and machine learning approaches to ensure accurate identification and extraction of data controllers, who typically coincide with the domain holder. Initially, policies are located through a crawling and scraping process that gathers potentially relevant texts from various sources. Since not all retrieved texts are actual privacy policies, a binary classifier based on Support Vector Machines (SVM) filters out irrelevant documents, ensuring that only valid privacy policies are analyzed. Once the privacy policies are isolated, they are processed using the SpaCy library to identify and extract the data controller specified in each policy.

Although SSL certificate analysis was considered as an additional data source for determining domain controllers, its limited accuracy led to its exclusion. Instead, ROI employs a cascading workflow that hierarchically integrates results from policy analysis and WHOIS queries, thereby enhancing the method's performance by reducing false negatives and maximizing precision in identifying data recipients.

Main Results

ROI was validated through a large-scale study involving 10,000 Android applications, yielding significant insights into the transparency of personal data handling. The method achieved an accuracy of 95.71%, surpassing similar tools such as WebXray, particularly for widely used domains. Of the 40,493 data flows detected from 3,526 apps, only 22% of those apps comprehensively disclosed recipients in their privacy policies, while approximately 61% failed to mention any recipients at all.

The study also highlighted how major technology firms, notably Google and Meta, dominate data reception, accounting for more than 50% of the analyzed flows. Furthermore, the datasets generated during the study—including information on domains, recipient organizations, and types of transferred data—have been released as open resources, facilitating new research endeavors in this field.

Connection to Thesis Objectives

The development of ROI aligns directly with the core and specific objectives of this thesis, making a substantial contribution to advancing automated regulatory compliance assessments for mobile applications. Within the area of technical behavior analysis, ROI precisely identifies the ultimate recipients of personal data transfers, a crucial step in determining how apps interact with third parties and in uncovering any undeclared data-sharing practices that might fall short of GDPR transparency requirements.

As for the practices disclosed in privacy policies, ROI automates the extraction and analysis of relevant information, enabling direct comparisons between stated practices and observed behavior. This functionality is vital for further automating compliance assessments, specifically with regard to data-sharing transparency.

In addition, the article highlights significant transparency gaps in the mobile ecosystem by identifying patterns that call for urgent attention from both developers and regulators. Hence, ROI not only contributes to documenting these inconsistencies but also reinforces the thesis's overarching aim by serving as a valuable tool for future advancements in automated compliance evaluations.

Impact

The article's impact is evident in both academic metrics and practical influence, as well as in its potential to spur additional research. According to Scopus, it has achieved a Field-Weighted Citation Impact (FWCI) of 1.98, indicating nearly double the expected citation rate for publications of similar scope. This measure underscores the work's importance within its disciplinary context, emphasizing its contribution to automating privacy evaluations.

Positioned in the 85th percentile of citations in Scopus, the article has garnered substantial recognition from the scientific community, particularly on a specialized topic such as data privacy. PlumX metrics reinforce this assessment, noting 10 captures in academic reference managers, reflecting initial interest among researchers in related domains. The article also includes an additional citation in academic indexes, suggesting its emerging role in broader

discussions on regulation and compliance. While these figures point to an early stage of impact, they nevertheless highlight the study's promise to shape the scientific literature.

Beyond raw citation counts, the article has sparked practical interest outside the immediate academic sphere. For instance, a graduate student at the University of Warwick contacted the authors to base her master's thesis on this research, exploring how to integrate new LLM-based techniques into ROI. This example underlines the method's relevance for future studies and its adaptability to evolving technologies.

Additionally, the intellectual property of the tool developed in this research has been officially registered. The "Herramienta para la identificación de la organización responsable de un dominio web" was granted a favorable legal qualification by the Registro Territorial de la Propiedad Intelectual de la Comunidad de Madrid under the registration number M-007861/2022, further reinforcing the work's originality and practical impact.

Although the total number of citations remains limited due to the relatively short time since publication, these metrics attest to its growing influence and capacity to lay a reliable groundwork for further research.

4.2.3 Comparing Privacy Label Disclosures of Apps Published in Both the App Store and Google Play Stores

Article Objective

This article aims to conduct a systematic comparison of the privacy statements provided in the privacy labels of mobile applications published on both the iOS App Store and the Google Play Store for Android. Specifically, the study explores discrepancies in privacy labels for the same apps across both platforms and examines potential reasons behind these differences. Through this analysis, the article identifies possible regulatory compliance issues related to transparency in data collection and handling, especially regarding sensitive data, location information, and other categories.

Background

Privacy policies have traditionally served as the primary mechanism for informing users about an application's data collection, processing, and sharing practices. These documents typically outline aspects such as the types of data collected, the purpose of processing, and whether data are shared with third parties. However, privacy policies are often lengthy, complex, and written in legal jargon, making them difficult for users to read and understand.

Privacy labels emerged as a simplified alternative, modeled after nutrition labels in the food industry. The goal was to enhance transparency by providing users with structured, easy-to-read summaries of an app's privacy practices. Apple introduced privacy labels in December 2020 as a mandatory feature in its App Store, requiring developers to disclose key information about data collection and tracking practices. Google followed suit in April 2022 by launching the Data Safety section in the Google Play Store, implementing a similar concept to inform Android users about an app's data handling practices.

These labels categorize information into structured sections, typically including (see Figure 4.3):

- Types of data collected (e.g., location, contact information, browsing history).
- Purpose of data collection (e.g., analytics, personalization, advertising).
- Data sharing practices (whether data are shared with third parties and for what purposes).
- Security and compliance (whether encryption is used and if users can request data deletion).

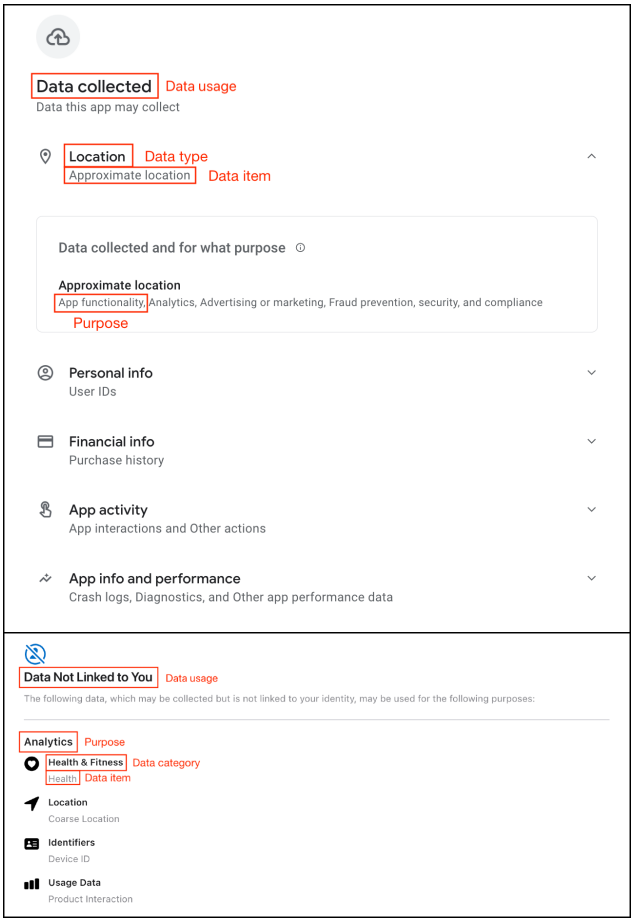


Figure 4.3: Example of Android (top) and iOS (bottom) privacy labels.

Technical Description of the Developed Artifact

The proposed approach employs a systematic method to analyze privacy labels and compare them to the actual behavior of the applications. Initially, 822 apps with available versions on both platforms were selected, and automated scraping techniques were employed to collect information. To ensure the accuracy of the selection process, a method based on Structural Similarity Index (SSIM) analysis of logos was implemented.

After the apps were selected, their privacy labels were gathered. In the case of iOS, the labels were retrieved using tools such as Selenium and BeautifulSoup, owing to the dynamic nature of the hosting web pages. On Android, the privacy labels were extracted directly from the "Data Safety" section using scraping techniques. Once the labels were collected, they underwent a conceptual mapping process to align the categories used by both platforms, facilitating the identification of any discrepancies between the respective declarations.

In addition, a static analysis of the code of 560 Android applications was performed to compare the privacy labels with the apps' actual behavior. Using JADX to decompile the APK files, the study investigated requested permissions, API calls, and other practices linked to personal data collection. This analysis revealed cases in which the labels did not accurately reflect the behavior observed in the source code.

Main Results

The study uncovered significant inconsistencies in the privacy labels of applications published on both platforms. Overall, 66.5% of the reviewed apps displayed substantial discrepancies between the statements provided on iOS and Android. Moreover, only 3.2% of apps that acknowledged data collection on both platforms offered fully consistent declarations. One notable finding was that 44.3% of the Android apps analyzed did not appropriately disclose the collection of location data in their labels.

Furthermore, the static analysis showed that a considerable number of apps accessed precise or approximate location data without acknowledging this behavior in their labels. In particular, 54 apps were found to gather such data without indicating this practice in their privacy labels, underscoring a lack of transparency in how developers or app owners handle data.

Connection to Thesis Objectives

This article directly aligns with the thesis's main and specific objectives, offering a novel assessment that focuses on privacy labels—an approach more recent than traditional privacy policies and specifically designed to enhance user transparency. The study analyzes how these labels, introduced by app distribution platforms, differ in their representation of data practices across two major ecosystems: the Google Play Store and the Apple App Store.

The research addresses the objective of identifying discrepancies between privacy statements and the technical behavior of apps, demonstrating considerable mismatches in the data reported through labels for the same apps on the two platforms. These inconsistencies pertain not only to variations in the amount and quality of information provided between platforms but also to divergences between declared and observed practices through technical analysis, thus contributing to the objective of comparing stated policies with real-world behavior.

Additionally, this examination contributes to the thesis's proposed compliance assessment methods by investigating a different transparency mechanism—the privacy label—and evaluating its reliability and accuracy compared to actual data-management practices. The methodology applied here and the patterns it documents—such as the lack of consistency in how collected data is disclosed—further support the objective of identifying and documenting gaps in transparency and regulatory compliance within the mobile ecosystem.

Impact

This article has made a significant impact within both academic and professional circles, as demonstrated by Scopus and PlumX metrics. With a total of seven citations on Scopus, it ranks in the 97th percentile, underlining its standing as a widely recognized contribution to the field of privacy and transparency in mobile applications. Its Field-Weighted Citation Impact (FWCI) of 6.09 is particularly noteworthy, indicating that the article has been cited over six times the average for comparable publications in this domain, a figure that highlights its importance and appeal within the scientific community.

In addition, PlumX metrics show that the work has drawn attention from eight readers on academic reference platforms, suggesting sustained interest from researchers and students investigating privacy labels and cross-ecosystem comparisons of mobile app distribution. Although the number of captures remains modest, these metrics are even more remarkable given the relatively short period since the article's publication, highlighting its ability to generate interest and achieve recognition in a short timeframe.

4.2.4 Sharing Is Not Always Caring: Delving into Personal Data Transfer Compliance in Android Apps

Article Objective

The primary goal of this article is to investigate how thoroughly Android applications comply with the GDPR's transparency requirements regarding personal data transfers to third parties. More specifically, it evaluates whether mobile apps adequately disclose the recipients of personal data, identifies data transfer practices, and explores the role of third-party libraries (SDKs) in potential transparency shortfalls. By examining a prevalent issue in the mobile ecosystem—the opacity surrounding data-handling practices—this research underscores the importance of rigorous regulatory adherence and proposes an automated method for compliance evaluation.

Background

The GDPR establishes transparency as a fundamental principle in the processing of personal data, ensuring that users are clearly informed about how their data is collected, processed, and shared. In the context of mobile applications, Article 13(1)(e) plays a particularly relevant role, as it mandates that controllers provide users with information about the recipients or categories of recipients of personal data, if applicable. This requirement aims to give individuals a clear understanding of who has access to their personal data and under what conditions, thereby fostering informed decision-making.

Ensuring compliance with this transparency requirement presents significant challenges in the mobile ecosystem, where applications frequently integrate third-party libraries (SDKs) to enhance functionality, such as analytics, advertising, authentication, or geolocation services. These SDKs frequently handle user data independently of the app developer's direct control, leading to a complex data-sharing landscape. Developers may unknowingly transfer user data to third-party entities, raising concerns about undisclosed data flows and the extent to which privacy policies accurately reflect real-world practices.

In this study, we present a new approach to map the personal data transfers sent by Android apps with the responsible SDK, gaining a better understanding of their data practices. Moreover, we present a method to assess apps' compliance with Article 13(1)(e).

Technical Description of the Developed Artifact

This study introduces multiple individual components—Recipient Analyzer, Disclosure Checker, and Library Analyzer—that jointly enable the assessment of GDPR compliance as it pertains to data sharing with third parties.

1. **Recipient Analyzer**: This module identifies the organizations receiving personal data and determines whether they correspond to the app's data controller or to a third party. It leverages advanced natural language processing techniques to extract relevant information from privacy policies and link it to the domains observed in data transfers. Internally, it relies on ROI to perform part of its functions.
2. **Disclosure Checker**: This component examines whether entities receiving personal data are adequately disclosed in the app privacy policy, as required by the GDPR's transparency principles. It automates the process of comparing observed data flows to the statements made in the policies, providing concrete evidence of potential violations.
3. **Library Analyzer**: Focusing on the software libraries responsible for personal data transfers, this module employs dynamic analysis tools like Frida, in tandem with a MiTM proxy, to monitor network connections and link execution traces to specific libraries. This examination is crucial for attributing transparency issues to third-party libraries, offering additional insight into data-transfer practices.

Each artifact's performance was validated individually. The artifacts were then combined in an experiment involving the analysis of 9,000 apps, demonstrating their capacity to detect large-scale discrepancies between observed practices and declared policies, as well as to quantify how third-party libraries contribute to these inconsistencies.

Main Results

Analysis of the 9,000 applications revealed concerning findings that underscore significant shortcomings in Android's regulatory compliance ecosystem. Over 81% of these apps fail to adequately disclose their personal data recipients in their privacy policies, falling short of the GDPR's transparency mandates. This result highlights a widespread lack of clarity in how data transfers are managed across the mobile environment.

During the analysis, a total of 206 distinct personal data recipients were identified, with Google, Meta, and Unity accounting for more than 70% of the transfers observed. Additionally, over 73% of undisclosed transfers were linked to third-party libraries. In many instances, these libraries handle data such as device IDs or location details without developers' full awareness, posing both technical and legal challenges for developers and regulators working to enforce GDPR compliance.

The study further identified a small yet concerning percentage of applications that transmit personal data via unencrypted connections, leaving user information exposed to major security

risks. This highlights the urgent need to implement minimum security standards in all data transfers.

Lastly, the article examined the role of privacy labels as a complementary disclosure mechanism. Although 54.48% of applications acknowledged data transfers in these labels, the analysis found that labels alone are insufficient to fulfill GDPR transparency requirements, as they do not clearly and precisely identify the data recipients.

Connection to Thesis Objectives

This article closely aligns with both the general and specific objectives of the thesis, contributing to the development and validation of advanced methods for automated regulatory compliance assessments in mobile applications. First, the work addresses the technical analysis of app behavior by identifying patterns of personal data transfers and the relevant recipients. Implemented via the Traffic module and interception tools, this technical approach documents data-sharing practices and characterizes the third-party libraries involved, advancing the objective of analyzing and recording practices that influence transparency and regulatory compliance.

Second, the article performs a detailed review of the practices disclosed in privacy policies, comparing them to the technical analysis results. By pinpointing discrepancies between stated policies and actual transfer practices, it furthers the goal of evaluating privacy policies and labels and highlights transparency issues that affect GDPR compliance.

Finally, this research culminates in an automated assessment of regulatory compliance, grounded in the GDPR's transparency principles, by systematically comparing declared practices to observed behavior. Demonstrating that third-party libraries are responsible for a substantial majority of undisclosed transfers, the article delivers critical insight into automating compliance evaluations. Moreover, the large-scale validation—analyzing 9,000 applications—illustrates the method's practical impact, showcasing its scalability and applicability to regulatory contexts.

Impact

This article has had a significant influence both within the academic community and in public outreach, reflecting its importance in the realm of privacy and data protection. According to Scopus, it has accumulated two citations, placing it in the 75th percentile for its field, with a Field-Weighted Citation Impact (FWCI) of 0.95. Although preliminary, these metrics point to emerging scholarly interest in the work. Additionally, PlumX Metrics records 11 captures, primarily from researchers who focus on privacy and regulatory topics.

From a media standpoint, the study has garnered attention in high-profile outlets, underscoring its broad relevance. Its results have been reported by *La Vanguardia*, a leading Spanish newspaper, and by *Computer Hoy*, Spain's most widely read technology magazine. Coverage in *Noticias de la Ciencia* and *Suwedi* further extended its reach among scientific and technological audiences. The findings were also featured prominently on "La Linterna" (a major program on the COPE radio network) and mentioned on TreceTV, demonstrating the general public's interest in these issues.

Moreover, the article was shared on *The Conversation*, a portal aimed at disseminating academic research in accessible language to the broader public, thereby extending its visibility beyond strictly specialized circles. This widespread media coverage emphasizes the article's practical significance and highlights the necessity of discussing transparency in the mobile ecosystem, reflecting a growing interest in the subject.

4.2.5 Large Language Models: A New Approach for Privacy Policy Analysis at Scale

Article Objective

The primary aim of this article is to investigate the capacity of large language models (LLMs), such as ChatGPT and Llama 2, to analyze and extract declared privacy practices from privacy policies. This work seeks to address the limitations of traditional NLP approaches to policy analysis—particularly their reliance on large manually annotated datasets for model training, which entail high costs and significant risks of errors. While LLMs may still require ground truth datasets for validation, they eliminate the need for extensive annotation efforts in training phases and simplify technical implementation by removing the need for model selection, hyperparameter tuning, and dedicated training processes. The study proposes and evaluates optimized LLM configurations for automating this analysis, demonstrating how they can surpass earlier methods in terms of accuracy, cost-effectiveness, and ease of implementation.

Background

LLMs have emerged as a transformative technology in natural language processing (NLP), offering advanced capabilities in text comprehension, information extraction, and content generation. These models, trained on vast datasets, can analyze and process complex documents, making them highly suitable for applications requiring text interpretation—such as legal and regulatory analysis. Their potential to automate privacy policy analysis represents a paradigm shift from traditional NLP approaches, which rely heavily on manually annotated datasets and predefined rule-based methods, both of which are costly and labor-intensive.

Historically, privacy policies have served as the primary mechanism for informing users about how their personal data is processed. However, these documents are often long, ambiguous, and difficult to understand, making automated processing particularly challenging. The introduction of LLMs, such as OpenAI's GPTs and Meta's Llama 2, provides a new opportunity to streamline the analysis of privacy policies. These models can extract and categorize privacy practices with minimal human intervention, potentially surpassing traditional methods in accuracy, efficiency, and adaptability. Unlike earlier NLP techniques, LLMs require little to no fine-tuning for many tasks, making them notably advantageous for assessing several GDPR requirements.

Technical Description of the Developed Artifact

The developed artifact consists of an optimized LLM configuration, specifically using ChatGPT, that integrates advanced techniques in prompt design, parameter tuning, and few-shot learning. Its design includes:

- **Prompt Segmentation and Design**: An iterative experimental design was employed to divide the prompts into three segments (data, task, and output format), enabling the models to effectively interpret declared privacy practices.
- **Parameter Optimization**: Settings such as temperature (fixed at zero to ensure determinism) and top-p were tuned to maximize result accuracy and consistency.
- **Validation with Reference Datasets**: The model was validated against the MAPP and OPP-115 datasets, which contain privacy policies annotated by legal experts. The final configuration demonstrated superior performance on key metrics such as the F1-score, outperforming traditional statistical and rule-based methods.
- **Generalization Capabilities**: The model was tested on identifying international data transfers and other complex practices, demonstrating a notable ability to manage tasks beyond those initially evaluated.

Main Results

Comparative analysis revealed ChatGPT to be the most effective tool for automated privacy policy processing, outperforming models such as Llama 2. In particular, it achieved F1-scores of 93.5% on the MAPP dataset and 93.0% on OPP-115, reflecting its ability to accurately and consistently identify and categorize privacy practices in various contexts.

The use of an optimized prompt enabled ChatGPT to maintain excellent performance across different datasets and tasks without requiring specialized fine-tuning or additional training. This significantly simplifies its deployment, lowering costs and improving accessibility. Its efficiency is exemplified by its capacity to process up to 45 policies per minute using GPT-4 Turbo, thus offering a fast and scalable solution for regulatory analysis.

Beyond accuracy and speed, LLMs stand out for their ability to adapt to different privacy practices and regulatory frameworks. This flexibility makes them useful for both mobile app compliance assessments and broader legal analysis. By simplifying privacy evaluations and making them more scalable, this technology can improve transparency and support regulators and developers in meeting compliance requirements.

Connection to Thesis Objectives

By introducing LLMs such as ChatGPT and Llama 2 for automated privacy-policy analysis, this article proposes and validates a paradigm shift. Although it does not directly address compliance assessment, it significantly contributes to the overall aims of this thesis by introducing an innovative tool for extracting declared practices from privacy policies—one of the fundamental steps in evaluating transparency within the mobile ecosystem.

The development and validation of optimized LLM configurations align with the specific thesis

objective of "evaluating declared practices in privacy policies." This approach demonstrates how LLMs can outperform traditional methods, offering a more scalable, accurate, and cost-effective solution for identifying complex practices, such as data collection.

Although the article does not explicitly compare GDPR requirements, the ability of LLMs to extract structured information from privacy policies opens the door to future research on compliance assessment, particularly in detecting inconsistencies between stated and actual practices. This approach also serves as a foundation for the methods developed later in this thesis, expanding its scope.

Impact

This article has made a notable impact in both academic and professional circles, as reflected in its citation metrics and the interest generated by its methodological proposal. According to Scopus, the article received one citation within three months of publication, placing it in the 76th percentile within its field. With a Field-Weighted Citation Impact (FWCI) of 0.99, it aligns with the average expected for similar publications; however, these figures should be interpreted in light of the relatively short time since its release, suggesting potential for more significant influence over the longer term.

Beyond quantitative metrics, the study has garnered broad academic and interdisciplinary interest. Three international researchers have expressed interest in the proposed method, recognizing its applicability and relevance to automated analysis in data protection. This enthusiasm has spawned new collaborations, including a joint investigation in progress with a researcher from ETH Zurich that applies the prompt design and methodology presented here to a legal context, extending its applicability beyond purely technical domains. Such multidisciplinary collaboration underscores both the versatility and broader usefulness of this approach in various data privacy fields.

The article has also laid the groundwork for further research, including evaluations of the method's multilingual capabilities in privacy policy analysis, expanding its potential application to different official languages of the European Union. Moreover, it has inspired at least three final-degree projects—in Telecommunications Engineering, Computer Engineering, and the Master's Program in Telecommunication Systems—reinforcing its impact on academic training and advancing knowledge.

Finally, the methodology developed in this article has been instrumental in the automated processing of privacy policies described in two subsequent articles of this thesis. Therefore, its impact goes beyond this work, serving as a foundation for further research on transparency and personal data protection.

4.2.6 Data Retention Disclosures in the Google Play Store: Opacity Remains the Norm

Article Objective

This article's primary aim is to assess the transparency and regulatory compliance of Android applications' privacy policies regarding personal data retention, as stipulated by Article

13(2)(b) of the GDPR. The study examines whether these policies offer clear information on data retention periods or on the criteria used to determine them, identifying any practices that are opaque or non-compliant. Additionally, the work underscores the need to improve data retention disclosure practices and proposes automated solutions based on advanced language models to address these shortcomings.

Background

The GDPR establishes strict transparency obligations for organizations processing personal data, requiring them to clearly disclose their data-handling practices. Among these obligations, Article 13(2)(a) specifically mandates that data controllers *inform users about how long their personal data will be stored or, if that is not possible, the criteria used to determine the retention period*. This requirement ensures that individuals can make informed decisions about engaging with digital services, knowing how long their data will be kept and under what conditions it might be deleted.

Organizations must determine data retention periods based on several factors, ensuring that personal data is not stored longer than necessary for the purposes for which it was collected. The GDPR does not prescribe specific time limits but requires controllers to justify their retention decisions. The most common criteria used to define retention periods include:

- Legal and Regulatory Requirements. Many sectors are governed by laws that mandate minimum retention periods for specific types of data. For example, financial and tax regulations often require organizations to retain transactional data for several years to comply with auditing and anti-fraud measures.
- Contractual Obligations. Data retention may also be governed by contracts between businesses and customers, employees, or partners. Service providers, for example, may need to store user data for a period outlined in terms of service agreements or warranty periods.
- Business Needs and Operational Purposes. Many companies justify long retention periods based on business continuity, fraud prevention, analytics, and service improvements.
- Security and Fraud Prevention. Retention periods can be defined based on cybersecurity and risk management needs.

Technical Description of the Developed Artifact

The proposed method relies on the use of LLMs—specifically ChatGPT—to automatically analyze and classify the data-retention statements within privacy policies. A ground-truth dataset was constructed, drawing from the OPP-115 corpus, which was manually re-annotated in collaboration with a legal expert specializing in data protection. The annotated dataset categorizes policies based on their level of GDPR compliance, ranging from those that do not specify retention periods to those that provide explicit information or clear criteria.

The method employs few-shot learning to guide GPT-4 Turbo in classifying privacy policies into six primary categories, from explicit declarations of retention periods to undefined or opaque practices. In addition, the system is designed to handle policies that describe multiple

retention periods, generating structured outputs in Python list format, which facilitates integration into large-scale automated workflows. The model validation, employing metrics such as precision, recall, and F1-score, achieved high performance, including an F1-score of 0.904 in identifying compliant cases.

To apply this method, a dynamic analysis platform was used to capture network traffic generated by applications running on physical devices, determining whether they process personal data subject to GDPR requirements. The associated privacy policies were then extracted and analyzed using the proposed method.

Main Results

The analysis covered 2,235 privacy policies from Android apps that were observed sending personal data. More than half of these policies were classified as potentially non-compliant due to either the absence of data retention periods or insufficient information about them. Among the remaining policies, 34.18% specified multiple retention periods, adding further complexity to evaluating their transparency.

The study also revealed a correlation between app popularity and the level of compliance: more frequently downloaded and highly rated apps tended to include policies more closely aligned with GDPR requirements. However, the same correlation was not statistically significant when examining user ratings of the apps in the Google Play Store.

Connection to Thesis Objectives

This article ties directly into the thesis's general and specific objectives by addressing a central aspect of regulatory compliance: transparency in personal data retention and its proper disclosure in privacy policies. The method integrates technical analysis with normative rule formalization, thus contributing to the thesis goal of translating regulatory mandates into programmable and verifiable artifacts.

Moreover, the work broadens the thesis's scope by incorporating a complementary approach to evaluating data-transfer practices and privacy-policy disclosures. By highlighting prevalent shortcomings—such as policy opacity and a lack of clear transparency criteria—the article underscores the importance of developing automated and scalable methods. This aligns with the thesis's aim of fostering a more transparent mobile ecosystem that adheres to current privacy regulations.

Impact

Although bibliometric metrics are not yet available due to the article's recent publication, its presentation at an academic conference garnered significant interest from both academic and industrial stakeholders. A member of HERE Europe B.V., a leading provider of navigation solutions, kindly asked for an evaluation of their application in the Google Play Store using the techniques introduced in this study. This practical application validates the proposed methodologies in a real-world setting and highlights their potential for promoting transparency and regulatory compliance in the mobile apps domain.

Furthermore, the Communications Office at the Universidad Politécnica de Madrid (UPM) contacted the authors to explore disseminating the findings in mainstream media. This interest underscores the work's broader relevance, demonstrating its significance not only to the academic community but also for knowledge transfer beyond specialized circles.

As part of its commitment to open research and transparency, the annotated dataset and method results have been publicly released. This allows researchers, regulators, and industry practitioners to replicate the study, validate the findings, and advance automated compliance assessments.

4.2.7 Privacy Settings of Third-Party Libraries in Android Apps: A Study of Facebook SDKs

Article Objective

This article aims to examine how mobile app developers configure privacy-related settings in the Facebook Android SDK and the Facebook Audience Network SDK. Specifically, it explores how default configurations influence both application privacy and regulatory compliance under frameworks such as the GDPR and the California Consumer Privacy Act (CCPA). In addition, the study addresses discrepancies between the practices observed in the app's behavior and the declarations made in privacy labels and policies, offering a critical perspective on how default third-party library settings impact user privacy.

Background

In the context of mobile applications, privacy-related behavior of Android apps is often dictated not solely by app developers but also by the third-party libraries (SDKs) that apps integrate, which provide core functionalities such as analytics, advertising, authentication, and location tracking. These libraries, developed by external providers, frequently come with default privacy settings that govern how user data is collected, processed, and shared, significantly shaping the privacy implications of the apps that integrate them.

One of the most widely used SDKs in mobile applications is the Android Facebook SDK, which enables developers to integrate features such as social login, analytics, and targeted advertising into their apps. Additionally, the Facebook Audience Network SDK facilitates monetization through personalized advertising, leveraging data collection mechanisms that track user interactions across multiple applications. The default configurations of these SDKs influence data transmission, user tracking, and consent mechanisms, often without developers explicitly modifying them.

From a regulatory standpoint, default privacy settings raise significant concerns under data protection laws such as the GDPR and the CCPA. The GDPR, in particular, establishes the principle of "privacy by default" under Article 25, requiring that systems and applications be designed to minimize data collection, processing, and sharing by default unless the user explicitly consents to additional processing. However, many SDKs do not adhere to this principle, and their default configurations may enable extensive data collection that developers may not fully understand or actively disclose in privacy policies or privacy labels.

This study seeks to analyze the prevalence of Facebook SDKs in Android applications, investigating how frequently they are integrated, what their privacy settings are, and what default values they use. Furthermore, it examines how often developers modify these settings and the impact of these configurations on regulatory compliance and transparency.

Technical Description of the Developed Artifact

The study implements a comprehensive approach to analyzing SDK usage in Android applications, combining both static and dynamic analysis techniques. Static analysis relies on LibScout to detect the presence and specific versions of Facebook SDKs, even in cases where the code is obfuscated. This stage identifies default settings declared in the app's manifest file, including the activation or deactivation of privacy-related parameters such as the automatic collection of in-app events and the user's advertising identifier (AdID).

One of the most remarkable elements of the developed artifact is a dynamic analysis method that employs Frida to monitor privacy settings in the SDKs while the applications are running. This approach not only captures initial configuration values but also tracks any changes that occur while the app is in use. Such real-time monitoring offers a detailed view of how SDKs interact with personal data and adjust their configurations in response to app activity. Additionally, Mitmproxy is used to intercept network traffic generated by the applications, identifying personal data transfers and tracing their origin.

Beyond technical data collection, the artifact also includes a regulatory compliance assessment component. This component correlates the information obtained through static and dynamic analysis with the privacy labels and policies provided by the apps, evaluating the consistency between what developers claim and what their apps actually do. For this purpose, the platform and modules described in Section 4.1 are utilized, enabling large-scale app analysis.

Main Results

The analysis revealed that a substantial proportion of developers do not modify Facebook's default privacy settings. For instance, over 83% of the apps studied maintained the automatic event collection (AutoLogAppEvents) setting enabled, allowing for the automatic collection of personal data and app events without explicit user consent. Only 6.79% of the apps disabled the advertising identifier collection (AdvertiserIDCollection), while the rest preserved its default activation. These choices emphasize functionality over privacy in the majority of cases.

The study also uncovered marked discrepancies between the observed technical practices and the statements found in privacy labels. For example, approximately 29.25% of apps that did not alter their default settings failed to report the corresponding data-collection practices in their labels. Such inconsistencies underscore a lack of transparency in how personal data are handled, as well as the need for more accessible tools and documentation for developers. Furthermore, the dynamic analysis showed that certain apps transmitted personal data via unencrypted connections, subjecting this information to additional security risks.

Connection to Thesis Objectives

This article encompasses and applies the objectives articulated in this thesis, forging a comprehensive link between the proposed goals and their practical implementation in the regulatory analysis of the mobile ecosystem. First, it integrates dynamic and static analysis methods to evaluate the technical behavior of apps. Particularly noteworthy is the development of a novel Frida-based technique that monitors SDK privacy settings during app execution, capturing both initial values and any subsequent changes. This work is complemented by tools such as LibScout for identifying SDK versions through static analysis and Mitmproxy to document personal data transfers in real-time. These advancements directly address the aim of identifying and documenting data-transfer patterns and assessing their alignment with regulatory practices.

Second, the article extensively investigates how the observed technical practices correlate with the statements contained in privacy labels and policies. This aspect employs LLM-based tools to automatically analyze privacy policies, playing an important role in identifying discrepancies between declared and actual practices—thereby fulfilling one of the thesis's most significant specific objectives. Moreover, the article merges data from various sources—dynamic analysis, static analysis, privacy labels, and policies—to conduct automated compliance evaluations, representing the main objective of this thesis.

Finally, the research moves beyond merely identifying issues; it proposes practical solutions for addressing transparency gaps. In particular, it recommends improvements to the default SDK settings and to the documentation provided for developers. These recommendations strengthen the goal of advancing automated compliance assessments and fostering a more transparent mobile ecosystem that aligns with data protection regulations.

Impact

The article's impact is evidenced by its acceptance at the prestigious Privacy Enhancing Technologies Symposium (PETS) 2025, as well as through the importance of its collaborations and the potential influence it may have on privacy practices. PETS is widely recognized as one of the most prominent conferences in the field of privacy and privacy-enhancing technologies, drawing academics, professionals, and regulators from around the globe. Acceptance at PETS 2025 underscores the quality and significance of the presented research.

Additionally, the co-authorship with notable figures in privacy and cybersecurity further underscores the article's importance. Norman Sadeh, a professor at Carnegie Mellon University, is internationally recognized for his contributions to privacy and cybersecurity, having substantially shaped both academic literature and practical applications. His involvement attests to the study's methodological rigor and relevance.

Meanwhile, Joseph A. Calandrino, who collaborated on this article while holding senior positions within key U.S. institutions, adds a valuable regulatory perspective. During this collaboration, Calandrino served as Research Director at the Federal Trade Commission (FTC)—a leading agency for consumer privacy and protection in the United States—and subsequently assumed the roles of Deputy Chief Science and Technology Advisor and Deputy Chief AI Officer at the U.S. Department of Justice. This trajectory highlights his influence

on both technological and regulatory policies, domestically and internationally. Moreover, at the FTC's request, the article's findings were shared with the agency.

4.3 Integrated Analysis of Results

This section synthesizes the key findings derived from the articles that constitute this thesis, directly linking them to the research's main and specific objectives. Each contribution addresses relevant aspects of the technical analysis of mobile applications and the statements made in privacy policies and labels, resulting in significant advances in the automated evaluation of regulatory compliance.

In particular, the developed methods and artifacts have enabled the identification of behavior patterns, transparency gaps, and discrepancies in compliance with the GDPR. These developments have not only been independently validated in each article but have also been integrated into an automated platform for legal compliance assessment.

4.3.1 Evolution and Cohesion of Contributions

The development, updating, and maintenance of the modules that form the basis of the analysis platform have been crucial for achieving the results presented. The platform facilitated the automated downloading of APKs, privacy policies, and labels, as well as the interception of network connections to analyze the technical behavior of applications. This infrastructure made it possible to generate up-to-date, reproducible data that underpin much of this thesis's contributions. Its first application was in the study *Automated GDPR Compliance Assessment*, where the updated *Traffic* module yielded current results that, together with previous developments, contributed to the article's acceptance.

One of the earliest challenges tackled in this thesis was the lack of information regarding the destination of connections established by applications. To resolve this, ROI was developed—an artifact that identifies organizations receiving transferred personal data. However, in the article *"Sharing is Not Always Caring,"* ROI was complemented by other components created explicitly for this study, including a ChatGPT-based method to determine whether recipients were third parties and another method capable of identifying third-party entities declared in privacy policies. Together with ROI, these components made it possible to assess whether transfers were properly disclosed by matching receiving organizations to statements in privacy policies. In addition, this article developed a dynamic, Frida-based method to discern which data transmissions originated from third-party libraries, revealing their role in observed regulatory noncompliance. The findings indicated that more than 81% of the applications did not adequately disclose personal data recipients, and that over 73% of undeclared transfers were carried out by third-party libraries.

In response to growing interest in new transparency mechanisms, another article was produced focusing on privacy labels—a more recent tool adopted by Google Play and the Apple Store to inform users. This work compared the privacy labels of applications published on both platforms, revealing significant inconsistencies in 66.5% of the cases and showing that only 3.2% were fully aligned. The study illustrated how cross-platform discrepancies could point

to structural issues in how privacy practices are represented, highlighting the need to improve these mechanisms so they can fulfill their intended purpose.

Research into LLMs for the automated processing of privacy policies continued after substantial promise was observed in their integration into the *"Sharing is Not Always Caring"* article. In *Large Language Models for Privacy Policy Analysis*, a ChatGPT-based method was introduced that achieved excellent performance across a range of privacy practices and datasets, establishing itself as a valuable and scalable tool. This approach was able to process a large number of privacy policies in a short time, reducing operational costs while achieving F1-scores of over 90% in all evaluated practices and datasets. Subsequently, this approach was validated in the context of data retention statements by applying an optimized prompt design to analyze this specific practice. Examining over 2,200 privacy policies, the study *Data Retention Disclosures in the Google Play Store* found that more than 50% of the applications failed to meet the GDPR's transparency standards for retention periods. The results confirmed the versatility of the LLM-method proposed, demonstrating its ability to adapt to multiple areas of regulatory compliance.

Finally, the article *Privacy Settings in Third-Party Libraries* synthesized the contributions of the previous works. This study introduced a novel Frida-based method for monitoring the privacy settings of third-party libraries and analyzing their impact on an application's regulatory compliance. The findings revealed that most applications depended on default configurations in third-party SDKs, and that incorrect or outdated configurations could account for a significant share of observed noncompliance. This work represents a milestone in the thesis, transitioning from automated evaluation toward a deeper understanding of the causes of noncompliance by documenting not only its prevalence but also the factors that contribute to it.

4.3.2 Identification of Cross-Cutting Patterns

Across these studies, recurrent themes highlight the structural deficiencies of the mobile ecosystem with regard to regulatory compliance. The results consistently demonstrate inadequate disclosure, whether in third-party data transfers, international transfers, or the declaration of data retention periods. These discrepancies reflect a persistent gap between actual data-handling practices and statements in policies or labels. Moreover, transparency issues observed across platforms, such as those identified in the privacy labels on Google Play and the Apple Store, underscore a systemic issue affecting the reliability of current mechanisms. These highlight the need for automated compliance solutions and reveal areas where regulators and developers must intervene to enhance transparency and foster alignment with privacy regulations.

4.4 Outputs of the Thesis

This section outlines the academic, industrial, technical, and practical contributions resulting from this doctoral research, highlighting both the impact and significance of the findings. These outputs include scientific publications, international collaborations, industrial and

regulatory applications, and extensive media coverage of the results.

4.4.1 Main Published Articles

Over the course of this thesis, seven scientific articles were accepted and published as direct outputs, each undergoing double-blind peer review. Four appeared in high-impact Q1 or Q2 journals, and another in one of the leading conferences in this field (Core A). Below is a description of these publications and their academic relevance:

1. **Automated GDPR Compliance Assessment for Cross-Border Personal Data Transfers in Android Applications** (Guamán et al., 2023). Published in *Computers & Security* (JCR Q1 in the "*Computer Science, Information Systems*" category, 2023), this article introduced a fully automated method for assessing compliance with the requirements for international transfers of personal data. It has garnered wide citation (FWCI of 4.13) and considerable attention in the scientific community.

2. **ROI: A Method for Identifying Organizations Receiving Personal Data** (Rodriguez, Alamo, et al., 2024). Published in *Computing* (JCR Q2 in the "*Computer Science, Information Systems*" category, 2023), this article introduces ROI, a tool capable of identifying organizations receiving personal data with over 95% precision. This study has been essential for evaluating GDPR compliance regarding data transfers to third parties.

3. **Comparing Privacy Label Disclosures of Apps Published in Both the App Store and Google Play Stores** (Rodriguez et al., 2023). Published in *IWPE23* (International Workshop on Privacy Engineering), this article compared the privacy labels of mobile applications across two major ecosystems, revealing considerable inconsistencies. With an FWCI of 6.09 and seven citations in Scopus, it has been widely cited and sparked debate regarding the effectiveness of privacy labels as transparency mechanisms.

4. **Sharing is Not Always Caring: Delving Into Personal Data Transfer Compliance in Android Apps** (Rodriguez, Del Alamo, et al., 2024). Published in *IEEE Access* (JCR Q2 in the "*Computer Science, Information Systems*" category, 2023), this article investigated the degree of transparency in data transfers carried out by mobile applications. Its findings—showing that more than 81% of apps do not properly disclose data recipients—have been highly impactful. The work received significant coverage in traditional and online media, including *La Vanguardia*, *Computer Hoy*, and *The Conversation*, extending its influence beyond academia and triggering debate about privacy practices in mobile apps.

5. **Large Language Models: A New Approach for Privacy Policy Analysis at Scale** (Rodriguez, Yang, et al., 2024). Published in *Computing* (JCR Q2 in the "*Computer Science, Theory & Methods*" category, 2023), this article signaled a paradigm shift in large-scale, automated processing of privacy policies using LLMs such as ChatGPT. Highly scalable and accurate, this approach has become the methodological foundation for subsequent studies in this thesis, including data retention analysis and SDK configuration analysis, confirming its central role in the field. Moreover, the

presented method prompted a collaboration with legal experts at ETH Zurich, applying it in various legal contexts, which underscores its interdisciplinary scope and impact of this research.

6. **Data Retention Disclosures in the Google Play Store: Opacity Remains the Norm** (Rodríguez et al., 2024). Published in *IWPE24* (International Workshop on Privacy Engineering), this article examined how more than 2,200 privacy policies disclosed data retention practices, finding that over 50% failed to comply with GDPR transparency requirements. This study highlighted the pervasive opacity around personal data retention practices, drawing attention from both academic and industrial stakeholders. Companies such as HERE Europe B.V. have requested audits based on the presented methods, illustrating the practical applicability and real-world impact of this thesis in enhancing transparency within the mobile ecosystem.

7. **Privacy Settings of Third-Party Libraries in Android Apps: A Study of Facebook SDKs** (Rodriguez et al., 2025). Accepted at the Privacy Enhancing Technologies Symposium (PETS) 2025—one of the world's most influential conferences in privacy—this article integrates most of the contributions developed throughout the thesis. Its significance lies in combining dynamic and static analyses to reveal how default SDK configurations often lead to regulatory noncompliance, offering both a critical perspective and concrete proposals for improving the mobile ecosystem.

4.4.2 Other Published Articles

The scientific and technical advancements stemming from this thesis have contributed to the following articles, produced in collaborative contexts:

1. **Reliability of IP Geolocation Services for Assessing the Compliance of International Data Transfers** (Cozar et al., 2022). Following the restructuring and reconstruction of the dynamic analysis module used for intercepting mobile connections, possible inaccuracies in the geolocation service came to light. This discovery led to the formalization of this study, which examined various geolocation services, measuring and comparing their accuracy. Presented at IWPE22 (International Workshop on Privacy Engineering) in Genoa, where the author of this thesis won the best presentation award, it has reached an FWCI of 3.97 in Scopus—four times above the expected citation average for similar publications. Its 10 citations and 9 PlumX captures place it in the 95th percentile, highlighting its recognized scientific impact.

2. **Identifying Organizations Receiving Personal Data in Android Apps** (Rodriguez et al., 2022). This proof-of-concept study preceded the "ROI" article and included preliminary results. Presented at SECRYPT 22 (International Conference on Security and Cryptography) in Lisbon by the author of this thesis, it served as the initial step toward the ROI development.

3. **The Court Speaks, But Who Listens? Automated Compliance Review of the GDPR** (Zac et al., 2024). The *Automated GDPR Compliance Assessment* study piqued the interest of a research group composed of legal scholars at ETH Zurich. This

collaboration focused on connection interception results at various time points and the analysis of international data-transfer practices in policies. An automated method was also developed to retrieve privacy policies from a specific point in time using the Wayback Machine. The resulting study remains open as a preprint, pending submission to a high-impact journal.

4. **ATLAS: Automatically Detecting Discrepancies Between Privacy Policies and Privacy Labels** (Jain et al., 2023). Developed during the initial months of collaboration with Professor Norman Sadeh and his group at Carnegie Mellon University, this work ran in parallel with "Comparing Privacy Labels," accepted and published at IWPE23 (International Workshop on Privacy Engineering), where the first author received the best presentation award. It has achieved an FWCI of 2.61 in Scopus, more than double the average expected for publications in this field, with 3 citations and 12 PlumX captures, placing it in the 90th percentile—a further testament to its scientific relevance.

5. **Hunter: Tracing Anycast Communications to Uncover Cross-Border Personal Data Transfers** (Pascual et al., 2024). Originating from the geolocation services analysis and the challenges in geolocating anycast IP addresses, this study builds on a substantial set of intercepted data from mobile applications, serving as the foundation for identifying anycast IPs and developing the corresponding geolocation method. Published in *Computers & Security* (JCR Q1 in "Computer Science, Information Systems" category, 2023), it has not yet received citations in Scopus due to its recent release, but Google Scholar lists two citations to date.

6. **Anycast and Third-Party Libraries: A Recipe for a Privacy Disaster?** (Pascual et al., 2025). Stemming from the "*Sharing is Not Always Caring*" study, which introduced an artifact for identifying third-party libraries in mobile applications responsible for collecting personal data, this research delves deeper into the potential GDPR violations in international data transfers when anycast addresses are employed. This paper has been accepted at *IEEE Communications Magazine* (JCR Q1 in "Telecommunications" category, 2023).

7. **GPT vs. Human Legal Text Annotators: A Comparative Study with Privacy Policies** (Cevallos-Salas et al., 2025). Resulting from an international collaboration with the Escuela Politécnica Nacional (EPN) in Quito, Ecuador, this article proposes a novel GPT-4o-based method for automated privacy policy annotation, achieving performance comparable to human annotators. The findings indicate that the model attains 80% accuracy at the segment level and 90% for entire texts. Additionally, a *logprob* analysis is introduced to enhance confidence in the generated annotations. This article is currently available as a preprint and is under review in a scientific journal.

4.4.3 International Collaborations

This thesis fostered interdisciplinary collaborations with prominent institutions such as Carnegie Mellon University (CMU), King's College London (KCL), and ETH Zurich. These partnerships fostered the generation of innovative results in the fields of privacy and data protection, combining technical and legal perspectives. Each collaboration had a direct

influence on the quality and scope of the articles presented, strengthening both the academic impact and practical applications of this research.

Collaboration with Carnegie Mellon University began in 2022, at the outset of the thesis, and continued throughout, playing a role in most articles developed in this thesis. This partnership integrated the technical expertise of the group led by Professor Norman Sadeh—internationally recognized in privacy and cybersecurity. A research stay at CMU in 2023 (Pittsburgh, USA) explored GenAI capabilities for the automated processing of privacy policies, yielding insights and knowledge that led to a scalable and efficient LLM-based approach and contributed to the thesis's international distinction. This stay resulted in the publication *Large Language Models: A New Approach for Privacy Policy Analysis at Scale*, as well as contact with Joseph A. Calandrino, then serving as Research Director at the U.S. Federal Trade Commission (FTC).

The Pittsburgh meeting with Joseph A. Calandrino led to a close collaboration that culminated in the article *Privacy Settings of Third-Party Libraries in Android Apps: A Study of Facebook SDKs*, accepted at the prestigious Privacy Enhancing Technologies Symposium (PETS) 2025. Calandrino's involvement, and his subsequent role at the U.S. Department of Justice, reinforced the regulatory significance of the findings, which were shared with the FTC.

The publication *Automated GDPR Compliance Assessment for Cross-Border Personal Data Transfers in Android Applications* prompted interest at ETH Zurich—particularly among a group of legal scholars—which resulted in the co-authored article *The Court Speaks, But Who Listens? Automated Compliance Review of the GDPR*, exploring the evolution of privacy policies following changes to the U.S. Privacy Shield. Additionally, *Large Language Models* garnered attention from another ETH Zurich group specialized in Swiss data protection law. This interest led to a new, ongoing collaboration focusing on how this law is reflected in privacy policies, thus demonstrating the multidisciplinary and global impact of this thesis's findings.

Collaboration with the Escuela Politécnica Nacional (EPN) in Quito stemmed from a previous doctoral student in the research group, now an Associate Professor at EPN, who is the first author of early studies this thesis builds upon—i.e., *Automated GDPR Compliance Assessment*. The subsequent re-establishment of contact resulted in an international collaboration culminating in a joint article (*GPT vs. Human Legal Text Annotators: A Comparative Study with Privacy Policies*), currently available as a preprint. In this article, a GPT-4o-based automated privacy policy annotation method achieves performance levels comparable to those of human annotators (approximately 80% accuracy at the segment level and 90% for complete texts). A logprob *analysis* was also introduced to evaluate and enhance confidence in automated annotations. The study is under review in a high-impact scientific journal, highlighting both the importance of the collaboration with EPN and the international resonance of these research outcomes.

Another noteworthy collaboration arose from Warwick University's interest in the ROI artifact. A master's student there integrated new LLM-based functionalities into the tool, receiving support and guidance throughout. This work exemplifies how the methods and artifacts developed in this thesis serve as a springboard for further academic research.

Finally, the partnership with King's College London was sparked by *Automated GDPR Compliance Assessment*. This relationship culminated in a three-month research stay in 2024, where automated compliance assessment efforts were extended to the domain of LLMs, specifically evaluating personalized chatbots to ascertain their alignment with usage policies and focusing on their security. This project produced preliminary results that will soon be published in a high-impact conference, expanding the thesis into new domains of research and application.

4.4.4 Industrial and Regulatory Impact

The findings from this thesis have had a significant impact in industrial and regulatory contexts, underscoring their relevance.

- **HERE Europe B.V.** The company requested an audit of its Google Play Store application after a member attended to the *Data Retention Disclosures in the Google Play Store* paper presentation at *IWPE24*. This request reflects the industrial sector's growing interest in applying automated assessment of privacy practices and GDPR compliance. The audit's results allowed the company to identify its current practices and potential improvements to enhance transparency.
- **Spanish Data Protection Agency (AEPD).** The author of this thesis maintained a close relationship with the AEPD throughout this research. An in-person meeting was held at an early stage to present and discuss methodologies and studies developed under the AutoGDPR project, aimed at automating data protection compliance assessments. The AEPD showed particular interest in the platform and proposed methods, especially the ROI artifact. Following this meeting, ROI was provided to the AEPD, which conducted internal tests on a set of web domains. Continuous communication with the AEPD was maintained via email, sharing subsequent study results, such as *Sharing is Not Always Caring*, and demonstrating the practical relevance of the findings for promoting transparency in the mobile ecosystem.
- **Intellectual Property Registration in Spain.** The ROI methodology, presented in *A Method for Identifying Organizations Receiving Personal Data*, describes an innovative solution for identifying domain controllers—an especially relevant challenge due to widespread efforts to obscure domain information. This methodology was formally registered at the Territorial Registry of Intellectual Property of the Community of Madrid on January 26, 2023 (registration number 16/2023/385, reference M-7861/2022). Formal recognition in the registry enhances its originality as an automated solution for identifying the organizations responsible for web domains.
- **Collaboration with the Federal Trade Commission (FTC).** Another regulatory milestone of this research resulted from international collaboration with Joseph A. Calandrino, who served as Research Director at the FTC during the collaboration. The FTC is a key authority for consumer privacy and protection in the United States, significantly shaping global policy in these areas. This joint work culminated in the publication *Privacy Settings in Third-Party Libraries*, accepted at PoPETs 2025—one of the most prestigious conferences in privacy. Additionally, the associated findings and

datasets were shared with the FTC, illustrating the agency's interest in the research outcomes. Toward the end of the collaboration, Joseph A. Calandrino took on the roles of Deputy Chief Science and Technology Advisor and Deputy Chief AI Officer at the U.S. Department of Justice.

These industrial and regulatory applications demonstrate that the methods and artifacts developed in this thesis extend beyond scientific advancement to offer solutions for improving transparency and regulatory compliance in the mobile ecosystem.

4.4.5 Open Data Generation

The scientific and technical output generated by this thesis has led to the publication of multiple open resources that advance research into automated compliance assessment and transparency in the mobile ecosystem. These datasets and tools, made publicly available, enhance the reproducibility of the presented methods, facilitate interdisciplinary collaboration, and enable the extension of the findings to new areas of study. The principal resources include:

1. **ROI Method's Contributions.** This resource includes the ground truth dataset (Rodríguez et al., 2023) used to validate the Receiver Organization Identifier (ROI) method. It provides manually annotated domains and organizations identified in captured network connections, enabling the evaluation of the method's accuracy in recognizing domain controllers and forming a basis for validating new techniques for detecting personal data transfers.

2. **Annotated Dataset on Declared Purposes of Data Processing.** This dataset (de Castro, 2024) gathers privacy policies manually annotated to identify declared purposes for personal data processing in compliance with GDPR requirements. The annotations follow legally established categories, offering a valuable resource for training and validating natural language processing (NLP) methods designed to automatically extract stated data processing purposes from privacy policies.

3. **Annotated Dataset on Data Retention Statements.** This dataset (Rodriguez, Fernández, et al., 2024) comprises privacy policies annotated to identify declarations related to data retention periods. The annotations allow the evaluation of whether the policies meet GDPR transparency requirements, documenting both explicit references and omissions of key information.

4. **Contributions from the Article "Privacy Settings in Third-Party Libraries."** This GitHub repository (Rodriguez, 2025) provides tools and scripts developed for analyzing privacy settings in third-party SDKs—specifically, Facebook SDKs used by Android applications. These contributions have been awarded the *Available Badge* at PETs after the artifact review and improvement. It includes:

    - **Frida-based Privacy Analysis Tool:** A dynamic analysis tool that monitors SDK privacy configurations during app execution on real devices, capturing both initial values and real-time changes during runtime.

    - **Maven Library Crawler and LibScout Profile Generator:** Tools to crawl

and download libraries from Maven repositories and generate profiles for static analysis detection via LibScout. These tools facilitate identifying specific versions of SDKs integrated into Android apps.

Publishing these resources in open-access platforms represents a central contribution of this thesis, promoting transparency, reproducibility, and further research in data privacy and protection. Each dataset and tool was designed, validated, and applied in the context of the studies presented here, and making them publicly available encourages their adoption in scientific, industrial, and regulatory communities.

4.4.6 Academic Training

This thesis has led to considerable academic training opportunities. Technical support was provided for two undergraduate theses in the Bachelor's Degree in Telecommunication Technologies and Services Engineering at UPM. These projects explored enhancements to certificate pinning bypass techniques using Frida, as well as NLP strategies for processing privacy policies.

Additionally, the author supervised three other bachelor's theses in the same program. These works expanded on third-party library data-connection identification, techniques for intercepting HTTP3-based connections, and the multilingual capabilities of ChatGPT for large-scale privacy policy processing. The latter led to the publishing of a multilingual dataset (Pantoja et al., 2024) in the data protection domain.

The supervision extended to a bachelor's thesis in Information Systems (ETSISI) and a Master's thesis in Distributed and Embedded Systems Software. These projects focused on applying LLMs to identify both the data controller (in the bachelor's thesis) and the declared purposes of data processing (in the master's thesis).

Altogether, seven fina-year projects were undertaken in connection with this PhD thesis.

4.4.7 Social Impact

The findings of this thesis have been extensively covered by national media outlets, raising public awareness about the importance of privacy in the mobile ecosystem and underscoring the relevance of the research results.

The article *Sharing is Not Always Caring: Delving Into Personal Data Transfer Compliance in Android Apps* attracted considerable media attention, featured in a diverse array of outlets. In February 2024, the UPM Research Portal and its Research Bulletin publicized the study, further disseminating it via the international AlphaGalileo platform. This facilitated coverage in prominent media sources such as *La Vanguardia*, *Noticias de la Ciencia*, *ServiMedia*, *AragonDigital*, and *Computer Hoy* (Spain's most widely read technology magazine), as well as online outlets including *Suwedi*, *DiarioGlobal*, *LaUnión*, and *Norte Informa*. Consequently, the article reached both specialized and general audiences across various contexts.

Its influence extended beyond print and online media, as it was also featured on radio and television. *La Linterna* on the COPE network—one of Spain's most popular radio

shows—devoted a special segment to the study's findings on February 27, 2023. This coverage was enhanced by mentions on TreceTV, evidencing the public's keen interest in the implications of personal data transfers and the transparency deficiencies in mobile applications. A Twitter thread further increased engagement and visibility, prompting discussion on social media as well.

The publication in *The Conversation* in May 2024 presented the findings in a more accessible and informative format, aiding comprehension among non-expert readers and encouraging public discourse on the importance of privacy in the technological context. In addition, *Computer Hoy* featured the study in its August 2024 issue (No. 677), reinforcing its influence through one of the most widely recognized technology outlets in Spain.

Moreover, *Data Retention Disclosures in the Google Play Store: Opacity Remains the Norm* has also attracted noteworthy attention. The Communications Office of the Universidad Politécnica de Madrid (UPM) plans to publicize the study's results, highlighting how pervasive opacity remains in the disclosure of personal data retention practices. This move further underscores the work's social relevance and its potential to inform public awareness of shortcomings in transparency in mobile applications.

The media coverage demonstrates the public's interest not only in each study's specific findings but also in the broader significance of the research agenda. Widespread attention from print, digital, and broadcast outlets suggests a growing social concern around privacy in the mobile ecosystem and affirms the importance of addressing these shortcomings.

4.5 Future Research Directions and Funding Opportunities

4.5.1 Future Research Lines

This thesis has laid a solid foundation for automating the regulatory compliance assessment of Android mobile applications, encompassing both the technical behavior analysis of apps and the evaluation of practices declared in privacy policies. However, the methodologies, artifacts developed, and published studies open up new and promising lines of inquiry that could extend and amplify the impact of this thesis in various directions.

Multilingual Evaluation of Privacy Practices Using LLMs

One of the most promising directions involves investigating the capabilities of LLMs to identify privacy practices in policies written in multiple languages. Currently, we are working on creating the first privacy policy dataset annotated in all 24 official languages of the European Union (EU), with the aim of evaluating ChatGPT's performance in identifying practices in non-English policies. This study would enable market-specific compliance evaluations across the EU, where linguistic diversity poses a significant challenge for conventional analysis methods.

By succeeding in this research, the scope and applicability of compliance evaluations would

grow substantially, allowing regulators and organizations in every EU Member State to adopt the methodology under GDPR. This extension would facilitate large-scale, multilingual analyses of privacy policies in a global context, addressing a mounting need in a widely global environment.

Subsequent research could compare multiple LLMs—both proprietary and open-source—to assess their ability to recognize privacy practices in multilingual policies. Such a comparative study would offer an objective appraisal of different models' performance, identifying which delivers the best balance of accuracy and efficiency. Moreover, regulators, researchers, and developers could reference these findings when selecting models tailored to their specific requirements.

As a matter of fact, this study is currently being conducted by a Master's student under a research grant. Preliminary results indicate that ChatGPT can process privacy policies across all 24 official EU languages with comparable performance, demonstrating its potential for multilingual compliance analysis.

Integrated Compliance Assessment Approach

Another potential research avenue involves unifying all privacy practice detections into a single automated method driven by LLMs. Currently, individual studies have focused on specific practices—such as international data transfers, data retention, or identifying third-party recipients. However, developing a comprehensive method capable of evaluating multiple practices in parallel would enable full legal compliance assessments of privacy policies under the GDPR.

This line of work could also evolve to include the automatic generation of improvement recommendations, bridging the gap between detecting shortcomings and advising on concrete steps for enhancing policy transparency and quality. Such an approach would benefit regulators and developers in meeting compliance obligations while offering users clearer, more precise policies aligned with data protection standards.

Extension to the iOS Ecosystem

To date, this thesis has centered on regulatory compliance analysis for Android mobile applications, focusing on advanced dynamic and static methods for examining technical behavior and declared practices. However, an enticing research direction involves applying these methodologies to the iOS ecosystem.

The automated processing methods for privacy policies can be readily adapted for iOS apps, permitting a direct comparison of compliance and transparency levels between the two ecosystems. In this regard, developing a platform to process and evaluate both Android and iOS applications simultaneously would mark a significant stride toward global legal compliance evaluations. Such an integrated platform would broaden the reach of the proposed methods and reveal any fundamental disparities between the two ecosystems, contributing to a more complete understanding of the state of mobile app compliance.

Automatic Compliance Assessment in Custom Chatbots

Collaboration with King's College London has enabled a natural progression of this research, transferring the methods developed for evaluating privacy policies to the automated assessment of usage policies for personalized chatbots. These policies outline the chatbots' operational and response guidelines, and their rapid adoption in recent years has presented new challenges in security and user safety.

Automating the evaluation of chatbot usage policies is essential for detecting noncompliance and behavioral deviations while also enhancing transparency and trust in these systems. This direction is especially relevant against the backdrop of the widespread adoption of generative AI by the general public.

4.5.2 Potential Funding Sources

Public Funding

Public funding is organized at three levels—regional, national, and international—each governed by strategic frameworks specifying priorities and focal areas. At the regional level, programs reflect strategies such as the *Estrategia Madrileña de Investigación e Innovación 2030 (EM2i)* and the *Estrategia de Especialización Inteligente (S3)*, which emphasize science and technology development in key areas for the Community of Madrid. At the national level, *Estrategia Española de Ciencia, Tecnología e Innovación (EECTI) 2021–2027* and the *Plan Estatal de Investigación Científica y Técnica y de Innovación (PEICTI)* are an integrated approach to advancing scientific excellence, knowledge transfer, and technological and social challenges. Finally, at the international level, the *Horizon Europe* program (2021–2027) is the EU's main instrument for R+D+i funding, supporting projects aimed at addressing global challenges and strengthening Europe's leadership in areas such as digital transformation and artificial intelligence.

These strategies define the research lines eligible for funding that address regional, national, and European objectives. Specifically, these public calls underscore a commitment to backing innovative solutions—such as those in this thesis—that contribute to security, privacy, data protection, and responsible technology assessments.

Regional Funding Programs The Community of Madrid has formulated a cohesive strategic framework for research, development, and innovation, embodied in three interrelated instruments:

1. *Madrid's Strategy for Research and Innovation 2030 (EM2i) - Estrategia Madrileña de Investigación e Innovación*,
2. *Smart Specialization Strategy (S3)* 2021–2027 - *Estrategia de Especialización Inteligente*, and
3. *VI Regional Plan for Scientific Research and Technological Innovation (VI PRICIT)* 2022–2025 - *VI Plan Regional de Investigación Científica e Innovación Tecnológica.*

These strategies aim to position Madrid as a leading R+D+i hub in Southern Europe, driving

competitiveness, economic development, and social well-being in the region.

- **EM2i** outlines a long-term vision through 2030, addressing digital and environmental transitions, territorial and social cohesion, and the need to integrate gender perspectives across research and innovation policy—rooted in the 2030 Agenda's Sustainable Development Goals (SDGs) and key European policies like *Horizon Europe* and the *European Green Deal.*
- The **S3** then specifies strategic priorities from 2021–2027, grounded in a thorough assessment of the *Madrid Research and Innovation System (SM2I).* This strategy addresses challenges such as post-pandemic economic recovery while meeting new EU directives in digitization and sustainability.
- The **VI PRICIT** operationalizes these goals over 2022–2025, providing tangible instruments to enhance scientific and technological infrastructures, foster research excellence, and promote collaboration among universities, research centers, businesses, and government agencies.

Within this strategic framework, the Information and Communication Technologies (ICT) sector figures prominently among the S3's six priority specialization areas, encompassing key disciplines—artificial intelligence, cybersecurity, big data, and emerging technologies—considered essential for promoting the region's digital transformation. In parallel, Disruptive Innovation offers an arena for developing high-impact technologies that can transform economic and social sectors.

This thesis's research—focusing on data protection and the automated assessment of GDPR compliance—directly aligns with these regional strategies. On the one hand, it contributes to the "Excellent Science" axis by proposing tools to stimulate applied research and the development of novel ICT-based solutions. On the other hand, its potential transfer to real-world settings fits the "Business Leadership" axis, providing technologies that enhance competitiveness and innovation in industry and technology sectors.

Additionally, the work spans the S3 priority areas of ICT and Disruptive Innovation by tackling AI-driven, automated regulatory analysis methods in digital contexts. More broadly, this research also supports meeting the Sustainable Development Goals (SDGs), specifically SDG 4 and SDG 16, aimed at advancing the digital society and building transparent, accountable institutions.

In the context of Madrid's EM2i, S3, and VI PRICIT, various public funding calls exist for research and innovation in priority areas relevant to the region. These calls, aligned with the S3's strategic pillars and the SDGs, represent a significant funding source for the present research agenda centered on data protection and automated GDPR compliance assessment.

- **Funding for R+D Activity Programs in Technologies**. One avenue for financing this work is the call for projects under *R+D Activity Programs in Technologies*, aimed at collaborative efforts among universities, research centers, and companies in strategic areas such as ICT. This research—featuring advanced techniques like LLMs and automated analysis—directly addresses the domain of data protection and transparency. The alignment with this call is reflected in its potential to reinforce the *Madrid Research*

and Innovation System (SM2I) by generating knowledge transferable to the tech and business sectors.

- **Industrial Doctorate Grants**. These grants support doctoral theses carried out in collaborative settings involving academia and the private sector. Given the practical orientation of this research—which could be implemented in regulated industries and the tech sector—this call appears especially relevant. In addition to advancing the digitization of regulatory processes, automated legal analysis appeals directly to businesses keen on ensuring their GDPR compliance or offering compliance services to others.
- **Synergistic R+D Project Grants**. Synergistic R+D grants back proposals addressing complex challenges through interdisciplinary collaboration. The nature of this research—combining NLP, software analysis, and regulatory compliance—ideally fits these goals. By integrating diverse approaches, this work can evolve into a practical tool for auditing apps and ensuring alignment with data protection regulations.
- **Equipment Acquisition Grants**. Finally, grants for Scientific-Technical Equipment could support the necessary infrastructure for conducting this research, especially for implementing and validating automated legal compliance platforms. Such funding would allow procurement of specialized hardware to locally run LLMs at a scale sufficient for undertaking the complex tasks associated with this study. The ability to compare local models with commercial alternatives also opens up the possibility of evaluating performance in terms of accuracy, efficiency, and domain specialization in data protection. Should these local LLMs demonstrate comparable or superior performance, they could become a new standard for automated legal compliance analysis, offering enhanced control, transparency, and technological sovereignty over the tools deployed in this field.

Such infrastructure support would substantially bolster the group's technical capabilities and stimulate disruptive innovation in ICT and applied AI, aligned with the strategic S3 goals and the demand for cutting-edge solutions that meet current technological and regulatory challenges.

National Funding Programs At the national level, Spain has established a strategic framework for research, development, and innovation, primarily articulated through:

1. The *Spanish Strategy for Science, Technology, and Innovation (EECTI) 2021–2027 - Estrategia Española de Ciencia, Tecnología e Innovación,* and
2. The *State Plan for Scientific, Technical, and Innovation Research (PEICTI) 2021–2023 - Plan Estatal de Investigación Científica y Técnica y de Innovación.*

Together, they aim to strengthen the *Spanish Science, Technology, and Innovation System (SECTI)*, aligning it with European policies, particularly *Horizon Europe* and the *European Green Deal*.

The EECTI sets out a long-term vision of advancing Spain as a leader in research and innovation, promoting scientific excellence, technology transfer, and industrial and digital skills development. This strategy tackles global and national challenges, focusing on digitalization, the green transition, and growth in the knowledge economy, contributing to the SDGs and a

resilient, inclusive economy.

Within this framework, the State Plan for Scientific, Technical, and Innovation Research is the main operational tool for funding R+D+i nationally. Organized around four state-level programs—*Knowledge Generation*, *Talent Development and Recruitment*, *Business Leadership in R+D+i*, and *R+D+i Oriented Toward Societal Challenges*—this plan seeks to bolster scientific and technical research, attract top researchers, foster public-private partnerships, and foster innovative solutions to strategic challenges.

This national strategic context is directly in line with the objectives and subject matter of this research. Building automated tools for GDPR compliance auditing through AI and NLP, along with the automatic evaluation of chatbots, aligns particularly well with the priorities in ICT. Automating legal analysis also meets the need for creating secure, transparent user-facing systems and drives knowledge transfer to industry, enhancing technological competitiveness and digital security.

Call for R+D+i Projects Oriented to Societal Challenges Managed by the *Agencia Estatal de Investigación (AEI)*, this call funds projects tackling high-impact social, economic, and technological challenges, with a focus on innovative and sustainable solutions. This research—centered on data protection and automated GDPR legal compliance—fully meets this program's aims by addressing one of the most pressing contemporary issues: the need to ensure privacy and develop secure, transparent personal data management systems. This concern, highlighted in the EECTI 2021–2027 and documents such as *Agenda Digital 2026*, underscores the importance of cultivating a reliable digital environment and reinforcing security in digital-enabling technologies.

Protecting fundamental rights such as privacy and transparency has become a strategic priority in the information era. The exponential growth in AI usage necessitates pioneering tools to audit and guarantee regulatory compliance within complex systems. Automating GDPR analysis through advanced models like LLMs offers a solution to this challenge, facilitating the deployment of transparent, secure systems in both public and private sectors.

Furthermore, evaluating chatbot behavior automatically addresses a developing concern highlighted in government reports on responsible AI. Official publications such as Spain's White Paper on Artificial Intelligence emphasize the importance of ensuring that AI technologies operate ethically and securely. Systematically auditing chatbots is thus a crucial step in mitigating associated risks—particularly in security, privacy, and potential misuse. As chatbots continue to gain traction in critical services, from customer support to organizational processes, validation frameworks are key to establishing user trust and responsible AI adoption.

Public-Private Collaboration Projects Another source of support is the *Public-Private Collaboration Projects* call, managed by the AEI, which encourages technology transfer and knowledge exchange among universities, research centers, and private-sector entities. One strand of research—the development of automated tools for chatbot assessment—directly aligns with this call's requirements by offering implementable, industry-focused solutions.

For instance, institutions such as BBVA have expressed interest in chatbots to optimize both

user experience and back-end processes. Ensuring that these systems are secure, transparent, and comply with regulatory and ethical standards underscores the relevance of this research. Automated tools that systematically audit chatbot compliance and reliability could benefit the financial sector and establish standards applicable to other technological domains.

Potential collaboration with private-sector actors like BBVA would highlight the project's applicability and potential for technology transfer, making it a promising approach to tackling the current challenges in responsible AI adoption. Hence, this call presents an opportunity to validate and consolidate results in a real-world context, showing their tangible impact on improving technological processes.

Industrial Doctorate Grants In the realm of training and attracting talent, *Industrial Doctorate Grants* represents another pertinent funding option. These grants support the execution of doctoral research in collaboration with industry partners through either industrial research or experimental development. The practical orientation of this research, which can be deployed in regulated industries and the tech sector, makes it an ideal fit, enabling the thesis to be conducted in a real-world setting that closely reflects market needs.

Scientific-Technical Equipment Grants Finally, *Scientific-Technical Equipment Grants* offers the possibility of outfitting research groups with specialized hardware needed to locally run LLMs. Such equipment would allow validating the platforms developed here, comparing the performance of local models against proprietary systems and potentially establishing new benchmarks in automated legal compliance assessments. This capability would encourage technological progress and enhance technological sovereignty and transparency in AI-driven solutions.

International Funding Programs At the European level, the European Union has established an ambitious and structured strategic framework to drive research, development, and innovation, led by *Horizon Europe* (2021–2027). Serving as the EU's flagship R+D+i funding mechanism, Horizon Europe aims to strengthen the European Research Area (ERA) and maintain Europe's position as a global leader in science, technology, and innovation. The program aligns with broader European initiatives such as the *Digital Decade 2030* and the *European Green Deal*, supporting the pursuit of an inclusive, digital, and sustainable economy and contributing directly to the *Sustainable Development Goals* (SDGs).

Horizon Europe is structured into three main pillars:

1. **Pillar I: Excellent Science**. Aims to reinforce Europe's scientific foundation by promoting frontier research, international researcher mobility, and cutting-edge scientific infrastructure.
2. **Pillar II: Global Challenges and European Industrial Competitiveness**. Addresses the most pressing societal and technological challenges via funding for strategic areas such as digital transformation, security, artificial intelligence, and data protection.
3. **Pillar III: Innovative Europe**. Seeks to foster disruptive innovation, facilitate technology transfer, and support high-impact innovation ecosystems with significant

economic and social benefits.

Digital transformation and the development of enabling technologies are strategic priorities for reshaping Europe's socioeconomic landscape. Documents such as the *European Strategy for Data* and the *White Paper on Artificial Intelligence* underscore the necessity of securing trust, transparency, and ethics in AI systems and data handling.

This European strategic framework resonates with the aims and focus of the present research. Developing automated tools for GDPR compliance auditing—combining AI and NLP—addresses priorities in Pillar II of Horizon Europe, specifically within Information and Communication Technologies (ICT). This line of inquiry responds to one of today's most pressing societal issues: ensuring privacy and establishing secure, transparent digital systems that safeguard fundamental rights within the European community.

Likewise, extending this research to automated chatbot behavior evaluations fits within the scope of responsible AI, an emerging challenge recognized in policy discussions. As emphasized in the AI White Paper, it is vital that these systems operate ethically, transparently, and securely in high-stakes contexts like user support or organizational processes. Deploying tools to audit and validate such systems lessens the risks associated with these technologies, fostering trust in digital solutions throughout Europe.

Funding Under Horizon Europe Through *Horizon Europe* (2021–2027), a variety of calls could finance the research lines proposed in this thesis, given their alignment with European strategic priorities in digital transformation, AI, and data protection. These calls target technological and societal challenges, promoting innovative solutions that foster digital sovereignty, security, and trust in automated systems.

A particularly relevant opportunity can be found under Pillar II: *Global Challenges and European Industrial Competitiveness*, especially Cluster 4: Digital, Industry, and Space. This cluster focuses on developing secure, advanced digital technologies to facilitate the digital transformation of European society and economy. Research centered on automating GDPR compliance and evaluating chatbot behavior aligns with the cluster's goals by helping create digital systems that are safer, more transparent, and more respectful of fundamental rights.

Specifically, the lines of research on trustworthy AI and secure data management tie into the following themes under Horizon Europe:

- **Trustworthy and Secure AI**. The program's calls in this area focus on advanced AI technologies that are transparent, reliable, and ethically aligned. This thesis's work—auditing and validating chatbots and normative analysis systems—directly addresses these concerns. By providing tools to assess and confirm these technologies' security, while reducing bias, the research fosters trust in their practical applications.
- **Digital Security and Privacy**. Horizon Europe also supports projects aimed at improving digital security and protecting personal data—key concerns in the European Data Strategy. Automated legal compliance with GDPR represents an innovative solution for ensuring privacy and promoting digital trust, meeting the need to safeguard European citizens' data in an increasingly digitized world.

Under Pillar III: *Innovative Europe*, the European Innovation Council (EIC) offers targeted calls for disruptive projects of high economic and social impact. Combining advanced NLP with automated auditing, this line of research has cross-cutting potential in a range of sectors, from finance to technology. Businesses can adopt these tools to ensure compliance and manage risks, bolstering both digital trust and operational efficiency. Consequently, the EIC Pathfinder and EIC Transition calls provide opportunities to validate and scale these innovations.

Lastly, the Marie Skłodowska-Curie Actions (MSCA)—part of Pillar I—offer another viable route for funding. These actions enhance researcher training and mobility through international and interdisciplinary collaborations. The applied, transferable focus of this thesis could benefit greatly from such a program, facilitating interactions with other centers of excellence in responsible AI and technological auditing and furthering its international outlook.

Private Funding

Beyond the public funding programs at regional, national, and international levels, private funding also constitutes a viable avenue for developing and reinforcing this research line. The practical and adaptable nature of these proposed tools makes them highly applicable to strategic sectors and positions them as innovative technological solutions with substantial market potential.

Industry Partnerships and Spin-Off Opportunities Major players in automated legal compliance solutions, such as Google Checks and similar enterprises, actively seek tools that guarantee transparency, security, and regulatory alignment in mobile apps and AI-driven systems. This research—combining AI and NLP—directly meets the needs of this sector by providing an automated, efficient approach to auditing compliance with regulations like the GDPR.

Additionally, the research could lead to the forming of a startup or spin-off venture, initially supported by public innovation grants such as the European Innovation Council (EIC) or national programs like NEOTEC. Such independence would foster the development of a company focused on auditing mobile applications, validating chatbot models, or offering related technological services. This approach would potentially create new business opportunities and accelerate the transfer of academic outcomes to market applications.

Moreover, technology providers for chatbots—like OpenAI or emerging companies in conversational AI—may be interested in funding this research. The capacity to audit and validate these systems automatically is necessary for ensuring safety, ethics, and transparency—criteria rapidly gaining importance among regulators and the broader public. Marketplaces such as the Google Play Store could also consider investing in this research line, given that it would bolster a platform that is more transparent and respectful of user privacy. Similar security-focused initiatives have already been integrated to evaluate apps published on Google Play, demonstrating the platform's interest in auditing its applications.

Companies adopting chatbots—exemplified by BBVA—might likewise find this research compelling. These organizations, which rely on chatbot technology in both customer interactions

and internal processes, require validated methods for ensuring compliance, security, and digital trust. In this sense, collaboration with the financial sector or with large corporations in areas such as e-commerce, digital healthcare, or public services could help secure funding and drive real-world implementation of these findings.

4.5.3 Future Funding Opportunities

The methodologies and artifacts developed open up new research avenues that could further expand the impact of this work. As already mentioned, two research directions stand out as particularly promising for obtaining funding: the automatic evaluation of chatbot compliance and the automated assessment of GDPR compliance in mobile applications.

Automated Chatbot Compliance Assessment: AI Governance and AI Safety

The rapid adoption of Generative AI and conversational agents has triggered growing concerns regarding AI safety, governance, and regulatory compliance. The European Union is at the forefront of regulating AI systems, as evidenced by the AI Act, which sets obligations for high-risk AI applications, including chatbots that interact with users in consumer support, healthcare, and financial services. Ensuring that these systems align with legal and ethical principles is a pressing challenge, as violations in chatbot-generated content can lead to legal liability, misinformation risks, and reputational damage. Regulators and policymakers are actively seeking solutions to establish trustworthy and accountable AI, making this research line highly competitive for financial support.

This research direction aims to develop automated tools for evaluating chatbot compliance, integrating methods to assess safety and fairness in conversational AI systems. The approach would build upon LLM-based techniques to analyze chatbot usage policies, the risk of hallucinations, and potential biases in AI-generated responses.

Given the strategic relevance of AI auditing and governance, this research line aligns well with Horizon Europe's strategic funding priorities, particularly under Pillar II, Cluster 4: Digital, Industry, and Space, which supports projects focused on trustworthy and safe AI. Future funding opportunities could emerge from calls similar to HORIZON-CL4-2023-HUMAN-01-01, which targeted the human-centric development of AI technologies, with the scope of making AI, data and robotics solutions meet the requirements of trustworthy AI, based on accuracy, robustness, safety, ethical principles and reliability. This project could contribute directly to the EU's AI regulatory framework by integrating compliance solutions into AI regulatory sandboxes, where high-risk AI applications undergo testing before deployment.

Beyond public funding, this research also aligns with industrial needs, as financial institutions, customer service providers, and technology companies seek automated methods to ensure their AI-driven services adhere to regulations and their own internal ethical frameworks. The demand for AI auditing solutions is increasing, especially as organizations face potential fines and reputational risks linked to AI outputs.

To apply for European funding, a consortium of partners is typically required, including research institutions, industry stakeholders, and real-world use cases that demonstrate the

project's feasibility and impact. Given the interest already expressed by the financial and energy sectors, a joint European project proposal could be formulated in collaboration with private institutions, allowing them to benefit directly from the research outcomes. These partners, such as BBVA, Endesa, and ING, could serve as real-world case studies, providing concrete applications for AI compliance assessment. Additionally, partnerships with AI providers could enhance the project's feasibility.

As an alternative, the research could be positioned within public-private funding programs, where collaboration between academia and industry is encouraged. These programs could provide a more agile path to implementation.

Automated GDPR Compliance Assessment in Mobile Devices

Mobile applications have become the primary digital interface for billions of users worldwide. While they offer unprecedented convenience and functionality, they also pose significant privacy and security risks, particularly due to the extensive reliance on third-party SDKs. These risks have been highlighted by European and international regulators, including the French CNIL and the U.S. Federal Trade Commission (FTC), both of which have issued warnings regarding the lack of transparency in data-sharing practices facilitated by third-party libraries.

Given the scale and dynamism of the mobile ecosystem, manual compliance assessment is inherently unfeasible. Automation is the only scalable solution to systematically evaluate mobile applications' privacy practices and detect non-compliant behaviors. This research proposes an automated compliance auditing framework, integrating dynamic analysis, privacy policy processing via LLMs, and privacy risk attribution to third-party libraries.

Since GDPR compliance applies uniformly across all 24 official EU languages, the development of multilingual analysis of privacy policies would enhance compliance assessment tools and would represent a substantial contribution to European regulatory enforcement. Moreover, extending this evaluation framework to the iOS ecosystem is a logical next step, ensuring that regulatory assessments cover the entire mobile device market. This would enable comprehensive, cross-ecosystem regulatory auditing across the mobile landscape.

To develop this framework at scale, securing funding through a European project under Horizon Europe would be a potential pathway. Given the regulatory nature of this research, a successful proposal would require the participation of data protection authorities (DPAs) such as the CNIL, AEPD, and potentially other European regulators, which could directly benefit from the project's results in their enforcement efforts. Additionally, collaborations with research institutions specializing in privacy engineering and legal informatics would be essential to ensure the interdisciplinary facet of the project. To maximize impact, the project could also involve non-profit organizations, consumer advocacy groups, or even industry partners interested in promoting compliance tools for app developers. Such a consortium would strengthen the proposal's alignment with European priorities on digital trust, privacy, and AI-driven regulatory enforcement.

A highly relevant funding opportunity would be a call similar to HORIZON-CL4-2024-DATA-01-01: AI-driven data operations and compliance technologies, which explicitly aims to support

companies and public sector entities in complying with existing and emerging regulations such as GDPR, the Data Governance Act, the Data Act, and the AI Act. The expected outcomes of this call highlight the need for solutions that enable citizens to trust that data-driven systems treat them fairly, respect their privacy and anonymity, and allow for better tracking of personal data use in a world where everything is increasingly digitalized.

An alternative would be to seek national funding through Spanish R+D+i programs, particularly with the participation of the AEPD as a regulatory partner. The AEPD has already expressed interest in automated compliance tools, and a national-scale project could serve as a pilot initiative that could later be expanded to a European level. Other potential collaborators at the national level include universities, research centers, and industry players in the mobile and cybersecurity sectors, which could contribute both technical expertise and real-world use cases.

Chapter 5

Conclusion

5.1 Fulfillment of the Thesis Objectives

The primary objective of this dissertation has been to develop automated methods and artifacts for assessing the regulatory compliance of mobile applications under the GDPR. This aim has been approached by comprehensively analyzing the technical behavior of applications, examining the practices declared in privacy policies and labels, and systematically comparing these two dimensions to detect potential non-compliance. Throughout the research, both the main objective and the specific goals have been achieved by implementing innovative tools and methods, validating them with ground-truth datasets, and employing them to generate empirical evidence.

5.1.1 Analysis of Mobile App Behavior

The thesis has developed and refined tools capable of capturing and analyzing mobile applications' actual behavior. One of the major milestones was reconstructing the traffic module, optimized for modern devices and enabling the interception and decryption of network connections—even in the presence of advanced security techniques such as certificate pinning. Integrated into a modular platform, this traffic module facilitated the collection of empirical evidence on actual transfers of personal data, thereby laying the groundwork for subsequent compliance-focused studies.

Additionally, the article *"ROI: A Method for Identifying Organizations Receiving Personal Data"* introduced an innovative solution for identifying organizations receiving personal data by correlating intercepted domains with external sources, such as WHOIS and automatically processed privacy policies. This served to precisely ascertain the recipient of personal data transfers.

The study *"Sharing is Not Always Caring"* significantly extended this analysis by employing a Frida-based method to determine which connections originated from third-party libraries. This progress provided valuable insights into third-party libraries' role in instances of noncompliance, demonstrating how developers' limited control over these libraries can result in unacknowledged

transfers of personal data. Indeed, the study revealed that around 73% of undocumented data transfers were attributable to third-party SDKs.

Taken together, these efforts satisfy the objective of investigating the behavior of applications by designing methods that both capture empirical evidence and identify technical patterns underlying noncompliance. The tools developed—such as the traffic module, ROI, and the dynamic Frida-based methods—have all been utilized in large-scale empirical studies, underscoring their applicability.

5.1.2 Evaluation of Declared Practices in Privacy Policies and Labels

Assessing the practices declared in privacy policies and labels has been central to this thesis, which has produced innovative automated methods for identifying and processing the information declared by developers in these documents. These methods range from conventional natural language processing approaches to advanced techniques leveraging LLMs, providing efficient, effective, and scalable solutions.

Initially, a supervised pre-trained classifier for automatically detecting valid privacy policies was presented in *"ROI: A Method for Identifying Organizations Receiving Personal Data."* Based on Support Vector Machines (SVM), this classifier filtered out irrelevant texts to ensure that only valid policies were analyzed in later stages. This first step was essential for guaranteeing the validity of documents used in compliance evaluations.

Subsequently, the thesis progressed toward automating the collection of privacy labels, a more recent transparency mechanism introduced on platforms such as Google Play and the App Store. In *"Comparing Privacy Label Disclosures of Apps Published in Both the App Store and Google Play Stores,"* a method was presented to automate the search and extraction of privacy labels using web-scraping techniques. These labels were collected and organized to facilitate systematic comparisons of declared practices across different mobile app marketplaces.

The most significant shift in privacy policy processing emerged with the use of LLMs, such as ChatGPT, to achieve automated, large-scale policy analyses. Consolidated in *"Large Language Models: A New Approach for Privacy Policy Analysis at Scale,"* this advancement presented an optimized, LLM-based method capable of identifying and categorizing various privacy practices with over 93% F1-score, validated on multiple ground-truth datasets. By employing an optimized prompt design, the model efficiently processed large volumes of policies, representing a major milestone in automating these analyses.

That LLM-based method was adapted and utilized in the study *"Data Retention Disclosures in the Google Play Store: Opacity Remains the Norm,"* where it was applied to identify statements related to data retention periods. In this work, the model automated the extraction and classification of information, evaluating whether policies provided precise details on retention periods or the criteria used to define them. This application demonstrated the flexibility of the LLM-based approach by effectively adapting to specific privacy practices while maintaining consistent performance.

Finally, in *"Sharing is Not Always Caring,"* additional methods were developed to pinpoint

which entities were declared in privacy policies as recipients of personal data. By merging policy analysis with automated techniques, the methods assessed whether the declared recipients matched the third parties detected in the applications' behavior. These methods laid a strong foundation for identifying discrepancies in declared practices, enabling automated regulatory compliance assessments.

Collectively, these studies fulfill the thesis's specific objectives related to the evaluation of declared practices in privacy policies and labels. The developed methods have automated the identification and extraction of key information necessary for GDPR compliance assessments.

5.1.3 Automation of Regulatory Compliance Evaluation

This subsection discusses the outcomes of automating regulatory compliance assessments in mobile apps, directly aligning with the overarching aim of this research. The first step entailed translating GDPR requirements into programmable, verifiable rules that could then be systematically applied via the designed artifacts and methods. By employing these rules, empirical evidence from both the analysis of application behavior and the declared practices in privacy policies and labels was integrated, enabling automated comparison and compliance evaluation.

The first work in this line is *"Automated GDPR Compliance Assessment for Cross-Border Personal Data Transfers in Android Applications,"* which focused on compliance with transparency requirements for international data transfers. Supervised classifiers were developed in a prior thesis to identify transfers disclosed in privacy policies and were correlated with the app's behavior captured via dynamic analysis, using the updated Traffic module. Approximately 32% of evaluated apps transmitted personal data internationally, and nearly half of these practices failed to comply with GDPR requirements due to inadequate disclosure of destination countries or implemented safeguards. These results were obtained using an automated method developed at the early stage of this thesis that compared the practices declared in privacy policies with each app's behavior, enabling compliance assessment.

In *"Sharing is Not Always Caring: Delving Into Personal Data Transfer Compliance in Android Apps,"* the appropriate disclosure of third parties in privacy policies was assessed. By integrating ROI and NLP techniques with ChatGPT, the study automatically mapped declared entities in policies to actual connections observed during dynamic analysis. The results were striking: over 81% of applications did not adequately disclose personal data recipients, while about 73% of undisclosed transfers stemmed from third-party libraries.

Evaluation of privacy labels was undertaken in *"Comparing Privacy Label Disclosures of Apps Published in Both the App Store and Google Play Stores,"* which analyzed discrepancies between privacy labels across the two ecosystems. This study revealed that 66.5% of apps showed notable inconsistencies between their disclosures on the two platforms, and that in many instances the labels did not accurately reflect actual practices, as observed through static analysis. These findings underscore the limitations of current transparency mechanisms and the need for more rigorous implementation.

Meanwhile, *"Data Retention Disclosures in the Google Play Store: Opacity Remains the Norm"*

deployed an automated, LLM-based method to assess whether privacy policies complied with data retention obligations. The approach validated an optimized prompt design to automatically analyze over 2,200 privacy policies, showing that more than 50% either failed to specify retention periods or to define relevant criteria, violating GDPR transparency standards. This study marks a significant extension of the method presented in *"Large Language Models for Privacy Policy Analysis,"* demonstrating its adaptability to new privacy practices and providing a robust tool for automated assessments.

Finally, *"Privacy Settings in Third-Party Libraries in Android Apps"* represents a high point in compliance evaluation by integrating all earlier contributions. This study introduced a dynamic, Frida-based method to monitor third-party SDK privacy settings during app execution, correlating these findings with developer-declared privacy policies and labels. The results revealed that most applications retained default SDK settings, leading to substantial discrepancies between declared practices and actual behavior. This work examined regulatory noncompliance and identified structural causes, such as developers failing to adjust the default privacy configurations of third-party libraries integrated into their apps.

Taken as a whole, these studies fulfill the thesis's primary objective: developing automated methods that translate and assess GDPR requirements by synthesizing evidence from both technical analyses and declared privacy practices. Large-scale evaluations have documented systematic noncompliance patterns in key areas such as international data transfers, data retention, and default SDK settings.

5.2 Limitations

Although substantial in its contributions, this thesis presents certain limitations stemming from its methodological scope and the tools employed. These constraints aim to offer a critical perspective on the work and delineate promising areas for future improvements.

1. **Coverage in Dynamic Analysis**. One principal limitation is that connection detection during dynamic analysis depends partly on how the applications are engaged with during experiments. The observed connections represent a lower bound of the problem, as additional data transfers may occur beyond what automated interactions via tools like Android Monkey can trigger. A more nuanced manual or advanced synthetic interaction might have uncovered further transfers, but such an approach would have compromised the focus on scalability essential for capturing a broad view of Google Play's ecosystem.
2. **Limitations of Statistical Models**. The methods used for analyzing privacy policies—chiefly supervised classifiers and LLMs—are fundamentally statistical. While these methods do not guarantee absolute effectiveness, they have been rigorously tested and validated with manually annotated datasets, ensuring high performance in precision, recall, and F1-score. For instance, models like ChatGPT achieved F1-scores exceeding 93% across multiple datasets. However, the statistical nature of these approaches introduces uncertainties; despite being mitigated by careful validation, these methods must be considered cautiously before drawing definitive conclusions about an application's legal compliance.

3. **Cost and Dependence on Proprietary Models**. Employing proprietary models such as ChatGPT proved highly efficient in automatically processing privacy policies, outperforming alternatives like LLaMA 2 in assessing specific practices. Nonetheless, reliance on proprietary models poses difficulties, including considerable operational costs and access restrictions, especially in academic or smaller-budget contexts. Although this thesis prioritized performance, the continuous development of advanced open-source models (e.g., new LLaMA or DeepSeek versions) suggests that regularly evaluating these emerging alternatives could be key to replacing proprietary solutions in the future. Such a transition would both reduce costs and bolster technological independence in implementing automated methods.
4. **Focus on the Android Ecosystem**. The thesis's evaluation concentrates on the Android ecosystem, particularly Google Play Store applications, justified by Android's global dominance as a mobile operating system. However, this leaves the iOS ecosystem unexplored, a comparison that would considerably expand the thesis's scope. Methods for analyzing privacy policies and labels are equally applicable to iOS apps, creating clear opportunities for pursuing an integrated examination and cross-ecosystem comparison in future work.
5. **Implications of Scalability and Generalization**. Finally, the methods were developed with scalability in mind, which can introduce constraints on customizing or adapting them for specialized use cases. Nonetheless, by prioritizing large-scale evaluations, the thesis offers an overarching perspective on the state of regulatory compliance in the mobile ecosystem, which remains its core objective.

5.3 Summary of Contributions

This thesis produced an extensive array of scientific and technical outcomes, combining both central research findings and supplementary contributions facilitated by the developed methods and tools. Among the direct outputs are four articles published in JCR-indexed journals—*Computers & Security (Q1)*, *Computing (Q2)*, and *IEEE Access (Q2)*—that established new metrics, methodologies, and approaches for evaluating the compliance of international data transfers and analyzing transparency in mobile app privacy policies. These direct efforts are accompanied by two contributions accepted at international workshops (IWPE23 and IWPE24), which dig into the effectiveness of privacy labels as transparency mechanisms and the analysis of personal data retention in the mobile ecosystem. Moreover, the acceptance of an article at the Privacy Enhancing Technologies Symposium (PETS 2025) underscores the global significance of this work in the privacy domain.

Indirect findings generated seven additional articles. These include studies published at IWPE22 and SECRYPT22, as well as one featured at IWPE23, all rooted in the analytical infrastructure of this thesis. In addition, a publication in *Computers & Security* (Q1) emerged from the analysis of anycast IP addresses on the platform, along with a publication in *IEEE Communications Magazine* (Q1), exploring the correlation between third-party libraries and international transfers through the use of anycast IPs. Furthermore, there are three more works: two in collaboration with ETH Zurich—one currently available as a preprint, pending

submission to a high-impact journal, and another underway (the eighth when completed)—and one co-developed with the Escuela Politécnica Nacional (EPN) in Quito, under review in *Artificial Intelligence and Law* (Q2) journal. These indirect studies, facilitated by the technological platform and methods created during the thesis, encompass everything from evaluating geolocation service reliability and detecting anycast IPs to analyzing third-party library behavior and applying LLMs to the automated annotation of policies. Collectively, they demonstrate the versatility and expansive capability of the thesis's contributions.

In addition to academic publications, this research fostered extensive international collaborations through research stays and joint projects with prestigious institutions. Continuous interaction with Carnegie Mellon University (CMU) throughout the thesis—culminating in a research stay in 2023—was key for applying LLMs to automated privacy policy analysis. Likewise, the collaboration with ETH Zurich led to cross-disciplinary studies merging legal and technical perspectives, and also gave rise to two ongoing works, one in preprint and another under development, focused on assessing how privacy policies adapt to different regulatory contexts. The research stay at King's College London in 2024 broadened the scope to automated evaluations of personalized chatbot systems, while the partnership with EPN resulted in a journal paper (under review) exploring LLM-based automatic annotation techniques comparable to expert human coders. These synergies, combined with interaction with the University of Warwick, shaped an international ecosystem of collaboration spanning technological, regulatory, and legal domains, thereby amplifying the research's impact and rigor.

The thesis demonstrated applicability in both industrial and regulatory spheres. HERE Europe B.V. requested audits based on the developed techniques to evaluate the transparency of its Google Play Store app, underscoring the capacity of these tools to guide tangible enhancements in GDPR compliance. Concurrently, the Spanish Data Protection Agency (AEPD) expressed early interest in the proposed platform and methods—testing them internally. Collaboration with Joseph A. Calandrino, then at the U.S. Federal Trade Commission (FTC), further emphasizes the work's regulatory dimension, highlighting its relevance for consumer protection authorities and international legal frameworks.

The academic-training aspect of this thesis has also been impactful. Supervision of seven dissertations (at both undergraduate and master's levels) aligned with refining the techniques presented, thereby contributing to a new generation of researchers with expertise in automated privacy policy analysis and related data protection questions. This teaching role expanded the adoption of the proposed methodologies and sparked fresh ideas and perspectives among emerging privacy and data protection specialists.

Widespread media coverage affirms the social relevance of these findings. Several major outlets, including *La Vanguardia*, *Computer Hoy*, and the global platform *The Conversation*, featured the results, fueling public debate on data handling transparency. Coverage extended to radio via mentions on *COPE* and television via *TreceTV*, with over a dozen additional media reports amplifying the thesis's reach and societal impact. Together with the open release of tools, annotated datasets, and methodologies in publicly accessible repositories, this media presence underscores the thesis's commitment to scientific dissemination, accountability, and promoting best practices in the realm of privacy.

5.4 Future Perspectives

This thesis has established the groundwork for automating compliance assessments in Android mobile applications, addressing both technical behavior and policy-level declarations. The developed methods indicate several promising future directions, such as implementing multilingual privacy practice evaluations using LLMs, aimed at analyzing policies across the 24 official EU languages. Beyond broadening geographic reach, these efforts will facilitate market-specific compliance checks under the GDPR.

Moreover, the proposed methods could readily extend to ecosystems like iOS, enabling comprehensive cross-platform comparisons and unified infrastructures for assessing mobile app compliance. Research on evaluating usage policies in personalized chatbots represents another compelling avenue, particularly in the context of rising generative AI, which raises new security and privacy challenges.

5.5 Final Reflection

This thesis constitutes a landmark in automating regulatory compliance evaluations for mobile applications by addressing hurdles in analyzing application behavior, scrutinizing privacy policies, and aligning these processes with GDPR requirements. Its contributions have documented patterns of noncompliance and illuminated underlying causes, proposing avenues to mitigate them. Taken together, these advancements position this research as a benchmark in privacy and transparency, with tangible impacts spanning academia, industry, and regulatory frameworks.

The significance of these findings lies in their potential to foster a more transparent mobile ecosystem aligned with data protection principles. The methods and tools developed offer viable solutions to today's challenges, but also open new possibilities for research and application in a global context. In an increasingly digital world, this thesis underlines the necessity of adopting technological solutions that uphold user trust and respect fundamental rights. It calls on researchers, developers, and regulators alike to embrace approaches and tools that promise a more transparent future—one that places privacy at the forefront.

References

Akash, A. R., Chithra, S., Vasuki, P., Shanmughapriya, T., & G., N. M. M. (2022). Towards privacy for android mobile applications. *2022 International Conference on Futuristic Technologies (INCOFT)*, 1–8. https://doi.org/10.1109/INCOFT55651.2022.10094512

Al-Sharafi, M. A., Al-Emran, M., Tan, G. W. H., & Ooi, K.-B. (Eds.). (2024). *Current and future trends on intelligent technology adoption* (Vol. 1161). Springer. https://doi.org/10.1007/978-3-031-61463-7

Balebako, R., Marsh, A., Lin, J., Hong, J., & Cranor, L. F. (2014). The privacy and security behaviors of smartphone app developers. *Symposium on Usable Security and Privacy*, 1–10. https://doi.org/10.14722/usec.2014.23006

Carrillo, A. J., & Jackson, M. (2022). Follow the leader? a comparative law study of the eu's general data protection regulation's impact in latin america. *ICL Journal*, *16*(2), 177–262. https://doi.org/10.1515/icl-2021-0037

Cevallos-Salas, D., Estrada-Jiménez, J., Guamán, D. S., & Rodríguez, D. (2025, January). Gpt vs human legal texts annotations: A comparative study with privacy policies [Preprint (Version 1) available at Research Square]. https://doi.org/10.21203/rs.3.rs-5799153/v1

Comunidad de Madrid. (2020). *Estrategia madrileña de investigación e innovación 2030 (em2i)* [Accessed: Feb. 25, 2025]. Dirección General de Investigación e Innovación Tecnológica.

Cozar, M., Rodriguez, D., Del Alamo, J. M., & Guaman, D. (2022). Reliability of ip geolocation services for assessing the compliance of international data transfers. *2022 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, 181–185. https://doi.org/10.1109/EuroSPW55150.2022.00024

de Castro, L. T. (2024). *Data analysis over gdpr compliance article 13 with chatgpt* (Version V1). Mendeley Data. https://doi.org/10.17632/3dkh7f7tnh.1

Eskhita, R., & Stamhuis, E. (2024). The influence of the brussels effect on the interpretation of data protection laws in the gulf. *Global Journal of Comparative Law*, *13*(2), 261–278.

European Commission. (2020). Horizon europe: The eu research and innovation framework programme 2021-2027 [Accessed: Feb. 25, 2025].

European Union. (2016a). Regulation (eu) 2016/679 of the european parliament and of the council of 27 april 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/ec (general data protection regulation) [Accessed: Feb. 25, 2025]. *Official Journal of the European Union*, *L119*, 1–88.

European Union. (2016b). Regulation (eu) 2016/679 of the european parliament and of the council of 27 april 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/ec (general data protection regulation) [Accessed: Feb. 25, 2025]. *Official Journal of the European Union*, *L119*, 1–88.

Federal Trade Commission. (2024, October). FTC Takes Action Against Marriott and Starwood Over Multiple Data Breaches [Accessed: Feb. 25, 2025].

GlobalStats, S. (2023). Mobile operating system market share worldwide [Accessed: Feb. 25, 2025].

Guamán, D. S., Rodriguez, D., del Alamo, J. M., & Such, J. (2023). Automated gdpr compliance assessment for cross-border personal data transfers in android applications. *Computers & Security*, *130*, 103262. https://doi.org/https://doi.org/10.1016/j.cose.2023.103262

Guamán Loachamín, D. S. (2022, January). *Contribution to software quality control techniques for assessing privacy and data protection* [Unpublished], Telecomunicacion. https://doi.org/10.20868/UPM.thesis.69641

Hao, X., Ma, D., & Liang, H. (2023). Detection and privacy leakage analysis of third-party libraries in android apps. *Security and Privacy in Communication Networks*, *462*, 569–587. https://doi.org/10.1007/978-3-031-25538-0_30

Horstmann, S. A., Domiks, S., Gutfleisch, M., Tran, M., Acar, Y., Moonsamy, V., & Naiakshina, A. (2024). "those things are written by lawyers, and programmers are reading that." mapping the communication gap between software developers and privacy experts. *Proceedings on Privacy Enhancing Technologies*, *2024*(1), 151–170. https://doi.org/10.56553/popets-2024-0010

Islam, M. T., Sahula, M., & Karim, M. E. (2022). Understanding gdpr: Its legal implications and relevance to south asian privacy regimes. *UUM Journal of Legal Studies*, *13*(1), 45–76. https://doi.org/10.32890/uumjls2021.13.1.3

Jain, A., Rodriguez, D., del Alamo, J. M., & Sadeh, N. (2023). ATLAS: Automatically Detecting Discrepancies Between Privacy Policies and Privacy Labels. *2023 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, 94–107. https://doi.org/10.1109/EuroSPW59978.2023.00016

Kelley, P. G., Cranor, L. F., & Sadeh, N. (2013). Privacy as part of the app decision-making process. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 3393–3402. https://doi.org/10.1145/2470654.2466466

Ketelaar, P. E., & van Balen, M. (2017). The smartphone as your follower: The role of smartphone literacy in the relation between privacy concerns, attitude and behaviour towards phone-embedded tracking. *Computers in Human Behavior*, *78*, 174–182. https://doi.org/10.1016/j.chb.2017.09.034

Khandelwal, R., Nayak, A., Chung, P., Fawaz, K., Bianchi, A., Celik, Z. B., & Hussain, S. R. (2024). Unpacking privacy labels: A measurement and developer perspective on google's data safety section. *33rd USENIX Security Symposium (USENIX Security 24)*, 2831–2848.

Lu, H., Liu, Y., Liao, X., & Xing, L. (2024). Towards privacy-preserving social-media sdks on android. *33rd USENIX Security Symposium (USENIX Security 24)*, 647–664.

Mahieu, R., Asghari, H., Parsons, C., van Hoboken, J., Crete-Nishihata, M., Hilts, A., & Anstis, S. (2021). Measuring the brussels effect through access requests: Has the

european general data protection regulation influenced the data protection rights of canadian citizens? *Journal of Information Policy*, *11*, 301–349.

marty0678. (2023). Google play unofficial python api [Accessed: Feb. 25, 2025].

Memon, Z. A., Munawar, N., & Kamal, M. (2024). App store mining for feature extraction: Analyzing user reviews. *Acta Scientiarum. Technology*, *46*(1). https://doi.org/10.4025/actascitechnol.v46i1.62867

Ministerio de Ciencia e Innovación. (2020). *Estrategia española de ciencia, tecnología e innovación 2021-2027* [Accessed: Feb. 25, 2025].

Pantoja, M. M., Rodríguez, D., & del Alamo, J. M. (2024). *Esit-101* (Version V1). Mendeley Data. https://doi.org/10.17632/3cy5cv56bv.1

Pascual, H., del Alamo, J. M., Rodriguez, D., & Dueñas, J. C. (2024). Hunter: Tracing anycast communications to uncover cross-border personal data transfers. *Computers & Security*, *141*, 103823. https://doi.org/https://doi.org/10.1016/j.cose.2024.103823

Pascual, H., del Alamo, J. M., Rodriguez, D., & Dueñas, J. C. (2025). Anycast and third-party libraries: A recipe for a privacy disaster? [Forthcoming. Accepted for publication in IEEE Communications Magazine.]. *IEEE Communications Magazine*.

Pollach, I. (2006). Privacy statements as a means of uncertainty reduction in www interactions. *Journal of Organizational and End User Computing (JOEUC)*, *18*(1), 23–49. https://doi.org/10.4018/joeuc.2006010102

Rodriguez, D. (2025). Privacysdksettingsanalyzer [Accessed: Feb. 25, 2025].

Rodriguez, D., Alamo, J. M. D., Cozar, M., & García, B. (2024). Roi: A method for identifying organizations receiving personal data. *Computing*, *106*(1), 163–184. https://doi.org/10.1007/s00607-023-01209-2

Rodriguez, D., Calandrino, J. A., Alamo, J. M. D., & Sadeh, N. (2025). Privacy settings of third-party libraries in android apps: A study of facebook sdks [Forthcoming]. *Proceedings on Privacy Enhancing Technologies*.

Rodriguez, D., Cozar, M., & Alamo, J. M. D. (2022). Identifying organizations receiving personal data in android apps [Poster presentation]. *Proceedings of the 19th International Conference on Security and Cryptography (SECRYPT 2022) - Poster Session*.

Rodriguez, D., Del Alamo, J. M., Fernández-Aller, C., & Sadeh, N. (2024). Sharing is not always caring: Delving into personal data transfer compliance in android apps. *IEEE Access*, *12*, 5256–5269. https://doi.org/10.1109/ACCESS.2024.3349425

Rodriguez, D., Fernández, C., del Alamo, J. M., & Sadeh, N. (2024). *Data retention period disclosures in privacy policies* (Version V2). Mendeley Data. https://doi.org/10.17632/c4x958pzpm.2

Rodriguez, D., Jain, A., Alamo, J. M. D., & Sadeh, N. (2023). Comparing privacy label disclosures of apps published in both the app store and google play stores. *2023 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, 150–157. https://doi.org/10.1109/EuroSPW59978.2023.00022

Rodriguez, D., Yang, I., Alamo, J. M. D., & Sadeh, N. (2024). Large language models: A new approach for privacy policy analysis at scale. *Computing*, *106*(12), 3879–3903. https://doi.org/10.1007/s00607-024-01331-9

Rodríguez, D., del Alamo, J. M., Cozar, M., & García, B. (2023). *ROI: A method for identifying organizations receiving personal data* (Version V1). Mendeley Data. https://doi.org/10.17632/3mdyg53c94.1

Rodríguez, D., Fernández-Aller, C., Del Alamo, J. M., & Sadeh, N. (2024). Data retention disclosures in the google play store: Opacity remains the norm. *2024 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, 19–23. https://doi.org/10.1109/EuroSPW61312.2024.00009

Sailaja, N., & Jones, R. (2017). Industry ideals and barriers in using alternative privacy policies. *Proceedings of the 31st International BCS Human Computer Interaction Conference (HCI 2017)*. https://doi.org/10.14236/ewic/HCI2017.12

Sipior, J. C., Ward, B. T., & Volonino, L. (2014). Privacy concerns associated with smartphone use. *Journal of Internet Commerce*, *13*(3–4), 177–193. https://doi.org/10.1080/15332861.2014.947902

Stempel, J. (2025). Lawsuit accuses amazon of secretly tracking consumers through cellphones [Accessed: Feb. 25, 2025]. *Reuters*.

Story, P., Zimmeck, S., Ravichander, A., Smullen, D., Wang, Z., Reidenberg, J., & Sadeh, N. (2019). Natural language processing for mobile app privacy compliance. *AAAI Spring Symposium on Privacy-Enhancing Artificial Intelligence and Language Technologies*, *2*(4), 4.

Trautman, L. J., & Ormerod, P. C. (2017). Corporate directors' and officers' cybersecurity standard of care: The yahoo data breach [Accessed: Feb. 25, 2025]. *American University Law Review*, *66*(5).

United Nations. (2015). Sustainable development goals (sdgs) [Accessed: Feb. 25, 2025].

Wang, H., Guo, Y., Ma, Z., & Chen, X. (2015). Wukong: A scalable and accurate two-phase approach to android app clone detection. *Proceedings of the International Symposium on Software Testing and Analysis*, 71–82. https://doi.org/10.1145/2771783.2771795

Zac, A., Wey, P., Bechtold, S., Rodriguez, D., & Alamo, J. M. D. (2024, February). The court speaks, but who listens? automated compliance review of the gdpr [Available at SSRN: https://ssrn.com/abstract=4709913 or http://dx.doi.org/10.2139/ssrn.4709913].