EDGELANDS

DISEÑAR ESPACIOS DIGITALES SEGUROS CON EMPATÍA: UNA GUÍA PRÁCTICA

CONTEXTO

Este documento representa la culminación de ocho semanas de trabajo colaborativo por parte de un equipo de investigadores, mentores y artistas, así como las contribuciones de ponentes invitados, como parte de la fase «Pop-Down and Beyond» del Edgelands Institute.

Este es el resultado del proyecto: una guía práctica que tiene como objetivo proporcionar recomendaciones para el diseño de espacios digitales seguros, basadas en el análisis de fuentes primarias, estudios de casos y datos de encuestas. Esperamos que ésta guía pueda ser utilizada por desarrolladores, investigadores, artistas y los propios usuarios en línea.

Si quieres leer más sobre el complejo contexto de la seguridad digital, las partes implicadas y las de iniciones de “espacios digitales seguros” y “seguridad en línea”, puedes acceder a la versión ampliada de esta guía práctica aquí (sólo en inglés).

AGRADECIMIENTOS

Esta guía práctica fue desarrollado por Pulkit Mogra, Tatiana Lysova, Lilian Olivia Otero, Catherine Keegan, Nina Martin, Mmabatho Oke, Jessica McClearn y Giovanna da Custódia, bajo la dirección de Nina Baranowska, Daniel Odongo, Virgginia Laborão, Vanessa Gathecha y Laura García Vargas.

Queremos dar las gracias a todas las personas que completaron nuestra encuesta y ayudaron a dar forma a las recomendaciones para crear un espacio digital seguro.

También queremos expresar nuestro especial agradecimiento a los ponentes invitados que participaron en las sesiones principales del Research Sprint y compartieron sus ideas e inspiración.

El diseño y la maquetación visual son obra de Flavia Lozano y Larissa Oliveira.

RESUMEN EJECUTIVO

Este documento sintetiza los conocimientos obtenidos a partir del análisis de fuentes primarias, estudios de casos y una encuesta cualitativa para ofrecer recomendaciones sobre el diseño de espacios digitales seguros que prioricen la empatía y las necesidades específicas de la comunidad por encima de los enfoques extractivos centrados en las plataformas.

➔ EL PROBLEMA

A diferencia de los espacios físicos seguros, los entornos digitales carecen de indicadores definitivos de seguridad. Tres actores clave determinan la seguridad en línea: las plataformas, los gobiernos y las comunidades de usuarios; sin embargo, los enfoques actuales siguen siendo inadecuados.

LAS PLATAFORMAS están diseñadas para un "usuario medio" universalizado (normalmente hombres cisgénero, heterosexuales, blancos y de clase media-alta del Norte Global), lo que hace invisibles las vulnerabilidades de las poblaciones marginadas. Los incentivos corporativos priorizan las métricas de participación por encima del bienestar de los usuarios.

LAS REGULACIONES GUBERNAMENTALES se centran exclusivamente en prevenir "daños tangibles" (material de abuso infantil, terrorismo, fraude), ignorando la naturaleza subjetiva y contextual de la percepción de seguridad en línea.

LAS COMUNIDADES EN LÍNEA ejercen una gobernanza de la seguridad crucial, pero frágil, a través de códigos de conducta y moderación voluntaria. Sin embargo, siguen estando estructuralmente subordinadas a la arquitectura de las plataformas en las que existen.

➔ PRINCIPALES RESULTADOS DE LA INVESTIGACIÓN

Nuestro análisis identificó cuatro condiciones fundamentales para la seguridad digital:

CONDICIONES RELACIONALES: incluyen respetar los límites personales sin necesidad de confrontación, una moderación eficaz del trabajo relacional (no sólo la aplicación de normas) y la reducción del coste emocional de la participación, en la que los usuarios deben defenderse constantemente.

CONDICIONES CULTURALES: incluyen normas sociales de dignidad y ausencia de juicios, inclusión lingüística entre diferentes idiomas y comunidades, y redes de personas que proporcionan una infraestructura protectora para grupos vulnerables.

CONDICIONES PROCEDIMENTALES: incluyen reglas claras y aplicadas de manera consistente, autonomía del usuario sobre la visibilidad y el seguimiento de los datos, y estructuras de gobernanza legítimas basadas en experiencias vividas, en lugar de en decisiones corporativas arbitrarias.

CONDICIONES INFRAESTRUCTURALES: abordan la fiabilidad técnica y la integridad de las plataformas, incluidas las prácticas transparentes en materia de datos, los mecanismos eficaces de notificación y la rendición de cuentas cuando fallan los sistemas.

A partir del análisis de estos temas, proponemos una lista de recomendaciones sobre las condiciones necesarias para crear espacios digitales más seguros, que hemos dividido en tres categorías: imprescindibles, deseables y señales de alerta.

➔ CONCLUSION

Este documento pretende ser una guía práctica y un llamamiento para reimaginar los espacios digitales como entornos gobernados por la comunidad, en los que la seguridad la crean conjuntamente quienes los habitan, en lugar de ser impuesta por arquitecturas corporativas optimizadas para el beneficio económico por encima del bienestar. Somos conscientes de las limitaciones de nuestra investigación, entre ellas, la necesidad de incorporar perspectivas más técnicas, una mayor participación de la comunidad y la traducción a lenguas indígenas. Finalmente, hacemos un llamado para explorar en mayor profundidad el tema de la seguridad en línea.

CÓMO IDENTIFICAR, CREAR Y GESTIONAR ESPACIOS SEGUROS EN LÍNEA: UNA GUÍA

A continuación, se presenta una guía sobre espacios seguros en línea basada en nuestro análisis de fuentes primarias, estudios de casos y resultados de una encuesta cualitativa. Puedes consultar el análisis completo en este documento (sólo en inglés).

➔ ELEMENTOS OBLIGATORIOS

NORMATIVA Y DOCUMENTACIÓN

Una documentación clara y bien formulada es crucial para crear una sensación de seguridad en los espacios digitales. La seguridad aumenta cuando las normas son claras, explícitas, accesibles y visibles antes de unirse a un grupo. Esta transparencia permite a las personas tomar decisiones informadas sobre si un espacio digital se alinea con sus valores y necesidades, y por tanto, si desean unirse a él o no.

Las normas deben especificar claramente los comportamientos aceptables, las normas de comunicación y las consecuencias correspondientes. Esto contribuye a la sensación de seguridad al reducir la incertidumbre y al demostrar que se desalentarán los comportamientos dañinos. Además, este lenguaje debe ser accesible para todos los posibles miembros de la comunidad.

Son buenas prácticas en normas comunitarias aquellas que abordan y prohíben explícitamente el acoso, el ciberacoso, el discurso de odio, el contenido extremista y las formas de abuso de identidad (por ejemplo, suplantación de identidad o uso de IA para crear imágenes inapropiadas de alguien sin su consentimiento). Si las normas son ambiguas, debe existir un espacio para su discusión, revisión y reformulación. Sustituir un lenguaje amplio o abstracto, por ejemplo, “comunicación

positiva", por expectativas explícitas sobre respeto, ausencia de juicios y transparencia, ayuda a que las normas sean más comprensibles y aplicables.

Las directrices de la comunidad deben elaborarse con la participación de las propias comunidades. Además, deben revisarse de forma continua e iterativa, de acuerdo con las experiencias de los miembros de la comunidad.

La gobernanza de los datos debe declararse claramente, especificando cómo se recopilan, utilizan, almacenan y protegen los datos personales. Las prácticas éticas de datos, los principios de privacidad desde el diseño y la comunicación transparente y accesible sobre las medidas de protección de datos son esenciales para generar confianza en los espacios digitales. Si una plataforma está dirigida a una comunidad multilingüe, es fundamental que todos los elementos, incluidos los relativos a la protección de datos, estén traducidos a todos los idiomas de dicha comunidad.

Por último, las normas sólo son efectivas cuando se aplican de manera constante e igualitaria a todos, incluidos moderadores y administradores. Cuando la aplicación de las normas es desigual o se percibe como arbitraria, la confianza se erosiona rápidamente y aumentan los sentimientos de exclusión o vulnerabilidad.

ELEMENTOS TÉCNICOS

Las herramientas de seguridad robustas y actualizadas son componentes vitales de la seguridad digital, pero al mismo tiempo deben respetar la privacidad, ser conscientes del contexto y no ser punitivas. Entre los elementos técnicos clave que contribuyen a generar una sensación de seguridad se encuentran los filtros de contenido, la eliminación automática de material perturbador o ilegal

y los mecanismos de denuncia. También son esenciales la encriptación, los mecanismos de anonimización, la prevención de fugas de datos y las medidas de seguridad actualizadas, como la autenticación en dos pasos o las protecciones contra filtraciones de información. Todo ello ayuda a generar una sensación de seguridad en los espacios digitales.

Aun así, se debe evitar una dependencia excesiva de estas herramientas técnicas. Por ejemplo, la moderación automatizada a menudo sobrepasa sus límites y elimina contenido inocuo debido a la falta de contextualización o a unas reglas demasiado estrictas. Esta situación puede limitar la participación, silenciar expresiones legítimas y erosionar la confianza en la plataforma. Por ello, la moderación automatizada debe complementarse con la supervisión humana. Además, debe ser posible ponerse en contacto con la plataforma para reclamar una moderación excesivamente estricta, y la plataforma debe disponer de un proceso para reconsiderar su decisión basándose en la explicación y el contexto del contenido proporcionados por el usuario.

Las protecciones técnicas también deben extenderse para prevenir abusos comunitarios, el uso indebido de datos privados y la infiltración hostil en comunidades sensibles o vulnerables. Las funciones de seguridad son más efectivas cuando limitan proactivamente estos riesgos, en lugar de aplicarse de forma reactiva una vez producido el daño.

Los usuarios deben tener el control sobre su privacidad en una plataforma. Dichas plataformas deberían ofrecer herramientas que permitan a los usuarios gestionar la visibilidad de sus perfiles, divulgar información personal de manera selectiva y elegir entre modos de participación públicos y privados. Estas funciones otorgan a los usuarios

un mayor nivel de control sobre su exposición según sus necesidades y los riesgos a los que están expuestos.

Las políticas de nombre real pueden resultar perjudiciales cuando el anonimato y la privacidad son importantes para la seguridad de la persona. Para muchos usuarios, especialmente para los pertenecientes a comunidades marginadas o en contextos políticamente sensibles, el anonimato no es una preferencia, sino una medida de protección esencial.

Finalmente, las plataformas deben incorporar de forma activa y preventiva ajustes y opciones para individuos o comunidades con necesidades específicas, convirtiendo la accesibilidad en la norma y no en un lujo o algo por lo que haya que luchar.

ELEMENTOS DE LA COMUNIDAD

En la ausencia de fronteras físicas, los usuarios confían en un conjunto de señales digitales informales, pero poderosas, para evaluar si un espacio es seguro. Pueden escanear marcadores visuales, "leyendo" efectivamente el "lenguaje corporal" de la plataforma: los pronombres en los perfiles, los símbolos inclusivos, las advertencias de contenido o los códigos de conducta fijados en la parte superior de un hilo funcionan como señales inmediatas de que existen límites y de que el espacio se está gestionando de forma consciente.

Estas señales visuales se refuerzan con indicios lingüísticos que moldean la atmósfera interpersonal de una comunidad. Los espacios seguros tienden a adoptar un lenguaje inclusivo de "nosotros" en lugar de una retórica adversarial de "nosotros contra ellos" y, con frecuencia, utilizan indicadores de tono para minimizar la ambigüedad. La inclusividad, incluso en el lenguaje, y el reconocimiento de

la diversidad son elementos importantes de las normas, reglas y códigos de conducta comunitarios. Estas prácticas son particularmente relevantes para los usuarios neuro divergentes, que se benefician de una comunicación explícita que reduce la ansiedad y la posibilidad de malinterpretaciones.

Para evitar infiltraciones no deseadas, las comunidades pueden establecer un proceso básico de verificación para los nuevos miembros. Por ejemplo, podría haber un proceso de incorporación en el que los recién llegados tengan que reconocer el propósito, los valores y los límites del espacio antes de poder participar y acceder completamente a él.

Sin embargo, la señal más decisiva es el comportamiento. Los miembros de la comunidad observan cómo se ejerce el liderazgo y la moderación en la práctica y juzgan la seguridad en función de la rapidez, la consistencia y la transparencia de las respuestas ante los daños y las violaciones de las normas comunitarias. Cuando el discurso no se aborda o cuando la aplicación de las normas parece arbitraria, las reglas escritas pierden credibilidad y la sensación de seguridad se disipa rápidamente.

En caso de conflicto dentro de una comunidad, en lugar de depender únicamente e inmediatamente de las herramientas de reporte y prohibición, las plataformas deberían promover que los miembros resuelvan primero los conflictos mediante la amabilidad y las conversaciones constructivas, reafirmando así el papel del individuo en la comunidad. No obstante, cuando la resolución no sea posible, las medidas disciplinarias deben aplicarse con prontitud.

➔ NUESTRA LISTA DE VERIFICACIÓN RECOMENDADA

"IMPRESCIDIBLES" / REQUISISTOS MÍNIMOS

- ◯ Directrices y normas comunitarias claras, explícitas y accesibles, con intervención en caso de infracción.
- ◯ Política de protección de datos sólida.
- ◯ Función de captura de pantalla restringida en comunidades privadas.
- ◯ Herramientas de reporte fiables para contenido dañino.
- ◯ Anonimato opcional.
- ◯ Supervisión para moderadores y administradores.
- ◯ Moderadores humanos (con acceso a apoyo psicológico y recursos si tienen que gestionar contenido traumático).
- ◯ Lenguaje inclusivo.
- ◯ Medidas reactivas y proactivas contra discursos o actividades nocivas.
- ◯ En las plataformas dirigidas a una comunidad multilingüe, todos los elementos están correctamente traducidos a todos los idiomas (especialmente los locales), incluidos los acuerdos de usuario, los códigos de conducta, etc.
- ◯ Capacidad interna dedicada a la ciberseguridad, especialmente para comunidades en riesgo.

“DESEABLES”

- ◯ Verificación durante el proceso de registro y admisión.
- ◯ Medidas de privacidad y protección integradas en la infraestructura de la plataforma.
- ◯ Recursos educativos sobre comportamiento y etiqueta en espacios digitales.
- ◯ Co-diseño participativo que incorpore las necesidades y preferencias de la comunidad.
- ◯ Agentes de soporte y/o moderadores que sean miembros de la comunidad usuaria o, al menos, que conozcan el contexto cultural y que dispongan del tiempo y la capacidad emocional para brindar un apoyo considerado.
- ◯ Soporte humano en tiempo real, por ejemplo, una línea de ayuda disponible en todas las zonas horarias.
- ◯ Herramientas algorítmicas dedicadas para aplicar normas locales específicas, incluidas las de lenguas minoritarias o locales.
- ◯ Higiene digital estructurada, por ejemplo, periodos de calma o nocturnos en los que los moderadores y administradores puedan utilizar herramientas que limiten a un mensaje por minuto.

“SEÑALES DE ALERTA”

- ◯ Inconsistencia en la aplicación de normas, tales como la censura de ideas diversas que no son dañinas.
- ◯ Falta persistente de respuesta ante daños o permitir daños bajo el pretexto de “libertad de expresión”.
- ◯ Exposición obligatoria de la identidad real.
- ◯ Seguridad performativa, por ejemplo, copiar y pegar reglas comunitarias por defecto sin adaptarlas.
- ◯ Infraestructura técnica desprotegida, por ejemplo vulnerabilidades que permitan infiltraciones y fugas de datos.
- ◯ Dependencia excesiva de algoritmos o términos de servicio genéricos para gestionar conflictos interpersonales.

EDGELANDS

edgelands.institute