ARTICLE 19

Queer resistance to digital oppression

Protecting MENA's queer communities: Recommendations for tech companies

July 2024

Part III

In collaboration with the De|Center

ARTICLE 19
72-82 Rosebery Ave
London EC1R 4RW
UK
www.article19.org

ARTICLE 19 is an international think–do organisation that propels the freedom of expression movement locally and globally to ensure all people realise the power of their voices.

Together with our partners, we develop cutting-edge research and legal and policy analysis to drive change worldwide, lead work on the frontlines of expression through our nine regional hubs across the globe, and propel change by sparking innovation in the global freedom of expression movement. We do this by working on five key themes: promoting media independence, increasing access to information, protecting journalists, expanding civic space, and placing human rights at the heart of developing digital spaces.

T:	+44 20 7324 2500
F:	+44 20 7490 0566
E:	info@article19.org
W:	www.article19.org
X:	@article19org
Fb:	facebook.com/article19org

ARTICLE 19 would appreciate receiving a copy of any materials in which information from this report is used. ARTICLE 19 bears the sole responsibility for the content of the document.

The research team that produced this report can be contacted at afsaneh@de-center.net and MENA@article19.org.

Table of Contents

Abbreviations

AI	Artificial intelligence
API	Application programming interface
DFM	Design From the Margins
DNS	Domain name system
iOS	iPhone Operating System (Apple).
IP	Internet protocol
LGBTQI+	Lesbian, gay, bisexual, transgender, queer, and intersex
MENA	Middle East and North Africa
NGO	Non-governmental organisation
NIST	National Institute of Standards and Technology (USA)
OWASP	Open Worldwide Application Security Project
PIN	Personal identification number
SMS	Short message service (text messaging)
VoIP	Voice over internet protocol
VPN	Virtual private network

Acknowledgements

These acknowledgements are drafted by Afsaneh Rigot. Afsaneh is the creator, principal researcher, and coordinator of the project on the tech-facilitated harms against lesbian, gay, bisexual, transgender, queer, and intersex (LGBTQI+) communities in the Middle East and North Africa (MENA) and its reports, working alongside the LGBTQI+ community. She led this project as part of ARTICLE 19 from 2016 to 2023 and has continued as a consultant senior adviser and principal researcher for the reports while transitioning into new positions outside ARTICLE 19. This research was built out of her deep love and admiration for the MENA queer community and her belief in the vision for their better futures.

Afsaneh and the De|Center team continued to collaborate in the writing of these reports, and their analyses provided necessary tech methodology practices under Design From the Margins.

The list of those who supported this work is too long to fit into this section, but we deeply thank every person involved in the project. These include the following teams:

In **Algeria**, we would like to thank the Algerian team (who will remain anonymous for their safety). We especially thank them for the data they gathered and for conducting focus groups in a context that had not been covered before due to its great risk.

In **Egypt**, we thank the partnership and work of the Bedayaa organisation whose work and support have been pivotal throughout this project, especially for gathering data and lived experiences via interviews and focus groups from Egypt’s queer community.

In **Iran**, we would like to thank our advisers and LGBTQI+ experts, but especially journalist and human rights defender Khosro Isfahani. Khosro conducted desktop research and in-depth interviews to gather vital insights from a particularly at-risk community.

In **Jordan**, we would like to thank Bin Amman (alias), Khalid Abdel-Hadi (editor-in-chief at *My Kali* magazine), and activist Hasan Kilani for their brilliant work in gathering context and conducting deep interviews in Jordan under increasing risks and difficulties.

In **Lebanon**, we would like to thank the many people that supported the work, especially researcher and sexual orientation, gender identity, and expression (SOGIE) expert Genwa Samhat, who was the lead coordinator and conducted deep interviews with some of the most marginalised community members, and Helem, an LGBTQI+ non-governmental organisation (NGO), for their pivotal support in conducting focus group discussions (with support from Sally Chamas) and gathering data for the surveys. We also thank the legal experts from Legal Agenda who provided invaluable legal and policy reviews and advice for this work.

In **Morocco**, we would like to thank Youba Darif and Roots Lab Morocco for their continuous and rich collaborations, impactful work, and leading data gathering in interviews and focus groups.

In **Sudan** we would like to thank Azza Nubi, Sam Adam, and Gamil for their deeply important work during one of the most painful periods of the country's history. They gathered contextual information and conducted deep interviews in a complex environment. Due to their meticulous work, we are able to present a small insight into the tremendous power of the Sudanese queer community, even during one of the worst humanitarian crises in modern history.

Finally, in **Tunisia**, we would like to thank Mawjoudin for their rich knowledge, interdisciplinary work, and support in this project, especially for gathering meaningful interviews and conducting focus groups.

We would also like to thank all the legal experts who provided reviews and insight for the complex legal labyrinths in each legal context. None of this would have been possible without the work of the community behind this report.

Further, we would like to thank our research and project assistants throughout the years: queer feminist expert Senda Ben Jebara; Ali Bousselmi, co-founder and executive director of Mawjoudin; and our Algeria expert, who will remain anonymous. They were a guiding light for this report and the main reason this project continued to function through all its complexities.

We would also like to thank our technical and international experts, many of whom have supported the work behind the scenes and attended our convenings to push for the implementation of the community's needs by tech companies. These include Kendra Albert, Mohammed Al-Maskati, Mahsa Alimardani, Sarah Aoun, Dia Kayyali, Moussa Saleh, and Apryl Williams.

We especially thank Nathan Freitas and the Guardian Project, Carrie Winfrey of Okthanks, and Norman Shamas, who were our technical experts co-leading the tech recommendations of these reports and many of its technical paths.

We also thank writer and editor Roja Heydarpour for her diligent, deeply attuned, and brilliant editing of texts of this project. We thank Manar Elharaké, the director of Astrum, for his excellent work coordinating the analysis of the thousands of complex survey data from many different dialects and languages and supporting the work to bring the findings to light. We thank our brilliant designer for the striking design work; as they are part of the community, they will remain anonymous. We thank Jessica Fjeld and Aymen Zaghdoudi for reviewing the report and their continuous support for the work.

Finally, we would also like to thank the numerous ARTICLE 19 teams and individuals who made this report possible, such as Maria Luisa Stasi who conducted our legal oversight, the ARTICLE 19 comms, editorial, and production team, and the Iran and MENA team.

This set of reports and the impact this work has had in the last eight years would not have been possible without the deep commitment, bravery, and genius of the LGBTQI+ community and experts on the ground fighting for better futures for all. This work is dedicated to them and their resistance. It is also dedicated to Palestine and the Palestinians, who, during the finalisation of this report, have been facing one of the most devastating and violent periods in modern history. Palestine might not be a focus of this report, but many of us learned the true meaning of living our truth, and living in resistance and resilience, from them.

Introduction

Between 2019 and 2024 we conducted research and investigation in **eight MENA countries** which covered **15 focus groups, 93 in-depth interviews, 5,000+ surveys,** and **two large community convenings** to highlight and show the methods of harm and human rights abuses faced by the LGBTQI+ community in the region, particularly through technology facilitated abuses. These were carried out in Algeria, Egypt, Iran, Jordan, Lebanon, Morocco, Sudan, and Tunisia. These years of work, and the years prior, have been condensed into three reports that outline the situation on the ground. This research and work behind it are part of in-depth community-led work we have been conducting on this topic since 2015–16.

In addition to the documentation and work of the project, the communities and individuals that took part in our interviews, focus groups, and surveys also provided insight into how and what technology companies can do to reduce risks when they come into contact with law enforcement, including protecting their identity, or making it harder to be tracked or reported for that identity. These actions are a step towards meaningful support for a community most impacted, but least consulted.

In our investigations and the documentation with the community, we wanted to find out what people are using to connect or for self-expression, what makes them feel safe or unsafe, and what they want companies to change to make them safer in their contexts. This community is acutely aware of the risks members undertake when using these technology platforms and thus can identify what is causing them harm and what change is needed. These final elements are vital in understanding and enacting meaningful change as part of tech companies' broader corporate responsibility to the community, and society at large. Implementing these changes will support the resistance and self-preservation methods already taken by the community.

We were able to gather this information and insight thanks to the courageous and diligent work of our country experts in all eight countries (see the important reflections from each of the researchers in Part I.

This report shows paths forward with an eye towards challenging how technology has been weaponised against the community. At the same time, technology has been providing support for the already savvy and ingenious LGBTQI+ communities of MENA, who have been forging groundbreaking ways to resist violence, abuse, and arrest in order to continue living and loving in community. Many of our recommendations include implementation proposals for engineering and developer teams at companies – from code bases to user experience design. It is important that the recommendations are implemented with the correct framing, and with justice, human rights, and privacy at their core.

The report outlines concrete and granular harm reduction changes for apps and platforms. With technical insight from our experts, we lay out the method of implementation for broad privacy requirements and harm reduction features to be implemented. **We provide 16 recommendations for privacy changes to existing infrastructure, 1 recommendation about hate speech and rapid response reporting systems, and 15 recommendations for feature changes, including suggestions for harm reduction methods against arrests and device searches.** The proposals and recommendations are curated and framed based on patterns and observations from arrests, prosecutions, and abuses the community faces. Their implementation is a step towards acceptance of accountability by big tech, as well as meaningful support for a community most impacted, but least consulted.

The community faces brutal harms and human rights abuses that are not only overwhelming, but complex and difficult to challenge without deep systemic, social, and legal changes at all levels of power in the focus countries, as well as globally. These complexities and methods are exponentially exacerbated by the reliance on technology by law enforcement and security apparatuses, especially communication tools. These are tools of connection and communication that the community uses that police and state actors are increasingly weaponising. Often, these tools fail the community through privacy breaches and harms, data harvesting, and security and safety gaps. In our research, we see how states and law enforcement weaponise these vital tools to surveil, target, entrap, and frame members of the LGBTQI+ community – not only for arrests, but also for extortion and various levels of violence. They operate with impunity and disregard for the rule of law, which has meant that non-state actors are also able to use the same technologies to harm and abuse LGBTQI+ people, leaving the victims with no social or legal protections against state or non-state abuses.

The level of complexity and coordination in the methods used to prosecute queer people by state sanctioned systems, as outlined in our reports, calls for and requires creative methods to mitigate the harm and trauma inflicted on these queer communities. Pushing for legal reform and decriminalisation are long term goals for which non-governmental organisations (NGOs), activists, and experts on the ground tirelessly advocate. But this work is not only theirs to shoulder. In the short term, we see a path forward that can provide meaningful safety and harm reduction for the community. The multimillion- and multibillion-dollar companies whose technology is used and implicated in these harms have huge roles to play here. This report urgently outlines a base of action for these companies, which are being used to form a new type of digitised prosecution. Technology companies need to address harms and rebuild trust with those impacted.

> *'As a trans woman and a sex worker, I find that the level of insecurity and violence allowed in these apps shows that these apps have no regard for us and our safety.'*
>
> – Interviewee in Algeria

> *'I feel like they don't care about the safety of their users, they just care to maximise their profit.'*
>
> – Interviewee in Jordan

Our focus countries are not the only countries to prosecute LGBTQI+ people, or to have abusive policing and law enforcement structures that allow for marginalised communities to be targeted.[1] It is certain that the use of digital platforms in criminalisation in such cases will not be unique to these countries or even to the LGBTQI+ community. The communities facing the same abuses must also be consulted and protected – these recommendations can also play a role in their safety.

The learning and recommendations outlined in this report are based on what we documented in the other two reports of this series, including the direct demands of thousands of people in the community. Our investigations have shown how adversarial actors weaponise gaps in safety, security, and privacy on apps and platforms, and how it is often **those with the least protection who are the most affected** first. Throughout the history of this work and similar investigations, we see that the patterns and methods used against highly marginalised communities are often then expanded and tested on more generalised and wider populations. Fundamentally, with this understanding, it should be seen that the recommendations of this report, if implemented, will impact and benefit people beyond this community who face similar risks, providing further safety and privacy for all who use these technology tools (see Design From the Margins).

This report looks at ways technology companies should work to:

1. **Support this impacted community** based on the harms they have reported and the requests they have outlined.
2. **Fulfil their due diligence and human rights obligations**.
3. **Make their apps and platforms more robust and secure**, and **introduce safeguards** to protect users from their misuse and weaponisation.

The issues the LGBTQI+ MENA community faces are varied and complex, enforced by large powers and oppressive might. We therefore require multilayered approaches to combat them, including support for systemic changes at both a governmental and a societal level. This report is one layer of the methods we can use. Our questions for this report are: How do we support those affected by vast harms and human rights abuses so that they can continue to connect and communicate more safely? And what can provide the community with more ways to exercise their freedom of expression, protect themselves, and navigate the risks while using technology tools?

In this final report of the series, we offer the first layer of urgent actions. In Parts I and II, we saw how individuals and organisations worked to challenge and navigate the modes of violence they face. From there, based on our documentation and the direct needs and wants of the community, we have formulated recommendations and priorities for technology companies implicated in these human rights abuses.

The changes we outline here might seem small, but in our years of work and experience in situations of severe risk, as we have seen in the investigations of this series, they will have massive impact. Our recommendations are practical for implementation, research-based, and also rooted in the direct wants and needs of the affected community. Bridging research, documentation, and how people use their technologies, we outline some of the main recommendations needed for change.

> *'I hope that applications and social media are more interested in our privacy and the fight against hate speech and work to fill the technical gaps that allow hijacking and surveillance and worse. User safety is more important than their profits.'*
>
> – Interviewee in Sudan

In regard to the responsibility of companies, we echo the notes from the Digital Crime Scenes report:

> *This burden [of safety] should not fully befall those using these technologies. In fact, this report would argue that the heavier reliance and necessity people have on these tools triggers the responsibilities of companies involved. Many of the apps, tools, and platforms were created for very different contexts, and the effects of their technologies are far reaching beyond any preformed contextual analysis. There is a real need to discuss this impact in contexts these technologies were not designed for and the effects of western centrism on vulnerable and/or hard-to-reach communities.*
>
> *The responsibility, therefore, of app developers and providers is key. The burden of protection should not be solely on users, and UN standards make clear that companies have human rights responsibilities. App companies must make the effort to understand their users' environments and experiences; sending security messages is simply not enough. As these technologies expand in use and importance, radically transforming lives and how we communicate, the application of human rights responsibilities to businesses becomes more and more crucial. They owe proactive protective, security, and safety measures to their users. … This is especially the case when these apps are functioning in countries in which there*

> *are higher risks to marginalised users. In fact, the UN Special Rapporteur on Freedom of Expression has clarified that for ICT companies, a proper process of due diligence requires considering the human rights impacts of 'design and engineering choices'.*

For this work we have adopted the Design From the Margins (DFM) methodology (see below) which remains at the forefront of this work. DFM is grounded in the knowledge that when those most marginalised are designed for, we are all designed for. It requires a radical reshifting of how we build our technology so that those most impacted by social, political, historical, and legal structures (the decentred) are the communities we build technology with and for from the onset of the technology design processes. We call on technology companies and platforms to adopt DFM methods and standards and further engage in a meaningful shift in how their technologies are built, scaled, and targeted.

> *'I wish we can get to a place where we can do a revolution on these platforms and on the management that they have – which is to force these companies to comply with what the people want. I will be very happy watching this.'*
>
> – Interviewee in Jordan

We are at a vital juncture. Radical change is necessary. There are increasing harms through technology, and the political impetus often prioritises fast band-aid options over structural thinking. We must design with those communities most impacted and left at the margins in mind. We must design based on those most harmed by security and privacy issues with full understanding of the contexts that impact those deemed vulnerable and/or hard-to-reach communities. What we often see are easy retroactive fixes to raise issues of harm that, in turn, harm those who are already the most marginalised.

DFM-based changes require a combination of community-based research, movement lawyering, harm reduction-focused technology interventions, and precise implementation details. These are the ingredients of this project's work and the method used. They are also what will translate the human rights documentation into vitally needed and direct technology changes.

Our overall demands are:

1. For companies to understand the context and harms experienced and to provide proactive measures for user safety through the application of human rights principles and DFM methods for their technologies.

2. For companies to build technology that can serve to protect and not become a tool in the toolbox of repression against LGBTQI+ people in MENA and around the world. **By implementing our recommendations in this report, companies must show accountability for harms and provide further safety for these at-risk communities.**

3. For companies to globally implement the learning and changes from this research, with a focus on safety for other communities facing similar harms and abuses.

'This made me feel really good that someone cares enough about making these apps, which are providing services in the third world, more secure. I hope that we will see some changes soon and we will not be disappointed.'

– Interviewee in Iran in reference to the research

Design From the Margins

One of the core elements of the report is the power of designing while centring the most marginalised and criminalised (the 'decentred'). This is the DFM methodology. This methodology pushes for just, equitable, safer technology that centres the most impacted and decentred users, from ideation to production. DFM is grounded in the knowledge that when those most marginalised are designed for, we are all designed for. Technology design needs to meet people where they are, with the systems they use, with a direction towards harm reduction.

Decentred users are the groups most at risk and under-supported in the relevant context. In the DFM methodology, technology design is not separate from the broader historical, political, social, and institutional contexts that surround and impact human interactions. Through understanding and establishing who is most impacted by existing power structures, we can also understand who would most likely be harmed when technology is weaponised.

Decentred groups are not just marginalised – they are also usually the most criminalised. They are the people the state not only fails to protect, but actively persecutes. Moreover, decentred communities are often located outside the USA and EU, and many of the harms and abuses embedded in our technology are rooted in capitalistic, heteronormative, racist paradigms that are byproducts of Western-centrism. The term 'decentred' plays with the notion that these users should not be at the margins but instead hold central power. It is a known truth that often the oppressed know more about the oppressor than the oppressors know about themselves. This knowledge is the fundamental tool in decentred communities' resilience, and with it, they generate their own power, adopting methods of self-protection in order to navigate the threats and risks they face.

Within the DFM methodology, once the most impacted ('decentred') cases are identified, they should be designed for directly – not through retrofitting – and their voices and needs should set the rules of engagement throughout the processes. Their experiences and needs for harm reduction are placed at the core of production, and ultimately influence the final product that is delivered to the general user base.

DFM interventions must acknowledge that technology is not neutral or an inherently positive force. When decentred communities' expertise about a technology's broader impacts and harms is ignored, that technology will perpetuate those harms and further become a tool of oppression.

This project should not be seen as one case study in order to only address the needs of the specific LGBTQI+ communities of MENA (who themselves are varied, nuanced, and complex). By centring these extreme cases, and building from these, safer and justice-oriented products are created. DFM and this work move to prove, empirically and practically, that centring the most marginalised creates better technology for all.

> *'Even in the West, many people are discreet and don't like everyone to know what they are doing. Freedom does not mean that people do not care about their privacy.'*
>
> – Interviewee from Iran

We used this method for this report. We have a detailed picture of how individuals are affected, which parts of these technologies are used against them, and how they become detrimental to users. By focusing on how to help lessen the access and power given to forces looking to prosecute queer people in these contexts, we can reimagine existing and future technologies for communication that are safer, more private, and centred with the needs of those at the margins.

Recommendations and technology changes

We have compiled data, oral histories, consultations, and the expertise of our team of technical experts in order to make the recommendations for changes needed from companies in this report.[2] ARTICLE 19 and the research lead have already pushed for many of these recommendations to be implemented by partnering companies in the past. Some of these have been successfully implemented. Since 2017, we have had variations on these recommendations for our behind-the-scenes advocacy with companies. While we have never made them public, readers may be familiar with them as some are commonly on your apps today.[3] We did not want to risk giving away these strategies before our communities could fully adopt them to stay safe. However, we are now making them more public for the first time, and with further resources. With technology harms increasing, we believe it is important to share this knowledge and encourage more companies to implement these changes and understand the needs of those most impacted.

Many experts have been working on this section of the work alongside ARTICLE 19. The main team working on these recommendations are the De|Center and Afsaneh Rigot (the principal investigator),[4] and technical experts Nathan Freitas of the Guardian Project and privacy and harm reduction expert Norman Shamas.[5] Please reach out to the teams for more information and guidance. These recommendations would not be possible without the collaboration and expertise of our partnering LGBTQI+ organisations, experts, and researchers. See Part I for their reflections on the work.

Our recommendations target dating apps, chat-based apps, and social media platforms. We also look at some important recommendations for operating system providers. The first section of recommendations looks at privacy changes for existing infrastructure and features of the apps/platforms. The second section focuses on hate speech and rapid response reporting systems. The third section focuses on the introduction of harm reduction features for the apps/platforms to implement.

Summary of recommendations

Privacy changes to existing app/platform infrastructure

Main recommendations

Recommendation 1: Ensure privacy and data security, especially for highly marginalised users.

Recommendation 2: Create and use contextualised and nuanced methods to authenticate and verify users, especially those in high-risk contexts.

Recommendation 3: Do not rely on phone numbers for discoverability and verification. Support marginalised users.

Recommendation 4: Delete all content from both devices after a user is blocked.

Recommendation 5: Remove real-name requirements and ensure rights to anonymity and pseudonymity.

Recommendation 6: Do not add more access barriers. Support communities to access apps and platforms.

Recommendation 7: Do not allow apps to save photos taken or received in the main device photo gallery by default.

Dating app specific recommendations

Recommendation 8: Dating apps should make vital safety features free, especially in high-risk contexts.

Recommendation 9: Dating apps should immediately disable options that allow for the non-consensual sharing of dating profiles, especially for high-risk context.

Recommendation 10: Safer and more private geolocation practices.

Chat-based app specific recommendations

Recommendation 11: Provide the option to have a proper separation between Facebook Messenger and Facebook profile.

Recommendation 12: Add safety measures for 'Stories' and conduct further research in emerging safety issues.

Social media app specific recommendations

Recommendation 13: More control and privacy with tagged photos.

Recommendation 14: Do not expose and out users through 'friend recommendations', and on app activities do not expose unintended information.

Recommendation 15: Add safety measures for 'Stories' and 'Lives' and conduct further research on the emerging safety issues.

Operating system specific recommendations

Recommendation 16: Conduct further research and work to challenge the use of jailbreaking or rooting and allow apps to opt out of operating system features, including new artificial intelligence features.

Hate speech and rapid response reporting systems

Recommendation 17: Combat hate speech and implement robust and contextualised reporting systems and direct lines of communication.

Harm reduction features needed

Main recommendations

Recommendation 18: Implement app icon 'stealth mode' options (app cloaking/discreet app icons) to hide the icon of the app in plain sight.

Recommendation 19: Implement stealthy self-destruct/panic button (or similar) options for emergency situations and blocking access to device content, especially for the most high-risk users.

Recommendation 20: All apps should have PIN and locking features, as well as added stealthy locked folders for the most sensitive content/chats.

Recommendation 21: All apps should provide ephemeral, delete for all, and 'view once' text and media messaging options, and implement them safely.

Recommendation 22: Provide methods and features to prevent non-consensual screenshotting and capturing of users' information.

Recommendation 23: Provide in-app video and photo blurring options.

Dating app specific recommendations

Recommendation 24: Remove distinctive sounds and notifications for queer dating apps.

Recommendation 25: Dating apps should have an option for incognito mode (a mode only to be seen by people who the user has verified).

Recommendation 26: Dating apps should have in-built video and voice call options.

Recommendation 27: Dating apps should provide contextualised information

Chat-based app specific recommendations

Recommendation 28: Chat-based apps need options to allow notifications to be paused for pre-set periods.

Social media app specific recommendations

Recommendation 29: Social media apps should have an option for incognito mode (a mode only to be seen by people a user wants to be seen by).

Operating system specific recommendations

Recommendation 30: Operating systems must have device-level ‘stealth mode’ or cloaking for apps and folders.

Recommendation 31: Operating systems should have a device-level and stealthy self-destruct or panic button option.

Recommendation 32: Operating systems should make certain content lists or contents hidden.

PRIVACY CHANGES NEEDED TO EXISTING APP/PLATFORM INFRASTRUCTURE

Main recommendations

Our first layer of recommendations is related to privacy gaps and safety deficiencies that may be present on many apps and platforms. Ensuring robust privacy and safety on the existing tools and infrastructure of the communication tools mentioned is vital for building more nuanced changes and improvements. The list of what would be needed in terms of strong, robust, and holistic privacy changes is long. However, here we have focused on the most urgent short-term actions to be taken based on our work, research, and the overwhelming desires and wants for change from those who participated in our research.

Privacy and data security are of vital importance to the community and are something that affects how they use and choose their platforms. This right must be protected by states and companies; as ARTICLE 19 has stated: 'Guaranteeing the right to privacy in online communications is essential for ensuring that individuals have the confidence to freely exercise their right to freedom of expression.' This idea was recognised in several reports by David Kaye, the Special Rapporteur on Freedom of Expression, in which he expressed concern about private actors monitoring and collecting information about individuals' communications and activities on the internet: 'These practices can constitute a violation of internet users' right to privacy, and ultimately impede the free flow of information and ideas online.'

The following pages detail the **16 recommendations for privacy changes to existing infrastructure, including their implementation proposals.**

We need privacy and data security

Recommendation 1: Ensure privacy and data security, especially for highly marginalised users.

The privacy of highly marginalised users must be prioritised with robust end-to-end encryption by default. This data should not be monetised and gathered through deceptive design patterns. In turn, shielding these groups will also protect the privacy of all users.

This recommendation echoes the long held concerns of privacy and digital rights advocates, in particular, who have long warned that the mass gathering of data and 'surveillance capitalism' are both manipulative and breaching users' human rights.[6] Apps and platforms should understand and respect the principle that transparency about what happens to private data creates a deep sense of trust with platforms – especially for those whose personal information and digital movements can be used to criminalise them.

Context and research behind the recommendation

Highly marginalised users require increased privacy needs and the preservation of their data on apps and platforms in order to build any trust in the technologies they use. The combined concept of 'privacy and data security' here refers to the user's overall privacy when using all aspects of an app or platform. For example, how their data is used, stored, and re-purposed, including with any third parties. This would undoubtedly include the proper handling of sensitive data, and the ethical and responsible use of any data collected. Here, it broadly refers to our respondents' request for 'data ownership' so as to prohibit acquisition and use of personal data without an individual's intentional consent. This would also encompass providing tools that provide data protection including robust encryption as a safety measure should there be unauthorised access from a state actor, for example.

A recurrent theme in all areas of our research is the community's call for privacy and concern for the nefarious use of their data. Many participants and interviewees saw this issue as intimately linked to safety, especially from threats of violence from the state.

General privacy and data security are one of the most mentioned wants identified in this research.

> *'These apps must provide the maximum protection possible in terms of privacy and security. All apps should do that. Maybe this should become law because the apps may skip this – not to lose customers who might move to other apps as they are easier and do not have these security barriers.'*
>
> – Interviewee in Iran

Interviews:	**34 out of 93** (37%) interviewees mentioned privacy and data security as what they want to see from platforms.
Surveys:	**305 out of 2,482** (12%) people mentioned that 'data security' (generally referring to data ownership so that it prohibits acquisition of personal data without individuals' intentional consent and privacy) was what they would ask from companies when addressing their safety and well-being. It was the second most mentioned concern. **392 out of 2,482** (16%) people pointed to 'more safety' and **120** asked for 'privacy features' (these are addressed in some of the changes and features outlined later).
Focus groups:	In **4 of the 6 countries** where we conducted focus groups, this issue was heavily mentioned.

One interviewee from Iran explained their stance on privacy and the issues of 'surveillance capitalism':

> *'The question is, is there anything that these giant companies do not know about us? And where is the red line? If the service being free means that users are seen as commodities, what implications does this have for not only my data but also my life?'*

Another interviewee in Egypt simply said:

> *'I think these platforms can help us by stopping tracking us and gathering our data.'*

Protection and transparency about what happens with private data have a very deep link to a sense of trust with platforms, and feelings of safety.

> *'I don't know if the apps have the right to use our data or not. It is very important to me that my data is not used. I don't know if the apps follow this policy or not. I am more worried that this data will be somehow given to the Islamic Republic. This is the worst thing that can happen.'*
>
> – Interviewee in Iran

In their responses, our participants and interviewees reported **using their own methods to enhance privacy**. We learned that people chose to use two-factor authentication, and other privacy features such as virtual private networks (VPNs), to protect themselves as best as they could, with the understanding that these alone did not maintain their safety or privacy. When introducing new safety and privacy features, we recommend consultation with LGBTQI+ communities in MENA and other impacted communities, as well as privacy experts, in order to ensure that their experiences are taken into account and that the new features do not harm the most vulnerable and marginalised users.[7]

Participants and interviewees also pointed to the importance of **using apps and platforms that have secure end-to-end encryptions.**

> *'[End-to-end] encryption should not be considered optional and should be a legal requirement for applications.'*
>
> – Interviewee in Iran, with the understanding that use of end-to-end encryption alone does not provide security on any platform

In our survey, this want was very important and at the top of the list:

***456 out of 2,482** (18%) who answered the open text box question in the surveys outlined wanting 'encryption & security of personal data', which was the top result from this question.*

***234 out of 2,482** (10%) who answered the question also outlined 'general privacy and security' as highly important to them.*

Further recommendation details

Although we do not dive deep into details here, as this idea has been well outlined by many privacy and human rights organisations, we provide a general baseline for platforms for the implementation of this recommendation.

How can it be implemented? (for app developers and security teams)

App developers play a critical role in protecting user data from malicious actors (e.g. ad libraries and data brokers who sell location data which can accurately identify individuals) and ensure that they provide proper controls to users. Companies should also ensure that they provide adequate privacy controls over sensitive data that are easy to understand, easy to access (preferably in a single location), and avoid the use of deceptive design patterns. Additionally, permissions and contextual privacy notices should be provided.

Outside of users' control, app developers should minimise the data collected directly and by third parties, such as ad libraries or researchers, to prevent that data from being abused. This should include preventing third-party libraries from collecting direct location data and sending that data over insecure methods, which could also introduce new security and privacy issues.

When it comes to encryption, robust **end-to-end encryption by default must be implemented**. Our emphasis here is that 'encryption' is more than a brand or term. How it is implemented and its robustness are vital. As privacy researchers in 'What is secure: analysis of popular messaging apps' from Tech Policy Press state:

> *Implementation is everything. The failure to implement end-to-end encryption by default, such as on Telegram and Meta's Messenger, illustrate this point. Users may not understand the distinction when presented with confusing options like 'secret chat' and 'private chat'.*

Thus the level of privacy is disingenuously presented and not protective without thorough implementation of end-to-end encryption by default on their apps. Yet, the use of deceptive language hints towards end-to-end encryption by default-level privacy.

It is also important to note the vital differences between encryption of network traffic (e.g. protect against internet server providers and other unintended parties from accessing the data), data storage (e.g. protect data on device from being accessed by an unauthorised person with access to the device or another installed application), and end-to-end encryption of messages in a chat (e.g. protect against the app from accessing content of messages). These require different approaches and robust implementation.

We need safer and nuanced authentication and verification

Recommendation 2: Create and use contextualised and nuanced methods to authenticate and verify users, especially those in high-risk contexts.

App companies should improve account verification and authentication processes with special consideration of the risks to highly vulnerable groups. Addressing fake profiles and entrapment accounts is vitally important. **However, this needs to be done with care. Due to risks faced by users, most require options to remain anonymous and keep only minimal data saved or added to their apps or profile.**

Authentication: A robust authentication process should be required to interact with the back-end application programming interface (API). Companies should also seek to reduce any exposure of details that might be leveraged for account identification or enumeration. We also urge all authentication processes or services to follow the general guidelines (laid out in the full recommendation implementation details).

Verification: Apps and platforms should push for safer and more creative verification methods, especially for highly marginalised users. Verification should not focus or rely on verifying the identity of a 'real person' or similar identity verification services, especially as many of those most at risk rely heavily on anonymity. It is vital to find more nuanced ways to make it harder for adversaries to create fake profiles to contact communities at risk.

Context and research behind the recommendation

In our research – as with all of our previous investigations – issues around fake accounts and entrapments (police using fake profiles and personas to lure LGBTQI+ people on dates, only to arrest or extort from them) are a huge cause for harm, violence, or arrest and are thus high priority. Our community participants and researchers felt there has been little to no work done to solve the issue. Researching and developing nuanced methods to **limit the ease in which fake profiles are made and weaponised** are highly important.

Due to the scale of this issue, unsurprisingly, **20 interviewees throughout the 8 countries** asked for better verification procedures in order to challenge fake profiles. This issue was also mentioned in **4 of our 6 countries** where we held focus groups (in Egypt, Lebanon, Sudan, and Tunisia).

In Part II, the high percentage of those experiencing arrests, police entrapments, and similar forms of abuse is reported:

***45%** of our survey and interview participants had **experienced arrests based on their identities**.*

***23%** had experienced **online entrapment** by police and state-affiliated actors.*

The most mentioned apps used against respondents for entrapment were **Grindr**, **Tinder**, **WhosHere**,[8] **Facebook**, **Instagram**, **WhatsApp**, and **Facebook Messenger**. When identified on a platform, they very often moved to other chat-based apps (if they were not already on one) such as **Telegram** or **Signal**. In the case of **Facebook Messenger**, people can be both identified (on Facebook) and continue the conversation on the chat-based platform (Messenger) in one place. Part II shows the increasing issues of security forces and police and fake profiles. A summary of statistics only seen in our research reports shows the extent of the issue.

The issue of fake accounts also expands to non-state-level outings and extortions via apps as outlined in the report. The most mentioned apps for fake account abuse are dating apps such as **Scruff**, **Grindr**, **Tinder**, and **Hornet**, but this occurs also through social media such as **Facebook**, **Instagram**, **Twitter**, and others.

For more details, see Part II, particularly the section on non-state outing, honey traps, violence, and extortion via apps.

Despite the pervasiveness of the issues, many of the methods used to challenge these abuses have harmed more than helped the community. Due to the legal and state-level targeting and surveillance of queer people in MENA, methods deployed in the West cannot be used (although queer communication app users in the USA, for example, are also sceptical of sharing certain information for a variety of reasons). Many create profiles without pictures that would identify them or show their faces. Thus, any verification of a user should be done with an understanding of this context. One interviewee exemplified this anxiety to verify, without exposing the information used:

> *'At least the app itself should allow a user to join only after it makes sure that this is a real person, and make sure that they are not connected to security forces. … Of course, they respect privacy and not share this [background] information with others, but assure us that the user is not fake.'*

In most dating apps, social media apps, and chat-based apps, the **use of methods to verify and authenticate users has caused concerns and increased likelihood of risks**. This case is especially true for methods that require users to provide photos of valid identification cards (IDs) or selfies to verify themselves, or ones that rely on third-party social media accounts or phone numbers to authenticate.[9] Most often, if this is asked, users will not use the app due to increased risk – this was verified in our research and discussions with local groups.

Further recommendation details

Methods to counter forms of entrapment and fake accounts that lead to arrests, evidence-gathering, and entrapment of individuals are vitally important. **However, this needs to be done with care. Due to risks faced by users, most require options to remain anonymous and keep only minimal data saved or added to their apps or profile.**

We understand that for many companies a priority is to address fake profiles of scammers and financial catfishing accounts. On top of these issues, in the regions we are working in, there are many cases of **fake profiles created to harass, persecute, arrest, or blackmail queer users**. The dilemma here is that the solution cannot require more personal, identifying information from these at-risk users. There must be more creative and privacy-preserving methods to address the need for authentication and verification.[10]

One of the common methods for verification, and the prevention of fake accounts, that we have observed and explicitly noted is companies using phone numbers for account sign-up. This practice puts users directly at risk (see Recommendation 3) and is not always a good signal for verification purposes. **Rethinking better privacy-preserving methods for authentication and verification would have concrete positive impacts on user safety.**

How can it be implemented? (for app developers and security teams)

Authentication and verification are two related processes which can be used to ensure a safe community on the app. Authentication focuses on ensuring a person is authorised to access information, for example logging into a dating application. Verification, on the other hand, focuses on ensuring the person is who they say they are, often focusing on identity verification. App companies should **improve account authentication and verification processes** with special consideration of the risks to highly vulnerable groups.

App companies should **improve account registration and verification processes** and **enforce API service to require valid authentication**. Having a proper account registration and verification process in place increases the difficulty for attackers to take over an existing user's account, and increases the difficulty of creating dummy accounts and for fraudsters and spammers to create bots en masse through automation.

Further, in countries where the use of LGBTQI+ dating services for same-sex sexual conduct is criminalised or persecuted, state actors may seek the ability to automatically enumerate users from a given country. Similarly, this could be done for those within a social group or family. By exposing, for example, whether a given phone number has already been used to register an account, the service unintentionally provides confirmation to anyone checking if the associated person is a user of the service.

A **robust authentication process should be required to interact with the back-end API** so that excessive information on users is not exposed publicly, and to increase the difficulty of creating malicious software to automatically interact with the API.

Companies should **seek to reduce any exposure of details that might be leveraged for account identification or enumeration.** Many services rely on an internal, undocumented API service to make clients interact with these APIs so that excessive information about users is not exposed publicly, and also to increase the difficulty of creating malicious software to automatically interact with the API.

We recognise that there is no single solution, and in some instances differing methods and options can be used by any app, with different options based on different regions or contexts.

In general, it is important to **provide authentication options other than social media or phone numbers**, and to **avoid verification processes that are linked to requiring user 'selfies' or real IDs**. Our implementation outline for each provides the initial paths:

Our recommendation for safer authentication

Any dating app authentication process or service should adhere to the following guidelines:

- Ensure the **authentication process is secure**. For specific guidance on how to secure authentication, see OWASP's (Open Worldwide Application Security Project) authentication cheat sheet.
- Ensure that all dating app service functionality and API endpoints have adequate authentication in place.[11]
- Provide authentication methods that do not require single sign-up through social media (especially platforms that pose even higher risks due to their real-name requirements) or using a phone number. In most cases, having a **registration option that relies on a username and email** will work.
- Consider **proactive abuse detection mechanisms** for rate-limiting creation of new accounts from specific internet protocol (IP) ranges, unexpected device types, or other unusual patterns.

- Utilise Google Play Integrity API or other '**apps device authentication**' technology to ensure the mobile app software has not been modified, to detect malware in-app, or to not allow it to run on rooted or unknown hardware. This approach is currently used by the gaming and banking industries primarily to stop cheating and theft.

Our recommendation for safer verification – a need for creativity

No form of authentication should replace a method for verification. It is relatively simple to create a fake social media account or fake email account, or to get an unverified phone number (e.g. through a SMS API service).[12] Any verification method should take this into account.

Verification does not have to focus or rely on verifying the identity of a 'real person' or similar identity verification services. Historically, queer people and queer communities do not always rely on information that can be verified by an ID (or similar ways that identity verification services operate), and instead rely on alternative methods to identify and verify other queer people. For example, slang languages (e.g. Gayle in South Africa) or specific locations (e.g. gay bars) have been used to identify members of queer communities.

We also know that there are very few platforms and options available that gather identity data without risking the exposure of this data or its monetisation and/or misuse. It is therefore vital to find more nuanced ways to make it harder for adversaries to create fake profiles to contact communities at risk.

There are no broadly accepted alternatives used by apps. However, there are methods[13] other apps and platforms have used to verify accounts that can be used for inspiration such as:

- Using **key types of engagement/activities** as a way to unlock additional features or functionality (e.g. gamification),[14] such as the method used by Ahwaa, a queer LGBTQI+ platform used in MENA.[15] (We can connect you with our research and technical team and have them advise you further. We can be contacted at afsaneh@de-center.net and MENA@article19.org.)

- **Adding hurdles through questions** is a method mentioned specifically to us throughout this research:
  - Around **10 participants** in our research directly mentioned the use of 'tests' or questions that are based on knowledge commonly known by the LGBTQI+ community. It would be a prerequisite for making an account on LGBTQI+-focused apps or groups/profiles. This method would not be foolproof, but it adds an extra level of difficulty for adversarial groups in making mass accounts to target the community.
  - These questions should of course be made with the engagement of community groups to make sure they are not harmful or exclusive.
  - This concept could also adopt the idea of badges or verified markers that show an individual's passing rate of the verification questions. 'There is an app [I use] that provides you with a psychological questionnaire right at the beginning. Such features can make these apps safer,' one interviewee said.
- **Optional verification**, as seen on many apps, is when the user decides to provide further verifying information themselves, with safety risks explained. The user can take additional steps to 'prove their identity' and display it on their profile, as the site Keybase does.
- **Verification through other dating app users**, such as 'verified profiles'. The queer dating app Romeo (previously known as PlanetRomeo) uses a unique and generally beloved version of this method. 'On PlanetRomeo one thing that is very useful and I would like to see it in other gay chat and dating applications. In Planet Romeo, you can't chat until you have 10 people who gave you a Recommendation. … It is very useful for security, especially in our context,' stated one interviewee in Morocco. However, this should be done without creating an

identifiable network of people connected on apps since that could create immediate risks to anyone arrested, as well as their network. We suggest more work and research into these inter-community and non-data-intrusive verification methods.

No matter which method of verification is explored, we recommend **transparency with users** in order to help them make informed decisions. Many of the current methods to prevent fraud and abuse use metadata and other data indicators, much of which can identify a specific device or person and pose privacy risks. We recommend against using those as the primary verification method.

Finally, since this is currently a design challenge with no agreed-upon core solution, we also recommend partnering together with our research teams to **implement a prototype** with funding.[16] We can use the needs and expertise of these users and our teams to create something industry-led and based on the wants and security needs of the most marginalised and targeted queer communities on these apps. Further advice for this can be obtained from our research and technical team. We can be contacted at afsaneh@de-center.net and MENA@article19.org.

We do not want to give you our phone numbers

Recommendation 3: Do not rely on phone numbers for discoverability and verification. Support marginalised users.

Our overall recommendation is the need for apps to move away from the continued reliance on phone numbers for discoverability and verification. This change must become an industry-wide change and we have seen its commencement. There must be efforts made to introduce added and alternative tools for discovery and verification. We ask that companies understand the added harm and marginalisation this reliance creates for highly marginalised and criminalised communities. The strategy used in its place should not add another layer of marginalisation for those already affected.

We also call for apps to fully refrain from using and sharing phone number information and **immediately halt the suggestions of profiles or 'friends' based on these collected numbers**. Further, apps and platforms must ban the use of design methods that deceptively nudge users to share phone numbers via dark patterns.

In the interim, while we wait for this move away from phone number reliance, we further recommend that **apps do not block the use of voice over internet protocol (VoIP) numbers** (a VoIP number is a real telephone number operating on an internet connection). For verification, **phone numbers must remain separate and disconnected from all other account information**, including account details and advertisement profiles.

We firmly recommend partnering with our research teams and/or other expert privacy and human rights research teams with adequate resources to investigate solutions for this contested area.

Context and research behind the recommendation

Participants and interviewees in this research have raised the issue of phone numbers and the reliance on them by communication platforms such as dating apps, social media, and messaging apps as an important privacy and safety issue they want to see companies change. We also see this issue as clearly linked to numerous arrests and risk factors we have documented – see Part II, especially the section 'Risk of using phone numbers'. This has been an ongoing privacy concern for years.

Even if the phone number is not visible or used for discoverability, the required process to register with a phone number through SMS verification raises many access and privacy concerns that have yet to be addressed. This issue is industry-wide. A key challenge is that the majority of countries have SIM card registration laws that require handing over sensitive data such as identification or even biometrics. They are retained in a trackable and searchable way by the governments. In those contexts, SIM cards act as a unique identifier for an individual and a tool for surveillance.

This explains why people place high importance and concern around phone number safety. We can see this in the rate that they raised the concern in our current research:

Interviews:	**12 out of 93** (13%) interviewees directly asked for a halt in the use of phone numbers as the main method to register and access social media and chat-based and dating apps due to risks to their safety and issues of access.
Surveys:	**1,472 out of 5,018** (28%) of our respondents said sign-up/login methods made them feel the most unsafe. This was the highest rated answer.
Focus groups:	In all **6 countries** where focus groups were held, these issues of real numbers were raised and a halt to this requirement was requested.

> *'People should be allowed to create an account on social networks with minimal information, that is, even without a phone number.'*
>
> – Interviewee in Iran

> *These applications confirm our accounts by the phone number, I would like to find another way of confirmation.'*
>
> – Interviewee in Algeria
>
> *'One of the problems with Signal is that it asks for a phone number, which is not necessary. Signal is a very good app, but I don't understand why it has this downside. Why not use an email address?'*
>
> – Interviewee in Iran
>
> *'WhatsApp is not secure. Especially since you can see the phone numbers very easily.'*
>
> – Interviewee in Iran

Our recommendation for phone number visibility

The risks associated with phone number registration and discoverability (meaning your phone number being used to search for you, for example, on a chat-based platform), and also the visibility of a phone number on a communication app, has been raised in this research. It has also been well documented in other contexts including our previous research.

In Part II, especially the section 'Risk of using phone numbers', we outline how phone number use has led to risks and even prosecution, identification, outings, and other abuses. We also outline how the reliance on them links individuals' legal identities to their dating apps, chat-based apps, or social media. This has, in turn, led to underground markets for LGBTQI+ people's numbers and, in some cases, led to selling them to police and state actors.

For example, authorities are able to link numbers to people's legal and official identity, leading to outing and prosecutions with digital evidence. In the case file analysis, it was clear that even if individuals used fake names in a conversation with police who were using entrapment profiles, the phone number from the chat was used with a screenshot from the entrapment account. When the individual was arrested in the entrapment and sting operation, the confiscated device and the SIM card of an arrested individual were used to match their phone number to conversations from informants or to polices' phone with the real name and identity of the individual. In many cases, this has led to years of imprisonment as it becomes nearly impossible to challenge in court. It is a very large security and privacy gap in design.

New methods such as the use of the platform **Truecaller** to reverse search and to link the identified phone number to the legal name and identity of individuals is another tactic that is being used increasingly. It is made more dangerous due to the prevalence and reliance on phone numbers.

> *'Some people can use your phone number on the Truecaller to find your real name and get some information they can use against you.'*
>
> – Interviewee in Tunisia

Researchers from Tech Policy Press outline the issue and its complexity with chat-based (messaging) apps:

> *Phone numbers are considered sensitive, personally identifying information, but almost every app we reviewed treats them slightly differently. … At present, only Telegram offers the option to hide one's phone number in favour of a username. While Signal and WhatsApp currently lack this feature, both appear to have plans to implement it soon.*

Signal released the highly requested feature for usernames in February 2024. This move is an important one that we hope to see on other apps such as WhatsApp. Signal is yet to remove the requirement for registration via phone numbers and SMS verification. This is the case for the majority of major social media, dating, and chat-based apps. Other platforms like Facebook Messenger still collect phone numbers without needing it for functionality, which is highly predatory.

Our recommendation for phone numbers for registration

Beyond these immediate risks of phone number visibility, phone numbers used for registration and SMS verification on apps and platforms also pose varied and complex risks that outweigh their utility as a registration tool:

- The discrimination through the reliance on phone numbers for registration has caused a lot of harm and disenfranchisement in countries like Iran and Sudan. Many of the most well-known communication and dating apps in Iran and Sudan exclusively allow registrations with text messages or phone numbers. Some apps do not even list countries such as Iran or Sudan as a country when requiring selection of country codes from a list. They do not accept their phone numbers, therefore effectively banning Iranian users.[17] See Part II, especially the section 'Impact of sanctions on the LGBTQI+ community: Iran and Sudan', to read more about the impacts of isolation and the harm users experience as a result of this practice, particularly as they are in the most high-risk contexts where technology tools can be life-saving.

  Furthermore, there is a massive risk posed to how phone numbers can become connected across accounts and create social graphs without consent or knowledge of individuals. Many social media platforms have used phone numbers[18] to suggest 'friends' that might be in an individual's phonebook but not on their social media, risking outing and other harms without the knowledge of the users. This issue is especially acute on Meta platforms, TikTok, and some chat-based platforms. As one interviewee in Morocco put it:

  'About Facebook and Instagram: do not use numbers to suggest friends; on Facebook and Instagram it puts me at risk.'

  In Iran, blocks have led to underground markets selling and price gouging for phone numbers so people can still access these apps. One interviewee in Iran said:

‘A very hot market has been created where they sell phone numbers. Now they are selling foreign phone numbers that you can use for Tinder. I can tell you the price range. This made me hate this app.’

- When both username and phone number are used, visibility of phone number can leak a new phone number to a contact without awareness or consent (e.g. the old and new phone numbers are linked and shared with the same profile which is transferred to the new phone number). This increases the burden for getting a new phone number and could require someone having to choose between phone number privacy and retaining their connections/contacts on a specific app.

- There are also major risks related to account takeover and access to SMS. State actors can block access to services that use SMS for registration by blocking those service’s SMS. While not covered in our research report, these risks are important to consider. In addition, our community consultants and our research showed that phone number-based verification has led to stolen verification codes over SMS as a method to hack or take over an account – this has been the case in Egypt, Iran, Iraq, and Lebanon.

- As mentioned earlier, phone numbers are linked to very personal information and identifiers such as a person’s legal name and home address. In most countries, governments require identifying information to register a SIM card and retain a database of phone numbers of registered persons. If the information gathered for registration is leaked, the safety and identity of highly vulnerable communities are most at risk, including for arrests and other harms. For example, this happened on the forum-based platform Clubhouse where users had to register with a phone number. Anyone that had the number could find a person on Clubhouse, which led to the outing of queer people in queer groups or rooms on the apps.[19]

- Often with phone number registration, only an SMS verification is needed to access a profile and its content on a platform. In these cases, there are risks of telecommunication providers providing access to such information such as the

SMS verification code (as they have power for visibility to SMS exchanges). State actors or individuals are then able to access accounts by gaining the SMS verification code that grants them full access to a profile.

- The lack of utility and reliance on phone numbers as a robust form of verification is highly questionable. For example, in Tunisia, and many other countries, if a person does not use their phone number for more than six months, it is sold to someone else, increasing the risk for lost phones. In another example, detainees without access to their devices for months on end could lose their phones.
- Even when someone makes a choice not to provide a phone number for registration, providing a phone number for SMS verification (e.g. multifactor authentication) can lead to that phone number being connected to an account without the user's awareness or consent.

We understand that many apps and platforms use SMS verification as they view it as one of the stronger verifiers and identifiers; however, there are major questions about their utility in this context, as well as their strength.

Other apps use it as a single point to avoid spam and provide ease through creating social graphs with contact lists. However, this tool is now an increasingly outdated one when considering all angles: safety, privacy, issues of hacking, and spam (as we saw in some of the earlier examples). Its reliability for tackling spam and hacking is now on the same level as methods such as email verification, which is less privacy intrusive (linking to legal information) and more inclusive of those in sanctioned countries.

Further recommendation details

The use of phone numbers is a complicated and long contested issue, and we continue to rely on it due to the lack of alternative options. However, it is now vital to expand our resources and invest in more secure, inclusive, and viable options that are not to the detriment of the most marginalised.

How can it be implemented? (for app developers and security teams)

Our recommendation for discovery

Apps and platforms should **use alternative ways to allow the discovery of other users**, which can be done simply through usernames or other more creative methods. Due to the popularity of this shift, we do not think more needs to be said for its implementation, rather it needs to be a prioritisation issue.

As we have seen, chat-based apps such as Wire, Telegram, and Signal have moved away from showing an individual's phone number in the profile of the user and do not require it for the discovery of an individual in differing ways. They now require usernames. We urgently recommend that other apps such as WhatsApp and iMessage follow suit.

As we wait for an industry change, there should be a **ban on the use of this information** and on direct reliance on processes that negatively nudge users to share phone numbers via dark patterns.

Apps should fully **refrain from using and sharing phone number information** (as seen on Meta platforms) and immediately **halt the suggestions of profiles or 'friends**' based on these collected numbers. (Unfortunately, more work needs to be done to reduce other data points from a user being used in other algorithms to 'suggest friends' without the consent of individuals involved; however, removing phone number-based recommendations is an important first step).

Our recommendation for verification

We need to look for added and **alternative tools for discovery and verification**. The strategy used should not add another layer of marginalisation for those already affected.

In our surveys, one of the top methods of verification for sign-up/login requirement purposes was via email. This method is used by Wire, which was an app mentioned directly by **5 of our interviewees** solely due to this feature.

One implementation method is through using a combination of emails and quizzes to weed out bot registrations. It can be an option to prevent abuse and also provide the privacy and safety needed by vulnerable users.

Due to the features of Wire and this ability to ensure that the conversations are not linked to government-registered phone numbers, many of our partners and activist groups, and members of the community in general, use and trust Wire for their communications. This must happen on other apps to ensure that people in the highest risk context such as Iran and Sudan are not kept out of using these essential tools.

While we wait for this move away from phone number reliance, apps should also **not block the use of VoIP numbers**.[20] Many users who are blocked from using SMS or registering their number are reliant on VoIP numbers, which security standards, such as those set by the USA's National Institute of Standards and Technology, view as an untrustworthy identifier that should not be used.[21] However, the bans often mean the verification SMS is not received.

In the case that phone numbers are needed for verification, they should **remain separate and disconnected from all other account information**, including account details and advertisement profiles. How the data is used and the methods for preventing connection, when they are present, should be transparently communicated to users.

Finally, since this issue is also currently a design challenge with no agreed upon core solution, we also recommend partnering together with our research teams to **implement a prototype** with funding.[22] We can use the needs and expertise of these users and our teams to create something industry-led and based on the wants and security needs of the most marginalised and targeted queer communities on these apps. Further advice for this can be obtained from our research and technical team. We can be contacted at afsaneh@de-center.net and MENA@article19.org.

After we block someone, we need all the content deleted

Recommendation 4: Delete all content from both devices after a user is blocked.

Many of our respondents stated the desire to have all of their chat history deleted when they blocked someone. It was a common need they expressed from platforms such as the chat-based platforms, and social media and dating apps. In many cases, individuals managed to block a police or state security honey trapping account, or an abusive account threatening to use the conversation to out, prosecute, or create a hate campaign against them. But they need to have some control over the content of their previous message, image, or video exchanges. This is vitally important.

Context and research behind the recommendation

In our investigation in Part II, we saw how much police/state actors rely on the contents of these conversations. This issue often came up in our focus groups in Algeria Egypt, and Morocco. As our interviewees said:

> *'Once the person is blocked, it will be more reassuring if the photos and videos exchanged with this person are deleted.'*
>
> – Interviewee in Algeria

> *'On Facebook, I would like that by blocking a person, the history of my discussions with them disappears too so that they can't use them against me anymore.'*
>
> – Interviewee in Morocco

> *'The system used by Grindr is effective especially when you block a person, the discussion disappears.'*
>
> – Interviewee in Algeria

How can it be implemented? (for app developers and security teams)

This recommendation is less complicated with an easily achieved outcome that can create harm reduction safety for our users. One participant in Algeria very tactfully outlined a methodology that provides a timeline for deletion and adds in a block on screenshotting in the allotted time (see Recommendation 22):

> *'Once the contact is blocked, all the discussion must disappear within 72 hours, for example, with a ban on capturing and saving.'*

Ideally, an **immediate deletion of the chat combined with adequate reporting mechanisms** (see below) can help combat abuse of the chat content. Further, the **deleted content should remain inaccessible in situations of forensic analysis**, for example. This again could be implemented as an option for the blocking party to decide if they want all the content to be deleted after blocking an individual. Reach out to our team for further information. We can be contacted at afsaneh@de-center.net and MENA@article19.org.

There are, of course, cases where users may want to retain the conversation. However, especially in a high-risk context, there is a need for victims to have the option to have control over their content and conversations after identifying and blocking a harmful user. This is a norm on **Instagram** and **Grindr**, for example. Our interviewees and participants pointed to Instagram as a good example of the implementation of this feature. With this outline, apps themselves should provide adequate methods for people to report an abuser who blocked someone and looked to have the chat deleted.

We need you to respect our right to anonymity

Recommendation 5: Remove real-name requirements and ensure rights to anonymity and pseudonymity.

Platforms should fully **move away from real-name policies and requirements**, including using names from linked platforms that require real names. Anti-fraud algorithms or account data validation at registration should **not incorporate or check name data**.

Context and research behind the recommendation

Anonymity and, more broadly, the right to use a pseudonym not tied to one's legal name have been contested through the years. They have become a target for states looking to challenge the right to anonymity online in a misled effort to challenge online abuse.[23] However, **anonymity and the ability to use pseudonyms online are often a life-saving tactic for many marginalised communities**.[24] Providing an option for LGBTQI+ communities, as well as other at-risk communities, to remain anonymous online is best privacy and security practice – and its absence can contribute to the silencing of oppressed groups.

Guaranteeing the right to privacy in online communications is essential for ensuring that individuals have the confidence to freely exercise their right to freedom of expression. As one of our interviewees in Iran put it: 'anonymity created a space for truer self-expression'. This is especially the case for communities such as the criminalised LGBTQI+ community whose physical safety and liberty depend on their right to remain safely anonymous.

The inability to communicate privately substantially affects individuals' freedom of expression rights. This was recognised in several reports of David Kaye, the Special Rapporteur on Freedom of Expression, in which he recommended that provisions must be in place to allow individuals to express themselves anonymously online and tech companies should refrain from using real-name registration systems. He also recommended that corporate actors reconsider their own policies that restrict encryption and anonymity (including through the use of pseudonyms). The real-world impact of the gap in such protections is always felt the most severely by highly criminalised and marginalised communities.

In our years of investigations, the LGBTQI+ community in MENA has overwhelmingly shown the need for the **preservation of anonymous communication without a real-name and age verification system**. Our respondents and interviewees pointed to this need from companies to respect their privacy and need for anonymity.

Surveys:	**349 out of 5,018** (7%) respondents answered the 'real-name' requirement as one of the main issues that had caused them the most risk.

Additionally, we also saw the **importance of data security and privacy for the community** when asked about what they wanted from platforms (see Recommendation 1).

We have had numerous accounts where our interviewees or participants highlighted that, when their real names did not match their social media account, the police and security forces were unable to prove a link, and in most cases they were released. They were often arrested for the activity on these accounts for crimes of queerness, or they were monitored for political activities or activities linked to queerness. This separation from their real names saved them.

For example, a Tunisian interviewee reported that after being monitored for being part of protests, and linked to a high-profile queer activist, the police monitored and searched for further information on them on their Instagram.

> *'My Instagram account was not with the name on my ID, and they couldn't find me.'*

Others have been able to deny connections to 'incriminating' accounts where they were linked to LGBTQI+ activity while in custody due to the inability to link their real names to their accounts (see Recommendation 3).

Although we have documented the numerous ways police and police 'consultants' use fake profiles to catfish and entrap queer people, we do not push for more uses of real names or ID verification. This is because we know that **verification does not keep people safe**.

Our interviewees pointed to Facebook and Tinder, which have features or policies that expose real names or enforce variations of their real-name policy. This is despite the harm these policies have caused and the fact that we have yet to see any research that shows these measures in curtailing anonymity, and the right to pseudonymity, have in any way created a safer online space. In fact, this report shows that fake accounts, entrapments, abuse, hate speech, and harms continue regardless of any of these policies.

In our research, our interviewees instead pointed to the risks these features caused, for example, on dating apps such as Tinder which requires your 'real' first name and only allows name changes through various complex methods:

> *'I logged in for Tinder with Facebook and I don't … Tinder does not allow you to change some personal information. For example, maybe you don't want your real name on Facebook to be shown there. It should let you change your name. When it was linked to Facebook, it copied my name from Facebook.'*

The interviewee continued to outline the risks involved in having the real name on something like Facebook. A person may have family and others who do not know the person is out on their Facebook account; they will then be linked to the dating app where the person is identifying as queer.

Others point to a deeper need for options around anonymity for remaining online and using the engagement features freely:

> *'You should have the possibility to create your own pseudonym when you want to keep anonymity especially in a group or a page of the community, or you want to leave comments and likes. This is for safety. Also for conversations with strangers or you just prefer to keep anonymity until you get to know someone.'*
>
> – Interviewee in Algeria

This is echoed in other higher-risk and conflict contexts such as Sudan:

> *'As a queer individual, I remain safe [when] my content does not contain any of my personal information such as photos and identifying information.'*
>
> – Interviewee in Sudan

How can it be implemented? (for app developers and security teams)

More creative methods for verification without increasing risk to those already at risk can exist. Prioritisation of finding better and more creative methods is key. See 'Our recommendation for verification'.

We need our access to the internet supported

Recommendation 6: Do not add more access barriers. Support communities to access apps and platforms.

Companies must **support communities to access their apps and digital platforms** and **remove added barriers** that block whole communities and contexts from accessing and using apps and platforms, especially for queer and other highly marginalised communities.

Companies must **support basic anti-censorship features** such as encrypted domain name systems (DNS) and configurable proxy support. More advanced features such as integration of Tor and pluggable transport technology, like Snowflake, should also be implemented. This can help keep users connected even in the most extreme network censorship situations.

Context and research behind the recommendation

Interviewees in Sudan who we talked to during the war asked for ways platforms can help during shutdowns as emergency tools for connection and information. They relied so heavily on platforms such as **WhatsApp**, **Facebook**, and **Signal** that access and connection were a matter of life and death. In places such as Iran, internet shutdowns are frequent as a method to thwart political speech and actions. This enables human rights abuses to occur during times of internet blackouts. Connection is a vital tool that allows people to gain support and show and document abuses to the world. This is especially important for highly marginalised communities.

Many of our participants and interviewees also pointed to experiencing internet access issues and censorship of LGBTQI+-related content.

Surveys:	**2,241 out of 4,275** (52%) respondents to the question about whether the apps or websites they used to connect to the LGBTQI+ community were censored answered ‘yes’.

There are several barriers that explain this response: state-level censorship, company compliance with queer content censorship requests, the cutting of company functionality in countries such as Sudan and Iran due to sanctions regimes (see ‘Impact of sanctions on the LGBTQI+ community: Iran and Sudan’ in Part II), and state-level internet shutdowns. With these barriers, the routes for connection, safety, information, community, and general ways to live and thrive are thwarted.

***7 out of 14** (50%) interviewees in Iran and **8 out of 16** (50%) interviewees in Sudan asked for companies to support them with access and to remove barriers placed on them accessing and using apps and platforms that link to the queer community and broader community.*

> *‘The most important thing that these apps can do is to find a way to bypass the censorship barriers.’*
>
> – Interviewee in Iran

Some asked for more direct support from platforms such as proxies and better VPNs:

> *‘I wish there was an ingrained VPN, not just a privacy, private browsing option, where it’s free … you know, like, open-source VPN, available for all human beings,*

> *that's part of the browser. You know, and allows you to automatically plug in. … And not everyone can afford to, or even know how to, you know, create [them].'*
>
> – Interviewee in Iran

How can it be implemented? (for app developers and security teams)

Companies should **support basic anti-censorship features** such as encrypted DNS and configurable proxy support. More advanced features such as integration of Tor and pluggable transport technology, like Snowflake, which can help keep users connected even in the most extreme network censorship situations, is also possible. When incorporating these technologies, apps should allow the user to add their own service or configuration instead of limiting the options for services. This will also require companies to have the expectation and contextual analysis that they will experience some level of blocking in some countries. This blocking may happen in a legal or ad-hoc manner, and may be temporary or more permanent. Proactively implementing resiliency in the network layer of the app is a better approach than reactively doing it after a block or shutdown. They should thus have preparation plans for these scenarios and how they will support user access.

Working with expert communities on building connectivity during shutdowns can be instrumental in supporting against such abuses.

We do not need our photos saved to the gallery

Recommendation 7: Do not allow apps to save photos taken or received in the main device photo gallery by default.

Simply, apps and platforms that allow for the sharing of photos (and equally other media) **should not have their default settings set to allow media to be directly saved on the main device photo gallery**. Having the photos in one place should be the default unless otherwise changed by the individuals.

At a minimum, photos and other media shared in timed or one-time messages should not be stored in the phone's gallery.

Context and research behind the recommendation

Findings in this research, as well as past research, point out that saving photos directly into the phone's main gallery, especially from explicit chats, has caused issues for many individuals who have been stopped and searched or interrogated. It adds risks and even further charges in their device searches and/or interrogation.

Further recommendation details

Having the photos in one place – and only on the app – helps limit access to them and reduces app logic for sharing one-time or disappearing messages with media. Photos and other media shared in timed or one-time messages should never be stored in the phone's gallery by default.

Maintaining this media on the main app/platform should be the default setting unless otherwise changed by the individuals in their settings. This also means that when an application allows a user to take a photo from within the application, the app should not save the resulting media file to the public phone gallery. When they are saved to the phone gallery as a default, the timed messages and other security features can become ineffective. Many people may not know that they can disable this feature and more should be done to either disable the feature or at least not have photos saved directly to the phone on the camera roll as the default.

This is implemented on a number of existing secure messaging applications. While storing photos in an encrypted way is ideal, even just storing them in internal private application storage would be sufficient.

Dating app specific recommendations

In addition to the previous points covered, there are specific issues relating to the unique uses of dating apps and needed social media platform changes. Dating apps can be very important tools for connection and community for LGBTQI+ people in MENA, thus the safety changes on them are vital. They are also aspirational for the future we want to see for our communities.

> *'Queer people have a community but this community is not very safe. I feel the need to be a part of a community. These apps can serve this purpose. Finding a safe space to socialise with my community is a need for me. The feeling of belonging is very important. Such an atmosphere can be created by these apps.'*
>
> – Interviewee in Iran

Our safety features must be free

Recommendation 8: Dating apps should make vital safety features free, especially in high-risk contexts.

All safety features on dating apps must be free, especially for the most high-risk users and contexts. Many dating apps have created safety features or features people use for safety – these features must all be free. **Access to safety should never be dependent on access to capital**. This is especially the case for dating apps that have immense privacy, safety, and security issues for LGBTQI+ communities and other marginalised communities.

Context and research behind the recommendation

A concerning effect of monetisation has been the increasing use of premiums paid for features on dating apps. We understand that dating apps, like most other corporate platforms, are for-profit ventures. However, with the immense privacy, safety, and security issues befalling LGBTQI+ communities and other marginalised communities through dating apps (as painstakingly outlined in this report series), it is imperative that vital safety features are made free – especially for the most at-risk contexts and users.

In our interviews and focus groups in Iran, Morocco, and Sudan, the issue of **paid safety features as a form of discrimination** was raised numerous times. This is especially a concern due to the socio-economic inequality faced by these communities, not to mention that, in some contexts, users are not able to even use app-based banking and payment systems.

> *'Premiums should disappear in the regions where LGBTQI+ are threatened. All the premium options should become free until the situation becomes better.'*
>
> – Interviewee in Morocco

Some pointed out specific vital features, such as disappearing messages being limited unless purchased with premium memberships:

> *'Grindr has set a limit for the number of disappearing messages that you can send, for example, you can send only three such messages in 24 hours. If you want to send more, you have to pay.'*
>
> – Interviewee in Iran

Based on our work with Grindr we know many of their safety features are open and free in context where LGBTQI+ users are at risk. However, not all relevant contexts are provided for this. We ask that Grindr, and all dating apps, take measures to ensure these vital safety features are free for most (as risks for LGBTQI+ people are increasing globally), but especially where LGBTQI+ people live with the highest legal and social risks.

Further recommendation details

Dating apps such as Grindr, Tinder, HER, OkCupid, Hornet, and others should make features (such as those in this report) created for safety free, especially for high-risk users. Access to safety should not be based on a person having financial means. This monetary barrier risks lives of some of the most vulnerable users who are often otherwise banned or unable to even purchase these features. To stay true to their missions of providing connection, love, and social good, this is **one of the more important and immediate**

actions apps should take, especially regarding features and changes introduced as part of engagements with at-risk communities and organisations such as ARTICLE 19.

Don’t allow shareable links for our dating profiles

Recommendation 9: Dating apps should immediately disable options that allow for the non-consensual sharing of dating profiles, especially for high-risk context.

As a general concept, the **‘share profile’ feature on dating apps should be disabled** with the exception of possible retention for profiles that have opted in to this feature. This feature has been used in mass outings as it allows for people outside of an app to see particular profiles.

Context and research behind the recommendation

Tinder and Bumble allow users to share profiles with people who may not have seen the profile otherwise, by allowing a user – even before matching – to generate a shareable link to said profile. This social feature’s original intention was to make it easier for people to get their friends’ and families’ opinions on their dating options. But in high-risk contexts of marginalised communities such as the LGBTQI+ community in MENA, it has added to bullying, outing campaigns, and increased risks, including access even after blocking someone. As one focus group participant in Egypt remarked:

> *‘Tinder shouldn’t let you share a link of other people’s profiles which facilitates smear campaigns like the ones that happened in Maghreb and Egypt. There are even accounts dedicated to making fun of “weird Tinder profiles” on other social media apps.’*

There is a further evolution of these features under Tinder’s newest Matchmaking feature, which lets people who are not even on the app look at other users’ profiles. These features can be fun and social in certain contexts, but they are clearly not thoughtfully designed for they lead to immense harm, outing, violence, arrests, and other types of abuse. This feature can also be used by extortion gangs and policing actors to use profiles to harm users, with even less control and review of how the app is being used and by whom.

How can it be implemented? (for app developers and security teams)

As a general concept, the **'share profile' feature should be disabled** with the exception of possible retention for profiles that have opted in to this feature. This is a dangerous feature that has already been used and documented as harmful without adding a significant amount to the dating experience of users in these regions.

We also recommend that **Tinder hold more consultations with at-risk and highly marginalised communities**, especially outside of US and EU contexts, to see how to roll out features like the Matchmaker feature to ensure a harm reduction approach. Reducing harm here does not mean fun and interactive features should not exist, it means they should be created intentionally with an understanding of how they can cause severe harm in broader contexts.

We need safer geolocation options

Recommendation 10: Safer and more private geolocation practices.

Dating apps **should follow our general guidelines** (laid out in the following recommendation implementation details) **in any implementation of geolocation matching** for dating apps. This includes a **reduction in the gathering and use of location data** in app functionality that does not require location data. Additionally, **requests for location data should come once**, or for a defined amount of time, to prevent the unintentional overcollection of location data. And finally, randomisation or otherwise **obfuscation of user location on the server prior to any matching** should be implemented.

As this is a complex topic, there must also be further user research to consider alternatives to direct point-to-point matching.

Context and research behind the recommendation

Geolocation is a desired feature of mobile dating apps, but inherently introduces risk. It is required in order to allow users to enjoy a sense of familiarity and proximity when using the app without immediately putting safety at risk. In our research, users suggested the use of geolocation as a method increasingly leading to risks, and potentially, to arrests.

> *'Sharing location was not an issue for me in the past, but recently it has become a concern. I used to be like, "It doesn't matter". Now I prefer not to share how far other users are from me or where I am. Your location is something that can get you in huge trouble if something goes wrong.'*
>
> – Interviewee in Iran

Most users rely on the use of geolocation and like it, but they do so with the knowledge of the risk it carries. Our participants and interviewees also pointed to increasing risk with this feature.

Interviews:	**10 out of 93** (11%) interviewees raised these issues as a direct ask from companies.
Surveys:	**1,283 out of 5,018** (26%) respondents answered 'location' in a question asking 'What information are you NOT WILLING to provide about yourself and why?'
Focus groups:	In all **6 countries** where focus groups were held, these issues were raised.

Our research shows that those who want location sharing want a **ballpark location that is not easy to triangulate** to their exact locations.

Further recommendation details

From our technical team and through the research, we know that there are a variety of risks here through the app infrastructure:

- **Triangulation**, where an attacker can simply play a game of hot-or-cold in order to see if they are moving further away from, or closer to, the victim.
- **Direct API calls**, which inadvertently reveal the exact longitude and latitude of users.
- **Sharing location data with third parties**, such as analytics services or ad libraries, which increases risks for users sharing location data.

While it is unlikely any implementation of geolocation within dating apps will remove all risk, the goal is to make targeting users through it as difficult as possible. Key points for tech company teams to consider are:

- Precision of GPS data collected and used.
- Granularity of match distance that is shown to users and available via the API (e.g. only respond to API calls and show whole miles/kilometres without fractions).

- Whether additional location information (e.g. city and state) is presented with location data.[25]

- How population density affects distance granularity presented.

- Whether distance can be presented in a generalised category instead of an individual measurement (e.g. 0–3 miles, 3–6 miles, 6–9 miles, 9–12 miles, 12–15 miles, 15+ miles).

How can it be implemented? (for app developers and security teams)

Dealing with this recommendation is complex and there is the need for further user research to consider alternatives to direct point-to-point matching. Current methods (of which we are aware) focus on **decreasing accuracy while introducing randomisation to a user's location**. For example, Tinder randomises the user's location to a set geographic polygon instead of basing distance on the exact location – and only returns integers for distance.

We recommend that any implementation of geolocation matching for dating apps ensures the following:

- Allow access to app functionality that does not require location data (e.g. messages) **without providing access to location data/service**.

- **Request location data only once or for a defined amount of time** while the app is in the foreground to prevent unintentional overcollection of location data.
  - For Android: Declare your location service as a foreground service and use the 'Access Coarse Location' permission unless it is absolutely necessary to use another method.
  - For iPhone Operating System (iOS): Declare the NSLocationWhenInUseUsageDescription key for Core Location. Consider using the CLAccuracyAuthorization Core Location property, which is currently in beta,

to limit the precision of the geolocation data collected by the app when it is implemented.

- Randomise or otherwise **obfuscate the user's location** on the server prior to any matching.
- **Round the distance matching to the nearest integer** on the server to ensure only integer distances are sent via the API.
- **Do not share city or other location data with the distance**, which could make triangulation attacks or other attempts to identify a user's location easier.

If these recommendations are implemented well, it should help mitigate many risks involved with location-based functionality. However, techniques such as triangulation can still be used to determine the general area of a user and put users at risk. In particular, dating app users in less dense environments (e.g. rural villages) are likely to remain at risk. The following are two options for alternatives that would require further user research prior to implementation:

- Instead of giving distance as a measurement, the server would provide a **category indicator for distance** (e.g. 0–5 km, 6–10 km, etc.).
- A distance indicator is not provided, but instead the **neighbourhood, city, or region is provided**, depending on the density of users, similar to how census tracts work. If a neighbourhood, city, or region is provided, **do not include a distance in the API response**.[26]

Further advice for this can be obtained from our research and technical team. We can be contacted at afsaneh@de-center.net and MENA@article19.org.

Chat-based app specific recommendations

In addition to the general points earlier (especially Recommendations 1, 2, 3, and 4), there are specific issues relating to the unique uses of chat-based apps and wanted social media platform changes. One specific example is added below.

Let us detangle Facebook Messenger from Facebook

Recommendation 11: Provide the option to have a proper separation between Facebook Messenger and Facebook profile.

Facebook Messenger **should have the optionality not to be immediately linked to an individual's Facebook profile**. It should have **optionality to have privacy controls over adding contacts** for communication on Facebook Messenger only, like other chat-based apps.

Context and research behind the recommendation

Our participants in Algeria specifically asked for a very practical safety change from Facebook Messenger:

> *'We would like the possibility to add contacts as friends on Messenger and not have them on Facebook.'*

This request is especially apt considering the popularity of Facebook Messenger as a chat-based messenger app in many of the focus counties. In the **8 countries** we conducted research, Facebook Messenger was the second most used chat-based messenger app**.**

1,165 out of 5,018 *(23%) respondents picked Facebook Messenger as their most used messenger app.*

However, it has created insecurity for individuals because it links to their Facebook profiles, which has led to risks, targeting by police, and outing campaigns (see Part II). For example, in entrapment cases where Facebook Messenger is often used, LGBTQI+ individuals can be identified on Facebook and continue the conversation on the chat-based platform, Messenger. Communications are in one place, later setting up the entrapment trap and using the information from the conversation, profile, and the connected Facebook networks and friends against them.

How can it be implemented? (for app developers and security teams)

Options to create and have separate Messenger accounts that do not immediately link to a Facebook profile can massively impact safety for those relying on the platform as a messaging platform. Alternatively, have the **optionality to have privacy control over adding contacts** for communication on Facebook Messenger separate to adding them to a Facebook account. **Controlling whether or not Messenger contacts can see or access the Facebook profile** could also impact safety. If Facebook Messenger moves to functioning similarly to other chat-based apps, it would exponentially decrease the risks and privacy gaps linked to it as seen in this report. Of course, it would also address overall privacy issues in the community.[27] Again, this can be an option, allowing those who prefer their Messenger and Facebook profile to stay linked to have that option.

We need safety guardrails on 'Stories'

Recommendation 12: Add safety measures for 'Stories' and conduct further research in emerging safety issues.

We recommend that apps look into **safety issues emerging from outing and hate campaigns conducted on 'Stories'**, and work with impacted communities in bringing harm reduction and safety guardrails to the feature.

Context and research behind the recommendation

Many chat-based apps such as WhatsApp, Signal, and Telegram have the 'Stories' feature we are often used to seeing on social media platforms. This investigation has shown that the ephemeral but also **public/mass announcement nature of 'Stories' has been used to out LGBTQI+ people** where their names, photos, and phone numbers have been blasted to an unknown number of people by state and non-state actors. This feature has been used to reach larger numbers of people outside the social media platforms. According to our participants, perpetrators often know it is less likely that their accounts will be reported or reprimanded on chat-based apps than on social media apps.

How can it be implemented? (for app developers and security teams)

As this is a newer, evolving issue and tactic, one of the first recommendations is for chat-based apps such as **WhatsApp, Signal, and Telegram to hold consultations with experts and impacted communities** about how newer functionalities such as 'Stories' can be made safer. Further advice for this can be obtained from our research and technical team. We can be contacted at afsaneh@de-center.net and MENA@article19.org.

In addition, **adding screenshot and capture blocking for 'Stories'** can add a layer of privacy and ephemerality for stories and can be an added barrier for these doxxing and outing campaigns, which happen through the screenshotting and mass sharing of stories.

Finally, it is vital to have robust reporting systems and direct lines of communication with impacted communities for rapid response efforts (see below).

Social media app specific recommendations

There are specific issues relating to the unique uses of social media apps and needed social media platform changes.

We need safety and privacy for photos we are tagged in

Recommendation 13: More control and privacy with tagged photos.

Platforms and apps must have **higher user controls for post and photo tagging**. Users can be tagged in content or media that can incriminate them unknowingly, and more controls would decrease the implication and potential risks. The names and **profiles of those tagged should remain private unless consented to** directly by the tagged individual.

Context and research behind the recommendation

Tagged photos have been a growing issue.

***5 out of 54** (9%) interviewees who had experienced arrest reported tagged photos had led to their arrests in Egypt, Iran, and Tunisia.*

In Tunisia, people associated with a high-profile political activist, whose LGBTQI+ identity was used to target her, were arrested based only on vague tagged photos. One of our interviewees said that they had been accused of being part of the protests and linked to the queer activist Rania Amdouni due to a tagged photo:

> *'Turned out they checked Rania Amdouni's Facebook profile, they found a post she posted and tagged our names in it; every single person that was tagged in that post was part of this case.'*

This included a queer friend who was summoned to court even though they were not in Tunisia at the time of the protest.

As outlined in Part II, outing campaigns from state and non-state actors with the use of data from social media profiles are highly prevalent, as seen in multiple sections of this report. In the case of a party in Iran, an individual was arrested due to tagged photos on Facebook and Instagram, leading to arrest.

Further recommendation details

We include this recommendation – connected to Recommendation 14 – to highlight the increasing risk of photo tagging on queer people in the region. For tagging, the **names and profiles of those tagged should remain private unless consented to directly by the tagged individual**. The tag should remain pending and without details of the tagged individual until the individual has accepted. The levels of privacy and visibility of the tagged person's profile information should also be in the control of users, especially those tagged. In the Amdouni case, those tagged were not physically in the photo or had not directly consented to being tagged; however, they were all arrested and detained, receiving varying charges.

Stop showing us and our activities to people who endanger us

Recommendation 14: Do not expose and out users through 'friend recommendations', and on app activities do not expose unintended information.

Apps must **enable privacy around 'friend recommendations' or 'people you might know' recommendations**, which can be done with easy and transparent user controls. Further, apps must **provide clear and easy privacy options** that allow users to make their activities and friend lists private, with granularity in user controls.

Context and research behind the recommendation

One important, practical, and simple change wanted from our interviewees and participants is **further protection and privacy around recommendations** and on app activities that can expose them, their networks, and their identity.

The request to remove recommendations, friend suggestions, and the showing of activities such as likes, follows, or comments was raised many times, especially in relation to TikTok, Instagram, Facebook, and Twitter.

Interviews:	**10 out of 93** (11%) interviewees directly raised these issues as a direct ask from companies.
Focus groups:	In all **6 countries** where focus groups were held these issues were raised.

These asks were especially directed towards Facebook and Instagram but apply to all platforms with similar functionality.

Further recommendation details

Do not recommend friends

We have already discussed the risks that come from the use of phone numbers from contact lists to suggest friends. It has led to safety risks and outing people's online profiles and their identity (see Recommendation 3) as seen on Meta platforms, TikTok, and some chat-based platforms. This is also the ask for location options that further the risk of physical danger. As one interviewee in Morocco mentioned:

> *'Facebook: I do not want Facebook to use my localisation to suggest friends. It puts me in danger. … Also, about Facebook and Instagram, I do not use numbers to suggest friends on Facebook and Instagram.'*

These options are implemented without the consent of users and out of their personal control. As a result, many at-risk LGBTQI+ people will further censor themselves online and not use these platforms for self-expression – even in contexts where they have selectively created safe friends lists – because the risk of their profiles being suggested to harmful accounts is too high. As one interviewee in Sudan said:

> *'The "Friends You May Know" feature I hope will be cancelled. It's very risky.'*

Many users might want this optionality. That said, the default for creating these recommendation frameworks and the gathering or linking of data to identify people in each other's networks should not be the default. Instead, it should be set to privacy, and only those who wish to be recommended to friends or want such recommendations should receive them. Privacy should be the default.

Keep friends lists private

This issue extends to presenting common or mutual friends. One interviewee outlines how easily this feature can expose the network and identity of an individual:

> *'I suggest not showing any of the common friends in the platforms which do it. They can only share the number of people in common to ensure the [safety of] users. Because when people see the names, they know your circle so they can track you easily. They can also know if you are from the LGBTQI+ community through tracking your friends.'*

Our interviewee above rightly argued that simply **showing the number of common friends, rather than exposing names and profiles**, would provide the security people need to feel trust in a shared community. In our research, we have seen how superficial links to prominent queer activists or groups have not only outed people but have led to their arrests. This **recommendation is for a fast and immediate change** that can lower risk.

Again, this can be an **opt-in option for those who do want to show their friends** (which should be limited to the friends that have opted in to be seen). Having the number of friends in common and not their exact profiles and usernames is an important middle ground that can help maintain trust and connection without exposure.

Keep activities private

Another high-risk issue that limits the activity and self-expression of the community is the visibility of activities, such as who the user follows or is followed by, what pages or posts they have liked or are a part of, and what a user has commented on. One of our interviewees described this issue and why they were rarely active, especially regarding their identity or content linked to their identity online:

> *'People can see what accounts you follow on Instagram. It was worse in the past and they could even see what you had liked. People can still see your name liked at the bottom of the post you have liked, which is painful.'*

They highlighted Instagram here, but it is the same on many of the social media platforms.

There need to be **clear and easy privacy options that allow users to make their activities private**. Likes and follows are self-explanatory, they can easily be made private by default, especially in high-risk contexts – unless a user wants to opt in to show them. Again, **privacy should be the default**. For comments, we suggest such users have their profiles or names obfuscated if the platform wants to still show engagement or comments, but it should not show their profile and username. In this way, at-risk users can still be part of communities and discourse without being put in harm's way.

These options can further be elaborated on and worked through based on the experiences and wants of the community. What is clear is the over-exposure of activities in what is seen as forms of ‘surveillance capitalism’ with techniques that seek to keep users engaged over keeping them safe in a healthy online environment.

We need safety guardrails on ‘Stories’ and ‘Lives’

Recommendation 15: Add safety measures for ‘Stories’ and ‘Lives’ and conduct further research on the emerging safety issues.

We recommend that apps look into safety issues that arise out of outing and hate campaigns conducted on ‘Stories’, and work with impacted communities to bring harm reduction and safety guardrails to the feature. See further recommendation details for more information.

Context and research behind the recommendation

As mentioned with issues of stories with chat-based apps, we see the ‘Stories’ features – as well as ‘Lives’ on Facebook, Instagram, and TikTok (or ‘Spaces’ on Twitter/X) – being used as part of mass outing and violent homophobia campaigns. There has been a rise in outing campaigns, and campaigns pushing for violence against the community, in places such as Egypt and Morocco. Those partaking in ‘Lives’ have also been targeted when these pages are being monitored – unbeknown to the owners. In June 2020, after several members of the Iranian LGBTQI+ community participated in Instagram Live videos of famous Iranian influencers, they were summonsed by security agencies. In 2023, our team witnessed calls for violence and outings in Egypt through ‘Lives’.

In rapid response work that ARTICLE 19 has conducted, we have documented and reported such cases (these are not public), including in Egypt where official police accounts were actively watching ‘Lives’ of outings and other outing campaigns on TikTok. As mentioned, this investigation has shown the ephemeral but also public/mass announcement nature of ‘Stories’ and ‘Lives’ that have been used to out LGBTQI+ people with their names, photos, and phone numbers. Even on social media platforms, due to the current nature and manner of reporting systems, reporting ‘Lives’ and ‘Stories’ can be difficult and not enacted in a timely manner. This gap is weaponised by **anti-LGBTQI+ groups that have learned to use these features to bypass moderation** of their posts.

How can it be implemented? (for app developers and security teams)

As this is a newer and evolving issue and tactic used, one of the first recommendations is for social media platforms – especially Meta, TikTok, and Twitter – to **hold consultations with experts and impacted communities** about how newer functionalities such as 'Stories' can be made safer. Further advice for this can be obtained from our research and technical team. We can be contacted at afsaneh@de-center.net and MENA@article19.org.

It is vital to have **robust reporting systems and direct lines of communication with impacted communities** for rapid response efforts so that violations on 'Stories' and 'Lives' are better and more methodically reported, especially as their use as a weapon against marginalised communities expands (see below).

In addition, adding **screenshot and capture blocking for 'Stories' and 'Lives'** can potentially provide support in these cases and the mass outing and doxxing content. However, more research needs to be done. This is due to the fact that without robust and rapid reporting mechanisms that can deal with the ephemeral nature of these tools, most victims can only use screen captures to prove that violence and outing occurred.

Operating system specific recommendations

We need our devices protected from forced access

Recommendation 16: Conduct further research and work to challenge the use of jailbreaking or rooting and allow apps to opt out of operating system features, including new artificial intelligence features.

Operating system providers and other entities should provide more support to allow for resources and research in **understanding how jailbreaking and rooting are being used against highly marginalised communities**. They should also create robust methods to support communities against non-consensual and enforced rooting/jailbreaking, especially by law enforcement.

Furthermore, operating system providers should also **allow apps to opt out of operating system features**, including new artificial intelligence (AI) features, in order to reduce harm from the risks of unauthorised access.

Context and research behind the recommendation

In Part II, especially the section 'Jailbreaking/rooting or hacking phones', we documented that police forces are physically accessing devices without a need for passcodes or permissions. It is currently unclear how this is happening; however, we are aware that it requires further research and investigation from the operating system side. A large proportion of the communities working with us are Android users (over 80%); however, there is still a significant number of iOS users in the community, which are increasing.

This is reflective of our previous research where individuals reported cases where, regardless of providing access to their devices or not, the police jailbreak or use rooting to access the contents of the device. Although this does not seem to be systemic, it is a concerning trend that can be used further against the community and their methods of self-protection. This was seen in **4 cases out of 54 (7%) interviewees who had experienced arrests**.

Depending on their devices, this is likely to be jailbreaking into phones with physical access through forensic data extraction technologies such as Cellebrite. It is likely that authorities specifically used Cellebrite technologies or similar technologies, seeing that both Morocco and Tunisia have been documented as having large training programmes in the use of Cellebrite.

As pointed out in Part II, these extraction methods were also previously documented in Egypt, Lebanon, and Tunisia. One Lebanese interviewee was threatened in 2021 with this method while detained when the investigator said:

> *'If you don't open the phone, we know how to open it. We will put you in a room on your own where you will rot.'*

One Lebanese trans interviewee saw her phone accessed by having it connected to a laptop. It is one of the few cases we have where the individual witnessed this forced access. Disturbingly, not only did police force access, but they distributed the data they uncovered to other officers, resulting in further abuse of the detainee.

How can it be implemented? (for app developers and security teams)

Further research and work to challenge the use of jailbreaking or rooting

Operating system providers should **work with communities and civil rights and privacy groups to investigate how these extractions are happening** on the ground (this can be done through further consultations with legal teams and prosecution team whistleblowers). They should determine how this access to the contents of confiscated devices is taking place. Further work also needs to be done to **block and counter the extensive use of technologies such as Cellebrite for data extraction**, as they are being used widely for human rights abuses and breaches.

Cellebrite and other tools that exploit vulnerabilities in the operating system will unfortunately continue to exist and increase in sophistication. **We require more resources and research to create robust methods to support communities against this non-consensual and enforced rooting/jailbreaking, especially by law enforcement**.

With this in mind, this recommendation is written with the aim to reduce unauthorised access and harm. **We also firmly recommend that operating system providers allow apps to opt out of operating system features, including new AI features.** Operating systems have features and designs that affect the apps on top of them. In an effort to reduce the amount of personal and sensitive data that is collected, new types of processing, including AI and machine learning functionality, are happening on the device. Solely processing data on a device does not reduce all privacy risks.[28] As operating system providers implement new features that affect all apps, which are not a requirement for basic operation or security, they should ensure apps and users have proper notice and choice about whether to use these features.

In particular, operating system features could collect and process data from apps, even if the data remains on the device or if there is some other method that otherwise should protect the data (e.g. end-to-end encrypted messages). Depending on the specific operating system feature, sensitive information across device users could be exposed, a copy of sensitive data could be stored in a less secure way, or other methods could weaken the security context and directly access or infer sensitive information about device users when there is direct access to the device.

HATE SPEECH AND RAPID RESPONSE REPORTING SYSTEMS

Extensive work and documentation have been done to show that the prevalence of hate speech on social media, dating apps, and chat-based apps can have very real-world impacts on people. Human Rights Watch recently launched a campaign with a focus on the failure of content moderation, especially on Meta platforms. In June 2020, 22 LGBTQI+ organisations, mostly from MENA, also urged platforms, especially Meta, to **address the rising levels of hate speech** on their platforms and **address issues concerning the lack of contextual, cultural, and linguistic understanding from their content moderators**. Though our work does not focus on this issue, Recommendation 17 (which looks at robust and contextualised reporting systems and direct lines of communication) is based on our research and consultation and includes working to improve rapid response challenges in moments of high crisis and need.

We need reporting systems that understand us with direct lines of communication

Recommendation 17: Combat hate speech and implement robust and contextualised reporting systems and direct lines of communication.

Companies should implement our general guidelines for the two types of reporting outlined:

1. Reporting **specific hateful content and comments**.
2. Reporting for the **suspension of accounts for safety after arrests** (see above).

Companies must also **implement changes and limitations on mass forwarding** (see below).

Context and research behind the recommendation

The issue of **ineffective reporting systems and a lack of support from platforms** is a concerning trend that has emerged from our work and this research. There are numerous reasons why maintaining reporting and rapid response systems with direct lines of communication with impacted communities is crucial. First, such systems are vital to address the vast and prolific amount of hate speech on the platforms and increasing doxxing and outing campaigns. The automated reporting systems have been inadequate not only in responding to these reports but also in understanding their content. Second, they are also vital in times of arrest. If the user is detained, their trusted networks should be able to immediately request that their accounts are 'secured' (or shuttered) in order to prevent authorities from accessing 'incriminating' content. The ability for detainees to have their support networks do this can prevent harassment, torture, questions, and intensifying charges or time in detainment.

In arrest cases or other forms of targeting the community, having swift action and rapid response between local groups, lawyers, and companies can be critical. This was also repeatedly found in our previous investigations based on the insights from defence lawyers protecting communities.

Interviews:	**11 out of 93** (12%) interviewees stated that they want companies to have better reporting and rapid response systems.
Focus groups:	This was echoed in **5 of the 6 countries** where we held focus groups.

Many assume there will be no support for them if and when they are arrested, and they are unfamiliar with resources available. There is no widespread knowledge of the organisations with rapid response networks with platforms, or awareness of platforms' obligations to respond to hate speech and violence online.

As reported in Part II under 'Prevalence of hate speech and the lack of support':

Interviews:	**70 out of 80** (88%) of our interviewees had directly experienced **hate speech online**. This was the most universal experience across all of our research countries.
Surveys	**235 out of 1,374** (17%) respondents who reported non-state-facilitated abuses in our surveys had experienced hate speech in the form of **harassment** (96 respondents), **threats** (127 respondents), and pure **homophobia** (12 respondents).

Despite these massive numbers, the work to counter hate speech and actions taken are concerningly inadequate: a huge number of people reported little to no remedial action and content moderation support from the platforms.

> ***41 out of 93*** *(44%) interviewees mentioned having no luck with their reporting. The majority felt platforms did not care about linguistically trained support for the region.*

Some reported that not only was the violent hate speech they had reported **not acted upon**, but they **received responses that the content was not hate speech at all**. People described experiencing transphobia, homophobia, racism, xenophobia, violent threats, doxxing, outing, revenge porn, and paedophilia, specifically.

Others mentioned specific features that allow for cross-platform hate speech, outing, and harassment such as those referred to in the previous recommendations.

There are various and continuous issues around how automation and machine learning algorithms versus human-level moderation are used to moderate or translate LGBTQI+ content. This content is often flagged by civil society groups, on top of the issues that link to the lack of contextual understanding of the community and languages (see further exploration in our research). In effect, the lack of nuance, contextual understanding, and co-designing moderation procedures with impacted communities has meant that hate speech is not addressed. This is exacerbated by the fact that queer accounts are instead penalised for using inter-community terminology.

As reported in Part II under 'Prevalence of hate speech and the lack of support', most respondents felt that platforms not only **do not care about harms from hate speech against LGBTQI people+** but are also generally **neglectful of hate speech in Arabic dialects or Persian**. Many point to **over-complicated or irrelevant reporting mechanisms** and options that do not address pressing needs within the community.

Further recommendation details

We echo the demands of organisations such as Human Rights Watch, whose campaign focuses on the failure of content moderation, especially on Meta platforms. We also echo calls from 2020 of 22 LGBTQI+ organisations, mostly from the MENA region, who urged platforms, especially Meta, to address rising levels of hate speech and address issues regarding the lack of contextual, cultural, and linguistic understanding from their content moderators.

How can it be implemented? (for app developers and security teams)

There are two main forms of escalation and reporting:

1. Reporting **specific hateful content and comments**.
2. Reporting for the **suspension of accounts for safety after arrests**.

Specific hateful content and comments

Dealing with specific hateful content and comments – especially through robust **on-app reporting systems** with an option for users to report harassment, entrapment, and/or arrest – is important; however, there needs to be a **commitment to the creation of better reporting systems**:

1. Initially, the improvement of these systems needs to be **co-created and co-designed with these highly at-risk and criminalised communities** in order to create impactful and meaningful systems.

2. These systems need to be **simple and accessible** in their functionality and options. They should provide specific pathways for reporting anti-LGBTQI+ content that are easy and swift to access. It is important that these escalations are dealt with outside automated systems, and have pathways that prioritise them, so they lead to faster action in countries where LGBTQI+ people are criminalised. This is vital as some of the content reported can lead to serious real-world risks and violence or arrests.

3. In detecting hate speech, much more work needs to be done in tandem with impacted communities to **ensure moderators have language and contextual skills** to deal with these reports. The correct lists of flagged classifiers of content must be created (as echoed by numerous civil society organisations, so we will not further elaborate on this here).[29]

4. There should be a commitment to work with civil society in researching and **moving away from over-reliance on** large language models via natural language processing used by the majority of platforms. Given the complex language and cultural context, as well as the inherent risks of bias, it is irresponsible to rely on just one language model.[30]

5. The reporting systems should also **provide links and contacts for NGOs** on the ground that can act and be connected to at-risk individuals in emergencies.

6. Companies should **develop best practices for targeted communities** based on the work already being done within civil society and internally at companies.

Suspension of accounts after arrest

For dealing with the suspension of accounts after arrest, there is an urgent need for **better direct lines of communication** with community members or civil society rapid responders who are directly communicating the actions that need to be taken.

1. When arrests or massive high-risk events occur, companies should be prepared and **have teams in place to rapidly respond and take immediate action** to remediate any potential harm. In cases of arrest, each moment is vital and can mean the difference between sentencing and acquittals depending on how much police access people's accounts before these are taken down or suspended.

2. Discussions should initially be held with the partners of this project to **discuss the best method of communication**, ensuring swift responses, and **setting up the point of contact** between the organisations/NGOs and the apps – including how to identify the right organisations and teams in the country.

3. **Identifying the right groups** in countries with laws criminalising LGBTQI+ individuals is a priority. This will include discussion with the app team about ensuring the deletion of the user's data and how to prepare government requests for data in such cases (e.g. data access subpoenas by governments).

4. **Implementation should not be done via an automated option within the app**: a direct point of contact is needed with relevant individuals, and **specific emails/communication lines** are needed for dealing with LGBTQI+ case escalations.

5. A **second point of contact** should also be identified within each organisation and app, in case a primary point of contact in the communication line is absent or inaccessible.

6. We also ask for **cross-platform connections and effort** since LGBTQI+ cases often impact multiple platforms that some responders may not be connected to.

Stopping or putting barriers on mass forwarding without consent

Combined with continued efforts to challenge mass forwarding and the widespread sharing of disinformation and hate speech,[31] further actions should be investigated. The proliferation of doxxing, hate speech, and outings happens with mass forwarding of photos, chats, and identifiers of individuals. Often, this happens through forwarding of videos or photos sent during conversation with adversaries (state or non-state anti-LGBTQI+ actors). There should be **warnings or blocks on the forwarding of content from a chat without the consent of the original sender/poster** (this is to be combined with Recommendation 22).

However, research needs to be done and resources invested into finding harm reduction methods for these issues.

HARM REDUCTION FEATURES NEEDED

Main recommendations

In this section, we look at very specific features that our participants have asked to be added to their apps and platforms. There are a total of **15 recommendations for feature changes and introductions** that provide for harm reduction against arrests and device searches – as well as added protective methods for users in these high-risk contexts. Technology is playing a role in the community's criminalisation and abuse, and while these methods do not address the core issues and risks users face, they serve to reduce these harms. In this report – as well as in Part II – our research shows that the features introduced by our previous and ongoing work have been used and are highly effective – including in some cases leading to release without charge from custody. In our survey, an overwhelming majority of our respondents said 'yes' to whether having these harm reduction features determined their use of apps and platforms, with only 20% saying 'no'.

***2,430 out of 4,128** (59%) survey respondents said that having harm reduction features determined their use of apps and platforms – only 7% said it didn't.*

The list of what would be needed in terms of harm reduction features is long. However, here we have focused on the most urgent short-term actions to be taken based on our work and the overwhelming desires and wants for change on this front from those who participated in our research.

REC

We need to hide our apps with app icon ‘stealth modes’: app icon cloaking/discreet app icons

Recommendation 18: Implement app icon ‘stealth mode’ options (app cloaking/discreet app icons) to hide the icon of the app in plain sight.

One of the main, very contextualised recommendations for feature changes is for a requirement for more apps to **implement discreet app icons or icon changes** that provide a ‘stealth mode’ or ‘app cloaking’. Such features hide an application in plain sight to help users protect their privacy and safety, especially if their device is searched by an adversary. We recommend that when implementing these features, **further modes of harm reduction are added** for optimal stealth, and that **user controls reflect the situational risk** for those most at risk, who deeply rely on the feature.

Context and research behind the recommendation

Our research has consistently shown that one of the most common, high-risk, and successful forms of persecution and targeting by law enforcement is confiscating a user’s phone and scrolling through its contents and the list of installed applications. When looking for certain apps – such as queer dating apps – they are looking for certain icons and app names. In these situations, **simply having certain apps can mean that a user has either outed themselves or raised suspicion** enough to justify looking further into the contents of the app. It is often the **contents on apps that lead to prosecution or further abuse and violence**.

This sort of identification occurs most often at checkpoints in the streets, but it can also take place when a user is stopped on the street by police or officers. It can also occur in detention and pre-trial investigations or interrogations. Separate to police abuse, this identification can also lead to abuse in familial or other social situations. When carried out in a prosecution setting (especially in the investigation phase), it leads to the discovery of conversations and connections/matches, leading to further criminalisation and exposure of others.

Interviews:	**47 out of 93** (50%) interviewees had experienced device searches in the 8 countries we studied. This is an important number. Nearly every interviewee who had had an altercation or interaction with police and law enforcement in these cases had had their devices searched or attempted forced access to devices.
Focus groups:	In **6** of the countries, it was also a very prominent issue, with every focus group bringing up the issue of device searches and forced access to devices.

Due to the pervasiveness of this tactic and its potential harm, individuals and networks have adopted tactics and security strategies to avoid exposing their devices' contents, especially on certain apps. When authorities and law enforcement discover an app like the queer dating app **Grindr**, it is seen as confirmation of an individual's sexuality. A generalised dating app like **Tinder** invites further inspection of the app to investigate the contents and potentially out an individual's sexuality (often by seeing the chats or profile). An app like **Signal** hints that the person is using a secure messaging app to potentially communicate higher-risk conversations away from surveillance. Authorities will then further inspect it with the possibility of outing an individual's sexuality and/or network. An app like **WhatsApp** invites further inspection in the countries of our focus groups. The main app used for communications is **WhatsApp**, and thus conversations and the main content, as well as broader networks, will be present.

'App cloaking', or Discreet App Icon, is currently a very popular feature on Grindr, which was implemented as part of a partnership with ARTICLE 19 and use of DFM. It is now available on Signal, with App Icon changes introduced as part of a collaboration with DFM and this work. Despite the limited number of apps and platforms that provide this option, it is still one of the most relied-upon features mentioned.

Interviews:	**9 out of 93** (10%) interviewees asked for more 'app cloaking' as a feature they wanted to see implemented by platforms.
Surveys:	**2,239 out of 5,018** (45%) respondents (only using responses from those who completed the full survey) overwhelmingly mentioned that 'app cloaking' was the main feature used. This was in reference to Grindr, which, at the time of writing, was the only dating app with this feature. **254 out of 2,264** (11%) respondents cited this strategy as their most used as a safety measure based on the available options to them on some dating apps we partnered with.
Focus group:	**3 of the 6** focus groups we held asked for 'app cloaking' options on more apps.

> *'I like the option to disguise the application on the phone as another application (a game, for example).'*
>
> – Interviewee in Algeria

> *'Changing the icon is also very good, as some policemen or security forces are not familiar with such features.'*
>
> – Interviewee in Iran

We are seeing the importance of such features. For example, one of our interviewees in Egypt used this safety feature on Grindr to successfully navigate device search and prosecutions:

> *'I used the feature to change the icons of the dating application and change its name, so they could not find it.'*

They were eventually released after authorities failed to establish a link to their identity or find the 'incriminating' app that was on their device. This case is only one very clear example of how these harm reduction features – which were created with human rights organisations and based on the lived experiences of those most at risk – lead to potentially life-saving impacts.

How can it be implemented? (for app developers and security teams)

There are four elements that make 'app cloaking' successful in these cases:

1. Changing the icon.
2. Changing the app name.
3. Silencing or hiding notifications.
4. Preventing the app's icon from appearing in the phone's 'recently used app' list.

Changing icon and app name

Changing the app's icon and name should operate together in such a way as to make the app look innocuous when scrolling through a phone's list of installed apps.

On the Android platform, these techniques can easily be implemented using standard and well-known configurations and APIs on all modern Android versions.[32,33] It is possible for an app to change its icon and name during runtime. The app should include a diverse set of typically available icons that would not stand out during a routine physical device inspection. The set of available alternative icons can be rotated and updated on a regular basis.[34]

Guardian Project has open source implementations available in several of its apps on GitHub. More information can be found on Guardian Project's website or by contacting the Guardian Project directly.

In recent versions of iOS, a user can dynamically change the launcher icon, so we should not rule out iOS implementations of this feature. Grindr and Signal currently have this feature available for their iOS app.

In cases where the app has not been configured to ensure the change in icon name to match the new 'stealth' icon, **advice should be provided** on how this can be done or advice on how to potentially hide the app name (e.g. by placing it in the app 'dock'). Either way, users should be made aware of the potential mismatch of the app icon and name (this was done well by Grindr).

Silencing or hiding notifications

Notifications are still an issue for apps that currently have 'app cloaking' options implemented, and they undermine some of its utility. Some of these limitations come from the operating system level (especially limitations on iOS where the name of the app is challenging to change on a notification). Some arise when the implementation of default measures is not added to reduce risks that arise through notifications.

> *'One of my problems is the notifications and the fact that they appear with the name of the app.'*
>
> – Interviewee in Iran

Specific **control of how notifications look and work in different app states** (i.e. cloaked vs uncloaked) is an easy feature to implement. In the design of the feature, we **recommend that the app's notifications are turned off automatically if the app is in stealth mode**. If the user cares about an app enough to cloak it (rather than simply delete it), it is likely they will use it regularly and check for notifications, voiding the need for notifications for those users.

If the app's notifications in stealth mode are not turned off automatically, it is **recommended that notifications not be displayed at all**.

That said, if showing notifications is a requirement for the user, care should be taken to change the icon, name, and content of the notifications to match the cloaking theme. The content of the notifications should also be masked or limited.

In a case where notifications cannot be turned off as part of ‘app cloaking’, it will be up to users to turn off in-app notifications. An **in-app reminder to the user to manually turn off notifications** should therefore appear when the ‘app cloaking’ option is selected.

We also recommend **removing an app from the ‘recently used app’ list**. This stops a specific type of privacy intrusion when a device is taken for ‘inspection’ at a checkpoint, for example by a teacher or family member, or other authority figure, especially law enforcement in similar contexts seen in our research. Often, a quick way to check what the user does with their device is to see which apps are open or recently used. By not having the app display itself in that list, you reduce the likelihood the app will be more closely inspected.

Finally, stealth mode can be taken a step further: a **personal identification number (PIN) can be required to access the app while in stealth mode**. The user interface for PIN entry can be simple or it can be further ‘cloaked’ by simulating a calculator or phone dialling user interface that matches the icon and app name selected. There are limitless possibilities for inventive cloaking measures. An app can also provide added users options and choose to disallow screenshots.[35]

It should be noted that **eventually police and security forces may become knowledgeable about the use of ‘app cloaking’ and ‘app cloaking’ icons, which could lead to a change in device-search tactics. Therefore, apps and support experts should continue to work together, and look into ongoing research and support on the issue to ensure that cloaking stays ahead of detection methods and efforts.** This is especially important due to how vital this feature has been for highly at-risk users and its success in supporting their safety. Further advice for this can be obtained from our research and technical team. We can be contacted at afsaneh@de-center.net and MENA@article19.org.

We need stealthy self-destruct buttons

Recommendation 19: Implement stealthy self-destruct/panic button (or similar) options for emergency situations and blocking access to device content, especially for the most high-risk users.

Apps should implement an **easy-to-use option/button that the app user can trigger to commence needed safety measures**, and to block access to the content of the app from adversaries who may have control of the device. This is often called a 'panic button' or a 'self-destruct button'. The **trigger must be easily accessible from the primary home screen of the app** and is usually a button. The available safety measures must be provided and implemented by the app developer, and the user is free to choose which safety measures best suit their risk level when choosing to implement the feature on their app.

Context and research behind the recommendation

As mentioned, our research has consistently shown that one of the most common, high-risk, and successful forms of persecution and targeting by law enforcement is confiscating a user's phone and scrolling through its contents. It is often the contents on apps that lead to prosecution or further abuse and violence. **This is why one of our more important recommendations is the implementation of panic buttons, or a method to wipe the content of apps or devices in a stealthy method.**

Interviews:	**11 out of 93** (12%) interviewees m**entioned** needing a wipe-all method with their devices and apps to support them in high-risk situations.
Focus groups:	In **3 of the 6 countries** where focus groups were held, participants mentioned needing a wipe-all method with their devices and apps to support them in high-risk situations.

The number of interviewees and participants asking for this type of feature was an unexpected outcome from this research. We did not expect our interviewees and participants to have this type of harm reduction feature at the top of their minds. For them to express this need so clearly is again a clear example of the savvy ingenuity of the community in navigating the harms that befall them. This feature is vital to provide users a method to protect their privacy and safety in times of danger or high stress – such as during device searches and/or interrogations, when providing access to the content of their apps can be detrimental to them and their networks.

> *'I don't know if this feature exists, but in emergency situations, there [should be] a possibility to press a certain button, say three times, and that makes me sign out or blocked from all platforms.'*
>
> – Interviewee in Sudan

> *'Now that they are arresting people for hijab, their phones are being searched. There should be a panic button or something to send sensitive data into a vault or a safer place in these situations.'*
>
> – Interviewee in Iran

In Part II under 'Risks taken to avoid providing access to devices', we noted:

15 out of 93 *(16%) interviewees reported having taken extreme measures to avoid giving access to their devices.*[36]

Methods included breaking their phones, hiding their devices, vehemently refusing to give their passcodes, claiming they did not remember their passcodes, or in a few cases using technological tools to wipe their phones before providing access.

Our cases show individuals **risking further criminalisation** or potential charges of tampering with evidence, or physical violence and abuse, **in order to refuse access to their device**. What we have seen in the research is that often when our interviewees have refused access to their devices, they have not faced further incrimination (potentially as the device searches are regularly done without warrants and illegally) – though they did risk further violence. On occasions when they have avoided both, there were few instances when they either had their phone with the high-risk content confiscated or provided a savvy excuse for why they could not provide access. Without further 'incriminating data' from their phones, the chances of lowered charges, sentences, and even acquittal/release increased.

> *'Maybe an immediate option to erase all the conversations and the affiliated files archived in the handset in case of arrest. We were detained in Somuah police station, and I refused to open my phone. At that time, the officer handcuffed me with handcuffs and forced me to open the phone with my fingerprint.'*
>
> – Interviewee in Egypt

> *'For example, if someone ends up in Evin prison or somewhere like that, the app itself will delete all the information.'*
>
> – Interviewee in Iran

These findings are important because they show that actions taken to block unauthorised access to devices, or ensuring that individuals have methods to wipe devices, are fundamentally important as a harm reduction tactic. It will support tactics people are already taking upon themselves.

They are also important as they show that they result in a **higher likelihood of reduced charges or even acquittal/release if they take these measures**. These are in cases when the 'full' extent of potential charges and networks have not already been revealed.

There is a **higher risk of incrimination/harm to people and their network** and community at large **if their devices are searched than if they provide access** in the first place.

Again, this pattern of 'by any means necessary' was much more prevalent than in our previous investigations, which is potentially due to police's increased reliance on digital evidence for prosecutions and using devices as ‘digital crime scenes’. Those arrested have taken it upon themselves to block access to their content by any means necessary because they know how detrimental a device’s contents can become for them and their networks.

For example, our interviewee in Sudan who successfully navigated these searches with a lesser version of our panic button recommendation showed its potential. They were arrested and released due to lack of evidence twice:

> *‘The first time I activated a feature called Kill Mode, it hid the content of my phone and showed it as if the phone was empty.’*

The second time they hid their phone:

> *‘Both times I was carrying my phone and both times I was able to protect my phone information, effectively.’*

‘Kill Mode’ seems to refer to the ability (though this may not be an option for many arrestees) to enable their devices’ remote kill switch to block access to their content after it was confiscated.

This feature is also extremely important as many of those at risk, especially as we have seen in our work and research, do not have access to rapid response teams externally who may be able to shutter their social media or other communication accounts in collaboration with app platforms in times of arrest and human rights abuses. To have this access, a connected network is required, as is the ability to trigger this contact after arrest or risk, which many may not be able to do in time.

Since the content of devices has become some of the most detrimental elements used against an individual, and since not all users have access to rapid response teams to shut down their apps externally, the power to protect the content on their app becomes vital. For example, on some secure apps that hold no backup metadata for privacy protective reasons (such as **Signal**), this support is not even an option (i.e. for them to remotely deactivate an account). Thus, providing users at risk the power to protect themselves and their communities can make a life-saving difference and becomes even more of an imperative for such apps.

Having privacy-respecting infrastructures such as **end-to-end encryption by default**, such as the ones seen on **Signal,** is highly important. However, it is the base for privacy preservation. This should not be seen as the pinnacle for safety and privacy, and added efforts should be built upon it that reflect lived realities. To provide the truest form of safety, **secure tools should also add harm reduction features that are attuned to the lived experiences of at-risk groups with stealthy user controls**, and be balanced with harm reduction approaches in order to allow users the optimal experience in privacy – and the power to protect themselves.

Of course, all strategies can, and will, have drawbacks. The level of police/law enforcement capabilities here is unknown, but we must remember that harm reduction is to reduce the amount of harm individuals are facing, and not a guarantee of foolproof safety (no strategy ever is). But they are still effective, considering the risks people are already taking,[37] and they must ensure people have the options to implement and use what they need.

How can it be implemented? (for app developers and security teams)

In addition to the work of this project, Guardian Project and Amnesty International have been doing research on this feature and its technical capabilities for nearly a decade. The results of their efforts are available to use in applications today.[38] Guardian Project has an open-source framework called PanicKit that can be dropped into an Android application to enable panic features. If using PanicKit, the developer must only implement the safety-measure actions; the 'panic button' itself is delegated to a dedicated panic button app, such as Ripple.[39] Guardian Project and ARTICLE 19 have also been collaborating on a physical safety check-in app called Círculo that implements panic features, 'app cloaking', PIN locks, and many other features. We **recommend these features to be implemented** as part of a combination of harm reduction safety features **alongside the use of a self-destruct or panic button mode**.

Developers are encouraged to reach out to this research team (afsaneh@de-center.net and MENA@article19.org), Guardian Project, or Amnesty International with questions about the self-destruct or panic button feature. App developers can also choose to implement all the features themselves; however, they should consult and collaborate with related communities and research teams for effective implementation, as well as the most useful user interface patterns, to ensure the user can properly trigger the panic button in times of stress. The following proposals are based on the experiences of criminalised, marginalised communities during high-risk periods of arrests and interrogations.

What would this mean on an app/device level?

The **trigger should be easily accessible from the primary home screen of the app** and is usually a button or configurable widget that can help it not stand out. This is very important, especially because they will be used in moments of high pressure – whether it be a police device check or family wanting to see the contents of a phone that may out someone.

The available safety measures must be provided and implemented by the app or operating system developer, and the **user should be free to choose** which safety measures best suit their risk level, if the developer has provided multiple options. Granularity is very important.

Data should be deleted from the device and server side. In the case that the device does not have connectivity, attempts should be made for some time to send a deletion signal to the server.

Strategies for triggering this mode

Taking into consideration the context within which this feature will be implemented, we should build around experiences and case study scenarios. In this research, we saw that device searches were often initiated by a forceful demand for phone access and passcodes. Instead of individuals saying they have forgotten passcodes or saying they will not comply (which is the route **10 of our interviewees** took), this mode can be triggered by the individual suggesting compliance while inputting an incorrect password – or a specific password or PIN – multiple times. This would then initiate the mode but also avoid further violence.

The **'self-destruct' PIN feature** would also be an addition to this mode. In this option (similar to 'We want more PINs and stealthy locked chats'), if the user turns on this feature and purposely inputs the PIN incorrectly three times, the app content will self-destruct and only provide an **'error' message**. As mentioned, in situations such as checkpoints, interrogations, and other forced access to the device, the user has the power to feign nervousness and forgetfulness in the face of coercion to also limit suspicion that, in fact, they are implementing the self-destruct feature. As we see in our research, this is what many of our interviewees (successfully and unsuccessfully) attempted.

It can also be **triggered stealthily by using external devices** like a headphone pause/play button being pressed in a pattern, or a Bluetooth device not being available. For instance, if the user had a Fitbit band or AirPods paired to their phone, the 'panic' state could be triggered to go into lockdown if the user was out of Bluetooth range from their phone. However, for safe and robustly implemented versions of this feature, we highly suggest developing teams to further research, experiment, and discuss with impacted communities. Our technical team can be reached for more information. We can be contacted at afsaneh@de-center.net and MENA@article19.org.

Ease of use

It is essential to present these options clearly with as little clutter as possible. The panic set-up should be on a devoted screen, not mixed in with other settings, and take up the full screen. Panic is a time of stress; therefore, the panic response should strive to avoid adding any stress.

This is an advanced recommendation that can be combined with many of the recommendations in this report. For example, **when the user triggers the panic button, applications could cloak the app** using methods previously outlined. Further advice can be obtained from our research and technical team. We can be contacted at afsaneh@de-center.net and MENA@article19.org.

Varied methods of implementation and granularity for safety measures

Examples of safety measures implemented in apps range from engaging cloaking mode or locking the app with a password, to more destructive actions like deleting users' messages and photos from the device – or disabling the app entirely (as outlined in the other features recommended in this section).

The **optimal self-destruct option should/could trigger switching to a second, 'clean' profile** for the interface of the device or app. This would provide the option for an individual not to lose all their data and device contents if they get their device back. It would provide a version of the device that does not contain 'high-risk' content that could be seen/searched. This would be useful for law enforcement situations as seen in our reports, but also if children or family members, for example, want to access the app on a user's phone. The user may want the app's content to remain child-friendly/free from private content.

A **remote wipe option triggered by a trusted device or contact** is also very valuable. Our interviewees asked for an added option that, in moments of need, would trigger the panic button and alert trusted friends or contacts.

> *'If I leave to meet someone, I would like my security contact to be informed of my last contact if I do not give a sign of life.'*
>
> – Interviewee in Algeria

This is already available in different devices and operating systems so we will not further elaborate for this report.

There are two important notes here on why the **default measure should not have the panic button connected to contacting law enforcement**:

1. In most high-risk contexts, any contact with law enforcement is risky because they are the source of the risks. This option should only be in the case that a user adds it themselves (which will potentially only be the case in some geographical contexts) – and if it makes sense with the response strategies of law enforcement.

2. The list of trusted contacts should be secured as it can cause further risk or outing of a network. Thus, it should not be readily discoverable.

We need more PINs and stealthy locked chats

Recommendation 20: All apps should have PIN and locking features, as well as added stealthy locked folders for the most sensitive content/chats.

Apps should all have PINs or locks as easy-to-enable options. These PINs should have their own sandbox to keep the data separate from other applications on the phone. Our **communication apps should implement methods used by other apps that store sensitive data** – like mobile financial applications that use this feature in conjunction with a password or PIN to help ensure only an authorised person can access the application's data.

We also recommend that apps implement further layers of safety and granularity within the app to add lock and PINs to specific high-risk individual chats, as well as further methods to hide the chats.

Context and research behind the recommendation

As with any sensitive app, or app with information users might not want others to access, an **app-specific PIN or password lock is optimal**. This is seen with banking apps to secure users' banking information on shared devices, even if an individual's device falls into someone else's hands.

App PIN locks can also provide security in contexts where the individual does not feel physically safe and needs extra security if their device is taken. In MENA, and countries with laws prosecuting LGBTQI+ people, this is amplified due to patterns at times of arrests. As we have seen in our research, device searches are one of the most prevalent and high-risk methods of policing and criminalisation.

***47 out of 93** (50%) interviewees had experienced device searches in the **8 countries** studied.*

Nearly every interviewee who had had an altercation or interaction with police and law enforcement in these cases had had their devices searched or attempted forced access to devices.

Interviews:	**12 out of 93** (13%) interviewees pointed to PINs and app locks as a feature they wanted to see from companies.
Surveys:	**48 out of 2,264** (2%) respondents mentioned PINs and app locks as a feature they used as a safety tool from the safety features that were available to them through the apps we have partnered with.
Focus groups:	In **3 of the 6 countries** where we held focus groups, participants pointed to PINs and app locks as a feature they wanted to see from companies.

Interestingly, **PINs** and **screenshot blocking** were also directly mentioned features: **68 respondents out of 2,264** mentioned using these features to protect themselves, even though they are not commonly available. In our research, we saw how interviewees and participants used this method, combined with the use of the protection of PINs, to refuse access to the data on their devices in high-risk interrogation situations (see Part II under 'Obfuscating data or no data in interrogations and searches'). In general, this is important for protecting information and data – someone with access to the device should have extra difficulty in accessing that data in other contexts too, whether due to general privacy and safety concerns, or dangers in the contexts in which they live.

In the Grindr app, we have seen the heavy combined use of the app cloaking features (discussed earlier) and PIN at the same time.

The implementation of a PIN makes unwanted access harder. In an investigation in collaboration with defence lawyers of LGBTQI+ people in MENA, our interviews and court file reviews included direct notes from officers in three Egyptian files. They reported that prosecutors were unable to access a specific device due to the PIN, where users refused or forgot the PIN, leading to lowered changes during prosecutions.

This method is not perfect, and in interrogations there are many ways a PIN can be drawn out of a person – but it is an extra layer of protection.

Importantly, this should not be entangled, or in any way used, with biometrics such as fingerprints or FaceID. We have seen the grave risks this raised for at-risk users in Part II under ‘Biometrics on devices’. The risks to individual privacy and rights are dramatically different (and often increased) when biometrics are used instead of manual passwords and PINs for locking apps.

Further recommendation details

For each app, there should be **an easy and practical option to enable a PIN code**. This would have granularity for how often users would like their PIN to be triggered (whether on each login, or after a period of inactivity, or options for users to enable it when they feel they will need it, for example when going to protest). Many apps currently have PIN reminders, which are very important.

The apps should also create an **in-app, password-protected camera roll** that photos/videos from messages in the app can be saved to (without saving to the phone’s general camera roll).

Importantly, this should not be entangled, or in any way used, with biometrics such as fingerprints or FaceID.

Double PIN/lock folders

Having a double PIN option is ideal where certain chats and groups have extra security. In most of the situations we have seen, police coerce individuals into providing access to locked devices and apps. So, having a **vault option or second PIN** that locks away more sensitive content or chats that the user has ‘hidden’ can be extremely important in these cases.

The content and data of these chats should be by default end-to-end encrypted.

WhatsApp example:

As we implemented the broader work of Design From the Margins, our team worked with WhatsApp to introduce Chat Locks[40] and Secret Code on the app. (At the time of writing this report, the feature was still entangled with biometrics, which we heavily advise against as it has made it less secure and private. If this entanglement is removed, this feature will be one of the best harm reduction-based, secure private chats/locked chat options available on any major chat-based app – and it would provide heavy protection in high-risk interrogation and device check situations.)

This feature was created with the scenarios and experiences of these communities in mind. It allows for a locked chat connected to a 'username', as WhatsApp advertises: 'You'll have the option to hide the Locked Chats folder from your chatlist so that they can only be discovered by typing your secret code in the search bar. If that doesn't suit your needs, you can still choose to have them appear in your chatlist. Whenever there's a new chat which you want to lock, you can now long press to lock it rather than visiting the chat's settings.'

This level of granularity supports situations where, for example, an interrogator forcefully searches a device and would: 1) need to know there is a locked chat; 2) need to know the username to type in the search bar of the locked chat; and 3) have the PIN for the locked chat. This also provides safety in a situation of access to a device by physical force.

However, despite its unique and important safety framing for high-risk contexts (and, in turn, for anyone who wants a safer chat), it requires biometrics for a user to be able to use it. This undermines the safety, data privacy, and security of the feature.

Due to its newness and entanglement with biometrics, we do not have examples of use by the community yet. We firmly continue to recommend its disentanglement from biometrics so it can be safer for those who will use it.

Telegram example:

Three of our interviewees in Iran mentioned Telegram's Secret Chats as a safe and ideal feature:

'For example, I know that Telegram Secret Chat is the best tool.'

– Interviewee in Iran

'I don't send a photo of my face in these hookup apps, even if the other person asks for it. If we want to go further, we move the conversation to Telegram, where we can send secret messages. I set a timer for the chat so that it will be automatically deleted later, and to see if, for example, the other person takes a screenshot or something like that.'

– Interviewee in Iran

'It would be great to have the features which Telegram has in other apps. For example, the secret chat features of Telegram.'

– Interviewee in Iran

As Telegram states: 'The secret chats on Telegram are different from the normal chats in that you cannot forward secret messages to any other contact. Moreover, the self-destructing feature is another key attribute of the Telegram Secret Chat. Besides, it will notify the sender about any screenshot.'[41] This does not address the situational specifics and obfuscation of the WhatsApp feature. It is, however, not connected to biometrics.

One of the areas of concern with Telegram is that they use a custom encryption protocol (MTProto) instead of an industry standard and verified cryptographic protocol. This had led to instances where researchers reviewing the security of Telegram have identified vulnerabilities in MTProto 2.0 (which have been fixed) that do not directly affect the confidentiality of messages. Concerns remain, however, with design choices that make it difficult to implement MTProto safely, which introduces higher risks within Telegram's robust third-party ecosystem. Here, we reiterate the importance of reviewing and testing secure encryption protocols and implementations, including whether it is future proofed against major changes in computing such as quantum computing, especially for locked chats and messages that can be easily captured and stored by governments or other bad actors.

No notification for locked chats

Sound or pop-up notifications add risks. This should **never be a default for locked chats**, but rather enabled by the user if they want these notifications (see below).

How can it be implemented? (for app developers and security teams)

Mobile apps have their own sandbox that can keep the data separate from other applications on the phone. Applications storing sensitive data (e.g. mobile financial applications) use this feature in conjunction with a password or PIN to help ensure only an authorised person accesses the application's data.

There are two key elements to this recommendation:

1. **Storing all of the application's data within the application sandbox**. For dating apps, this should include keeping photos from messages within the application sandbox, instead of storing them on the phone's photo library.

2. **Implementing an application password or PIN to access the application**. Use of biometric access, such as fingerprint or FaceID, is not recommended since a user can be physically forced to unlock their device under duress.

For the implementation, a developer can use these following elements:

- **Standard file APIs** to implement storage of media with the local application storage. Important to note, on Android specifically, do not use 'External Storage', which any app with 'Read External Storage' permissions can read from.
  - Android now supports file-based encryption built-in, which can be used for additional data protection.
  - SQLCipher is a cross-platform encrypted database that can be used to store data securely on Android and iOS and is currently used by Signal and

other apps (this was partly developed and is fully endorsed by Guardian Project).

- **Standard password/PINconfigure**, lock and unlock user interface. Biometrics should not be used.[42] A password option should be given to enable stronger security than a numeric PIN can offer alone. It is important to inform the user that this password/PIN will be **stored in the app only and cannot be reset**. They must be able to remember it.
- The user should be able to **immediately lock** (i.e. require a PIN to re-enter) the app or to have it lock after a configurable time-out.
- The app should be **locked upon opening after a first-time reboot**.
- **Code**: Both Android and Apple offer system mechanisms for storing credentials.
  - PFLockScreen-Android is a simple open-source library for implementing PIN lock feature and user interface.

Further advice for this can be obtained from our research and technical team. We can be contacted at afsaneh@de-center.net and MENA@article19.org.

For **locked chats**, the very contextualised and important framing used in the WhatsApp Locked Chat (above) is a top standard for this; however, as we have mentioned, we **strongly oppose safety features which are entangled with biometrics**. Thus, here we recommend the methodology used in this chat lock without a link to biometrics (maintaining the use of a unique PIN or passcode for the locked individual chats should be the go-to method). The locked chat should also of course be end-to-end encrypted by default.

We need more ephemeral/timed and deletable messages and photos

Recommendation 21: All apps should provide ephemeral, delete for all, and 'view once' text and media messaging options, and implement them safely.

Apps that allow the sending of photos, video, voice notes, and messages must have **options for timed messages** (also known as ephemeral or disappearing messages) as well as 'unsend messages',[43] 'delete for everyone', and 'view once' text and media messages. These features should have granularity for their options, allowing for their settings to be controlled by the user.

Context and research behind the recommendation

In our research, we found that timed (ephemeral) messages, as well as 'unsend messages', and 'delete for everyone' messages, were some of the most used and wanted safety features. We are also seeing an increasing reliance on, and desire for, 'one-time view'/'view once' messaging, which **incorporates screenshotting and forwarding disabling**. The high demand and reliance on these types of features are not surprising with the level of risk people are under solely for communicating while LGBTQI+.

Having **options to communicate without building a backlog and history of conversations** (whether visual, audio, or written), and a method that allows for them to be safely removed, provides an extra layer of safety for people – not only for people they are talking to, but also if their device is taken or it falls into the hands of adversaries. During device searches, this feature adds protection so that there is less chat history on both the send and receive points that can be used against individuals, even if they had not been on top of regularly clearing their chats.

Our research in countries in the MENA region (and beyond) clearly shows levels of targeting, arrests, and entrapment of users of social media, chat-based, and dating apps. They are targeted by extended entrapment operations that use messages against them in criminal proceedings, where the evidence presented in court are messages and photos that the police-run fake account screenshotted and printed after the defendant sent them. In device searches, we see law enforcement read through messages and use search bars to search for specific terms and words in order to find chats they should look for, and then use them as evidence from the retained chats and data on their apps (see Part II). People have overwhelmingly asked for features that **allow for ephemeral and non-permanent conversations** in an attempt to lessen some of these risks.

Every defence lawyer in our Digital Crime Scenes report, directly or indirectly, mentioned the idealised concept of 'if the evidence did not exist in the first place' when asked how to decrease the use of digital evidence in these court cases. Secondary to that concept, **23 of the interviewees** directly talked about the need for timed/ephemeral messages, or options to delete messages or photos/videos and voice notes.

Interviews:	**23 out of 93** (25%) interviewees directly mentioned timed and/or unsend messages as the feature they wanted more of from companies.
Surveys:	**600 out of 2,264** (26%) respondents answered that they used "media and chat deletion, and disappearing messages' used for safety. **191 out of 2,264** (8%) respondents also directly mentioned 'unsend message'.
Focus groups:	In **3 out of 6** country focus groups, participants directly mentioned timed and/or unsend messages as the feature they wanted more of from companies.

In our surveys, we asked participants about the features they used to provide safety for themselves. We specifically focused on safety features we had worked on with partnering companies. The top response was the use of '**timed pictures**' or **ephemeral photos**, with **600 out of 2,264 people** from our survey selecting this.

‘**Unsend messages**’ was second with **191 people.** Here, we view these features under the umbrella of **‘ephemeral messaging’**, and a large percentage of respondents rely on it for their safety. This feature is available on some dating apps and a number of chat-based apps. This optionality has always been highly requested and relied upon by the community, as we have seen and recommended in previous work.

Many of our participants and interviewees linked these features to the combined use of screenshot blocking as the safest option for them, and many pointed to the Snapchat options here (see Recommendation 22).

Other apps combine these features, for example on Grindr:

> *‘Before it was risky to send photos, but now with the new options of ephemeral photos that stay only for seconds and the fact that you can’t screenshot ephemeral photos, I feel more safe.’*
>
> – Interviewee in Morocco

This recommendation does not assert that timed message deletion, unsend messaging, or ‘view once’ messaging options can stop targeting entirely. Rather, **timed message deletion limits the window in which the antagonistic party can copy and save the conversation logs and photo history**. This strategy relies on raising opportunity costs and making it financially and technically infeasible to carry out such attacks (as with all of our recommendations). ‘View once’ options that are implemented by many social media and messaging apps like Snapchat, Instagram, WhatsApp, and Signal are also often combined with screenshot blocking and bans on downloading and forwarding. This again raises the costs to make it financially and technically infeasible to carry out such attacks.

There is a need for ‘an option that you can delete your conversations’, so even ‘if you forget deleting it, it auto-deletes’.

> *'Delete all messages immediately after closing the chat or the app. Many people are careless. For example, they spend too much time on chats. When the danger is so high, there is no place for risk.'*
>
> – Interviewee in Iran

In our current research, many respondents and interviewees pointed to Snapchat for the 'view once' features. Not only does it expire immediately after being seen, making it very hard for an adversary to record anything in that time, but it also provides screenshot notifications to the sender.

> *'In Snapchat: One of the reasons for the app's popularity is that the maximum time you can keep a chat is 24 hours. If you take a screenshot the other person would be notified. It has become a norm not to take screenshots. The origin of the sent photos in the app is also clear and you can understand from the logo whether the photo was taken at the moment or in advance.'*
>
> – Interviewee in Iran

The 'norm' here is important as many people point to very real ways around this issue, for example by using a second phone to take photos, which we have seen in case files in contexts of post-arrests and when a device has been confiscated. This is why timed and unsend messages and other harm reduction methods for the device search situations are vital.

Further recommendation details

Timed messages (also known as ephemeral messages)

Apps that allow the sending of photos, video, voice notes, and messages should have the **option to delete them from both parties' devices after a specified period of time**.

This technique, also known as 'disappearing messages', exists in popular messaging applications such as Snapchat, Instagram, and Signal. Grindr and WhatsApp implemented these features based on recommendations from the work of this project, and they reported a lot of positive feedback from users.

When sending a photo or text message, the application should allow the user to choose a time period after which the message/photo will be deleted from both the sender's and the recipient's device. This can also be set at the conversation or group chat level.

Timed message and deletion options also need to be varied to provide for different scenarios. The ideal version of this would allow for the user to customise the exact timings for deletion, as well as having predetermined times (for example, the granularity seen on Signal). Currently, most of the apps with this feature lack timing customisation options.

The **user interface should make it clear that disappearing messages are enabled** and either recipient should be able to **force disappearing messages** for both parties.

Additionally, it should be possible to **enable the timed deletion of an entire conversation** with a recipient, as well as deletion on a per-message basis.

Notably, for photos, this recommendation relies on the fact that photos taken from the app are not saved in the camera roll (see Recommendation 7).

This should be combined with optionality to enable screenshot blocking (see Recommendation 22).

Unsend messages

'Unsending' of messages is more about a **user correcting a mistake or changing their mind** about sending a message, either before or after the message has been sent. For a message 'undo', a **simple 'hold' time buffer in the app** can make it possible for the user to have a chance to intervene before the message or photo is actually sent. If the message has been sent, after being sent with edit/deletion, then it can also be possible to edit or remove the message from both devices. However, a notice should be added ‘this message was removed’ in case they have seen a notification. Signal allows for any sent message, photo, voice note, etc. to be deleted at any time, either just on your device or ‘for everyone’.

Interviewees pointed to the helpfulness of being able to unsend messages without the other person being notified, as seen on Instagram:

> *‘You can unsend your message on Instagram. This is very good. Or, for example, look at all the features on Telegram. You can delete all the information. This can be very useful. It would be great to have the features which Telegram has in other apps. For example, the secret chat features of Telegram.’*
>
> – Interviewee in Iran

‘View once’ photos

However, there is so much more that can be done with timed and or unsend messages:

> *‘I especially like the ephemeral photos option; I would like the option for discussions using ephemeral messages which disappear after reading.’*
>
> – Interviewee in Algeria

This interviewee was describing the feature often called ‘view once’ as seen on Instagram, Signal, WhatsApp, and Grindr, for example.

When discussing ephemeral features and the community’s desire for more options and implementation of this feature, many mentioned options for **messages to be deleted after they had been viewed or listened to** in the ‘view once’ options. This is an option many used for their most intimate and high-risk content, from nudes to a passcode. This feature is important. On most apps that have it, it is linked to either screenshot blocking or notifications, which is the most stealthy and popular version of the feature. All of these features, combined with the short timeframe, raise opportunity costs and make it financially and technically infeasible to carry out such attacks.

> *'Before, it was risky to send photos, but now with the new options of ephemeral photos [on Grindr] that stays only for seconds and the fact that you can't screenshot ephemeral photos, I feel safer.'*
>
> – Interviewee in Morocco

However, there is so much more that can be done with timed and/or unsend messages:

> *'I especially like the ephemeral photos option; I would like the option for discussions using ephemeral messages which disappear after reading.'*
>
> – Interviewee in Algeria

This should be possible for many types of media – not only photos or videos, but also audio and text.

Finally, one lawyer specifically highlighted the need for an extra method to be able to **delete conversations or messages/photos when not in the possession of the device**. In a scenario where officers are already in possession of a device, there can be a way for the contents of that device to be wiped of information on the queer individual that would be used against them.

How can it be implemented? (for app developers and security teams)

It is important to note that all following suggestions are best implemented in addition to secure end-to-end encryption of messages by default.

For all of the features, it is recommended that **apps utilise the Matrix.org specification and audited open-source tools**[44] to build private, secure messaging and communication features instead of rolling their own solutions. This is used by the Convene ephemeral message app and Círculo safety check-in app, which focus both on sending messages for limited times, and later redacting or deleting them.

Matrix.org provides an open-source specification driven approach to solving these problems and it allows for the building of both private group chat features and custom application logic on top of it.

Timed messages

There is no generally applicable implementation for unsending messages, as each app’s back-end and front-end systems are unique. It is important that **messages and photos are fully removed from the devices of the sender and recipient**, and not just merely hidden. Using encryption for messages and data storage can also help ensure they are fully deleted on the device. Ideally, **messages and photos would also be deleted from the service provider's back-end servers**; however, that point should not impede the implementation of client-side message deletion.

‘Timed messages' should be linked with screenshot blocking or notifications for these selected conversations and timed periods. The extra security can be enabled for the whole duration or time period when the ephemeral messaging has been enabled.

Further advice for this can be obtained from our research and technical team. We can be contacted at afsaneh@de-center.net and MENA@article19.org.

Unsend messages

Unsending messages can be implemented in a number of ways, such as through a user experience mechanism of **delayed sending**, or through **actual remote deletion**. For the user experience approach, a message can be shown as sent, but delayed from leaving the user’s device for a certain, but short, amount of time. However, this kind of artificial delay could cause usability issues for those expecting ‘instant' messages.

For a full unsend, the server and any client apps that received the message must **support a verifiable, authenticated mechanism for accepting a remote deletion request**. Again, **using encrypted messaging** can support this capability, since it also provides for a cryptographic identity key that can be used to verify that the sender of a delete request matches the original message. For services without message encryption support, a simpler mechanism using **sender user ID matching** could be acceptable for a basic unsend/remote delete feature**.**

‘View once’ photos

This feature should be possible for many types of media – not only photos or videos, but also audio and text. Any sent item should be able to have a **defined number of views set for it on receiving clients, with a typical default of once**. After the item is viewed the specified number of times, the client device must delete the item from its local storage. As stated before, the use of encrypted storage facilitates ensuring this deletion is total, as does the implementation of screenshot blocking.

We need ways to stop non-consensual screenshotting

Recommendation 22: Provide methods and features to prevent non-consensual screenshotting and capturing of users' information.

Apps should have methods and features to **challenge the increasing use of non-consensual screenshotting** and capturing of users' information and messages. This should be done through **further research and consultations with impacted communities** for the best methods of implementation. **User controls** on how this is done with adequate granularity so users can implement this safety method to apply to their risk model are very important. The way this is implemented does differ based on the type of application (social media, dating app, or chat-based app).

Context and research behind the recommendation

The issue of screenshots and capturing of profiles or the content of conversations without consent has been one of the most highly raised issues in the research in relation to safety changes needed. There is no consensus on strategy regarding how we should be dealing with non-consensual screenshotting and capturing of content; however, there is wide consensus about the problem in general.

Through our years of research, we have seen that **screenshots from profiles, conversations, and pictures are a risk to users** and can be used directly against them (see Part II). These screenshots are often used in courts as evidence, especially through **entrapment by law enforcement accounts** that use the screenshots to link the individual to the conversation or content. For example, we saw this in our analysis in the Digital Crime Scenes report in 2022:

> *Screenshots are generally defined as a visual information capture of all or a selected part of what is seen on a screen and may be taken from either the phone of the accused or an informant. They can be produced through a built-in function of most phones and laptops, and offer a low-tech method to create hard evidence for court proceedings. Screenshots are significant in*

> *evidence-gathering and are frequently presented in court by prosecuting teams. They are popular because they are wrongly viewed as irrefutable and concrete, and have a visceral impact on judges. They are also an alternative to technical data extraction of the evidence, which may be required by the laws of evidence but in practice, might not be enforced.*

This method is combined with the risks associated with phone numbers, for example. As we have documented in many cases, the entrapper screenshots conversations with the sender's phone number linked to the account. These are often used in courts to create 'solid evidentiary' links between the communications and the individual's legal identity. They have also been heavily used as a tool for blackmail and harassment and heavily linked to outing campaigns (see above recommendations, especially Recommendation 3).

Due to its prevalent harms and layers of harm, **screenshotting has been raised as an area of real concern by the community**. Companies involved should have more informed and contextualised ways to address this issue, with the understanding that it is a complex and difficult issue with no one fix.

Interviews:	**33 out of 93** (35%) interviewees (one of the highest in interviews) stated that they wanted platforms to ban non-consensual screenshotting or support against non-consensual screenshotting.
Surveys:	**144 out of 2,482** (6%) respondents directly raised that they wanted platforms to ban non-consensual screenshotting or support against non-consensual screenshotting.
Focus groups:	**4 of the 6 countries** where we held focus groups stated that they wanted platforms to ban non-consensual screenshotting or support against non-consensual screenshotting.

This feature should also be tailored for granularity in options and settings as there are many reasons people will want to use a screenshot for safety or security. One of our Jordanian interviewees raised this:

> *'Screenshots are important because I was using them for … they're important when the management of these platforms makes peace with racism and misogyny and homophobia and transphobia that are in the bios of these accounts.'*

Although this was only mentioned once in our research, it is an important issue and explains why granularity is important in how this option would be implemented. Examples of our methods to provide for this are outlined later. It is also important for all safety features to be considered alongside better reporting mechanisms.

Further recommendation details

The way this is implemented **differs based on the type of application** (dating app, chat-based app, or social media). It is also possible to block screenshots on a specific screen of an app, or just for a limited time, or for a specific user experience workflow.

On dating apps

On dating apps, there have been multilayered risks linked to screenshotting and data capturing of profile data and conversations.

Profile

- **Profile information on a dating app can be very dangerous** for individuals and has been used for outing, harassment, and arrests. On a generalised dating app, this is highly risky when someone's identity as queer is identifiable. On a queer dating app, solely being on the app can be enough for these risks.

- **Profile screenshotting and downloading of profile content should be blocked** (see above).

In private chats

- **Conversation screenshotting and downloading of content should be blocked**, especially in high-risk contexts and where queer people are criminalised. This should be a feature that can be turned off by users who may want to keep this option.
- This should be **combined with timed/unsend messages and photos** (see above).

On dating apps, conversations are often very intimate and personal. These are voice notes, photos, videos, or messages. It is overwhelmingly asked that the **conversations on dating apps should be maintained as private** with a mutual understanding of their private nature. This has been tested with Grindr, and was implemented via the partnerships between our projects and the Grindr for Equality team – and with much popularity.

> *'Grindr has a nice feature where if someone tries to take a screenshot of a picture of a conversation it'll come out black.'*
>
> – Interviewee in Jordan

In Morocco, one of our interviewees echoed many others to ask for other apps to adopt this:

> *'I share photos, especially nudes, only on Grindr because it is not scannable. So I take advantage of this option. I would love all the applications like Snapchat to show me if something is screened or saved.'*

It is our opinion that **this option should be a default for chats on dating apps in high-risk contexts and contexts where queer people are criminalised**. More research and collaboration should be done with communities for options on how reporting and documentation can be maintained for documenting abuse content on chats without removing the safety option.

If this option is not provided by a dating app, the app should allow options for individuals to **enable notifications to be sent if someone is non-consensually screenshotting or downloading content**. This should be easily enabled as an option when a private conversation has started between two people, for example a pop-up to ask both users if they want to enter this conversation that has notification for screenshots and/or screenshot blocking). If both are not in agreement on this, this would allow for the users to decide not to remain in the conversation.

Importantly, a **warning or note should be provided** to users to inform them that screenshots can still be obtained by perpetrators in other ways to avoid a 'false sense of security'.

For notifications-only options, although this may mean that the conversation or photo has already been captured, the **notification of an unsolicited or non-consensual screenshot** can provide the user with an indication that the **person they are in contact with may not be trustworthy** and could help prevent further information being shared.

On chat-based apps

> *'For me, it is above all the removal of the screenshot, in order to protect the privacy of conversations between app users and to prevent messages and photos from being shared with a greater number of people.'*
>
> – Interviewee in Morocco

Here, the needed options and issues are very similar to dating apps around the capturing of conversation content such as voice notes, photos, videos, or messages. These issues with chat-based apps like WhatsApp, Telegram, Signal, and Facebook Messenger were raised many times in our interviews, with some mentioning options on some apps that have helped towards this harm reduction measure on certain apps:

> *'They can delete the screenshot option, like on Signal even on a video call you can't record screen or screenshot.'*
>
> – Interviewee in Tunisia

> *'For example, you can send a picture on WhatsApp that won't last a lot, which is good and helpful since you can't screenshot them.'*
>
> – Interviewee in Tunisia in reference to 'view once' photos or videos)

On messenger apps, the issue of downloading and forwarding the audio and visual content is very much interlinked not only to arrests and criminalisation but also to mass outings (see Part II).

> *'First thing, screenshots, they need to do something so that people cannot do that anymore, sometimes you are living your life and then find your face in group and group chats!'*
>
> – Interviewee in Morocco

However, default banning of screenshotting will not always make sense or be appropriate for all chats on chat-based apps and their threat model and issues.

The consent element here is very important and thus these options should be thought about with granularity.

In private chats

- For chat-based apps, it is, in fact, more important to allow for the **option to turn on screenshotting and/or blocking of screenshots in chats** so both or all parties are aware and in agreement in this set-up for the chat – similar to how this is transparent and set up in chats with disappearing messages.
- This should be **combined with timed/unsend messages and photos** (see above).

These can be for the most sensitive chats or groups rather than a default for all chats. It is especially important as many in high-risk contexts are moving to chat-based apps as dating app options due to the risks of dating apps.

Again, we suggest a **pop-up to ask both users if they want to enter this conversation that has notifications for screenshots and/or screenshot blocking**.

> *'We should be informed if someone shares your photos, videos, publications and especially your voice memos or takes a screenshot.'*
>
> – Interviewee in Algeria

In the case of downloading and sharing of data, this must be combined with an **option to add** (e.g. using a 'tick' or 'enabled' for uploading) to provide for whether they allow/consent for what they are sharing to be downloaded or shared by the receiver(s). Only then should the sender be allowed to share.

> *'Screenshots on profiles and discussions must be done with permission of the other.'*
>
> – Interviewee in Algeria

On Telegram's Secret Chat feature, there is an element of this where messages and content cannot be forwarded or screenshotted.

Importantly, a **warning or note should be provided** to users to inform them that screenshots can still be obtained by perpetrators in other ways to avoid a 'false sense of security'.

On social media apps

Many of our participants and interviewees raised the issue of screenshots, especially regarding social media. Many of these needs are very similar to the previous options.

Profile

- We suggest the same methods for the protection of profile data and not allowing them to be downloaded or screenshotted as mentioned with dating apps.

In private chats

- We suggest the same methods for the protection of private chats as mentioned with chat-based apps.

Several people mention their reliance on **Snapchat** due to the screenshot blocking feature and the safety it provides.

> *'On **Facebook**, if someone will take a screenshot or screen recording something from my profile, I want to have a notification saying they did it, for me to know, like **Snapchat**. If anyone tries to take a screenshot, I will directly receive a notification letting me know about that.'*
>
> – Interviewee in Tunisia

> *'Like **Snapchat** when someone screenshots or records you (photo, video, discussion, voice memo) I would like this system to be used by **Facebook**.'*
>
> – Interviewee in Algeria

> *'Prohibit all captures of photos, videos, discussions especially on Facebook.'*
>
> – Interviewee in Algeria

> *'For **Facebook** for example, I would like them to remove the option to download other people's photos. Anyone who enters your profile can download your photo or screenshot it and use it. I don't know which platform has removed this option and if someone tries to take your photo even as a screenshot, we will inform you directly and tell you who it is.'*
>
> – Interviewee in Algeria

Considerations for this recommendation

The option of screenshot blocking is also not always foolproof as adversaries can use a second phone to capture data. However, harm reduction links to the idea that the changes are never going to be foolproof, but rather are intended to add an extra layer of difficulty to challenge the harms these actions can cause. Though many security experts have disregarded this highly needed feature as promoting 'a false sense of security', this is often unreflective of the knowledge and intelligence of affected communities. They are of course aware (and this has come up many times in our interviews and focus groups) of the ways around such features; however, they are asking for support in their security to add difficulty to the increasing prevalence of these issues. The popularity of these options when available show why. We have no documentation of harm that has occurred through capturing of data that was protected by screenshot blocking through other methods. This is also why adding notifications provides extra security.

Sending screenshot notifications can also be a double-edged sword as it can harm a user who is trying to document abuse. Thus, having multilayered and granular optionality for this should be adopted.

How can it be implemented? (for app developers and security teams)

Screenshot notifications

There are a number of ways you can do this. We recommend the following:

- On Android, as of OS 14, you can use a Screenshot Detection API. You can also use a ContentObserver to detect new screenshot images being created in the photo storage service.
- On iOS, the solution is as simple as using Notification Center to add an observer to 'UIApplication userDidTakeScreenshotNotification'.

If a notification is received while the app is the front, open app, then a user knows it is a screenshot of their app. This information could then be stored and transmitted to the remote user whose photo or personal information was displayed on the screen when the screenshot happened. Further advice for this can be obtained from our research and technical team. We can be contacted at afsaneh@de-center.net and MENA@article19.org.

Screenshot blocking

As stated above, there are operating system specific programming interfaces that can be utilised to receive a callback notice in the application code that a screenshot was captured. Requesting and receiving permission for these features can and should be required to use the app, and in fact, may not require the user to approve any extra permissions.

Blocking content downloading

Enabling of the ability for users to download content from the app to their local storage is a feature implemented by the application itself. The technical work is in adding more fine-grain control and granularity to when and how this capability can be used. This includes adding support for sending a notification into the chat when a piece of content has been downloaded, as well as blocking all downloading of content, if that is desired. In addition, implementing local data encryption ensures any content that is stored by the app cannot be copied or extracted using an external means, such as rooting or jailbreaking of the operating system.

We need options for video and photo blurring

Recommendation 23: Provide in-app video and photo blurring options.

Apps that provide photo and video capturing and sharing should implement features to allow users to auto blur or obfuscate faces/identifying features.

Context and research behind the recommendation

4 of our interviewees and 1 focus group directly asked for options to be able to alter photos and videos on the apps and for blurring so people can share intimate or risky photos without providing too many identifying features or too much information about themselves.

> *'I'd love tools for editing photos in messengers and dating apps. For example, let me easily crop my photos before sending them.'*
>
> – Interviewee in Iran

These types of obfuscations and blurring options can be the difference between sentencing and release of individuals. In our research, we have seen a number of cases and court proceedings where, due to blurry or unidentifiable images (photos or videos), charges had been dropped. For example, in Egypt an individual was reported based on an explicit queer video, but their face was not identifiable:

> *'I was taken to the Student Department, and it turned out that one of my colleagues at work made a report and sent the video to the police. I was interrogated and transferred to the Public Prosecution. I denied my connection with the video, and it was fortunate for me that my face was not clear in the video, and it was not recorded clearly. I was released!'*

How can it be implemented? (for app developers and security teams)

Guardian Project first implemented automatic face detection and blurring in their ObscuraCam app back in 2010, almost 15 years ago. This is not a difficult technical problem. However, there are a variety of approaches, complexities, and other things to consider.

- **Manual vs automatic:** Automatic face detection, which finds faces in an image without knowing 'who' the face is, is a privacy-respecting approach to simplifying the process of face blurring. Offering a first 'automatic' step to find faces, and then allowing the user to tap to add or remove faces to blur, is a possible positive user experience. Beyond faces, allowing the user to use their finger to quickly tap or move over an area to blur it is an effective solution.

- **Blurring vs pixelating vs redacting:** There are aesthetic differences between the type of visual changes you can make on an image to remove faces. Blurring provides a smudge version of the original pixels; pixelating uses larger blocks of colour based on the original pixel; and redaction uses fully black pixels. The most private and impossible to reverse is full redaction because all underlying pixels are removed. However, block pixelisation can also be made nearly impossible to reverse.

- **Photos are easy, video is hard(er):** Blurring an area of a still image is quite straightforward. Blurring moving video is much harder, especially if the targeted area to blur is moving in the frame. Given processing capabilities of modern mobile phones, it is technically possible to blur selected areas of videos fairly quickly. Tracking the selected area can be implemented both manually, allowing the user to drag their finger to follow along, or automatically, using a more advanced object detection system built on machine learning.

  - iOS offers the 'Vision' framework for object detection in video.

  - Android supports this through its Machine Learning Kit (ML Kit) library.

Apps like Signal have in-built blurring tools for photos as part of their image editor and we encourage more apps to implement them.

Dating app specific recommendations

We need stealthy notification sounds

Recommendation 24: Remove distinctive sounds and notifications for queer dating apps.

This recommendation is simple: dating apps, especially queer dating apps, should **remove their distinctive sounds and notifications**. These should be replaced with generalised notification sounds of general chat-based apps. Distinctive options can be provided, but they should not be the default of the app.

Context and research behind the recommendation

In some cases, the notification sound of particular dating apps which are distinctive have led to outing and further searches, as seen in Egypt, Lebanon, and Morocco.

5 interviewees and 2 of our focus groups mentioned the risks they faced due to the distinctive notification sound on the Grindr app:

> *'One time, I was sitting with friends and I received a notification. I forgot to remove the app, honestly. … When I received that notification one of my friends looked at me and was altered. He didn't say anything. He went to my friends and told them. They started offending me. They were saying: "You are gay", "You are a faggot", "Turns out you are dirty."'*
>
> – Interviewee in Lebanon

> *'Once someone heard the notifications in a public place, with a loud voice he said, "Who is the one who has Grindr?"'*
>
> – Interviewee in Sudan

> *'I found myself out of work because of a sound!'*
>
> – Interviewee in Morocco

We are aware that other dating apps also have unique notification sounds. Grindr is the most well known and currently most likely to cause risk. However, this is a present risk with all queer dating apps – apps can out individuals if an app creates unique notification sounds by default that a user may forget to snooze.

How can it be implemented? (for app developers and security teams)

As mentioned, this recommendation is simple: dating apps, especially queer dating apps, should **remove their distinctive sounds and notifications**. We do not think this requires further implementation details.

We need accessible incognito modes

Recommendation 25: Dating apps should have an option for incognito mode (a mode only to be seen by people who the user has verified).

Dating apps should provide a feature that **allows a user to explore the app without exposing their identity**. This should only be seen by those they have 'liked' or 'swiped right' on as an extra layer of security so that they are not easily visible for all users on the app. This is seen as an extra vetting process for highly at-risk users especially.

Context and research behind the recommendation

The 'incognito mode' is an important feature available on a number of dating apps such as Bumble, OkCupid, Tinder, and Feeld. This effectively means no one will see a user's profile unless they 'liked', 'swiped right', or accepted them. Only profiles that have been OK'd/accepted are able to see the profile. This would allow many queer users in the region to have dating app accounts that are only seen by people users want to be seen by, reducing levels of unsuspected monitoring by state and non-state actors. It would also lower the chances of being approached for a connection to their profiles by fake accounts for scams, entrapment, extortion, or outings.

There have been an increasing number of homophobic 'outings' using social media and dating apps based on the patrolling and manual surveillance of the apps (see Part II). In many of the countries studied, these profiles are targeted by extended entrapment operations that use messages against them in criminal proceedings. Screenshots of people just on the apps are also used against them. On social media, police are known to patrol profiles, add unsuspecting users, and entice them into meetings before arresting or further abusing them.

These risks can be lowered with the incognito-type option to only be seen by profiles an individual has consented to being seen by. Here, the most vulnerable users can have an option to protect themselves and only talk to and be seen by people they have accepted and ‘liked’. This option **should be 100% free and not a paid feature**, especially in high-risk contexts.

How can it be implemented? (for app developers and security teams)

The current industry standard implementation of this feature is already very versatile and implementable, so we do not need to go further into implementation details. However, when implemented it should be a free feature and seen as a vital safety mechanism, especially for at-risk and vulnerable users.

We need in-app video and voice calls

Recommendation 26: Dating apps should have in-built video and voice call options.

Dating apps should allow for safe and private video and voice calls in-app so users do not have to share links to their external profiles to communicate and create further trust with their connections.

Context and research behind the recommendation

In **4 of the 6 countries** where we held focus groups and in **4 interviews**, our participants and interviewees asked that dating apps **implement video and voice calling options in-app**. This would help them avoid providing extra information to their chat-based apps and/or social media before having more options to verify an individual as real on videos and voice calls. These features should of course also be privacy respecting and end-to-end encrypted by default.

> *'I like that there is now the option of voice messages. Because I can tell if a person is good or bad just from the voice. And same goes for the option of video.'*
>
> – Interviewee in Morocco

> *'Receiving the voice message reassures me that at least this is a girl I am talking to. At least I don't need to be worried about [them] being a guy.'*
>
> – Interviewee in Iran

> *'The fact that you can have a video call without a phone number or personal information to see the face of the other person and go on your date with confidence can help a lot. … This video chat should be very safe so that the information is not saved or shared anywhere, and be end-to-end encrypted like Signal.'*
>
> – Interviewee in Iran

> *'In Badoo, it is possible to ask someone for a video chat right away. This is very good.'*
>
> – Interviewee in Iran

How can it be implemented? (for app developers and security teams)

The current industry standard implementation of this feature is already very versatile and implementable, so we do not need to go further into implementation details. However, when implemented, immediate care should be taken to ensure end-to-end encryption by default and options for wiping metadata from any media stored.

We need more in-app advice and notices

Recommendation 27: Dating apps should provide contextualised information

Dating apps should provide contextualised and informative **prompts, messages, or notifications that remind users about privacy issues and risks**, as well as needed information and advice. This should be done with local organisations and experts in regions and countries, especially in the most high-risk contexts.

Context and research behind the recommendation

There is a need for prompts, messages, or notifications that remind users about privacy issues and the risks they face in identifying themselves or their locations in their country of activity. These would be short messages – avoiding alarmism – to users which provide information on how to avoid putting themselves at risk of arrest, harassment, or other forms of harm from homophobic entities.

***18 out of 93** (19%) interviewees asked for advice and messages to be provided to users containing contextualised information.*

Areas of risk to users, including entrapment, cannot be mitigated solely through app or local organisation interventions. However, sharing necessary information can be of vital support.

Users are already savvy at manoeuvring around dangers they may face, but providing essential information will ensure that they have all the tools they need to continue to enjoy dating apps more securely. Therefore, **disseminating up-to-date and contextualised information on legal, digital, and health matters** – created or sourced directly from local groups – through the apps will be of fundamental value.

Sharing essential and relevant sexual health information is also important: in some of these countries, this information can be difficult to obtain due to the risks of being open about one's sexuality. This means LGBTQI+ dating apps have an important role in helping to close that gap. Also, importantly, research conducted and input from local organisations confirm that such information is not only needed, but desired by users of the apps.

Many dating apps provide some sort of warnings; however, they are sporadic and often not contextualised or updated, thus many find them unhelpful even though this feature is one they want. Making such messaging contextual and relevant to the country is the most helpful approach. It is important to note that in countries like the ones in our research, **some such messaging has already been successfully provided on dating** – but it has been led by local groups who have limited capacity and resources to maintain their campaigns.

Special priority should be given to countries with laws criminalising LGBTQI+ people. Due to the sensitivities and variance of contexts in each country, the informative materials and/or messaging **need to be created in collaboration with experts and local groups who work directly on these issues**. At present, the information most needed is about legal, digital, and sexual health, as we have seen overwhelmingly in our research. This includes applicable advice for arrests and detentions, digital security advice based on harm reduction when using apps, and sexual health advice. The information should contain details of trusted local groups which can provide support.

> *'Like on Grindr they send messages saying you are in a country where you are not safe.'*
>
> – Interviewee in Tunisia

> *'I know Grindr for a while, they worked with Mawjoudin, and you can get the emergency phone number as a message when you use the app. Tinder: when you put your identity in the settings and who you are interested in, according to that they give you a warning and also show a message about the queer community and the laws in the country where you are located.'*
>
> – Interviewee in Tunisia

> *'Grindr has made a fundamental change, for example, sending awareness messages and sharing news about the queer community, safety messages, and tips that help make you remain safe while you are on the app or in reality.'*
>
> – Interviewee in Sudan

How can it be implemented? (for app developers and security teams)

This is an easy recommendation to implement. The most important part will be **collaboration with local groups and organisations familiar with the exact situation** on the ground.

The text for such messaging can be curated directly with local groups that are aware of the risks for LGBTQI+ persons in each country. These can simply come in the form of **introductory slides** of the app when first installed, or in **notifications** on the app when the app is opened. Alternatively, they can be provided at regular intervals for countries where same-sex relations are criminalised or where LGBTQI+ persons are highly targeted.

For users to have an option to look further into the details of the issues and the actions they can take, it will be important to link the messaging to further information. This **messaging should avoid being alarming and be simple**, but it must remain **contextual and relevant to the country** concerned.

There is a two-part system for this. The first is regularly scheduled safety messages – some countries get them daily while others get them weekly – as well as (second) a holistic security guide. Language access is also critical for both of these.

Further advice for this can be obtained from our research and technical team. We can be contacted directly at afsaneh@de-center.net and MENA@article19.org.

Chat-based app specific recommendations

We need options to ‘snooze’ notifications

Recommendation 28: Chat-based apps need options to allow notifications to be paused for pre-set periods.

Chat-based apps should allow for **more options for timed snoozing of notifications** that allow for customisations and granularity of options.

Context and research behind the recommendation

Allowing for options to not only turn off notifications, but also have timed snoozing of notifications, would be very useful and practical for chat-based apps. Due to how often people check their chat-based apps, it is more likely they would not want the notifications turned off fully. However, having an option to snooze or pause notifications for a period of time will be an added layer of safety – for example, a period of three hours when an individual will be at a protest and know they will be at risk. One of our interviewees had been outed by a notification, arrested, and detained. They asked for a feature to help with this type of situation.

> *‘I want WhatsApp to add a feature to stop receiving messages for a predefined time. … [I was] at the entrance to Tahrir Square and [they] forced me to open and searched my phone, but they did not find anything on it. When the police secretary wanted to return my phone and ask me to leave, something happened … one of the people I knew through the Grindr application, and we exchanged WhatsApp numbers, sent me a message containing sexual suggestions, and then pictures of his penis.’*
>
> – Interviewee in Egypt

This interviewee was then arrested and detained for days. He was only released when he convinced interrogators that he did not know the account messaging him.

Signal Messenger currently has support for this feature through its ‘Mute’ feature for each conversation group, with 1 hour, 8 hours, 1 day, 7 days, or ‘always’ options. In addition, both Android and iOS now support various kinds of ‘do not disturb’ or ‘focus’ features for hiding or muting notifications.

Social media app specific recommendations

We need incognito modes on social media

Recommendation 29: Social media apps should have an option for incognito mode (a mode only to be seen by people a user wants to be seen by).

Social media apps should have **options for users to make their profiles only visible or searchable to profiles they have OK'd/accepted**, similar to incognito options on some dating apps.

Context and research behind the recommendation

Our focus groups in Egypt and Sudan pointed to a need for an ‘incognito mode’ (linked to Recommendation 14 and activity privacy). This is transferable to social media as it is currently seen on dating apps only. Unfortunately, this is often a paid subscription feature on a number of dating apps rather than a free vital feature. It effectively means no one will see a user’s profile unless they ‘swipe right’ or accept them. This is seen on Bumble, OkCupid, Tinder, and Feeld. Only profiles that have been OK’d are able to see the profile. If this verification and user control model on visibility is implemented, it would allow for many queer users in the region to have social media accounts that are reflective of their true lives without worry of unsuspected monitoring by state and non-state actors. It would lower the chances of being approached for a connection to their profiles by fake accounts for scams, entrapment, extortion, or outings.

There have been an increasing number of homophobic ‘outings’ using social media and dating apps (see Part II). In many of the countries in this research, these profiles are targeted by extended entrapment operations that use messages against them in criminal proceedings. Screenshots of people just on the apps are also used against them. On social media, police are known to patrol profiles, add unsuspecting users, and entice them into meetings before arresting or further abusing them.

These risks can be lowered with the incognito-type option, which allows profiles to only be seen by those an individual has consented to being seen by. Here, the most vulnerable users can have an option to protect themselves and only talk to and be seen by people they have accepted and 'liked'.

This feature is a new and emerging request for social media, thus more research should be done with impacted communities and technical experts to fine-tune its application and implementation.

How can it be implemented? (for app developers and security teams)

The current industry standard implementation of this feature by dating apps is already very versatile and implementable, so we do not need to go further into implementation details. However, when implemented it should be a free feature and seen as a vital safety mechanism, especially for at-risk and vulnerable users.

Operating system specific recommendations

It would be hugely beneficial if many of the previous recommendations were also made available on the operating systems. This would also support advocacy groups such as ours at ARTICLE 19 and De|Center as we push for these changes, knowing we cannot reach all of the implicated companies. There are very few teams from the operating system providers that are engaging with these conversations around human rights and harm reduction needs. However, changes to the operating systems can be some of the most needed and fundamental.

We urge more operating system providers to understand the harms befalling highly marginalised communities – and to provide support in implementing some of the previous recommendations on the operating system side.

We need options to hide our apps and folders

Recommendation 30: Operating systems must have device-level ‘stealth mode’ or ‘cloaking’ for apps and folders.

Operating systems must provide options for users to have **stealthy cloaking/hiding features**. They must also provide ways to **hide sensitive apps and folders** on the device with added safety measures and user controls.

Context and research behind the recommendation

Implementing cloaking options and ways to hide certain apps and folders would be a massive safety support from the operating system providers. The reliance on photos and videos as evidence against individuals for ‘crimes’ of queerness – as well as for other forms of harassment – have been outlined in Part II:

> ***Thirteen [out of 93** (14%)] of our interviewees who had been arrested directly mentioned the search and identification of photos on their devices, either from the galleries of the devices or collected from chats within different apps. **Four other interviewees** also mentioned videos.*

> *Photos, for the police, investigators, and other state actors, are seen as 'hard evidence' and proof of sexuality. Photos or videos of individuals engaged in intimate acts that are considered to be criminal by the prosecution teams are the most damaging. In **10 of these cases**, the photos were used as evidence against them in court cases or for adding additional charges.*

We see here the role devices in general and as such their operating systems play in the risks against the community. We have also seen that when a device search occurs, the photo gallery is one or the most searched folders of the device, along with other apps. Thus, allowing for obfuscation and app cloaking options for sensitive folders and apps (see Recommendation 18) on an operating system level can be fundamentally important – and add a huge layer of harm reduction and safety when a device falls into adversarial hands.

How can it be implemented? (for app developers and security teams)

Here, we recommend **operating system teams develop app and folder obfuscation options** and features similar to our app cloaking/discreet app icons feature outlined earlier in Recommendation 18. In June 2024, Apple's iOS launched new features for hiding and locking apps with its iOS 18. We are very encouraged by these first steps and see them as part of the success of our teams' advocacy on this issue. However, they are interlinked with FaceID biometrics that make the features both less protective of privacy and more unsafe (see outline of biometrics risks, including risks of police violence, under Recommendation 20). For robust implementation of these features by those who most need them, our guides for implementation of app locks (in Recommendation 18) should be used along with guidance to avoid biometric-based authentication.

Linked to this feature, we also recommend that **operating system developers enable photo metadata removal options** for photos taken and saved in the galleries. See Samsung's Secure Folder capabilities as an example.

We need stealthy self-destruct or panic buttons

Recommendation 31: Operating systems should have a device-level and stealthy self-destruct or panic button option.

Operating systems should provide methods for users to have an **easy-to-use option/button that triggers the commencement of needed safety measures**. It would block access to the content of the device should it fall into the hands of an adversary who may have physical control of the device.

Context and research behind the recommendation

The number and layers of data and apps on a device that can be detrimental to the safety of individuals when their phone falls under the control of law enforcement and other adversarial actors can be high. Therefore, self-destruct/panic button features at an operating system level will be one of the safest and most important methods to support individuals who may have 'incriminating data' about their identity in many areas of their phone.

> *'It doesn't matter how and in which app, everything should be deleted by one app. And it should be just one button and it should be very fast.'*
>
> – Interviewee in Iran

Recent updates of Android and iOS offer Emergency SOS features to trigger alerts to government emergency services, which is appreciated. However, many users, including those who have been the focus of our research, face the most risk from state and law enforcement.

Samsung does allow a device-wide 'secure folder' feature that can be used to hide entire apps and their data.

How can it be implemented? (for app developers and security teams)

Here, we recommend **operating system teams develop panic button features** similar to our panic button feature outlined in Recommendation 19. Specifically, the existing Emergency SOS features available in Android and iOS should be **expanded to support a configurable set of actions beyond the current limit of contacting governmental emergency services**. Similar to the approach of Guardian Project's PanicKit, a user should be able to **choose from a set of available configured actions** they would like to take. This approach would make it simpler for app developers to integrate into the operating system-level API and then for the user to decide how to use the feature.

We need safety features for our phone contact lists

Recommendation 32: Operating systems should make certain contact lists or contents hidden.

Contact lists can cause many risks and expose networks and family members. Thus, operating systems should provide ways to **hide or obfuscate certain contacts** so they are not easily findable in forced device searches.

Context and research behind the recommendation

> ***7 out of 93*** *(8%) interviewees explained how during device searches and inspections, officers had called people on their contact lists or in their messages to gain more information.*

While it might seem inconsequential, having in-built options to make certain contact list entries more private can be very helpful in situations when a device is confiscated and searched for the networks of the arrestee. Often, the officers look for particular contacts such as their parents, or simply the last people with whom they communicated (see Part II under ‘Device contacts lists’).

Endnotes

[1] For example, these exact patterns are being documented across sub-Saharan Africa by AccessNow.

[2] For the methodology for the gathered data and how to understand the data from the interviews, focus groups, and surveys, please see the methodology section in Part II.

[3] This work, with Afsaneh Rigot and ARTICLE 19, pushes changes such as Grindr's numerous safety changes, including discreet app icons, unsend and disappearing messages, PINs, and direct lines of communications. The same team also gained changes to WhatsApp's disappearing messages, and the introduction of their Lock Chat feature and its added privacy features (though we are waiting for its detangling from biometrics for it to be fully aligned). This also led to Signal's app icon changes. Many other changes and shifts have happened that we will not be adding here for privacy reasons. However, these were all possible due to the time, expertise, and direction from the impacted communities. All of the work was continuously based in the belief that we must Design From the Margins.

[4] ARTICLE 19's Senior Researcher from 2015 to 2023 and Founder and Principal Researcher at the De|Center: https://www.de-center.net/

[5] Norman Shamas can be contacted at Internet of Post-Colonial Thoughts.

[6] See e.g. Amnesty International (2019) Surveillance giants: How the business model of Google and Facebook threatens human rights; ARTICLE 19 (2021) Watching the watchmen: Content moderation, governance, and freedom of expression.

[7] For example, the detrimental impacts of company efforts to combat child sexual abuse imagery (CSAM) by reducing privacy and impacting LGBTQI+ communities in the name safety. See York, J.C. (2021) How LGBTQ+ Content is Censored Under the Guise of 'Sexually Explicit', Electronic Frontier Foundation, 18 August.

[8] Rigot, A. and ARTICLE 19 (2022) Digital Crime Scenes: The Role of Digital Evidence in the Persecution of LGBTQ People in Egypt, Lebanon, and Tunisia, p. 121, on the prevalence of WhosHere for entrapments and their lack of safety measures.

[9] Many dating apps that only allowed Facebook for authentication have moved to alternative methods at the request of their users. The new option is often to authenticate with a phone number, and most dating apps do not provide clear information on who owned the SMS authentication and verification service. When Facebook offered the now deprecated AccountKit as an SMS authentication service, dating apps like Tinder advertised it as an option to log in without Facebook, but did not adequately notify their users that AccountKit was a Facebook service. In the experience of the research team, most participants in talks and discussion groups related to the app research were unaware that Tinder's SMS authentication service was operated by Facebook and would not have used the service if they had been aware of that.

[10] Researchers have also begun to explore how gamification can be used for authentication; see also GitHub, SeedQuest: A 3D Mnemonic Game for Key Recovery.

[11] The communication between the application and the platform should be secure and should limit access to sensitive information. These limitations should follow the basic principle of not exposing more information than is strictly necessary. This should include using proper best practices around: application security should be built into the software development lifecycle. This typically includes threat-modelling, static analysis, dynamic analysis, and security testing. While it can take time to develop a robust security program, the OWASP Top 10 and CIS Cybersecurity Best Practices provide great resources to detect and remedy common security issues. Further advice for this can be obtained from our research and technical team.

Ideally, no information about user profiles should be available without having logged in. For example, while the back-end platform may provide images of nearby profiles to a non-logged-in user, it should not make available usernames or location information, or even rough estimations of such information. When this internal communications request is made from the application to the platform, displayed profiles should not be based on precise locations when there is not a logged-in user, and should instead offer individuals within a larger area.

The application should also put in place, at the least, basic monitoring of the requests made by IP addresses and accounts in order to detect abuse of the platform. Developers should also have logical limitations on such requests (i.e. someone can't move around a city in seconds, which would be used to triangulate another user). Further advice for this can be obtained from our research and technical team.

[12] US National Institute of Standards and Technology (NIST) and other security standards recognise this risk and have taken varying stances. Back in 2016, NIST deprecated SMS from the standard for use as an out-of-band (e.g. multifactor authentication) identifier. But the most recent version of the same standard allows for use of SMS as long as it is associated with the Public Switched Telephone Network and not Voice over Internet Protocol (VoIP). Even though this is a recommendation from security standards, we do not think it is the best approach because of the effects of blocking VoIP numbers. See 'We do not want to give you our phone numbers'.

[13] Such as: 1) key engagement/activities is a good way to help block bots and try to ensure 'real' people verification, or 2) verification through other app networks is about helping verify inclusion within a specific community.

[14] Researchers are exploring how gamification can be used for authentication. See also Ebbers, F. and Brune, P. (2016) The Authentication Game – Secure User Authentication by Gamification?; Kroeze, C. and Olivier, M.S., Gamifying Authentication; GitHub, SeedQuest. Even though this gamification research has focused on authentication and not verification, there are likely creative approaches that can be applied to verification as well.

[15] See Ahwaa: Serving the Arab LGBTQ community; and a detailed report: IdeasTed.com, The Smart Strategy That One LGBTQ Forum Uses to Keep Out Trolls And Bullies.

[16] Through Supporting Internet Freedom Worldwide or a similar mechanism.

[17] From the research briefing provided between 2022 and 2023 to ARTICLE 19 by Iran expert Khosro Isfahani on political, social, and legal issues of Iran.

[18] This includes as part of algorithms.

[19] From reports to ARTICLE 19 in 2023.

[20] Williamson, B. (2023) What Is a VoIP Number?

[21] NIST and other security standards recognise this risk and have taken varying stances. Even though this is a recommendation from security standards, we do not think it is the best

approach because of the effects of blocking VoIP numbers. See Part II, especially the section 'Risk of using phone numbers'.

[22] Through Supporting Internet Freedom Worldwide.

[23] Knight First Amendment Institute (2023) Artur Pericles Monteiro; Anonymity, Identity, and Lies; Electronic Frontier Foundation, Anonymity.

[24] Rigot, A. (2021) Why Online Anonymity Matters; and Electronic Frontier Foundation (2023) The Growing Threat of Cybercrime Law Abuse: LGBTQ+ Rights in MENA and the UN Cybercrime Draft Convention.

[25] As seen recently on Bumble: Heaton, R. (2021) Vulnerability in Bumble dating app reveals any user's exact location.

[26] Bumble provides both the name of the city and distance in its matching. This can be used to narrow the search range when determining a user's location. For example, if there is a match for a user in Upper Manhattan (New York), knowing if the user is in Manhattan, Brooklyn, Queens, Bronx, or another city (e.g. Jersey City) would reduce the potential area that a user is located and make it easier to identify their location.

[27] Hendrix, J., Quintin, C., Sinders, C., Wagner, L., Bernard, T., and Mehta, A. (2023) What is Secure? Analysis of Popular Messaging Apps.

[28] A prominent example of this is when Apple announced plans to scan photos on devices for unlawful material. Many prominent privacy advocates rightfully critiqued this approach and highlighted how implementing scanning on devices breaks other privacy guaranties, such as end-to-end encryption. See Portnoy, E. (2019) Why Adding Client-Side Scanning Breaks End-To-End Encryption.

[29] See, for example, ARTICLE 19, Content moderation and freedom of expression handbook; Business for Social Responsibility, A human rights-based approach to content governance; Human Rights Watch, Questions and answers: Facebook, Instagram, and digital targeting of LGBT people in MENA.

[30] Stop Silencing Palestine, Tell Meta: Stop Silencing Palestine.

[31] Sophos News (2020) Facebook Messenger may ban mass-forwarding of messages, 24 March; Hern, A. (2020) WhatsApp to impose new limit on forwarding to fight fake news, *Guardian*, 7 April.

[32] Android O has been developing adaptive icons. See: Android Developers, Adaptive icons.

[33] Biometrics should not be allowed when app cloaking is enabled under the assumption that additional security is needed.

[34] Note from our experts: For Progressive Web Apps (aka web apps that look like native apps and that can install a launcher icon on your phone), you can potentially have a custom icon and app name generated dynamically for each user. There is also the potential of releasing a native app with a generic icon ('settings' gear) and having it create fully unique and dynamic shortcut link icons or home screen widgets. There are app shortcuts that are fully dynamic on Android. See Petrotta, M. (2013) How can I add my application's shortcut to the homescreen upon app installation?.

[35] Regarding how to prevent Android from taking a screenshot when the app goes to the background, see Sting Ray (2012) How do I prevent Android taking a screenshot when my app goes to the background?.

[36] In Part II, under 'Risks taken to avoid providing access to devices'.

[37] As seen in Part II, under 'Risks taken to avoid providing access to devices'.

[38] Also refer to 1Password's Travel Mode: 1Password Support (2024) Use Travel Mode to remove vaults from your devices when you travel.

[39] See Guardian Project: Ripple; Google Play (2016) Ripple: respond when panicking.

[40] WhatsApp (2023) Introducing Secret Code for Chat Lock; *Filo News* (2023) Novedades en WhatsApp: la aplicación lanza una opción para proteger chats usando un código secreto.

[41] Elsa (2023) Telegram Secret Chat: Everything Parents Should Know, Airdroid, 28 April.

[42] Note: If biometrics are allowed, a warning about their limitations should pop up when enabled. Further, biometrics should not be allowed when app cloaking is also enabled under the assumption that additional security is needed.

[43] This is when you undo a message after it is sent. This is often available only after a short period between sending a message.

[44] Redaction of events by the user/client: Matrix Specification, Client-Server API and also moderator: Matrix (2023) Community Moderation.