A. PROPUESTA PROYECTO DE INVESTIGACIÓN

1. TIPO DE PROYECTO:

Interno	**X**	**Grupal**	
Semilla		**Multidisciplinario**	

2. TIPO DE INVESTIGACIÓN:

Básica		**Aplicada**	**X**

3. UNIDAD EJECUTORA *(Departamento, Instituto o Estructura de Investigación)*:

1. Departamento de Electrónica, Telecomunicaciones y Redes de Información.

4. LINEA(S) DE INVESTIGACIÓN:

1.Privacidad y seguridad

5. TÍTULO DEL PROYECTO *(mínimo 10 palabras)*:

Mejora de las técnicas de evaluación de la *integridad contextual* de la privacidad en las aplicaciones móviles.

6. RESUMEN *(máximo 200 palabras)*

La ubicuidad de las aplicaciones móviles permite que ciertos datos personales de los usuarios sean fácilmente recogidos y transmitidos globalmente hacia diferentes proveedores de servicios en un ecosistema complejo. Esto supone grandes riesgos para la privacidad de los usuarios, como el tratamiento no autorizado de sus datos personales para generar perfiles de comportamiento. La *privacidad como integridad contextual* es un marco de referencia para evaluar potenciales violaciones a la privacidad, detectando aquellos flujos de datos personales que no se ajustan a los flujos normativos definidos en un contexto específico. Aunque se han realizado importantes esfuerzos de investigación en este ámbito, la mayoría se han enfocado en evaluar a la *privacidad como confidencialidad* (p.ej. fugas de datos).

Para avanzar hacia la evaluación de la *privacidad como integridad contextual*, este proyecto proveerá un conjunto de componentes de software (artefactos y herramientas) de soporte para diseñar pruebas que permitan evaluar las aplicaciones móviles más allá de las meras fugas de datos. Nos enfocamos en (1) las técnicas para extraer flujos normativos desde las políticas de privacidad, que son el principal medio para divulgar las prácticas de privacidad de las aplicaciones y (2) las técnicas para la generación automática de eventos de usuario en las aplicaciones.

7. PALABRAS CLAVE *(4-6)*

Privacidad, protección de datos, calidad de software, pruebas de software, validación, verificación.

8. OBJETIVOS

8.1. OBJETIVO GENERAL

Desarrollar artefactos y herramientas para el diseño de pruebas de evaluación de la integridad contextual de la privacidad en las aplicaciones móviles.

8.2. OBJETIVOS ESPECÍFICOS

a. Analizar las técnicas de análisis de políticas de privacidad escritas en lenguaje natural que han sido reportadas en el estado del arte.
b. Generar un conjunto de artefactos y herramientas para el análisis y extracción de flujos normativos relevantes que han sido definidos en las políticas de privacidad y que aún no han sido abordadas en el estado del arte.
c. Desarrollar una herramienta para la generación automática de eventos de usuario sobre las aplicaciones móviles.
d. Validar los artefactos y herramientas desarrollados a través de la evaluación de la integridad contextual de los flujos normativos seleccionados en un conjunto de aplicaciones móviles.
e. Divulgar la temática investigada y difundir los resultados científicos obtenidos.

9. HIPÓTESIS (opcional)

N/A

10. DETALLE DE LOS RESULTADOS ESPERADOS *(con relación a los objetivos)*

a. Informe sobre los diferentes enfoques de análisis de políticas de privacidad escritas en lenguaje natural que han sido reportadas en el estado del arte. Este incluirá un análisis en profundidad de las técnicas (p.ej. las técnicas basadas en aprendizaje automático y procesamiento de lenguaje natural), las prácticas de privacidad o flujos normativos abordados (p.ej. recolección de datos personales o compartición de datos personales), el objetivo del análisis (p.ej. evaluación de cumplimiento de una política de privacidad con una regulación) y el enfoque de evaluación de la técnica (p.ej., qué métricas de efectividad se han usado).
b. Un conjunto de artefactos y herramientas que permitirán el análisis y extracción de flujos normativos que todavía no han sido abordados en el estado del arte. Por un lado, los artefactos generados incluirán, por ejemplo, el o los corpus de anotaciones de los flujos normativos abordados. Por otro lado, las herramientas incluirán los componentes de software que facilite la anotación de los flujos normativos en las políticas de privacidad escritas en lenguaje natural.
c. Un prototipo de herramienta para la generación automática de eventos de usuario sobre las aplicaciones móviles Android.
d. Un reporte sobre la validación de los artefactos y herramientas desarrolladas a través de un experimento controlado realizado sobre un conjunto de aplicaciones populares y gratuitas de la tienda oficial de aplicaciones Android.
e. El envío de un artículo a una revista indexada en Scopus o WoS reportando los resultados científicos alcanzados.
f. La divulgación de la temática, su necesidad e importancia, y de los resultados alcanzados a través de charlas abiertas.

11. IMPACTO DE LA INVESTIGACIÓN (*científico, social, económico u otros*)

Este proyecto de investigación contribuye al área de las técnicas de control de calidad de software, enfocándose en la evaluación de atributos relacionados a la privacidad y protección de datos personales en las aplicaciones móviles. Por un lado, las técnicas de control de calidad de software son requeridas por las organizaciones para evaluar si

diversos artefactos de software cumplen con requisitos de privacidad exigidos por ciertas legislaciones vigentes en materia de protección de datos. Estas técnicas también son necesarias para las autoridades de control, quienes se encargan de garantizar el cumplimiento de estos requisitos. Por otro lado, los teléfonos inteligentes y las aplicaciones móviles se han convertido en sistemas de software del lado del usuario que han ganado tremenda popularidad en los últimos años y que son usados como soporte de un sinnúmero de servicios y negocios. No obstante, la ubicuidad de estos dispositivos plantea diversos riesgos a la privacidad de las personas. En este contexto, el impacto de este proyecto de investigación es:

- a nivel de las personas, proveyendo y validando técnicas de evaluación para proteger su derecho a la privacidad y protección de datos,
- a nivel económico, proveyendo a las organizaciones de técnicas de control de calidad para evaluar aspectos de privacidad con el fin de evitar violaciones a la privacidad y potenciales penalizaciones y multas,
- a nivel de ingeniería, proveyendo a los ingenieros artefactos y herramientas que puedan incorporar en sus procesos de evaluación de aplicaciones móviles,
- a nivel científico, con la mejora de las técnicas de diseño de pruebas de evaluación de privacidad y protección de datos, con la conformación de nuevas líneas de investigación en la EPN, con la colaboración con investigadores internacionales, y con el afianzamiento de redes de investigación en el área,
- a nivel docente, con la divulgación de charlas a nivel de pregrado y posgrado.

La privacidad sienta las bases para que los individuos tomen el control de sus datos personales (Westin, 1990). Por un lado, los usuarios deben poder tomar decisiones sobre sus datos personales basados en sus intereses (Clarke, 2006); por otro, los proveedores de servicios deben respetar esas decisiones. Este último punto es llamativo, ya que, por ejemplo, los proveedores de aplicaciones tienen intereses comerciales en procesar el mayor número de datos posible. Teniendo en cuenta la relevancia económica de los datos y la emergente economía de los datos (Deloitte, 2020), es crucial **garantizar el derecho a la privacidad y protección de los datos personales de las personas**. Esto incluye tener en cuenta los aspectos de la privacidad a lo largo del diseño y desarrollo de una aplicación móvil (privacidad desde el diseño), así como ofrecer métodos y técnicas de evaluación para verificar su cumplimiento.

Esta investigación precisamente contribuye en las técnicas de control de calidad de software para evaluar el cumplimiento de aspectos de privacidad y protección de datos en las aplicaciones móviles. El uso de los smartphones y de las aplicaciones ha experimentado un crecimiento constante en los últimos años; por ejemplo, en Ecuador el Instituto Ecuatoriano de Estadísticas y Censos (INEC) reporta que el uso de los smartphones se ha incrementado en aproximadamente 32% en el período 2014-2019 (INEC, 2019). En ciertas regiones, como la Unión Europea (UE), el ecosistema móvil ha generado un impacto total en su economía en alrededor de 187.200 millones de Euros (Deloitte, 2020). Dada la relevancia económica de los smartphones y aplicaciones para las organizaciones, también es esencial asegurar el cumplimiento de requisitos de privacidad para evitar caer en penalizaciones o multas establecidas por ciertas regulaciones, como el Reglamento General de Protección de Datos (RGDP) (European Commission, 2016). El RGDP tiene alcance global, por lo tanto, podría aplicar incluso a organizaciones ecuatorianas ofreciendo algún servicio dentro de la UE. Además, ampliamente alineado con el RGDP, el proyecto de Ley de Protección de datos Personales del Ecuador (Asamblea Nacional del Ecuador, 2019), presentado el 19 de septiembre de 2019, también establece un conjunto de requisitos cuyo incumplimiento

podría significar sanciones económicas de hasta el 17% del volumen total del negocio. En este sentido, esta investigación podría tener un gran **impacto a nivel económico** local y globalmente.

Esta investigación también tiene un **impacto en el proceso de ingeniería de software**, concretamente en el desarrollo de aplicaciones móviles. El ecosistema de las aplicaciones móviles es una mezcla de requerimientos formales establecidos por regulaciones de protección de datos vigentes y la formación informal de los desarrolladores. Los desarrolladores de aplicaciones son desde aficionados hasta profesionales con experiencia que trabajan en grandes organizaciones (Stack Overflow, 2020). Como señala Balebako et al. (2015), mientras que las grandes empresas son capaces de formar equipos multidisciplinares para asegurar el cumplimiento de los requisitos legales, los desarrolladores de pequeñas empresas pueden tener dificultades para entender las implicaciones de su código en materia de privacidad y protección de datos. Así pues, disponer de técnicas y herramientas que puedan integrarse en los procesos de evaluación supone una ayuda para los desarrolladores a la hora de evaluar el cumplimiento de los requisitos de privacidad y protección de las aplicaciones móviles antes de su distribución.

El estado del arte en el dominio de las técnicas de control de calidad del software, orientadas a la evaluación de los aspectos de privacidad y protección de datos, reporta avances significativos, aunque la mayoría de ellos se centran en la evaluación de la privacidad como confidencialidad. Estas técnicas son insuficientes para evaluar aspectos que son dependientes del contexto. Este proyecto de investigación persigue la creación de ciertos artefactos y herramientas necesarios para evaluar la integridad contextual de la privacidad en las aplicaciones móviles, yendo un paso más allá del estado del arte y generando así un **impacto a nivel científico**. Además, nuestras contribuciones están orientadas a sumar a los esfuerzos de investigación llevados a cabo por investigadores internacionales con quienes hemos colaborado, para así crear y fortalecer estos lazos de colaboración. Esto, junto con las publicaciones en revistas de alto impacto, permitirá la consolidación de nuevas líneas de investigación en el campo de la ingeniería de la privacidad en la EPN.

Finalmente, se espera tener un **impacto en la docencia** a través de la divulgación del marco conceptual y resultados obtenidos a través de sesiones en las asignaturas de pregrado y postgrado y de charlas magistrales abiertas.

12. ESTADO DEL ARTE, E INVESTIGACIONES PREVIAS DEL EQUIPO *(máximo tres carillas)*

Estado del arte

Los smartphones se han convertido en un elemento esencial tanto para los usuarios (para organizar su vida privada y profesional) como para las organizaciones (para proveer una interfaz de acceso a sus servicios), aunque paralelamente plantean importantes riesgos a la privacidad de los usuarios y riesgos de incumplimiento a las organizaciones. Por un lado, los smartphones mejoran enormemente la comodidad del usuario permitiéndole acceder a los datos de, por ejemplo, su agenda o calendario. Adicionalmente, pueden extender la funcionalidad de su dispositivo mediante un sinnúmero de aplicaciones ("apps") provistas por diversos desarrolladores y

proveedores en un ecosistema generalmente abierto[1]. Además, las apps forman parte de los procesos de negocio en muchas industrias, como la de los viajes, el comercio, la banca, la salud y el entretenimiento (Deloitte, 2020). Por otro lado, la naturaleza portable de los smartphones, así como las particularidades de los ecosistemas de desarrollo y distribución de las apps son factores que indicen para que surjan ciertos riesgos de violación de la privacidad de los sus usuarios y de incumplimientos por parte de las organizaciones (Article 29 Data Protection Working Party, 2013; ENISA, 2017).

La evaluación de cumplimiento de aspectos de privacidad y protección de datos es una prioridad para los proveedores de aplicaciones, ya que determinadas regulaciones conllevan importantes sanciones económicas en caso de que se produzcan violaciones de la privacidad. Por ejemplo, en el Ecuador, el Proyecto de Ley Orgánica de Protección de Datos Personales (PLOPD) plantea sanciones económicas de hasta el 17% del volumen total del negocio (Asamblea Nacional del Ecuador, 2019). Sin embargo, evaluar las apps con respecto a los requisitos legales es un reto para las organizaciones y desarrolladores. Los flujos de datos personales de las aplicaciones suelen ser opacos, es decir, los tipos de datos personales que se transfieren, los destinatarios y los países de destino no son fácilmente visibles (Razaghpanah et al., 2018), ni siquiera para los desarrolladores cuando se llevan a cabo mediante bibliotecas de terceros (Balebako et al., 2015). Además, sigue existiendo un gran desfase entre los requisitos legales y la traducción de estos requisitos en soluciones prácticas, y también se necesitan herramientas para probar, verificar y auditar las aplicaciones, bibliotecas y servicios existentes (ENISA, 2017).

Estudios recientes del estado del arte (Guamán et al., 2020; Sadeghi et al., 2017) señalan que se han realizado importantes esfuerzos de investigación sobre técnicas y métodos de evaluación de la privacidad y protección de datos en las aplicaciones móviles. Se puede distinguir que estos esfuerzos se han centrado en dos paradigmas distintos de la privacidad: la *privacidad como confidencialidad* y la *privacidad como integridad contextual,* aunque la mayoría se han centrado en el primero.

La *privacidad como confidencialidad* se basa en el criterio binario de que cualquier divulgación de un dato personal es una violación de la privacidad (Gürses, 2014). Es decir, se asume que no hay otros actores con distintos intereses o necesidades de recoger datos personales. Algunas de las técnicas incluidas en esta categoría detectan una violación de la privacidad cuando una aplicación simplemente accede a datos personales en el dispositivo móvil, aunque no se envíen necesariamente fuera del dispositivo (Tiwari et al., 2019), otras lo hacen cuando los datos salen efectivamente del dispositivo pero no tienen en cuenta quién es el destinatario (Alkhattabi et al., 2020). En el caso de Android, Google Play Protect (GPP) también se alinea con este paradigma y persigue detectar posibles comportamientos perjudiciales en el ecosistema Android, incluida la divulgación de datos personales fuera del dispositivo a través de Spyware (Android Developers, 2020). En general, estas técnicas por sí mismas son útiles para advertir de (posibles) fugas de datos personales, pero fallan cuando determinados datos personales son realmente esperados en un contexto particular.

Por otro lado, la idea de que los individuos no siempre actúan de forma aislada, sino que realizan transacciones en una variedad de contextos sociales se conoce como

[1] Android es la plataforma abierta dominante en este mercado con una cuota cercana al 85% (IDC, 2021).

integridad contextual (Nissenbaum, 2004). Cada uno de estos contextos tiene un conjunto diferente de normas que equilibran los múltiples intereses, prescribiendo para ello flujos de datos personales apropiados o inapropiados en función del contexto. A estos flujos se los conoce como flujos normativos de datos personales o simplemente prácticas de privacidad. Por ejemplo, en un servicio basado en la localización como Google Maps, el proveedor de la aplicación debe recoger la ubicación del individuo para que funcione correctamente y le proporcione el servicio. En ese contexto particular, el flujo normativo de datos personales desde el dispositivo móvil un individuo hacia el proveedor de la aplicación es apropiado. Sin embargo, la revelación de la ubicación del individuo a otras partes que no participan en la prestación del servicio (por ejemplo, los proveedores de publicidad) no es apropiada y puede dar lugar a una violación de la privacidad. Hemos observado que algunos trabajos recientes se basan en las políticas de políticas de privacidad (Andow et al., 2020), en la normativa (Eskandari, et al., 2017; Zimmeck et al., 2019) o en ambas (Guamán et al., 2021) como fuente para extraer los flujos de datos personales (in) apropiados que luego son usados como oráculos en las pruebas de una aplicación. Así, posteriormente, estas técnicas extraen los flujos de datos personales desde las aplicaciones analizando diferentes artefactos de software a través de técnicas de análisis estático (Zimmeck et al., 2019), análisis dinámico (Andow et al., 2020; Guamán et al., 2021) o híbridas (Eskandari et al., 2017) . Finalmente, se evalúa la alineación de los flujos normativos con los flujos de las aplicaciones para detectar violaciones de la privacidad en caso de discrepancia.

A pesar de los esfuerzos de investigación en la evaluación de la integridad referencial de la privacidad, estos son apenas recientes en el ecosistema de las aplicaciones móviles y aún quedan desafíos por tratar (Guamán et al., 2020). Por un lado, las técnicas de extracción de algunos flujos normativos desde las políticas de privacidad usan enfoques manuales y no sistemáticos (Hatamian et al., 2019; Papageorgiou et al., 2018), todavía no han cubierto ciertos flujos normativos relevantes, o no han cubierto prácticas de privacidad escritas en otros idiomas diferente al inglés (Gallé et al., 2019). Por otro lado, la extracción de los flujos de datos personales reales desde las aplicaciones requiere la generación de eventos del usuario que disparen los flujos de datos relevantes. Para ello, los enfoques actuales reutilizan soluciones orientadas a la evaluación de aspectos funcionales o de seguridad de una aplicación (Kong et al., 2019), más que a la evaluación de privacidad y protección de datos.

En este escenario, se necesitan contribuciones para avanzar en la evaluación de la integridad contextual de la privacidad en las aplicaciones móviles. Este proyecto de investigación contribuye en esta línea a través de mejoras en las técnicas de extracción de flujos (normativos) de datos personales tanto desde de las políticas de privacidad como desde las aplicaciones móviles.

El área por explorar va en línea con nuestras investigaciones y conocimiento previo, es decir, la evaluación de la calidad de software enfocándose en los aspectos de privacidad y protección de datos, y se centra además en las aplicaciones móviles que es un tipo de software esencial en un sinnúmero de servicios del ecosistema digital actual.

Publicaciones previas relacionadas con el proyecto

Publicaciones previas en revistas de alto impacto (JCR Q1):

- Guamán, 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. https://doi.org/10.1109/ACCESS.2021.3053130
- Guamán, D. S., Del Alamo, J. M., & Caiza, J. C. (2020). A Systematic Mapping Study on Software Quality Control Techniques for Assessing Privacy in Information Systems. IEEE Access, 8, 74808–74833. https://doi.org/10.1109/access.2020.2988408
- Caiza, J. C., Martin, Y.-S., Guamán, D. S., Del Alamo, J. M., & Yelmo, J. C. (2019). Reusable Elements for the Systematic Design of Privacy-Friendly Information Systems: A Mapping Study. IEEE Access, 7, 66512–66535. https://doi.org/10.1109/ACCESS.2019.2918003

Publicaciones previas en congresos internacionales de relevancia (Ranking CORE[2] e indexadas en Scopus)

- Caiza, J. C., Del Alamo, J. M., & Guamán, D. S. (2020). A framework and roadmap for enhancing the application of privacy design patterns. The 35th ACM/SIGAPP Symposium On Applied Computing, 1297–1304. https://doi.org/10.1145/3341105.3375768
- Guamán, D. S., del Alamo, J. M., Veljanova, H., Reichmann, S., & Haselbacher, A. (2019, July). Value-Based Core Areas of Trustworthiness in Online Services. In IFIP International Conference on Trust Management (pp. 81-97). Springer, Cham.
- Guamán, D. S., del Alamo, J. M., Veljanova, H., Haselbacher, A., & Caiza, J. C. (2019). Ranking Online Services by the Core Areas of Trustworthiness. Revista Ibérica de Sistemas e Tecnologias de Informação, (E19), 465-478.
- Colesky, M., & Caiza, J. C. (2019). A System of Privacy Patterns for Informing Users: Creating a Pattern System. Proceedings of the 23rd European Conference on Pattern Languages of Programs - EuroPLoP '18, 1–11. https://doi.org/10.1145/3282308.3282325
- Colesky, M., Caiza, J. C., Del Álamo, J., Hoepman, J.-H., & Martín, Y.-S. (2018). A system of privacy patterns for user control. 33rd Annual ACM Symposium on Applied Computing, 1150–1156. https://doi.org/10.1145/3167132.3167257
- Del Alamo, J. M., Martín, Y. S., & Caiza, J. C. (2018). Towards organizing the growing knowledge on privacy engineering. In IFIP Advances in Information and Communication Technology (Vol. 526, pp. 15–24). https://doi.org/10.1007/978-3-319-92925-5_2
- Caiza, J. C., Martín, Y.-S., Del Alamo, J. M., & Guamán, D. S. (2017). Organizing design patterns for privacy: A taxonomy of types of relationships. Proceedings of the 22nd European Conference on Pattern Languages of Programs - EuroPLoP '17, 1–11. https://doi.org/10.1145/3147704.3147739

Proyectos de investigación relacionados en los que se ha colaborado:

- **TRUESSEC.eu (2017-2019): TRUst-Enhancing certified Solutions for SEcurity and protection of Citizens' rights in digital Europe (Acuerdo No.731711)**. Programa de investigación e innovación Horizonte 2020 de la Unión Europea. Este proyecto exploró la situación, las barreras y los beneficios de la certificación y el etiquetado de los atributos de confianza en los productos y servicios de las TIC, enfocándose en los aspectos de privacidad y protección de datos desde una perspectiva multidisciplinar.
  Se colaboró con el equipo de la Universidad Politécnica de Madrid en los estudios de soporte del proyecto desde la perspectiva de ingeniería. El resultado fue la elaboración de 5 entregables y la publicación de dos artículos en conferencias previamente citadas, así como también el capitulo de un libro.

- **PrivApp (2018-2019): Privacidad y protección de datos en el tratamiento de datos personales en las aplicaciones móviles.** La Agencia Española de Protección de Datos

[2] Ranking de conferencias internacionales http://portal.core.edu.au/conf-ranks/

ha financiado este proyecto para investigar la recolección y uso de datos personales por parte de aplicaciones móviles educativas y salud.
Se colaboró con el equipo de la Universidad Politécnica de Madrid en la definición de los criterios para evaluar las prácticas de recolección de datos personales en las aplicaciones móviles. El resultado fue la elaboración de un reporte sobre el estado de las aplicaciones estudiadas.

- **CLIIP (2020-2022): Investigación en Ingeniería de Privacidad.** Este proyecto aborda la investigación en aspectos de ingeniería de privacidad y protección de datos, incluyendo un conjunto de soluciones que ayuden a las organizaciones a introducir consideraciones de privacidad y protección de datos personales en sus procesos de ingeniería de una forma sistemática y eficiente.
  Se ha colaboró con el equipo de la Universidad Politécnica de Madrid en la revisión del estado del arte de las técnicas de control de calidad de software enfocándose en los aspectos de privacidad y protección de datos. El resultado ha sido publicado en uno de los artículos de revista de alto impacto previamente citado.
- **PDP4E (2018 - 2021): Methods and tools for GDPR (General Data Protection Regulation) compliance through Privacy and Data Protection Engineering (Acuerdo No.787034)**. Programa de investigación e innovación Horizonte 2020 de la Unión Europea. Este proyecto busca proporcionar a los ingenieros con métodos y herramientas para la creación de sistemas que cumplan con el RGPD, en particular en lo referente a la privacidad y protección de datos desde el diseño.
  Se colaboró con el equipo de la Universidad Politécnica de Madrid en la revisión y análisis del estado del arte de los elementos de diseño (p.ej. métodos, modelos, herramientas). El resultado ha sido publicado en uno de los artículos de revista de alto impacto previamente citados.

13. DESCRIPCIÓN DETALLADA DEL PROYECTO, INCLUIDO METODOLOGÍA *(máximo tres carillas)*

Este proyecto de investigación contribuye a la línea de evaluación de la calidad del software, centrándose en los aspectos de privacidad y protección de datos de las aplicaciones móviles. Para ello, se proporcionará un conjunto de artefactos y herramientas para avanzar en el diseño de pruebas para evaluar la integridad contextual de la privacidad en las aplicaciones móviles.

La estrategia general a usarse en esta investigación estará basada en estudios sistemáticos de literatura, guiados particularmente por las directrices propuestas por Petersen et al. (2008), y desarrollos incrementales de artefactos y herramientas que permitan ir un paso más allá del estado del arte. Concretamente, se llevará a cabo cuatro tareas (T1-T4) que se describen a continuación.

T1 (M1-M3). En esta tarea se llevará a cabo un estudio sistemático de literatura (ESL) sobre los diferentes enfoques propuestos en el estado del arte para la extracción de flujos de datos normativos desde las políticas de privacidad escritas en lenguaje natural. La búsqueda de información será agnóstica al dominio, buscando ganar conocimiento sobre los diferentes enfoques usados en el ecosistema de las aplicaciones móviles, pero también en otros dominios como la web o el Internet de las Cosas. Se excluirán del estudio los enfoques de análisis totalmente manuales, no sistemáticos o no validados, ya que son precisamente una carencia que hay que cubrir. Los enfoques relevantes (p.ej., determinados por el número de citas) serán analizados en profundidad y se **generará un informe exhaustivo que incluya**:

- ***las técnicas aplicadas***, como por ejemplo aquellas basadas en aprendizaje automático, procesamiento de lenguaje natural o tecnologías semánticas;

- ***los flujos normativos abordados***, tal como la recolección de datos personales, compartición con terceras partes, retención de datos personales, etc.;
- ***el objetivo del análisis***, por ejemplo, la evaluación de cumplimiento de la política de privacidad con una regulación particular, o la evaluación de cumplimiento de un sistema software con la política de privacidad, etc., y;
- ***el nivel de madurez de las técnicas***, determinado, p.ej. por el tipo de evaluación, tipo de referencia usada (ground truth o corpus), etc.

T2 (M4-M8). A partir de los resultados obtenidos en la tarea T1, identificaremos y caracterizaremos aquellos flujos normativos que aún no han sido abordados en el estado del arte, que son relevantes en el ecosistema digital actual y cuyo cumplimiento es susceptible de ser evaluado automáticamente. A continuación, de ser necesario, se realizará la anotación de los flujos normativos seleccionados sobre un conjunto de políticas de privacidad, que serán usados como referencia (ground truth) tanto para el desarrollo como para la evaluación de las herramientas. Esta tarea aprovechará, de ser posible, los corpus disponibles sobre anotaciones de flujos normativos en dominios cercanos, p.ej. para la web (Wilson et al., 2016). Posteriormente, se realizará un desarrollo incremental de componentes de software (artefactos y herramientas) para el análisis y extracción automática de los elementos de estos flujos normativos esperados de una aplicación y declarados por su proveedor en las políticas de privacidad. Para ello, se analizará y seleccionará las técnicas de extracción adecuadas usando como insumo los resultados de la tarea T1. Los flujos normativos extraídos mediante las técnicas seleccionadas se usarán como oráculos para la evaluación de cumplimiento de las aplicaciones (Tarea T4). Se usarán las métricas estándar (precisión, sensibilidad, medida F y exactitud) para evaluar el rendimiento de la extracción de flujos. Esta tarea **generará los artefactos y herramientas necesarias para la extracción automática de los flujos normativos seleccionados**.

T3 (M7-M9). Esta tarea explorará las técnicas para mejorar la generación automática de eventos de usuario y el análisis de comportamiento de las aplicaciones móviles con respecto a los flujos de datos personales. Ya que el estado del arte principalmente hace uso de enfoques aleatorios (p.ej. Exerciser Monkey), esta tarea investigará como proporcionar una orientación que mejore la generación de flujos relevantes para la privacidad y protección de datos desde una aplicación bajo evaluación, extrayendo para ello información de alto nivel sobre el estado de la aplicación móvil bajo evaluación (p.ej., identificadores de los elementos accionables de la interfaz gráfica) y usando diferentes algoritmos de recorrido (p.ej., Depth-first search) que permitan guiar la generación de eventos de usuario. Se usarán métricas, tal como el tiempo de ejecución y el número de flujos de datos detectados, para evaluar los componentes propuestos respecto a Exerciser Monkey que es el mecanismo de vanguardia en el estado del arte. Esta tarea **generará un prototipo de herramienta para la generación automática de eventos de usuario**.

T4 (M9-M11). Finalmente, las técnicas y herramientas para extracción de flujos de datos personales tanto de las políticas de privacidad como de las aplicaciones serán integradas en un proceso de evaluación. Para ello, se llevará a cabo la evaluación de la consistencia (cumplimiento) entre los flujos de datos personales de las aplicaciones y los flujos normativos seleccionados sobre un conjunto de aplicaciones populares gratuitas de la tienda oficial de Android. Esta tarea **generará un reporte sobre los resultados de la validación**.

Bibliografía (Normas APA)

Alkhattabi, K., Alshehri, A., & Yue, C. (2020). Security and Privacy Analysis of Android Family Locator Apps. In *Proceedings of the 25th ACM Symposium on Access Control Models and Technologies* (pp. 47–58). New York, NY, USA: ACM. https://doi.org/10.1145/3381991.3395612

Andow, B., Mahmud, S. Y., Whitaker, J., Enck, W., Reaves, B., Singh, K., & Egelman, S. (2020). Actions speak louder than words: Entity-sensitive privacy policy and data flow analysis with policheck. In *Proc. 29th USENIX Security Symposium* (pp. 985–1002). Online-Event.

Android Developers. (2020). Google Play Protect. Retrieved December 28, 2020, from https://developers.google.com/android/play-protect/phacategories

Article 29 Data Protection Working Party. (2013). *Opinion 02/2013 on apps on smart devices. October*. Brussels, Belgium: European Comission. Retrieved from https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2013/wp202_en.pdf

Asamblea Nacional del Ecuador. (2019). Proyecto de ley orgánica de protección de datos personales.

Balebako, R., Marsh, A., Lin, J., Hong, J., & Faith Cranor, L. (2015). The Privacy and Security Behaviors of Smartphone App Developers. In *Proc.Workshop on Usable Security* (pp. 1–10). San Diego, CA, USA: Internet Society. Retrieved from https://www.ndss-symposium.org/ndss2014/workshop-usable-security-usec-2014-programme/privacy-and-security-behaviors-smartphone-app-developers

Clarke, R. (2006). Introduction to Dataveillance and Information Privacy. Retrieved January 20, 2021, from http://www.rogerclarke.com/DV/Intro.html

Deloitte. (2020). The App Economy in the European Union: a review of the mobile app market and its contribution to the European Economy, (June).

ENISA. (2017). *A study on the app development ecosystem and the technical implementation of GDPR*. European Union: Union Agency for Network and Information Security (ENISA). https://doi.org/10.2824/114584

Eskandari, M., Kessler, B., Ahmad, M., Oliveira, A. S. de, & Crispo, B. (2017). Analyzing Remote Server Locations for Personal Data Transfers in Mobile Apps. *Privacy Enhancing Technologies*, 118–131. Retrieved from http://content.sciendo.com/view/journals/popets/2017/1/article-p118.xml

European Commission. (2016). General Data Protection Regulation. Retrieved October 28, 2020, from https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=EN

Gallé, M., Christofi, A., & Elsahar, H. (2019). The case for a GDPR-specific annotated dataset of privacy policies. In *Proc. AAAI Symposium on Privacy-Enhancing AI and HLT Technologies* (Vol. 2335, pp. 21–23). California, USA.

Guamán, D. S., Alamo, J. M. Del, & Caiza, J. C. (2020). A Systematic Mapping Study on Software Quality Control Techniques for Assessing Privacy in Information Systems. *IEEE Access*, *8*, 74808–74833. Retrieved from https://ieeexplore.ieee.org/document/9069219/

Guamán, 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. https://doi.org/10.1109/ACCESS.2021.3053130

Gürses, S. (2014). Can you engineer privacy? *Communications of the ACM*, *57*(8), 20–23. Retrieved from https://dl.acm.org/doi/10.1145/2633029

Hatamian, M., Momen, N., Fritsch, L., & Rannenberg, K. (2019). A Multilateral Privacy Impact Analysis Method for Android Apps. In *Proc. 7th Annual Privacy Forum* (Vol. 1¡, pp. 87–106). Rome, Italy.

IDC. (2021). IDC - Smartphone Market Share - OS. Retrieved January 5, 2021, from https://www.idc.com/promo/smartphone-market-share/os

INEC. (2019). Encuesta multiporpósito - TIC 2019. Retrieved March 3, 2021, from https://www.ecuadorencifras.gob.ec/tecnologias-de-la-informacion-y-comunicacion-tic/

Kong, P., Li, L., Gao, J., Liu, K., Bissyandé, T. F., & Klein, J. (2019). Automated testing of Android apps: A systematic literature review. *IEEE Transactions on Reliability*, *68*(1), 45–66. https://doi.org/10.1109/TR.2018.2865733
Nissenbaum, H. (2004). Privacy as contextual integrity. *Wash. L. Rev.*, 101–139. Retrieved from http://heinonlinebackup.com/hol-cgi-bin/get_pdf.cgi?handle=hein.journals/washlr79&section=16
Papageorgiou, A., Strigkos, M., Politou, E., Alepis, E., Solanas, A., & Patsakis, C. (2018). Security and Privacy Analysis of Mobile Health Applications : The Alarming State of Practice. *IEEE Access*, 9390–9403.
Petersen, K., Feldt, R., Mujtaba, S., & Mattsson, M. (2008). Systematic Mapping Studies in Software Engineering. In *12Th International Conference on Evaluation and Assessment in Software Engineering* (p. 10). Bari. Italy.
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 *Proc. Network and Distributed System Security Symposium* (pp. 1–15). San Diego, CA, USA. Retrieved from https://www.ndss-symposium.org/wp-content/uploads/2018/02/ndss2018_05B-3_Razaghpanah_paper.pdf
Sadeghi, A., Bagheri, H., Garcia, J., & Malek, S. (2017). A Taxonomy and Qualitative Comparison of Program Analysis Techniques for Security Assessment of Android Software. *IEEE Transactions on Software Engineering*, 492–530.
Stack Overflow. (2020). Developer Survey 2020. Retrieved January 8, 2021, from https://insights.stackoverflow.com/survey/2020#developer-profile
Tiwari, A., Grob, S., & Hammer, C. (2019). IIFA: Modular Inter-app Intent Information Flow Analysis of Android Applications. In S. Chen, K.-K. R. Choo, X. Fu, W. Lou, & A. Mohaisen (Eds.), *Lect. Notes Inst. Comput. Sci. Soc. Telecommun. Eng* (Vol. 305). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-37231-6
Westin, A. (1990). *The Ethics of Privacy Protection* (1st ed.).
Wilson, S., Schaub, F., Dara, A. A., Liu, F., Cherivirala, S., Leon, P. G., … Sadeh, N. (2016). The Creation and Analysis of a Website Privacy Policy Corpus, 1330–1340.
Zimmeck, S., Story, P., Smullen, D., Ravichander, A., Wang, Z., Reidenberg, J., … Sadeh, N. (2019). MAPS: Scaling Privacy Compliance Analysis to a Million Apps. *Privacy Enhancing Technologies*, *2019*(3), 66–86. Retrieved from https://content.sciendo.com/doi/10.2478/popets-2019-0037

14. INFRAESTRUCTURA Y EQUIPOS

- *Indicar la infraestructura y equipos **disponibles** para la ejecución del proyecto, con la ubicación actual de los mismos*

N/A

Dada la naturaleza y alcance del proyecto, se usarán los equipos portátiles y terminales con los que ya cuentan los investigadores.

Infraestructura	Equipos	
Laboratorio	**Nombre del Equipo**	**Ubicación del Equipo**

15. MONTO REQUERIDO

Este proyecto es sin financiamiento.

15.1 Monto y justificación del equipo requerido

N/A
15.2 Monto y justificación del personal requerido
N/A
15.4 Monto y justificación de los investigadores invitados
N/A
15.5 Monto y justificación de los viajes y salidas del campo requeridos
N/A

16. FONDOS ADICIONALES

- *Otros fondos de otros organismos (si los hubiere)*

N/A

DATOS INFORMATIVOS

1. INFORMACIÓN DEL DIRECTOR, CODIRECTOR, COLABORADORES Y COLABORADORES TÉCNICOS

Apellidos y nombres	No. de Cédula	HSS*	Departamento	Rol	Título de mayor nivel y mención.
Guamán Loachamín Danny Santiago	1717529562	8	Departamento en Electrónica, Telecomunicaciones y Redes de Información	Director	Máster Universitario en Ingeniería de Redes y Servicios Telemáticos.
Caiza Ñacato Julio César	1717824450	4	Departamento en Electrónica, Telecomunicaciones y Redes de Información	Colaborador	Doctor en Ingeniería de Sistemas Telemáticos.

* HSS =Horas Semana Semestre: Es el número de horas que se dedica por semana a la investigación. Este número de horas se mantiene para todo el semestre

B. DECLARACIÓN FINAL
DECLARACIÓN DEL DIRECTOR DEL PROYECTO

El equipo de investigadores, representado por el Director del Proyecto declara lo siguiente:

- Que el presente proyecto es una creación original de mi autoría y del equipo de investigadores, y por tanto asumimos la completa responsabilidad legal en caso de que un tercero alegue la titularidad de los derechos intelectuales del proyecto, exonerando a la EPN de cualquier acción legal que se derive por esta causa.

- Que el presente proyecto no ha sido presentado en ninguna convocatoria de otra institución pública o privada. El incumplimiento será causal para que el proyecto no sea tomado en consideración.

- Que si el proyecto genera algún producto o procedimiento susceptible de obtener derechos de propiedad intelectual, de los cuales se deriven beneficios, aceptamos que éstos serán compartidos entre los investigadores y la institución o las instituciones participantes en el proyecto, conforme a lo establecido en el COESC.

- Que el equipo de investigadores y/o instituciones participantes se comprometen a mantener la confidencialidad de la información si ésta podría ser susceptible de protección por patentes, y solicitar la valoración de propiedad intelectual respectiva previa a cualquier publicación o difusión.

- Que para el caso de derechos de autor otorgamos una licencia de uso exclusivo con fines académicos para la o las instituciones participantes en el proyecto.

- Que aceptamos conocer y cumplir con la normativa vigente para la gestión de proyectos.

Firma del Director del Proyecto
Nombre: Danny Santiago Guamán Loachamín
C.I.: 1717529562

ESCUELA POLITÉCNICA NACIONAL

VICERRECTORADO DE INVESTIGACIÓN, INNOVACIÓN Y VINCULACIÓN

Proyecto de Investigación Interno

CRONOGRAMA DE ACTIVIDADES DEL PROYECTO

Título del Proyecto: **Mejora de las técnicas de evaluación de la integridad contextual de la privacidad en las aplicaciones móviles.**

AÑO 1																																															
Nº	**Actividad**	**Mes 1**				**Mes 2**				**Mes 3**				**Mes 4**				**Mes 5**				**Mes 6**				**Mes 7**				**Mes 8**				**Mes 9**				**Mes 10**				**Mes 11**				**Mes 12**			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1	**Analizar las técnicas de análisis de políticas de privacidad escritas en lenguaje natural que han sido reportadas en el estado del arte.**																																															
1,1	Planificación de la búsqueda en la literatura de las técnicas de análisis de políticas de privacidad																																															
1,2	Búsqueda exhaustiva en la literatura de las técnicas de análisis de políticas de privacidad y selección de las técnicas relevantes.																																															
1,3	Análisis en profundidad de las técnicas de análisis de las políticas de privacidad.																																															
1,4	*Informe sobre los enfoques de análisis de políticas de privacidad que incluya las técnicas aplicadas, flujos normativos abordados, objetivo del análisis y nivel de madurez de las técnicas.*																																															
2	**Generar un conjunto de artefactos y herramientas para el análisis y extracción de flujos normativos relevantes, que han sido definidos en las políticas de privacidad y que aún no han sido abordadas en el estado del arte**																																															
2,1	Caracterización de los flujos normativos seleccionados																																															
2,2	Anotación de los flujos normativos seleccionados																																															
2,3	Desarrollo de herramientas para la extracción de los flujos normativos seleccionados																																															

2,4	Evaluación de las herramientas (actividad iterativa)
2,5	*Artefactos y herramientas necesarias para la extracción automática de los flujos normativos seleccionados.*
3	**Desarrollo de una herramienta para la generación automática de eventos de usuario sobre las aplicaciones móviles.**
3,2	Caracterización de la información de alto nivel que se puede extraer desde las aplicaciones móviles para guiar su ejecución y generar eventos
3,3	Análisis de los algoritmos para guíar la ejecución de una aplicación móvil en base la información de alto nivel previamente caracterizada.
3,4	Desarrollo de una herramienta para extraer la información de alto nivel y ejecutar una aplicación móvil
3,5	Evaluación de la herramienta (actividad iterativa)
3,6	*Herramienta para la generación automática de eventos*
4	**Validar los artefactos y herramientas desarrollados a través de la evaluación de la integridad contextual de los flujos normativos seleccionado en un conjunto de aplicaciones móviles.**
4,1	Integración de las herramientas en un proceso de evaluación
4,2	Ejecución de la evaluación
4,3	Análisis de cumplimiento
4,4	*Informe sobre los resultados de validación*
5	**Divulgar la temática investigada y difundir los resultados científicos obtenidos.**
5,1	Escritura de artículo reportando resultados.
5,2	Elaboración de presentaciones para charlas abiertas

5,3	*El envío de un artículo a una revista indexada en Scopus o WoS reportando los resultados científicos alcanzados.*																																																
5,4	*La presentación de la temática, su necesidad e importancia y de los resultados alcanzados a través de charlas abiertas.*																																																