



Exploring at-risk users' mental models of Apple Lockdown Mode in Latin America and the Caribbean

OXCHE

DECEMBER 2025

Table of contents

Executive summary	3
Key Findings	4
About this report	5
Acknowledgements	5
About authors	5
Introduction	6
Background	7
The spyware landscape in Latin America and the Caribbean	8
Security properties of Apple Lockdown Mode	8
Related work	10
Methodology	11
General mental models of spyware mitigations	14
User's perception of their digital security	14
User's understanding of attack surface	20
Usage of spyware mitigation techniques	22
Expectations using spyware mitigation features	26
Discussion	29
Limitations	31
Future work	32
Appendix	33
References	34

State-sponsored spyware tools like Pegasus have been extensively deployed across Latin America and the Caribbean.

This report exposes how at-risk users in the region protect themselves and their communities against these threats through twelve in-depth interviews with journalists, activists, and human rights defenders.

We found that human rights defenders view digital security holistically—inseparable from physical safety and collective protection networks. In contexts where surveillance and device seizure are immediate realities, technical solutions cannot be evaluated in isolation.

It was also elicited that adoption is driven by trust. Users enable protective features like Apple’s Lockdown Mode primarily through recommendations from peers and trainers who translate technical capabilities into accessible guidance. However, Lockdown Mode faces a critical usability paradox—it remains invisible until it disrupts essential workflows, leading users to toggle it on only during self-perceived threats rather than maintaining continuous protection.

Effective spyware protection must integrate seamlessly into the trust-based, context-specific security practices these communities depend on. This requires more precise system feedback, better communication about protection trade-offs, greater user control, and improved compatibility with tools defenders rely on for their work. Technical features succeed only when they support—rather than disrupt—the broader ecosystems of peer support and collective defense that sustain high-risk users.

1. Digital security is understood as a broad, holistic concept. Human rights defenders (HDRs) from Latin America and the Caribbean did not separate digital risks from the rest of their lives. Security, as they understand it, includes **emotional well-being, mental health, physical safety, and the collective environments in which they work.** Family members, and children in particular, were repeatedly seen as part of the “inner circle” that must be protected.

2. The understanding of the attack surface was directly shaped by how they imagined LDM working. Participants' explanations of the attack surface are directly anchored in their mental model of how the LDM feature works.

3. LDM usage is context-aware and rooted in interpersonal trust. Most people turned on LDM because someone they trusted, being a peer, a digital security trainer or their organization, told them it was worth enabling. **Breakage or difficulty on tasks they had to perform often led them to turn it off at least once.** Some used LDM only in moments of heightened risk such as border crossings, sensitive communications, travel or politically tense periods, while others kept it on by default. Those who understood the tradeoffs clearly were the ones most likely to keep it enabled constantly.

4. As for the trust factor counts, digital security trainers played a central bridging role in LDM adoption. Practitioners were the people who explained how LDM works, helped participants think through their threat models, and supported them when problems appeared. **This intermediary layer of trainers, helplines, or local threat labs helped bridge the gap between HDRs and industry-designed security features.** The metaphors participants shared were also highlighted by practitioners as essential teaching tools, enabling them to turn complex technical ideas into explanations that non-technical users could relate to.

5. LDM is seen as fallible and “too invisible until too opaque”. Participants saw LDM as not bulletproof and often too transparent in problematic ways, to the point that they couldn't tell whether it was enabled at all. When everything worked normally, LDM felt absent. Then, suddenly, it would appear **“in action” only through breakage that disrupted urgent work.** Without any clear indicator of its status, participants realized LDM was active only when something stopped working.

About this report

Mental models for understanding Apple Lockdown Mode usage by at-risk users in Latin America and the Caribbean project aims to expose the case of human rights defenders, journalists, lawyers, and other high-profile personas targeted with mercenary spyware by state actors and other perpetrators, through their understanding of the attack, and the defenses and tactics they put in place to mitigate them. The report's findings, analysis, and recommendations seek to assist developers of mitigation software in identifying, prioritizing, and improving current initiatives to support at-risk users on the ground.

Acknowledgements

We are deeply grateful to our participants for their time, openness in sharing their insights, and willingness to talk about their impactful work. We would also like to thank Ruba Abu-Salma for her insightful guidance throughout this project, Michael Brennan, Pablo Bofelli, Gia and our anonymous reviewers for their efforts to help improve this manuscript.

This work was funded in part by the Spyware Accountability Initiative and individual contributors. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funders.

About the authors

antonela is a design leader and researcher focusing on usable security and privacy. She is interested in critical internet infrastructure, open-source communities, decentralization, and feminism as an intersectional practice. In recent years, she has actively contributed to open-source communities and decentralized networks, effectively bridging technical and democratic demands for security and privacy with human-centered design principles. Currently, as part of the 0xche collective, she continues to push the boundaries of how design can empower users—particularly those in high-risk contexts—through thoughtful, participatory, and open design practices.

0xche is a horizontal and self-organized collective of technologists committed to strengthening the information security of activist organizations and civil society in Latin America. Our identity combines a commitment to social transformation with solid technical expertise.

Suggested citation

0xche, (2025), Exploring at-risk user's mental models of Apple Lockdown Mode in Latin America and the Caribbean.

Spyware has been widely used by state actors, such as governments and law enforcement agencies, to monitor civil activity worldwide. Several commercial vendors operate in this space, developing surveillance technologies to access and monitor mobile devices. Vendors, including NSO, Intellexa, Paragon, Candiru, and Memento Labs, among others, have become well known for commercializing these surveillance technologies, frequently in contexts associated with misuse and human rights abuses. For instance, research has shown that two of the most prominent spyware examples, Hacking Team's Galileo or NSO Group's Pegasus, have reportedly been deployed in several Latin American and Caribbean countries, such as Mexico, Brazil and Colombia.

The growing harm caused by spyware has made the security needs of human rights defenders in the region urgent. Meeting those needs requires listening closely to how they understand security and protection in the context of their work. **Effective protection against sophisticated attacks grows out of a whole network of trust, care, and shared labor, including grassroots peers working on the ground, self-organized collectives, local and international NGOs, helplines, threat labs, and the family support networks that sustain defenders.** Within this interconnected landscape, the tech industry also plays a decisive role. Companies are part of this chessboard too, and the tools they design can either widen vulnerabilities or meaningfully strengthen the protection of those most exposed by tech-facilitated attacks.

We conducted twelve in-depth, semi-structured interviews with at-risk users —journalists, lawyers, digital human rights defenders, and security trainers—to examine how they perceive spyware, understand the strategies they employ on their devices, and their operations to protect themselves and their work against it, with a focus on their mental models, that include perceived security and privacy threat models, and their use of Apple's

Lockdown Mode (LDM), a feature offered by Apple to reduce the attack surface of its users, which has proven to be effective against zero-click attacks (Citizen Lab, 2023).

Our research found that **human rights defenders in Latin America and the Caribbean understand digital security as inseparable from their physical safety, emotional well-being, and the collective networks they work within**—particularly because threats such as civil monitoring, robbery and device seizure are immediate realities in their context. **Their mental models of how LDM functions directly shape how they perceive its protective value and the attack surface it addresses.** Participants' explanations of the attack surface are directly anchored in their mental model of how the LDM feature works. LDM adoption decisions are deeply relational: **most participants enabled LDM because a trusted peer, trainer, or organization recommended it, and digital security practitioners serve as crucial intermediaries who translate technical features into accessible explanations and ongoing support.** However, LDM's effectiveness is undermined by a paradox of visibility—it remains "*too invisible until too opaque*," offering no clear indication that it's active until it breaks critical workflows, leading many users to toggle it on only during high-risk moments rather than maintaining it as a constant protection. This tension between trust-based adoption and friction-based abandonment reveals that for LDM to be sustainable in collective, high-stakes environments, it must better communicate its presence and integrate more seamlessly into the relational, context-specific security practices these communities depend on.

Based on these findings, we made recommendations to improve the **adoption, usability, and comprehension of the LDM feature by improving system feedback visibility, how risk management is communicated to users, user control, and compatibility** with native and third-party apps.

High-risk individuals in Latin America and the Caribbean, including frontline activists, journalists, land and environmental defenders, feminist activists and digital security trainers who support them, live in an environment where surveillance takes many different forms. Their everyday risks are not limited to digital attacks. The presence of sophisticated spyware along with simple phishing and long-standing physical intimidation tactics shows that these threats do not replace one another but coexist. In many countries, this creates a context where HRDs must deal with highly complex digital attacks and simple forms of intimidation at the same time, a pattern that can be seen in other parts of the world as well. Depending on the country, they could face state monitoring, political persecution and invasive spyware attacks, along with problems such as social media monitoring, phishing attempts and the possibility of having their devices taken when crossing borders or during detentions, and kidnapping.

Moreover, the information they carry often places them at even greater risk, since it could include sensitive investigative material, internal organizational documents, alternate identities they rely on for protection, location details and communications with sources.

The adversaries behind these harms are varied, and their motivations often align. According to documented cases, state institutions, including intelligence agencies, migration offices and municipal authorities, could operate alongside organized crime, anti-rights groups, troll centers and private corporations. The range of adversaries listed shows that in some contexts HRDs could be targeted by a mix of state and non-state actors whose motives frequently overlap, who deploy a wide array of intimidation tactics, including the use of spyware.

2.1 The spyware landscape in Latin America and the Caribbean

The current landscape in Latin America and the Caribbean reveals a long term crisis on expanded surveillance practices and increasingly repressive security protocols, often implemented beyond the bounds of legality. In fact, certain governments in the region have abused surveillance powers to target journalists, human rights defenders, political opposition and civil society actors. When it comes to spyware, some of the most alarming examples come from Mexico, Colombia and Brazil (for example, Rodríguez et al. 2024; Suárez, 2014). Sophisticated tools like Pegasus are often deployed under the pretext of addressing national security threats. In practice, they are used to suppress dissent by the fear of being watched, with little or no accountability by each country's law, and operating in a climate of impunity.

Drawing on what has been documented so far, the region can be divided into two categories: countries where the use of spyware has been publicly exposed and those where such tactics remain unconfirmed, likely reflecting a more stealth deployment. The latter probably illustrates the difficulty of conducting investigations without visibility rather than the absence of abuse, as it is confirmed that most LAC countries have negotiated with main spyware vendors (Pérez de Acha, 2016; Domínguez Rubio, 2026).

To highlight just a few examples of the most significant Pegasus cases in the region, in Mexico, across multiple administrations, from Calderón to Peña Nieto and later López Obrador, spyware has been repeatedly misused against journalists, human rights defenders and people connected to politically sensitive cases such as the Ayotzinapa crime investigation. Amnesty International reported that more than 15,000 phone numbers were targeted for potential surveillance following spyware purchases by the Peña Nieto administration (Amnesty International, 2022). Moreover, the investigation carried out by R3D, SocialTIC and Article 19 revealed a wave of Pegasus attacks against journalists and activists. Building on this evidence,

these organizations later confirmed more than 76 additional infection attempts against ten journalists, several human rights defenders and even a minor, among others exposing the scale of Pegasus abuse in Mexico (Scott-Railton et al., 2017a; R3D et al., 2017; Scott-Railton et al., 2017b). Besides in 2019, 456 Mexican phone numbers were compromised using WhatsApp based Pegasus attacks, representing the largest number of victims globally (WhatsApp v. NSO Group, 2025).

In Colombia suspicions about Pegasus had circulated for years but concrete proof only surfaced in 2024 (Fundacion Karisma, 2024), when President Gustavo Petro announced that the Financial Intelligence Unit had identified two cash payments, for a total of \$11 million USD, made to NSO Group in 2021. Two months after Petro's televised speech, the Biden administration acknowledged that the United States had actually financed Colombia's purchase of Pegasus. US officials told media outlets that the delivery supposedly took place without Petro or Duque's knowledge and claimed the tool was intended strictly for anti-narcotics investigations (Valdés & Torres, 2024). In El Salvador, Citizen Lab and Access Now confirmed that Pegasus was used extensively against journalists and human rights defenders between 2020 and 2021 (Scott-Railton et al., 2022). Finally, in the Dominican Republic, Pegasus was confirmed through forensic analysis of the investigative journalist Nuria Piera's phone, which was targeted and infected three times between 2020 and 2021 (Amnesty International, 2023).

Taken together, these cases show a region where sophisticated spyware and long-standing surveillance practices have become a constant part of the landscape in which activists and human rights defenders operate. The same pattern appears across countries, with confirmed infections, weak or nonexistent independent investigations and an environment where journalists and human rights defenders work under the persistent threat of being monitored.

2.2 Security properties of Apple Lockdown Mode

The Pegasus Project investigation was published by an international consortium of media organizations in July 2021. A year later, Apple launched Lockdown Mode (LDM) as a direct answer to these threats. According to the company, "Lockdown Mode offers an extreme, optional level of security for the very few users who, because of who they are or what they do, may be personally targeted by some of the most sophisticated digital threats, such as those from NSO Group and other private companies developing state-sponsored mercenary spyware" (Apple, 2022).

Apple designed LDM as a system-wide configuration that intentionally reduces the attack surface by restricting or disabling entire categories of functionality known to be abused in sophisticated exploit chains. Introduced with iOS 16, LDM has expanded across Apple's ecosystem and continues to evolve as new vectors emerge, although the company warns that only a specific subset of users are the ones at risk of sophisticated and state-sponsored attacks and suggested only for them to use LDM.

This section is drawn from three principal sources. The first is Apple's official documentation, which has the support article "About Lockdown Mode" (Apple, 2025), aimed at end users and is referred to when the feature is enabled. This support article is brief and does not describe every change made to the system under LDM, as shown by the next two sources. The second source is Blacktop's 2023 static analysis of iOS 16, which remains to be the most detailed public piece of Lockdown Mode's technical internals on iOS. Finally, the third is a new macOS focused research presented at the OBTS conference by Marie Fischer, where both static and dynamic methods were used to study the macOS 26 implementation of LDM, covering platform specific mitigations.

Apple's stated protections. Apple's official documentation, describes Lockdown Mode as an “optional, extreme protection for the very small number of users who face grave, targeted cyber threats.” It emphasizes that enabling LDM “further hardens device defenses” and “strictly limits certain apps, websites, and features” to reduce the potential attack surface (Apple, 2025).

According to the documentation, in Messages, most attachment types are blocked and only basic media such as images, video and audio are allowed, while link previews are removed entirely and some images may appear as missing icons. Web browsing is similarly constrained, as LDM blocks “complex web technologies”, web fonts may not load and parts of a site may fail to display. FaceTime makes its rules more strict so that only people you've recently communicated with can establish a call, while features like Live Photos and SharePlay are unavailable. Invitations to use Apple services, such as Home Sharing or Game Center, are blocked unless the sender is already known and Focus may not behave reliably. Shared Albums are removed from the Photos app and new invitations are blocked, while any photo shared has its location metadata automatically stripped.

Apple states that these restrictions extend to the device's connectivity and system configuration. A locked iPhone or iPad will not connect to accessories or computers and, in addition, devices will disconnect from non-secure WiFi networks and will no longer join them automatically. Similarly, support for 2G and 3G networks is disabled entirely. Finally, configuration profiles cannot be installed and the device cannot be enrolled into mobile device management while Lockdown Mode is enabled.

Blacktop's research in iOS. In his “*Anatomy of Lockdown Mode*” presentation at 0x41con 2023, Blacktop walked through his static analysis of iOS 16 and 17 that shows how LDM appears to behave internally based on his reverse engineering efforts

and the internal system messages that are logged when features are blocked due to LDM being enabled. With his analysis, he publicly confirmed many Apple documented restrictions and also uncovered additional features that are disabled. His work remains the most comprehensive technical research to date of how Lockdown Mode operates under iOS.

Blacktop provides further detail on the types of complex web technologies that are disabled when browsing in LDM. Among others, these include MathML, the Speech Recognition API, WebAssembly, ServiceWorkers and the ability to display PDFs or SVG fonts, in addition to disabling just-in-time compilation (Blacktop, 2023).

Additional restrictions affecting communication and collaboration features also emerged from his analysis. These limitations are inferred by Blacktop directly from the internal log strings used to log blocked events. FaceTime audio calls from unknown contacts are prevented (“*An unknown contact attempted to FaceTime Audio call*”, “*An unknown contact attempted to Group FaceTime*”) and iCloud collaboration requests are similarly rejected (“*An unknown contact attempted to iCloud collaboration*”). Lockdown Mode influences crash report behavior as well. According to Blacktop, when LDM is active and a crash occurs, the system may flag it as requiring urgent submission to mark it for “*additional scrutiny*” (Blacktop, 2023). USB Restricted Mode is also enforced (“*Lockdown Mode blocked a Device Inquiry*”), developer mode becomes unavailable (“*Lockdown Mode blocked Developer Mode*”) and PDF parsing is disabled across the system (“*Lockdown Mode blocked PDF parsing*”), removing a common exploit delivery path.

Several elements of Apple's ecosystem are restricted in the same way according to this work. Activity sharing invitations (“*Lockdown Mode blocked Activity Sharing invitation*”), Home access and related invitations (“*Lockdown Mode blocked Home access*”), AirDrop transfers (“*Lockdown Mode*

blocked AirDrop”) and interactions through Find My (“*Lockdown Mode blocked Find My*”) are all blocked when initiated by unknown contacts. The system also prevents the rollback of revoked updates (“*Device is in Lockdown Mode. Ignoring revoked update*”), ensuring that compromised or weakened versions cannot be reinstalled.

Taken together, these restrictions show how LDM is disabling additional subsystems, limiting automated content handling and preventing unsolicited interactions across Apple’s services and communication channels. These conclusions come from a reverse engineering methodology, which may have its own inherent limitations but still provides a clearer view into how the feature operates in practice.

Marie Fischer research in macOS. Recent work from Marie Fischer presented at the OBTS v8.0 conference in October of 2025, in a talk titled “*What’s New in Lockdown Mode?*” has also examined how Lockdown Mode works on macOS. Using both static and dynamic analysis of the Lockdown Mode framework and its internal components, she was able to confirm most of the restrictions Blacktop had identified on iOS and found additional system level changes. She observed that when LDM is active on a Mac, the system performs safe rendering for attachments. Also, several Game Center features are further restricted, such as chat and multiplayer functionality. Finally, Fischer discovered that PDFs stop being recognized as a valid image type in certain processes, and that macOS adjusts its power management behavior by switching to Safe Sleep mode instead of Hibernation mode whenever LDM is active.

Blacktop and Fischer’s work provide the most complete technical picture available, although significant parts of LDM inner workings still remain opaque. Nevertheless, these findings help fill important gaps left by Apple’s scarce and end user oriented documentation.

3. Related work

In the Human-Computer Interaction field, mental models have been used to support usable security and privacy interface design and tool communication because they suggest human-friendly ways to visualize complex system components or user interaction with them. Mental models describe how a user thinks about a problem or a system. There is not a “good” or “bad” mental model; it is the model in the person’s mind of how things work (Rouse and Morris, 1986).

3.1 Mental models for usable privacy and security.

In usable security and privacy, mental models have been used to understand how people explain privacy (Oates, 2018), how users get phished (Van Oorschot, 2021), to gain insights on how users process and respond to warnings (Bravo-Lillo et al., 2011), to learn what people know about how Internet works and how it affects their responses to privacy and security risks (Kang, 2015), to categorise, name and assess type of attacks to improve risk communication (Camp, 2009), to improve SSL indicators in browsers (Felt et al., 2015) and to learn how to prevent users to avoid sending sensitive information over non E2E encrypted secure channels (Abu-salma et al, 2018).

Having clear understanding of user’s mental models is effective in reducing the gap between how the system works and what users understand on how the system works and how they will be affected by it. In prior work, Wash (Wash, 2010) describes how attacks with botnets cleverly take advantage of gaps in these models so that many home computer users do not take steps to protect against them. Mental models can provide predictive and explanatory power for understanding the interaction (Norman, 1982). Mental models also influence trust and acceptance of technology: an incorrect mental model can make users mistrust secure technologies (Volkamer, 2013).

Our approach to usable security and privacy is rooted in end users (Nottingham, 2020). Instead of paternalizing (Dodier-Lazaro et al., 2017) users or demising their own sensitivity or response capacity, we approach the design of usable security and privacy tools understanding first users' contextual threat model and second what users' beliefs are about the system they have in hand or the attack they are about to receive. In this line, the research of mental models is offered as a promising approach, although incomplete (Norman, 1982) but with a solid track record in the cognitive sciences. Furthermore, end users are more concerned with security and privacy problems than with general computer failures (Gross, 2017).

3.2 At-risk users. Our work might be framed as prioritizing users from the margins (Hooks, 1984), we prefer instead to describe it as focusing on at-risk users from the vast majority of the world. We define users as being at-risk if they face an elevated likelihood of an attack to their digital safety, if they have factors that influence or exacerbate their chances of being targeted and/or experience heightened harm as a result of a digitally-mediated attack. Moreover, if they are influenced by social factors (e.g. legal situation or political affiliation, marginalization), relationship factors (e.g. relying on a third party for digital support, access to other at-risk users) and personal circumstances (e.g. having access to sensitive sources or sensitive data handling) (Warford, 2021). While designing for at-risk users, it is crucial to prioritize the development of security and privacy mechanisms that effectively protect them on the frontlines.

For this research, we propose a novel and bottom-up approach. We intentionally interviewed at-risk users with privacy and security needs, with a special focus on our demographics and regional threat models to identify, collect, and shape understanding of spyware infections, especially in the context of state-sponsored attacks.

4. Methodology

We conducted in-depth semi-structured interviews with twelve at-risk users working and living in the LAC territory during July, August, September and October 2025. In this section, we describe the protocol for secure recruiting and communication, participants' demographics, the interview procedure, the strategy for working with at-risk communities, our positionality as researchers, and the details of our data analysis methodology.

4.1 Consentful privacy-preserving communication with interviewers. We developed a consent privacy-preserving process to convene and keep communication with research participants. We relied on our grassroots network of digital rights defenders and civil organizations to connect with the first cohort of participants which enabled us to reach a second group through direct recommendations. The first contact was made through direct reachout or by a trusted person, and via a secure channel. The second contact was conducted using E2E encryption, chosen by the interviewee (such as Signal or GPG-encrypted email).

4.2 Participant selection. For initial contact we distributed an online preliminary survey. Our questions included a first attempt to learn about their daily activities and operations and also their self-registered technical background. Next, we asked about previous attacks or experience with Apple threat notifications. Finally, we asked them to choose a pseudonym, a secure electronic mail contact and their PGP keys or Signal handle to continue our communications through encrypted channels. After that, we collected signed consent via the secure channel. .

4.3 Participants demographics. TABLE 1 summarizes the demographics of our participants. Seven participants self-identified as male and five as female. TABLE 2 shows their current location.

Participants were based in Costa Rica, Mexico, El Salvador, Guatemala, Brazil, and Peru. All of them were Apple users, with one or more devices in daily use.

4.4 Participants profile. This study focused on a specific group of at-risk individuals, including activists, journalists and whistleblowers, who face the risk of targeted digital surveillance. The regional scope covered Latin America and the Caribbean, and the project deliberately excluded general or everyday users to center on those most directly exposed to such threats.

Within this community, participants were broadly distinguished by two intersecting factors: their level of technical expertise and their previous exposure to digital threats.

In terms of technical expertise, one group consisted of frontline activists such as land and environmental defenders, feminist activists and journalists whose work is primarily field-based rather than digital. While they rely on digital tools in both their personal and professional lives, digital security is not the central focus of their work. The second group included digital security practitioners, whose activism is deeply rooted in online and technical spaces. Many of them are members of regional threat labs, helplines or security support networks, lead digital security training and provide guidance on privacy-enhancing technologies. They often act as trusted points of contact when others in their communities face digital threats, offering assistance with threat modeling, forensic analysis and organizational security planning.

Participants also differed in their degree of exposure to threats, a distinction that cut across both groups rather than aligning within one or the other. On one side, some had experienced confirmed infections or targeting attempts, having received official threat notifications from platforms like Apple, Google or WhatsApp, or having a confirmed

attempt or positive infection by a forensic threat lab. This group also included people working in organizations where colleagues had been directly targeted. A second group was composed primarily by highly exposed profiles, who had not received explicit alerts but were still considered at risk because of their activism, visibility, or professional roles, working in contexts with a higher likelihood of surveillance.

4.5 Interview pilot. We conducted 5 pilots to test our interview script, the cadence of the conversation and the impact of sensitive questions on interviewers. We iterated our original script based on the pilot's feedback. The pilot interviews were tested with expert users. We engaged with experts earlier to ensure that potential problems were caught and corrected before they lead to harm to the rest of our interviewers (Bellini, 2023).

4.6 Interview protocol. Our interview protocol included an initial verbal summary of the consent protocol participants had previously signed, a brief reminder of what the study is about, and a warm-up trust-building set of initial questions about their regular workflow and operational security. Next, we asked about their personal threat model and how it matched or did not match their embedded organization one. The second part of the interview began with exploring their understanding of what an attack looks like, what they consider a sophisticated and targeted attack, and how they think these attacks occur. Then, we specifically asked about LDM, their knowledge of its protections, how they felt about it, their adoption of the feature, their onboarding process, and their daily usage. Other questions included Apple's protection sentiment and LDM engagement. We also asked when, how, and why they enabled LDM for the first time, and -when it applied- what their rationale was for turning it off. Finally, we closed with participants' exploratory ideas about secure mobile states and their wishes for security features in mobile devices. Each interview took between

60 and 90 minutes. Participants received \$100 gift cards as fair compensation for their time. Interviews were run in Spanish. The transcripts were analyzed in Spanish, and we manually translated the quotes used in this report.

4.7 Strategies for safe at-risk research. After our safe first contact and our privacy-preserving protocol to obtain formal consent for our interviews, we ensured our operational actions were safe for us as researchers and for participants, preventing re-victimization. All the interviews were done by secure end-to-end encrypted video chat. We used encryption for research data in transit and at rest. Our data retention period ended 90 days after collecting it, and we did not collect any PII during our interviews. Interviews were transcribed by an open-source, local-first audio transcription and subtitling suite and manually reviewed. We used encrypted and local file vaults for our thematic analysis.

4.8 Data analysis. The qualitative nature of this research aims to highlight existing threats and cases in the field rather than rely on numbers. We analysed our data doing Reflexive Thematic Analysis (Braun & Clarke, 2021), which involved several stages: by using a local-first audio transcription and subtitling suite, we transcribed audio interviews, went through the data manually, developed a set of codes which we introduced by hand in our transcripts, collating these codes into themes, revisited and consolidated them, and then analyzed patterns of these themes. Reflexive Thematic Analysis is critical in a usable security and privacy design field. Instead of treating security as a commodity, we conducted this analysis recognizing that usable security tools are essential for human rights defenders. Two bilingual researchers conducted the analyses, with technical backgrounds in private and secure enhancement development and usability, until the themes were saturated.

4.9 Author's positionality. Our feminist intersectional approach was informed by our

research and past work on underrepresented populations and on users facing imminent risks associated with their contextual threat models and the power of their adversaries. We are situated researchers (Singh et al., 2025). We share the historical heritage of being part of a continent that has been colonized and has ever since been emancipated. With the same tenure, we produce knowledge with a certain faith that voices pushed to the margins can shape how big tech companies produce software. Thus, for this research, we focused on the LAC region because users at the exact location where the risk is faced are rarely represented, and this approach does not render the same insights as a controlled study in a laboratory. We understand that our methodology is not neutral.

4.10 Research questions. This study aims to explore the following research questions:

RQ1: What are at-risk users' mental models of spyware?

RQ2: What are the protection mechanisms they use against spyware?

RQ3: What are the technical design recommendations that should be prioritized to balance usability, privacy, and security needs?

5. General mental models of spyware mitigations

The qualitative analysis of participant responses allowed us to elicit various mental models from overlapping or repeated ideas and/or metaphors expressed by participants with non-similar technical backgrounds and threat models in their daily activities. In the following section, we will outline participants' perceptions of their digital security and how they relate it to the collective and organizational work in which they are embedded. Then, we will explore how participants explain high-technical concepts and how these concepts are linked to the metaphors they use to describe the attack mitigation technique. We will see how relevant the explainers are to the attack surface, as they directly anchor on their mental models of how the LDM feature works.

5.1 User's perception of their digital security

The first step in understanding how participants think about their digital security is paying attention to how technology mediates their daily activities and how they assess their device setup. All of our participants were Apple users with more than one Apple device in their setup. Remarkably, most of them said their operating systems were up to date. Participants with technical roles in their organizations, such as system administrators or DevOps engineers, also had Android devices, including Pixel devices hardened with GrapheneOS for specific activities.

It is relevant to consider how participants perceive their safety, as this will inform their approach to digital attack-mitigation technology. And it will illustrate the extent to which they can tolerate usability degradation. We will continue with this analysis in Section 5.3.

As participants talked about what “*digital security*” meant to them, the term quickly broadened. They linked it not only to their specific digital security strategies and device configurations but also to emotional and psychosocial support, to their

physical safety, especially in environments where robbery or kidnapping were real and persistent threats on the horizon, and, repeatedly, to the collective and organizational grounding for any security approach. This expanded understanding aligns with holistic security approaches that emphasise the deep interconnection of digital security, psychosocial well-being, and organisational security processes (TTC, 2016).

5.1.1 Multilayered digital security strategies.

Participants described a broad set of tactics that, when combined, shaped their personal, collective, and organizational security posture. Many relied on compartmentalization, using separate devices, SIM cards, and Apple IDs, or keeping travel-only phones to divide personal life, work, and activism. Anonymization was a common strategy, such as using SIM cards not registered under their own name, avoiding linking personal phone numbers to Apple IDs, or maintaining secondary identities to separate different spheres of their lives.

Some participants emphasized the importance of having a specific mental state as part of their security strategy. For some, staying aware, remaining skeptical, or even adopting an “*assuming compromise*” mindset was a way to remain cautious. As P09 put it:

“Una cosa que tengo clara es que tenemos un alto nivel de espionaje. (...) En momentos donde tenemos un caso o una confrontación con el estado, es como que ya por supuesto que nos están vigilando. Por supuesto, pues el nivel de vigilancia y espionaje que tenemos es mucho y de varios lados.”

“One thing I'm certain about is that we're under a high level of espionage. (...) In moments when we have a case or a confrontation with the state, it's like, of course they're watching us. Of course. The level of surveillance and espionage we're under is high, and it comes from many different directions.”

Others focused on minimizing data or exposure. This included turning devices off during border crossings, carrying nearly empty travel phones, relying on data-only numbers that cannot receive calls, or keeping sensitive information off devices entirely by storing it in a cloud service they trusted. For many, this made sense in a context where theft, raids, and loss of equipment were always in the background.

Participants with technical roles described even more hardened setups. Some used GrapheneOS on a Pixel device for additional security, encrypted everything by default or ran strictly separated networks for different types of traffic. But even among highly technical users, security was never understood as a purely technical matter. People stressed the importance of combining layers. One participant described their approach as *“la reducción de la superficie ataque, combinado con los controles operativos, combinados con minimización de información y mitigación de impacto / reducing the attack surface, combined with operational controls, combined with minimization of information and impact mitigation”*.

Another recurring theme was security as a continuous process. Participants talked about monitoring traffic, performing regular forensic checks on their phones for peace of mind, and treating digital security as an ongoing learning process rather than a one-time configuration.

Finally, many referred to *“levantar barreras / raising defensive shields”* by relying on encrypted apps, enforcing 2FA and coordinating collective alerts when suspicious signals appeared within their networks.

Taken together, these practices show that participants did not rely on any single feature or setting. Their security strategies were built through multilayers of reinforcing defenses, adapted to their personal realities, their organizations and their perceived level of risk.

5.1.2 Physical safety as part of digital security. For many participants, digital and physical risks were inseparable. One indicator of this may be the factor that pushes physical safety as a high threat in the region and people at-risk threat model. P02 and P06 said:

“Allanaron la casa y se llevaron todos los dispositivos. Pero no tuvieron acceso a nada de la información porque estaba todo cifrado. (...) Más bien afortunadamente la parte tecnológica era la que más estaba protegida pero no así la seguridad de la casa.”

“They raided the house and took all the devices. But they couldn’t access any of the information because everything was encrypted. Fortunately, the technological side was the one that was best protected, unlike the security of the house itself.”

“(Los dispositivos electrónicos) me dan una falsa sensación de seguridad. Por las capacidades que tienen (mis atacantes) no necesitan monitorearme, con solo decir que soy sospechoso de cualquier cosa, me llevan. En el mejor de los casos, sigo. En el peor de los casos, ya no vuelvo a aparecer.”

“(Electronic devices) give me a false sense of security. With the capabilities (my attackers) have, they don’t need to monitor me; just by saying I’m suspicious of anything, they take me away. In the best case scenario, I’m still here. In the worst case scenario, I will never appear again.”

“A nivel personal pienso en los agentes en aeropuertos, agentes migratorios a los que no puedes negarte a darles equipos. Eso sí es algo que me preocupa y es algo por lo que también viajo con dispositivos medianamente vacíos. Y veo al robo común también como parte de los adversarios.”

“On a personal level, I think about airport agents, immigration officers, since you cannot refuse to hand over your devices to them. That

is something that worries me, and it is also why I travel with devices that are mostly empty. I also see common theft as part of the adversaries."

Even in a presumably solid operational security setups with strategies like devices and account compartmentalization, two-factor authentication, disk encryption, network privacy through a third party VPN, the physical threat is still a big fear. This makes features that account for physical access, such as those that restrict USB connections, even more critical.

5.1.3 Emotional and mental well-being. Another important aspect of the perceived security approach is the visibility of attention to mental health issues:

"Una cosa que es importante también mencionar, es la parte más integral de la cosa, más holística. A veces estamos con mucho estrés, y pasamos por casos muy difíciles o por situaciones personales. En toda la parte de la seguridad también hay que estar siempre atentas a eso."

"Another thing that I think is important to mention is the more comprehensive, more holistic side of it. Sometimes we are under a lot of stress, and we go through very difficult cases or personal situations. Throughout all aspects of security, we always have to stay attentive to that."

5.1.4 Security as a collective effort. Participants also emphasized that digital security is not something they approach individually. Their comments repeatedly framed security as a collective process, built together across the different groups they belong to. As one participant put it:

"(La seguridad) es una estrategia colectiva, la seguridad no es individual. Vos no vas a desconectar tu teléfono o tu computador y vas a estar más seguro porque te vas a una isla. La cuestión es habitar los espacios digitales

de manera segura y bueno, hay estrategias para eso."

"(Security) is a collective strategy, it is not something individual. You're not going to disconnect your phone or your computer and suddenly be safer just because you isolate yourself on an island. The point is to inhabit digital spaces safely, and there are strategies for that."

Others described security as something shaped through *"procesos de reflexión tanto institucionales como colectivos / institutional and collective reflection processes"*, often informed by exchanges within their own organizations and the wider community of groups they coordinate with that had already been under attack and were willing to share lessons learned.

For most of the participants, it wasn't easy to set boundaries between their primary job, their volunteer or activist work, and their personal lives. Participants whose primary activities were embedded within organizations rely on the organization's structure to define their operational security. In large organizations, this setup is defined within hierarchical structures, where dedicated teams provide guidance and training on configuring devices to support safe operations. On the other hand, participants with labor in the ground, on a volunteer basis, in more informal settings, or in less mature organizational security models discuss through collective decision-making and iterative processes to define the operational security of the collectives they run. P07 expressed how the differences between the two structures shape his operational security setup:

"En mi trabajo remunerado es bien centralizado: hay un equipo que se dedica a eso, hay una persona que es el coordinador de ese equipo y el resto le damos la implementación. En mi trabajo no remunerado, es un esfuerzo más colectivo, más disperso, complejo y al final las decisiones se van tomando. Tenemos menos

capacidad de centralización, en el sentido de que cada uno tiene dispositivos muy distintos, capacidades tecnológicas muy distintas entonces luego la aplicabilidad es otra cosa.”

“In my paid job everything is very centralized: there is a team dedicated to that, there is one person who coordinates the team, and the rest of us carry out the implementation. In my unpaid work, it’s a much more collective and dispersed effort, more complex, and decisions are made along the way. We have far less capacity for centralization, in the sense that each person has very different devices and very different technological capabilities, so then the applicability becomes something else entirely.”

Another layer of this collective approach emerged when participants framed digital security to be approached within a broader family perimeter. Several described how securing their own devices was not enough if people closest to them remained vulnerable. For many, protecting their families, especially children, became part of their operational security thinking.

“Generalmente, siempre los ataques (...) van hacia un grupo de personas. Creo que podría ser más probable que le entren a personas ya sean familiares, cercanas, conocidas o partes de la colectiva.”

“Generally attacks always (...) go toward a group of people. I think it might be more likely that they get to people who are family members, close friends, acquaintances or part of the collective.”

“Faltó hablar del sistema familiar, porque también yo con mis hijos definí que no podían tener un teléfono no tan asegurado. Ellos son como el núcleo más cercano (...) Tengo una creencia de que Apple tiene un poco más de seguridad. En la parte de la cercanía de mis hijos, también hubo una definición de que había que ahorrar y comprarles teléfonos Apple.”

“It’s also important to mention the family system, because I also decided that my children couldn’t have a phone that wasn’t well secured. They are part of my closest inner circle (...) I believe Apple has a bit more security. And given how close they are to me, we also decided that we needed to save and buy them Apple phones.”

In this sense, digital security was consistently framed not as an individual task but as something embedded in networks of trust, care and shared responsibility whether within organizations, informal collectives or the family circle.

5.1.5 Digital security setup. Participants with a deeper technical background scored themselves higher on perceived safety. When we asked participants to rate the security of their devices from 1 to 10, most placed themselves at 7 or 8. One of the reasons they mention is their ability to make their own security configuration choices. The feeling of being in control of their setups shaped that score. P08 acknowledged his tech-savvy status:

“Me gusta pensar que con las medidas que he tomado es un poquito mejor que el promedio, entonces voy a decir siete.”

“I like to think that with the measures I’ve taken, it’s a little better than average, so I’m going to say six, seven.”

Another reason for higher scores was the organization they work with and its operational security protocols. P05 relies on the organization’s operational security. Its enforcement makes them feel safe:

“Yo diría que son bien seguros, porque tengo muchas prácticas y políticas de seguridad que tengo que cumplir para todas las organizaciones que participo. (...) No daría diez, porque todos los sistemas pueden ser hackeados, entonces, pienso que por ahí nueve, ocho, siete.”

"I would say they're quite secure, because I follow many practices and security policies for all the organizations I'm part of. (...) I wouldn't give it a ten, because any system can be hacked, so I think maybe a nine, eight, or seven."

Participants with a less technical background raised their scores to 6. Even with a less technical background, given the daily activities they do, they realize they lack confidence when setting up their operations and devices for security and privacy.

5.1.6 Attack vectors. We wanted to learn which types of attacks participants perceived as more or less effective against them, to understand their awareness focus points. After talking about participants' operational security, we asked them which was the easiest and most difficult way to get attacked. Not surprisingly, the easiest way to get attacked for most of the them was physical control through robbery:

"Yo me imagino (un ataque a través de) un control físico de mi casa. Que yo haya dejado encendida la computadora y salido y entonces que (los atacantes) entren en ese momento. Creo que sería lo más fácil."

"I imagine (an attack through) physical control of my house. That I might have left the computer on and gone out, and then (the attackers) come in at that moment. I think that would be the easiest."

"Para mi, si lo que quieren es sacarme de las canchas, lo más fácil es atacarme físicamente."
 "For me, if the goal is to take me out of the game, the easiest thing would be to attack me physically."

Interestingly, when referring to the digital sphere, some participants mentioned phishing as both the easiest and most challenging way to be attacked. It is difficult because they have the background to identify malicious emails or messages. At the same time, it is easy to get phished by trusted sources.

Since most participants were working with sensitive information, in a trusted circle of peers, and with high exposure, it may be possible that a close peer was infected and, as a consequence, they were infected as well. P06 was explicit on default trust while working with peers:

"La forma más fácil de atacarme es por el correo electrónico. Normalmente no estoy pensando si una comunicación que esté relacionada a mi trabajo pueda ser un correo con phishing u otro tipo de aplicación para extraer información. Si tiene que ver con cosas relacionadas al trabajo, difícilmente puedo dudar de alguien por las temáticas en las que nos movemos."

"The easiest way to attack me is through email. I'm usually not thinking that a message related to my work could be a phishing email or some other kind of app to extract information. If it has to do with work related matters, it's very hard for me to doubt someone because of the topics we deal with."

For P09, the easiest way to get attacked was through password stealing: "*The easiest way is like my password settings, which I'm telling you are almost the same but changed a little bit / La forma más fácil es como la configuración de mis contraseñas que te digo que de repente las casi es la misma pero cambiadas tantito*".

This set of questions was not only helpful in understanding participants' awareness of risks, but also a safe path to explore previous attacks.

5.1.7 Previous attacks. Most of our participants received digital attacks in the past. From phishing attempts, account hijacking, and doxxing to sophisticated spyware infections, the landscape of confirmed threats can take many forms. When we asked whether they had changed anything in their digital security operations after the attacks, we received varied responses. Some participants increased their attention on their security

operations after the attacks were confirmed.

On changes in their operational security after an attack, P6 shared their improvements:

“Después de los ataques hemos intentado ser mucho más disciplinados con los protocolos de seguridad que tenemos. En mi caso personal, desde marzo me deshice de los teléfonos físicos, números y cuentas de teléfonos. Tengo nuevos chips, nuevos equipos.”

“After the attacks, we’ve tried to be much more disciplined with the security protocols we have. In my case, since March I got rid of all my physical phones, numbers and phone accounts. I have new SIM cards, new devices.”

For P10, the confirmation of a Pegasus infection didn’t translate into fear of her phone. She understood the harm as coming from the actors behind the operation, not from the technology and her frustration was directed at the government she believed was responsible.

“Más bien dije: ay que pedo, cómo es que este pinche Peña Nieto(...)Me enojé mucho como diciendo: ¿qué onda con esta gente? Pero no lo viví como un tema de la tecnología que me ataca. No lo asocié con un nivel de vulnerabilidad de la tecnología o el aparato en sí sino más bien cuestioné la estrategia del estado de estar jodiendo.”

“What I actually said was: what the hell, how is this damn Peña Nieto(...)I got really angry, like, what’s wrong with these people? But I didn’t experience it as technology attacking me. I didn’t associate it with a vulnerability in the tech or the device itself, instead I questioned the state’s strategy of messing with us.”

A participant with a strong track record in defending digital human rights was less impressed by the confirmation of state-sponsored attacks, which did not lead to inaction but to a strategic

pause. When monitoring became obvious to her, she encouraged her team to pause, rest, and return to “*fight the fight*” the next day:

“Cuando mi teléfono se calienta se que me están escuchando bastante. O cuando el internet está muy inestable les digo: ya muchachas, no se desesperen hay alto nivel de intervención y entonces ya dejen de batallar. Y ya pues un poco nos relajamos. (...) Entonces ahí me tienes escribiendo a la antigüita porque me bloquearon la efectividad de acciones con celulares o de la tecnología misma. (...) Entonces cómo que te regresas un poco a las formas antiguas. O me resigno y digo: chicas no hay condiciones ya no sacamos nada, o sea no da y no vamos a estar batallando. Que descansen, ya no hay que dar la batalla. Por ahora. Mañana vemos. Hoy podemos descansar.”

“When my phone heats up, I know they’re listening to me a lot. Or when the internet becomes very unstable, I tell the girls: don’t get frustrated, there’s a high level of intervention, so stop struggling with it. And then we relax a bit. (...) So there I am, writing the old fashioned way because they blocked the effectiveness of anything we could do through our phones or the technology itself. (...) So you kind of go back to the old ways. Or I resign myself and say: girls, the conditions aren’t there, we’re not getting anything done, it’s just not happening and we’re not going to keep fighting. Rest, we don’t have to fight the fight today. For now. Tomorrow we’ll see. Today we can rest.”

5.2 User’s understanding of attack surface

Participants’ mental models varied according to the nature of their daily activities and the people they interact with. Some participants were already doing mentoring/educational work as digital security trainers, so explaining high-tech concepts in their own words was easy for them. For others, it felt like

a slightly unfamiliar exercise. After the warm-up on learning about their technology-mediated activities, regional threat modeling, and self-perceived digital security, we asked participants whether they knew or had heard of the term “attack surface” and, if so, how they would explain it. Our findings reveal that participants’ explanations of the attack surface are directly anchored in their mental model of how the LDM feature works. Let’s see the models in detail.

5.2.1 House. When discussing the attack surface, some participants used a domestic metaphor. P03 was explicit on sharing how the presence of a front yard and an exterior fence prevents someone from immediately approaching the front door of a house:

“A veces conversando con periodistas les digo imagínate que tú tienes una casa pero tu casa tiene un patio delantero, y ese patio a su vez tiene una reja. La persona que quiere entrar a tu casa va a lograr llegar a la reja de afuera, pero no va a poder llegar hasta la casa o tocar directamente la puerta de tu casa”
 “It’s like when I’m talking to journalists, I tell them, imagine you have a house, but your house has a front yard, and that yard has a fence. The person who wants to enter your house will be able to reach the outside fence, but they won’t be able to reach the house or knock directly on your house door”

This model about the attack surface was relevant and an anchor when she explained how LDM works:

“Lo que yo entiendo es que (LDM) es como un filtro, una cortina que impide que algunas imágenes o archivos por ejemplo, puedan llegar al nivel o a la capa donde la persona es la que toma la decisión (de abrirlos). En general en muchos ataques hay una actuación por parte de la persona, del humano frente a la agresión. Que hace que la agresión se transforme en un ataque, de pasar de amenaza o riesgo a ataque directo.”

“What I understand is that (LDM) acts like a filter, a curtain that prevents certain images or files, for example, from reaching the level or layer where the person is the one who makes the decision (to open them). In many attacks, there is an action taken by the person, the human, in response to the aggression. That is what turns the aggression into an attack, moving it from a threat or risk into a direct attack.”

When participants explained how attacks happen and what an attack surface is, they consistently rely on the concept of doors. P03 described it very simply, the more apps installed on a phone, the more “doors” an attacker can try to enter through, since each app becomes a potential way in. P10 echoed the same idea saying that disabling certain phone features is essentially a way of “cerrar esos portillos / closing those little doors.”

P07 used a similar framing when defining the attack surface itself, describing it as “*the sum of the options an attacker has to try to enter or exploit a device or system / la suma de las opciones que tiene un atacante para tratar de ingresar o explotar un dispositivo o un sistema*” adding that it is directly tied to those “*entry doors/puertas de entrada*” that remain open or exposed.

Doors, windows, gateways were common in participants’ narration of their attack surface. The notion of security layers is also elicited on the house theme through the idea of layered protections that need to be secured.

5.2.2 Body. Noticeably, three participants mentioned “two sides” while thinking about the attack surface: the digital and the physical. While the digital belongs to electronic devices and operations like messaging or exchanging files, the physical scope is directly related with their bodies, including their mental health:

“Las superficies de ataques son todos los dispositivos y la operación, de todo lo que hacemos en las colectivas, tanto la parte digital como la parte física. La parte física es donde yo a veces siento que podría ser un poco más sensible el asunto. Cada dispositivo, cada cuenta, cada interacción nueva que tenemos entre nosotras con personas externas, para mí todo eso es parte de la superficie de ataque. Incluyendo, por ejemplo, la salud mental, verdad.”

“The attack surfaces are all the devices and all the operational work we do in the collectives, both the digital and the physical parts. The physical side is sometimes where I feel things can be a bit more sensitive. Every device, every account, every new interaction we have among ourselves or with external people, for me, all of that is part of the attack surface. Including, for example, mental health.”

A more fragile mental health is also called as an effective way to infringe the attack surface. P10 explained it in her threat model:

“Una de las estrategias más efectivas de los anti-derechos siempre ha sido mantenernos bajo la lupa de la ilegalidad, bajo la lupa del acoso, bajo la lupa del asedio constante, de la culpa. (...) Entonces en esos momentos, yo siempre pauso, y digo: hay alguien que no se tiene que meter en esta ola de desesperación, ¿verdad? ¿Por qué? Porque esos son los momentos en los que atacan, bueno, entonces la salud mental es muy importante, ¿por qué? Porque entonces te empiezan a pasar enlaces, te empiezan a pasar cosas y uno empieza a compartir y ahí, falla.”

“One of the most effective strategies of anti-rights groups has always been to keep us under the microscope of illegality, under the microscope of harassment, under the microscope of constant siege, of guilt. (...) So in those moments, I always pause and say: someone here

cannot get caught up in this wave of despair, right? Why? Because those are the moments when they attack. So mental health becomes very important, because that's when they start sending you links, they start sending you things, and you start sharing them, and that's where things go wrong.”

One participant referred to the attack surface as something that could be hurt, like skin. P02 made a parallelism between different types of attacks and health issues:

“Yo creo que es un poquito como la piel. (...)La superficie de ataque, básicamente yo la entiendo, como todas las posibilidades: me puedo raspar al salir de casa o aquí en casa hasta me puedo quemar cocinando, puedo puedo desarrollar un cáncer de piel por estar mucho tiempo expuesto bajo el sol o sólo puedo ensuciarme las manos y luego tener una infección gastrointestinal. Siento que básicamente los peligros que vienen de estar vivo son un poco como la superficie de ataque. Luego lo puedes acotar: cáncer de piel, no, no me va a pasar, o me quiero cuidar de eso. Pero siempre (la superficie de ataque) es como todas las posibilidades y luego lo puedes precisar un poquito más a tu contexto y a tus condiciones de vida.”

“I think it's a bit like skin. (...)I understand the attack surface basically as all the possibilities: I can scrape myself when I leave the house or even here at home, I can burn myself while cooking, I could develop skin cancer from spending too much time exposed to the sun, or I could just get my hands dirty and later end up with a gastrointestinal infection. I feel that the dangers that come with being alive are a bit like the attack surface. Then you can narrow it down: skin cancer, no, that's not likely to happen to me, or that's something I want to be careful about. But (the attack surface) is always all the possibilities, and then you refine it a

little based on your context and your living conditions.”

If we situate our analysis in local contexts, it seems clear that the relationship between the physical threat and the body metaphor is intrinsic.

5.2.3 Hardening. Participants with a deep technical background directly mirrored LDM to the known concept of hardening. This association comes with certain expectations. P01, a technical expert, was extremely clear comparing LDM to a traditional hardening process:

“LDM te hace un hardening a las opciones de tu sistema operativo del teléfono o computadora o tablet. (...) Entonces, lo que hacen es restringir las áreas que son más sensibles a ser vulneradas, para reducir lo que es su superficie de ataque por medio de eso.”

“LDM does a hardening to the options on your phone, computer, or tablet’s operating system. (...) So, what it does is restrict the areas that are most vulnerable to being compromised, in order to reduce their attack surface.”

While hardening in other OSs is a controlled process carried out through a series of steps, the single toggle for enabling LDM appears insufficient to control granular settings. Indeed, granular control was the most requested feature to extend LDM by tech-savvy participants.

5.2.4 Car safety. One interesting approach to explain reducing the attack surface was related to car safety. P02 was directly relating computing with driving a car:

“Es un poco como cuando pegas el freno y el cinturón te detiene, ¿no? Entonces dices, uy, qué bueno que tenemos el cinturón. Algo casi pasa, ¿verdad?”

“It’s a bit like when you hit the brakes and the seatbelt stops you, isn’t it? Then you say, oh, good thing we have the seatbelt. Something

almost happened, right?”

Across interviews, metaphors like houses with doors and fences, bodies with skin, or cars with seatbelts emerged as intuitive ways to explain concepts such as the attack surface or LDM. Participants who work as digital security practitioners even emphasized how central these metaphors are in their own training, where they help make technical ideas feel more approachable to non-technical folks. This makes unpacking and understanding these narratives more relevant than ever, as they are not just descriptive but pedagogical tools actively shaping how digital security is taught, which directly influences how non-technical users adopt it.

5.3 Usage of spyware mitigation techniques

Most of the digital security trainers we interviewed enabled LDM for the first time right after iOS 16 was released. They were motivated to enable LDM because they wanted to test how the feature worked before spreading the word to their communities, but many also knew from the beginning that a one-click toggle would not be sold as a magical solution. Less technical participants, on the other hand, enabled LDM for the first time at the recommendation of a trusted peer or through the enforcement of security operation protocols. Almost everyone had disabled it at least once after trying it for the first time.

The usage of LDM in our participants was a spectrum. For many, LDM is something they use only during periods of heightened risk:

“En situación de riesgo aumentado o alto riesgo lo habilito. En el día a día se me dificulta un poco porque los servicios de Apple no funcionan bien con LDM”.

“In situations of increased or high risk, I enable it. Day to day it’s harder because Apple services don’t work well with LDM.”

Border crossings were the most cited example of a “*high risk moment*”. Participants felt that the likelihood of authorities gaining physical access to their device was significantly higher at airports and checkpoints:

“Activo el Modo Hermético cuando hay posibilidades que me tope con una autoridad que requiera acceso físico al dispositivo. En aeropuertos y controles fronterizos, activo el Modo Hermético y apago dispositivos por si se lo iban a confiscar.”

“I enable Lockdown Mode when there’s a chance I’ll run into an authority who might require access to my device, at airports or border checkpoints. I turn on LDM and power off my devices in case they’re going to confiscate them”

LDM tended to stay enabled among digital security trainers and among users who had previously gone through a risk analysis process and understood the tradeoffs well.

Participants consistently described practical barriers when using LDM. The first difficulty involved handling files essential to their work such as photos, spreadsheets, collaborative documents and other investigation materials. They often become inaccessible or unusable under LDM restrictions. Participants also reported difficulties with routine activities such as online banking, government and transport apps and even basic communications like making calls or receiving messages.

The second issue was notification overload. Several participants said notifications accumulated and persisted even after they had been read, degrading the battery and the device’s overall performance. P04 also explained how LDM raises anxiety levels:

“Creo que lo primero tiene que ver con la gestión de notificaciones. No sé por qué el Modo Hermético, en mi caso, deja todas las notificaciones acumuladas. Incluso aunque ya haya abierto la aplicación y haya revisado

la notificación, cuando está bloqueado el dispositivo, las notificaciones están ahí. Entonces, si yo por ejemplo no elimino las notificaciones manualmente, puede pasar todo el día y puedo acumular mil notificaciones en un día. Y eso es confuso, porque por ejemplo, ahorita estoy en esta reunión y veo que hay notificaciones, pues no las puedo abrir porque estoy ya teniendo la reunión. Pero comienza a generar como cierto estrés o ansiedad de si me están no me están hablando, o no estoy contestando, o qué pasa que tengo tantos mensajes, y no, simplemente son las notificaciones acumuladas de cosas que ya atendí. Eso es como lo primero y es súper fastidioso.”

“I think the first issue has to do with notification management. I don’t know why Lockdown Mode, at least in my case, leaves all the notifications piled up. Even if I’ve already opened the app and checked the notification, when the device is locked, the notifications are still there. So, for example, if I don’t manually clear them, a whole day can go by and I can easily end up with a thousand notifications in a single day. And it’s confusing because, for instance, right now I’m in this meeting and I see notifications, but I can’t open them because I’m already in the meeting. It starts to create a kind of stress or anxiety, like, are they talking to me? Am I not responding? Why do I have so many messages? And no, they’re simply accumulated notifications of things I already handled. That’s the first thing, and it’s incredibly annoying.”

Participants also mentioned social side effects such as being excluded from everyday social interactions because “*nothing ever works*” on their phone. This included trying to buy concert tickets or participate in group plans:

“Mis amigos, que no son técnicos ni nada, siempre me vacilan mucho con: ‘Es que usted

y su teléfono no funcionan'. Y es cierto. Es decir, te voy a poner un ejemplo. Cuando hay que comprar boletos para un concierto todo el mundo intenta que no sea yo el organizador porque siempre tengo problemas. Mis amigos que no están en el mundo del activismo ni en el trabajo ni tal, cada vez que me han enviado algo es como: 'Ay qué pereza a usted no le va a funcionar.' Porque a mí nunca me funciona nada."

"My friends, who aren't tech savvy or anything, always tease me, saying: 'It's because you and your phone never work.' And it's true. For example, whenever we have to buy concert tickets, everyone tries to make sure I'm not the one in charge because I always run into problems. My friends who aren't in the activism world or at work or anything, every time they've sent me something it's like: 'Ay what a bore, this is not going to work for you.' Because nothing ever works for me."

Another pattern participants mentioned firsthand or by watching their trainees perform, was that users blamed LDM by default whenever a technical problem appeared. While giving support to a human rights lawyer, P03 explain how she was experiencing unrelated issues and directly pointed to LDM:

"Tenía un antivirus en su iPhone y le estaba causando muchos problemas. No sé si era por el Modo hermético o porque el antivirus en un iPhone simplemente no funcionaba bien. Pero siempre tenía ese problema y lo atribuía al Modo hermético."

"She had an antivirus installed on his iPhone and it was causing a lot of problems. I don't know if it was because of Lockdown Mode or simply because the antivirus wasn't working well on an iPhone. But she kept having that issue and blamed it on Lockdown Mode."

Across interviews, people emphasized that they did not know when LDM was enabled.

For them, there was no clear or persistent indicator, so participants realized LDM was active only when something suddenly broke: *"Creo que el dispositivo nunca me notificó que estaba habilitado el lockdown mode, hasta donde recuerdo. Fueron cosas que se rompieron. / I don't think the device ever notified me that lockdown mode was enabled, as far as I remember. Things just broke"*

In a similar line, participants described that notifications were not functioning as indicators of system state. Several said they had no way of knowing whether Lockdown Mode was enabled and instead of any explicit signal, users only realized LDM was active when certain features suddenly stopped working. As P04 explained, the only signs were *"things that broke"* not any notification related to LDM itself. At the same time, the notifications they did receive were unrelated to LDM, often accumulating in large numbers due to what they perceived as an app notification bug. This mismatch created confusion and stress, due to an overload of irrelevant notifications while lacking the single, meaningful notification they actually wanted, the device telling them clearly when LDM was on.

The use of the Apple ecosystem of devices while trying to raise defenses was confusing. Participants who used multiple Apple devices or whose children were part of their Family group often worried that infections could spread across devices. P09 was concerned about her daughter and son, as part of her Family subscription, and how they can be protected with LDM:

"Faltó hablar del sistema familiar, porque también yo con mis hijos definí que no podían tener un teléfono no tan asegurado. Ellos son como el núcleo más cercano (...) Tengo una creencia de que Apple tiene un poco más de seguridad. En la parte de la cercanía de mis hijos, también hubo una definición de que había que ahorrar y comprarles teléfonos Apple."

"It's also important to mention the family system, because I also decided that my

children couldn't have a phone that wasn't well secured. They are part of my closest inner circle (...) I believe Apple has a bit more security. And given how close they are to me, we also decided that we needed to save and buy them Apple phones."

Others explained that whenever a feature stopped working as expected, the quickest solution was simply to turn LDM off entirely:

"Bueno, mayor parte de las veces (deshabilite LDM) por los bloqueos de mensajes en línea. Sí. Fue un poco difícil tener los problemas que tenía. O algunas aplicaciones de gobierno que no lograba usar. Algunas cosas no me acuerdo, porque fueron varias veces que lo deshabilite, pero siempre era por alguna función que se me estaba tornando más difícil."

"Well, most of the time (I disabled LDM) because online messages were blocked. It was a bit difficult dealing with the problems I was having. Or some government apps that I couldn't get to work. I don't remember everything, because I disabled it several times, but it was always because some function suddenly became too difficult to use."

"En el uso cotidiano (LDM) es poco práctico, pensando en perfiles como periodistas que reciben todo el tiempo fotos y archivos, se hace poco funcional y difícil de sostener. La periodista dice que lo desactivó hace un mes porque estaba haciendo una investigación y ya no la dejaba avanzar."

"In everyday use (LDM) it is impractical, especially for profiles like journalists who constantly receive photos and files; it becomes dysfunctional and difficult to maintain. The journalist says she deactivated it a month ago because she was doing an investigation and it was preventing her from making progress."

In specific cases, when certain tasks become impossible, instead of disabling LDM altogether,

workarounds were found. P03 was explicit on how a prominent human rights lawyer she was accompanying in her digital security work ended up circumventing LDM protections simply by moving between devices:

"Ella pues obviamente decía: '¿Qué es esto? No me sirve porque no puedo recibir las imágenes y los archivos y entonces la desactivé porque estoy justo en una investigación donde mis informantes me están mandando mensajes'. (...) Lo que hacía ella es que se metía a la misma aplicación pero en la computadora."

"She would say, obviously: 'What is this? It's useless to me because I can't receive images or files, so I disabled it because I'm right in the middle of an investigation and my informants are sending me messages.' (...) What she ended up doing was opening the same app, but on her computer instead."

P10 mentioned also using the computer as a workaround or even asking her peers for doing the task that was blocked for her in LDM:

"Sí, es decir, como yo trabajo todo el día en la computadora eso es mi workaround también. Es decir, hay cosas que yo solo hago en la compu. Si estoy fuera de la computadora, lo que hago es que le pido a alguna amiga, a alguien cercano que haga algo por mí."

"Yes, I mean, since I work on the computer all day, that's my workaround too. I mean, there are things I only do on the computer. If I'm away from the computer, what I do is ask a friend, someone close to me, to do something for me."

Within the group that left LDM enabled permanently, they mentioned they understood the trade-off and they decided to leave LDM enabled forever. P07 explained:

"Claramente estas en la página streaming y no se me renderizan los videos, esta bien. O no

se me ven las fuentes en algunas aplicaciones como el banco. Pero a mi la verdad me importa muy poco. Ya de por si la aplicación es bien fea y me genera estrés solo verla. En general yo no tengo mayor impacto. Con respecto al uso de batería yo creo que evidentemente hay impacto pero yo no diría que es tan significativo al menos en mi caso que tengo un iphone más nuevito, con una vida de batería un poco más larga.”

Obviously, when I’m on a streaming page the videos don’t render for me, and fine. Or some fonts don’t display in certain apps, like the banking app. But honestly, I don’t really care. The app is already ugly and just looking at it stresses me out. Overall, it doesn’t have much impact on me. As for battery usage, I think there’s clearly some impact, but I wouldn’t say it’s that significant, at least in my case since I have a newer iPhone with a slightly longer battery life.”

For digital security trainers, the experience showed that when assessing at-risk users’ threat models and deciding to use LDM together, the resulting LDM-enabled setting lasted longer. Understanding the trade-offs before using LDM was critical to improve retention. P08 explained the process in detail:

“Pues por el contexto en el que trabaja la mayoría de las personas a las que se lo recomendamos, no se quejan (del LDM) porque ya hicimos el paso de análisis de riesgo. La mayoría de las quejas vienen de personas que lo configuraron sin haber hecho el análisis (de riesgo). Entonces después vienen a quejarse: ‘¿Por qué? ¿Por qué necesito esto? No veo que estoy ganando nada’.”

“Due to the context surrounding the work of most of the people we recommend (using LDM), they don’t complain (about LDM) because we already went through a risk analysis with them. Most of the complaints come from people who enabled it without having done that (risk) analysis. Then they come back complaining:

‘Why? Why do I need this? I don’t see what I’m gaining’.”

We analyse and provide recommendations for the limitations on using LDM in the Discussion section.

5.4 Expectations using spyware mitigation features

LDM is well-known as a spyware mitigation feature. In part because Apple’s perceived safety is outstanding, and it has made it easier for Apple device owners to enable it with just one swipe. Some participants noted that LDM is not a third-party extension or add-on, which they considered a safe path for adoption.

Expert participants also mentioned trust in the LDM design choice. Reducing the attack surface while downgrading the user experience is a path taken by other security-enhancing tool developers, like the onion routing application layer, the Tor Browser. Furthermore, at-risk users are more resilient and are willing to experiment with reduced usability to keep their communications and work running (Bellini et al., 2023).

When discussing LDM specifically, expert users were able to describe how it works, recalling their prior experience with hardening techniques on other devices or operating systems, as explained in the previous section.

Besides the trust factor in deciding to use LDM, when we asked whether the device would be compromised with LDM enabled, most participants replied affirmatively. All of them said that the devices could be compromised even with LDM enabled.

When we asked about LDM expectations after they first enabled it, the responses focused on reducing the chance of arriving at a point where a user action would lead to a malicious path, or, more commonly, on not having any expectations. P03 was explicit:

"Mi expectativa es, bueno, por lo menos el ataque no va a llegar hasta este punto donde la decisión es de la persona"

"My expectation is, well, at least the attack won't reach the point where the decision is up to the person"

P10 and P02 were aseptic:

"No, no tenía, o sea, no tenía expectativas así, como decir: 'Oye, esto va a ser...' No, era más como: Quiero ver qué pasa, qué se rompe."

"No, I didn't, I didn't have expectations like, 'Hey, this is going to be...' No, it was more like: I want to see what happens, what breaks."

"No tenía expectativas. Yo nada más dije: 'Bueno, a ver, ¿qué deja de funcionar?'"

"I had no expectations. I just said: 'Well, let's see, what stops working?'"

One of the reasons is that they did not have something similar to compare in the past for their iOS devices.

When we asked whether LDM was enough to mitigate the different ways a person can be attacked, the answers were diverse. Some participants like P06 said yes, LDM offered enough protection.

Other participants were less optimistic about the coverage, like P05, who stated that "*todos los sistemas pueden ser hackeados / all systems can be hacked*". Or participants like P03 that flagged that even with LDM enabled, an attacker could rely on social engineering (using an urgent message or file as a bait) to get a journalist or lawyer to disable LDM:

"Hay otras maneras de lograr el objetivo (de atacar) como que elimine el Modo Hermético o lo desactive. Piensa en periodistas o abogados, perfiles a los que les cuesta

trabajar seguridad digital, porque les cuesta sostener cuidados. Incluso con Modo Hermético habilitado, (un atacante) siempre va a encontrar una manera de hacer que la persona termine diciendo: 'Yo si quiero acceder a esto'. Pero como no puede hacerlo por el dispositivo por LDM: o lo desactiva o va a la compu donde Lockdown Mode no está y puede acceder a esos mensajes."

"There are other ways to achieve the goal (of attacking), such as getting the person to remove or disable Lockdown Mode. Think of journalists or lawyers, profiles that struggle with digital security because it's hard for them to maintain care practices. Even with Lockdown Mode enabled, an attacker will always find a way to make the person eventually say, 'I do want to access this.' And since they can't do it on the phone because of Lockdown Mode, they either disable it or switch to the computer, where Lockdown Mode isn't present, and can access those messages."

5.4.1 Expected breakage. A significant trade-off of relying on security-enhancing features is a degraded user experience. As we showed in previous sections, the worst path for security is seeing participants circumvent the breakage by moving from an LDM-enabled device to a device without LDM enabled to perform the task. Even worse, disabling a global setting like LDM to perform one task (e.g. open a file) with the risk of not raising global shields again was heard more than once.

The human factor in deciding whether LDM is effective relates to the capacity to execute the tasks they aim to perform.

5.4.2 Mismatches. More interestingly, there was a notable discrepancy between participants' perceptions of the feature's function and what it actually offers. In Section 2.2 we listed the features LDM is offering to the userbase. It allowed us to detect the differences between what the feature was offering and what the users were expecting. P6 was extremely confident in stating that LDM, when

enabled, introduced device communications into an incognito mode:

“Lo que hace, entiendo yo, (es que) te restringe los puertos de entrada de las aplicaciones del celular. Eso hace que de alguna manera estés más incógnito en la red de telefonía y en la red de internet. Es por ahí el funcionamiento del Modo Hermético.”

“What it does, as I understand it, is restrict the entry ports of the phone’s apps. That makes you, in a way, more hidden on the phone network and on the internet. That’s basically how Lockdown Mode works.”

Another interesting mismatch concerned the perception of infection signals when LDM was enabled. P10 mentioned how she felt signals caused by LDM behaviour appear identical to infection alerts:

“Si vieras que al tiempito de estar lo usando, era interesante porque eran como todas las señales de compromiso. Cuando uno decía, ay, el teléfono se está calentando. Y yo decía, ay, es el Lockdown Mode. Ay, el dispositivo se me puso lento. No, pero yo creo que es el Lockdown Mode”.

“If you only knew, after using it for a little while it was interesting because it felt like all the signs of compromise. When you’d say, oh, the phone is heating up. And I’d go, oh, it’s Lockdown Mode. Oh, the device got slow. No, I think it’s Lockdown Mode.”

It is relevant for developers of spyware mitigation tools to reduce the gap between what the mitigation offers and what users understand it is doing, to reduce the false sense of security.

5.4.3 False sense of security. While the adoption of LDM was partly because digital security trainers has been doing immense on-the-ground work educating users about the benefits of spyware-mitigation features, they are also critically aware of

the limitations of a “one-click” security shield:

“Entonces, si yo lo estoy usando sin haber hecho análisis de riesgo, lo que hemos visto es que la gente eventualmente empieza a buscar atajos o saltarse la forma de quitarse ese inconveniente de encima. Entonces eso nos puede crear un falso sentimiento de seguridad, lo cual para mí es peor.”

“So if I’m using it without having done any risk analysis, what we’ve seen is that people eventually start looking for shortcuts or ways to get around the inconvenience. And that can create a false sense of security, which to me is even worse.”

Our analysis provides a detailed understanding of LDM mitigation capabilities of spyware attacks by HRD and other users at-risk. On the one hand, participants have specific ideas about how LDM works, and they also believe that appropriate threat modeling when deciding to use it, is critical to achieving retention. On the other hand, our participants were transparent about the issues they faced in making LDM their default choice.

In this section, we provide design recommendations to improve the user experience and user interface of spyware mitigation defenses in LDM, based on our study's results. We sorted the recommendations by relevance and impact on the at-risk end-user experience.

6.1 System feedback visibility. Users don't know when LDM is enabled. P08 shared his experience enabling LDM for first time:

"Fue muy decepcionante la primera vez que lo habilité. No noté ningún cambio al inicio. Yo pensé que iba a haber un downgrade visual, por ejemplo, como en Windows."

"It was very disappointing the first time I enabled it. I didn't notice any changes at first. I thought there would be a visual downgrade, like in Windows."

P04 detected LDM enabled after something broke unexpectedly:

"Creo que el dispositivo nunca me notificó que estaba habilitado el Lockdown Mode, hasta donde recuerdo. Fueron cosas que se rompieron."

"I don't think the device ever notified me that Lockdown Mode was enabled, as far as I remember. Things just stopped working."

This may be a feature, but the issue arises when they infer that LDM is enabled when something breaks or becomes unusable. We recommend using a persistent indicator when LDM is enabled. Another option would be to display disabled feature icons in

the Status Bar.

After LDM is enabled for the first time, the user interface provides no feedback. We recommend a walk-through of the feature, setting expectations about why something may not work, and offering an anchor for the trade-off.

It was common to hear participants explain how LDM disrupted their regular tasks. The main issue is the opaque experience when something isn't loading, a message isn't sent, or it simply doesn't work. Users don't know why something is broken, so they blame LDM in any case and deactivate it preemptively during a task. P03 explained how a peer in her organization was blind to discerning why something was not working:

"Tenía un antivirus en su iPhone y le estaba causando muchos problemas. No sé si era por el Modo hermético o porque el antivirus en un iPhone simplemente no funcionaba bien. Pero siempre tenía ese problema y lo atribuía al Modo hermético."

"She had an antivirus installed on his iPhone and it was causing a lot of problems. I don't know if it was because of Lockdown Mode or simply because the antivirus wasn't working well on an iPhone. But she kept having that issue and blamed it on Lockdown Mode."

Again, we recommend anticipating which features will be disabled or the expected experience degradation at first-time use of specific apps or in-context. The communication about these changes to the system and user experience should be trustworthy and empathetic, especially for Apple-native apps.

6.2 Informative feedback. When LDM is enabled, people lower their guards. There is a common belief that LDM is effective against specific threats or attacks. Altho it raises a false sense of security, as described in Section 5.4.3. P06 was transparent:

“(después de habilitar LDM) aquí no entra nada, no importa, yo descargo de lo que me haga la gana. La falsa certeza de seguridad es uno de los más eminentes riesgos. Entonces para mí es importante pero no es suficiente. Tiene que complementarse como otras medidas operacionales y de uso y de concientización del usuario”

“(after enabling LDM) nothing comes in here, it doesn't matter, I download whatever I want. The false certainty of security is one of the most imminent risks. So for me, it is essential, but it is not enough. It has to be complemented like other operational and usage measures and user awareness”.

Participants think they can follow bad security practices because LDM will keep them secure. We recommend communicating effectively that the use of LDM requires to be accompanied with complementary operational security and good digital security practices.

6.3 User control. Participants with advanced technical background, familiar with manual hardening, want granular control over LDM settings. P02 compared the feature with GrapheneOS:

“Por eso me gusta más Graphene, porque sé, ellos te dicen exactamente qué es lo que están cubriendo y cómo lo están cubriendo, y tú puedes desactivar ciertas cosas o activar ciertas cosas. Tienes como un control granular de las cosas. No solo es un swipe donde, bueno, y ahora no me funciona el zoom y no sé por qué”.

“That's why I prefer Graphene, because they tell you exactly what they're covering and how they're covering it, and you can turn certain things on or off. You have granular control over everything. It's not just one swipe and suddenly zoom doesn't work, and you have no idea why.”

We recommend allowing granular control over

LDM Advanced Settings. Granular control over general settings may help in reducing breakage in specific scenarios. Another option would be to set the rule of three: block everything, block everything untrusted, or soft block. Ideally, we aim to prevent the user from downgrading the global, OS-level security setting to perform a single task.

6.4 Compatibility. Issues arose from third-party apps that do not work with LDM enabled. It motivates users to switch devices to overcome breakage barriers and perform sensitive tasks. A commonly described flow is that a user with an LDM-enabled device cannot open media in iMessage, so they open the iMessage app on their MacBook.

Compatibility is particularly relevant because LDM is designed for high-risk users, who already operate with a multilayered digital security strategy and rely on a broad ecosystem of privacy focused apps. For this group, if security enhancing apps break under LDM, users are forced to use workarounds that weaken their overall protection. While it is hard to be compatible with the entire app ecosystem, we recommend partnering with the most widely used apps in the security and privacy ecosystem to ensure LDM works for those who need it most. Reducing breakage in these kinds of security-enhancing applications is critical to improve adoption. P010 shares how doing a sensitive task is not possible with LDM enabled: “*ProtonPass tiene la opción de compartir de manera segura contraseñas. Eso no funciona / ProtonPass has the option to securely share passwords. That doesn't work*”. Again, allowing users to set up a one-time contextual setting may prevent a general security downgrade. We also recommend flagging to the developer community that LDM is enabled. In that case, apps can gracefully downgrade their usability when needed.

Nine of twelve participants described their use of VPNs in their operational security setup to raise the standard on their network traffic privacy: “*Sólo usamos VPNs, por ejemplo, o sea, está estrictamente*

prohibido en el teléfono de seguridad, está, inclusive por configuración del teléfono, si la VPN no está activa, no hay conexión de internet. Es decir, obligatoriamente tiene que estar la VPN / We only use VPNs, for example; that is, it's strictly prohibited on the security phone. It's even configured on the phone so that if the VPN isn't active, there's no internet connection. In other words, the VPN must be enabled". We recommend to offer iCloud Private Relay while LDM is enabled to both cover users from using a third party solution and also perform consensual private traffic analysis on that at-risk device.

7. Limitations

We ran our study with acknowledged limitations. First, focusing on Apple's LDM inevitably narrowed who could take part in the study. Apple devices account for only 15.45% of the mobile market in Latin America, while Android represents about 84.39% of active devices (StatCounter, n.d.). In a region where average wages are comparatively low and iPhones are priced far beyond local purchasing power, the cost barrier puts iOS devices out of reach for most at-risk users. Brazil, for example, was cited by Forbes as one of the most expensive countries in the world to buy an iPhone (Buchholz, 2025). This geographic and platform limitation restricts how far our findings can be generalized. The majority of human rights defenders in the region depend on Android devices and therefore face different constraints and protection opportunities.

Second, our participants included digital security trainers as well as HRDs that mentioned they had previously participated in security training and risk assessments. Being part of those training environments probably influenced the way they described threats, protection strategies and even LDM itself.

Third, our research focused primarily on individual

device usage and did not fully explore the implications of Apple ecosystem synchronization when Lockdown Mode (LDM) is enabled across multiple devices, nor the impact of remote LDM configuration. Understanding how LDM settings propagate and function across synced devices (e.g., iPhone, iPad, Mac) remains an area requiring further investigation, particularly regarding potential vulnerabilities or usability challenges that may arise from cross-device implementation. Furthermore, we did not examine family sharing dynamics and reactions when one member enables LDM. Given that Apple's Family Sharing features allow multiple users to share purchases, subscriptions, and device management, activating LDM by one family member may increase awareness among other members.

Finally, the sensitivity required to avoid re-victimizing participants presented ongoing ethical challenges throughout the research process. Discussing experiences of targeted surveillance, harassment, or other digital threats requires careful, trauma-informed approaches that may have limited the depth or breadth of data we could ethically collect.

Several promising directions emerge from this research. First, future studies should investigate **cross-device security implications within the Apple ecosystem, examining how LDM functions when enabled on one or multiple synced devices, and whether users understand the scope and limitations of protection across their connected devices.**

Second, **research into family and shared device contexts would provide valuable insights into the social dimensions of security tool adoption.** Studies could explore how LDM activation affects family members who share accounts or devices, and how to better support collective security decision-making in household and organizational contexts.

In this line, **expanding research to bridge users at-risk experiences with major development companies by developing a consensual, non-extractive, with sensitive and deep understanding of who the subjects of study are, a feedback cycle process, where civil society organizations, sharing with spyware mitigation developers insights about usability, comprehension, and case of use, to inform further developments.** These cycles of feedback help technical teams prioritize their development to have a significant impact on communities in need and increase adoption among those who need it most.

TABLE 1

#	DESCRIPTION	GENDER	TECH-PROFICIENCY
P1	Digital security trainer, operational security auditor, working on capacity building	Male	Expert
P2	Security researcher, forensic analyst, working with lawyers, journalists and grassroots organizations	Male	Expert
P3	Technologist, longtime independent infrastructure developer, security trainer providing support for women and LGBTIQ+ journalists, activists and gender, land and labor rights, and freedom of expression defenders	Female	Expert
P4	Security researcher, working with lawyers, journalists and grassroots organizations	Male	Expert
P5	Digital security trainer working with women and LGBTIQ+ journalists	Female	Expert
P6	Journalist and social media administrator	Male	Novice
P7	Principal lead in helpline, working with land and labor rights defenders	Male	Expert
P8	Helpline incident operator, working with human right defenders	Female	Novice
P9	NGO Director providing privacy and security support and companion to women through the peace program	Female	Novice
P10	Principal technical person in women-led grassroots organizations on popular education and reproductive rights	Female	Expert
P11	Journalist working with political affairs in major national media outlet	Male	Novice
P12	Helpline incident responder, working with human right defenders	Male	Expert

TABLE 2

COUNTRY	PARTICIPANTS
Costa Rica	3
Mexico	3
El Salvador	2
Guatemala	2
Brasil	1
Peru	1

- Abu-Salma, R., & Livshits, B. (2019). *Evaluating the end-user experience of private browsing mode*. arXiv. <http://arxiv.org/abs/1811.08460>
- Abu-Salma, R., Redmiles, E. M., Ur, B., & Wei, M. (2018). *Exploring user mental models of end-to-end encrypted communication tools*.
- Altaf, W., & Tiefenau, C. (2018). *Folk and expert mental models of security and privacy*.
- Amnesty International. (2022). *The Pegasus Project: How Amnesty Tech uncovered the spyware scandal*. Amnesty International. <https://www.amnesty.org/en/latest/news/2022/03/the-pegasus-project-how-amnesty-tech-uncovered-the-spyware-scandal-new-video/>
- Amnesty International. (2023). *Dominican Republic: Pegasus spyware discovered on prominent journalist's phone*. Amnesty International. <https://www.amnesty.org/es/latest/news/2023/05/dominican-republic-pegasus-spyware-journalists-phone/>
- Apple. (2022). *Apple expands industry-leading commitment to protect users from highly targeted mercenary spyware*. <https://www.apple.com/newsroom/2022/07/apple-expands-commitment-to-protect-users-from-mercenary-spyware/>
- Apple. (2025). *About Lockdown Mode*. Apple Support. <https://support.apple.com/en-us/105120>
- Bellini, R., et al. (2023). *SoK: Safer digital-safety research involving at-risk users*.
- Bravo-Lillo, C., Cranor, L. F., Downs, J., & Komanduri, S. (2011). *Bridging the gap in computer security warnings: A mental model approach*. IEEE Security & Privacy Magazine, 9(2), 18–26. <https://doi.org/10.1109/MSP.2010.198>
- Braun, V., & Clarke, V. (2021). *Thematic analysis: A practical guide*. SAGE Publications Ltd.
- Buchholz, K. (2025). *How many hours of work pay for an iPhone 17?* Forbes. <https://www.forbes.com/sites/katharinabuchholz/2025/09/12/how-many-hours-of-work-pay-for-an-iphone-17/>
- Camp, L. (2009). *Mental models of privacy and security*. IEEE Technology and Society Magazine, 28(3), 37–46. <https://doi.org/10.1109/MTS.2009.934142>
- Citizen Lab. (2023). *BLASTPASS: NSO Group iPhone zero-click, zero-day exploit captured in the wild*. Citizen Lab. <https://citizenlab.ca/2023/09/blastpass-nso-group-iphone-zero-click-zero-day-exploit-captured-in-the-wild/>
- Coopamootoo, K. P. L., & Groß, T. (2014). *Mental models for usable privacy: A position paper*. In T. Tryfonas & I. Askoxylakis (Eds.), *Human aspects of information security, privacy, and trust* (pp. 410–421). Springer International Publishing. <https://doi.org/10.1007/978-3-319-07620-136>
- Dodier-Lazaro, S., Abu-Salma, R., Becker, I., & Sasse, A. (2017). *From paternalistic to user-centred security: Putting users first with value-sensitive design*.

- Dominguez Rubio, L. (2026). *Inteligencias intrusivas: tecnologías de interceptación y monitoreo en Argentina, Chile y Uruguay*. SocialTIC. [En prensa]
- Felt, A. P., Ainslie, A., Reeder, R. W., Consolvo, S., Thyagaraja, S., Bettes, A., Harris, H., & Grimes, J. (2015). *Improving SSL warnings: Comprehension and adherence*. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (pp. 2893–2902). ACM.
- Fundación Karisma. (2024). *La compra de Pegasus en Colombia demuestra la necesidad reformar la Ley de Inteligencia y Contrainteligencia*. <https://web.karisma.org.co/la-compra-de-pegasus-en-colombia-demuestra-la-necesidad-reformar-la-ley-de-inteligencia-y-contrainteligencia/>
- Gentner, D., & Stevens, A. L. (Eds.). (1983). *Mental models*. L. Erlbaum Associates.
- Gross, J. B., & Rosson, M. B. (2007). *End user concern about security and privacy threats*. College of Information Sciences and Technology.
- Hooks, B. (1984). *Feminist Theory: From Margin to Center*. Cambridge, MA: South End Press.
- IDEO. (2016). *The little book of design research ethics*.
- Jaffer, J. (2022). *Why We're Suing NSO Group: Dada v. NSO Group*. Knight First Amendment Institute. <https://knightcolumbia.org/blog/why-were-suing-nso-group>
- Kang, R., Dabbish, L., Fruchter, N., & Kiesler, S. (n.d.). *'My data just goes everywhere.' User mental models of the Internet and implications for privacy and security*.
- Kitroeff, N., & Lopez, O. (2022). *Key evidence in report on missing Mexican students cannot be verified, experts say*. The New York Times. <https://www.nytimes.com/2022/10/31/world/americas/mexico-43-missing-students.html>
- Knight First Amendment Institute. (n.d.). *Dada v. NSO Group: A case challenging the use of spyware against journalists*. <https://knightcolumbia.org/cases/dada-v-nso-group>
- Lin, J., Amini, S., Hong, J. I., Sadeh, N., Lindqvist, J., & Zhang, J. (2012). *Expectation and purpose: Understanding users' mental models of mobile app privacy through crowdsourcing*. In Proceedings of the 2012 ACM Conference on Ubiquitous Computing (pp. 501–510). ACM. <https://doi.org/10.1145/2370216.2370290>
- Marczak, B., Scott-Railton, J., Abdul Razzak, B., & Deibert, R. (2023). *Triple Threat: NSO Group's Pegasus spyware returns in 2022 with a trio of iOS 15 and iOS 16 zero-click exploit chains*. Citizen Lab. <https://citizenlab.ca/2023/04/nso-groups-pegasus-spyware-returns-in-2022/>
- Megiddo, G. (2024). *\$13m Cash on a Private Jet: How Colombia Paid for Israeli Spyware*. Haaretz. <https://www.haaretz.com/israel-news/2024-03-26/ty-article-magazine/.pr/13m-cash-on-a-private-jet-from-colombia-a-nonissue-for-israeli-head-of-defense-export/0000018e-7689-d706-a39f-f7f93fa10000>
- Norman, D. (1982). *Some observations on mental models*. 7-14.
- Nottingham, M. (2020). *The Internet is for end users* (RFC No. 8890). Internet Engineering Task Force. <https://datatracker.ietf.org/doc/html/rfc8890>

Oates, M., Ahmadullah, Y., Marsh, A., Swoopes, C., Zhang, S., Balebako, R., & Cranor, L. (2018). *Turtles, locks, and bathrooms: Understanding mental models of privacy through illustration*. Proceedings on Privacy Enhancing Technologies, 2018(4), 5–32. <https://doi.org/10.1515/popets-2018-0029>

R3D. (2023). *Periodista de República Dominicana fue espiaada con Pegasus, denuncia Amnistía Internacional*. R3D. <https://r3d.mx/2023/05/05/periodista-de-republica-dominicana-fue-espiaada-con-pegasus-denuncia-amnistia-internacional/>

R3D, SocialTIC, & Article 19. (2017). *Gobierno espía 2017: Documento del caso NSO en México*. R3D. <https://r3d.mx/wp-content/uploads/GOBIERNO-ESPIA-2017.pdf>

Raja, F., Hawkey, K., Hsu, S., Wang, K.-L., & Beznosov, K. (2011). *Promoting a physical security mental model for personal firewall warnings*. In CHI '11 Extended Abstracts on Human Factors in Computing Systems (pp. 1585–1590). ACM. <https://doi.org/10.1145/1979742.1979812>

Renaud, K., Volkamer, M., & Renkema-Padmos, A. (2014). *Why doesn't Jane protect her privacy?* In E. De Cristofaro & S. J. Murdoch (Eds.), *Privacy enhancing technologies* (pp. 244–262). Springer International Publishing. <https://doi.org/10.1007/978-3-319-08506-713>

Revista Raya. (2024). *Virus espía Pegasus lo compró el gobierno de Duque con dineros del narcotráfico*. <https://revistaraya.com/virus-espia-pegasus-lo-compro-el-gobierno-de-duque-con-dineros-del-narcotrafico.html>

Rouse, W. B., & Morris, N. M. (1986). *On looking into the black box: Prospects and limits in the search for mental models*. Psychological Bulletin, 100(3), 349–363. <https://doi.org/10.1037/0033-2909.100.3.349>

Sambasivan, N., et al. (2018). *'Privacy is not for me, it's for those rich women': Performative privacy practices on mobile phones by women in South Asia*.

Scott-Railton, J., Marczak, B., Claudio Guarnieri, & Masashi Crete-Nishihata et al. (2017). *Bitter Sweet: Supporters of Mexico's soda tax targeted with NSO exploit links*. Citizen Lab. <https://citizenlab.ca/2017/02/bittersweet-nso-mexico-spyware/>

Scott-Railton, J., Marczak, B., Abdul Razzak, B., Masashi Crete-Nishihata, & Deibert, R. (2017, June 19). *Reckless Exploit: Mexican Journalists, Lawyers, and a Child Targeted with NSO Spyware*. Citizen Lab. <https://citizenlab.ca/2017/06/reckless-exploit-mexico-nso/>

Scott-Railton, J., Marczak, B., Nigro Herrero, P., Abdul Razzak, B., Aljizawi, N., Solimano, S., & Deibert, R. (2022). *Project Torogoz: Extensive Hacking of Media & Civil Society in El Salvador with Pegasus Spyware*. Citizen Lab. <https://citizenlab.ca/2022/01/project-torogoz-extensive-hacking-media-civil-society-el-salvador-pegasus/>

Singh, A., Dechant, M., Patel, D., Soubutts, E., Barbareschi, G., Ayobi, A., & Newhouse, N. (2025). *Exploring positionality in HCI: Perspectives, trends, and challenges*. In Proceedings of CHI '25: CHI Conference on Human Factors in Computing Systems. Association for Computing Machinery.

StatCounter. (2025). *Mobile operating system market share in South America*. StatCounter GlobalStats. <https://gs.statcounter.com/os-market-share/mobile/south-america>

Tactical Technology Collective. (2016). *Holistic security: Trainers' manual*.

Valdés, I., & Torres, M. (2024). *EEUU confirma que financió la compra del 'software' Pegasus para su uso en Colombia y que lo suspendieron por supuesto uso indebido*. CNN Español. <https://cnnespanol.cnn.com/2024/11/08/pegasus-uso-colombia-eeuu-confirma-financio-compra-orix>

Van Oorschot, P. C. (2021). *Computer security and the internet: Tools and jewels from malware to bitcoin*. Springer. (pp. 269-275)

Volkamer, M., & Renaud, K. (2013). *Mental models—general introduction and review of their application to human-centred security*. In *Number theory and cryptography* (pp. 255–280). Springer.

Warford, N., Matthews, T., Yang, K., Akgul, O., Consolvo, S., Kelley, P. G., Malkin, N., Mazurek, M. L., Sleeper, M., & Thomas, K. (2022). *SoK: A framework for unifying at-risk user research*. In *2022 IEEE Symposium on Security and Privacy (SP)* (pp. 2344–2360). IEEE.

Wash, R. (2010). *Folk models of home computer security*. *Proceedings of the Symposium on Usable Privacy and Security (SOUPS)*. <https://doi.org/10.1145/1837110.1837125>

WhatsApp Inc. v. NSO Group Technologies Ltd., No. 4:19-cv-07123-PJH (2025) *Exhibit 2 to Declaration of Micah G. Block in Support of Plaintiffs' Opposition to Defendants' Motion for Summary Judgment or Partial Summary Judgment, Deposition of Carl Woog*. <https://storage.courtlistener.com/recap/gov.uscourts.cand.350613/gov.uscourts.cand.350613.679.6.pdf>

Pérez de Acha, G. (2016). *Hacking Team: Malware para la vigilancia en América Latina*. *Derechos Digitales*. <https://www.apc.org/es/pubs/hacking-team-malware-para-la-vigilancia-en-america-latina>

Suárez, A. (2024). *EEUU admite que financió compra del software espía Pegasus en Colombia*. AP News. <https://apnews.com/article/colombia-eeuu-pegasus-software-espia-714dc3e016d74325a917a7aee2b576de>

Rodríguez, A., Recamier, M., Souza, D., & Franco, D. (n.d.). *Vigilad@s: Jalisco invierte millones en tecnologías para espiar y monitorear*. *Zona Docs*. <https://www.zonadocs.mx/vigilads-jalisco-invierte-millones-en-tecnologias-para-espiar-y-monitorear/>

