



January 22, 2023

Ms Lama Fakih
Director, Middle East and North Africa
Human Rights Watch
350 Fifth Avenue, 34th Floor
New York, NY 10118-3299

Dear Lama,

Thank you for your letter of January 9th and for letting us know about your upcoming campaign to protect the rights of LGBTQIA+ people in the MENA region.

We deeply appreciate the work of Human Rights Watch to document violations and advance human rights, often in the most challenging of circumstances. We've welcomed the opportunity to engage with you in 2023 in your reporting on how governmental security and police forces employ digital targeting as an adversarial tactic to support prosecutions and harassment of LGBTQIA+ people.

First, it's important to note that our [Human Rights Policy](#) states:

We pay particular attention to the rights and needs of users from groups or populations that may be at heightened risk of becoming vulnerable or marginalized. We are committed to identifying relevant such groups for each context, undertaking meaningful engagement to hear their hopes and concerns, and to protecting and promoting their rights when using our products.

In our policy, we also clearly recognize human rights defenders are a high-risk user group.

We strive to offer specific measures to protect their safety and well-being. On social media, these risks can include digital security risks; online attacks against individuals or groups; surveillance; and censorship demands from governments or their proxies. More importantly, these online risks may have the potential to lead to offline harms, including violence, arrest, and termination of employment. We proactively engage with human rights defenders to understand their needs and the heightened human rights risks

they face. We strive to offer specific measures to protect their safety and mitigate such risk

These are the foundational policy commitments that frame our work across Meta to foster safer online spaces for vulnerable groups, including LGBTQIA+ communities.

This work includes:

Our safety tools

The safety of our users is of utmost importance to us. We have and continue to develop a number of tools focused around increasing user safety and security on our platforms. Many of these tools are designed to protect communities and individuals who are most vulnerable and are driven by assessments on human rights risks. In particular:

- To increase privacy across the platform we are working towards rolling out End to End Encryption (E2EE) across all our messaging platforms. As such, we commissioned an [independent human rights impact assessment on](#) our plans to expand end-to-end encryption by default to Messenger and Instagram direct messages. This assessment found that expanding end-to-end encryption enables the realization of a diverse range of human rights and recommended a range of integrity and safety measures to address unintended adversities.

Specifically, the recommendations encourage us to look at marginalized communities around the world, who may benefit the most from end-to-end encryption and are often disproportionately affected by positive and adverse impacts. Rather than prioritizing rights or offsetting one right for another, we're advised to identify feasible, effective solutions that would address adverse impacts to maximize all rights.

We have committed to implementing many of the recommendations, some of which touch on the points in your letter including on improving reporting mechanisms and investments in user safety and education. We urge you to read our [response and commitments](#). You can also find more details about the roll-out of E2EE by default on Messenger [here](#).

- We're building on our [existing work to protect defenders' accounts](#) — efforts that include [combatting advanced threat actors](#) targeting them, protecting them from incorrect content removals using [Cross-Check](#), offering advanced [security options](#) such as [Facebook Protect](#), taking steps to thwart unauthorized access to the accounts of defenders who are arrested or

detained, and partnering with human rights organizations on outreach and training.¹

- In 2023, we took down more than 7 covert influence operations in [Türkiye](#), [Iran](#), Togo and [Burkina Faso](#) spanning hundreds of pages. You can also see more about our work on surveillance for hire, [here](#).

Our policies protecting vulnerable groups

We care very deeply about LGBTQIA+ communities and have built dedicated resources in our [Safety Center](#) that set out the tools, policies and partnerships we have in place to help create safe online spaces for these communities, which we seek to update regularly.

We also take a comprehensive approach to safety, designing policies, building safeguards, and developing resources and tools in partnership with key LGBTQIA+ safety and advocacy organizations around the world - including in MENA - to foster a safer online environment. Relevant policy areas in [Facebooks' Community Standards](#) and [Instagram Community Guidelines](#) include hate speech, bullying and harassment, and non-consensually shared intimate imagery among others, which cover some of the issues outlined in your report.²

Specifically:

- Under our [Coordinating Harm and Promoting Crime Policy](#), to prevent and disrupt offline harm we prohibit people from facilitating, organizing, promoting or admitting to certain criminal or harmful activities targeted at people, businesses, or property. This includes “outing,” which we define as content that exposes the identity or locations affiliated with anyone who is alleged to, among other things, be a member of an outing-risk group.
- As per our [Hate Speech Policy](#), Hate speech against women, LGBTQIA+ and all other protected groups is prohibited. We believe that people connect more freely when they don't feel attacked based on their identity. We define hate speech as a direct attack against people - rather than concepts or institutions - based on what we call our protected characteristics: race, ethnicity, national origin, disability, religious affiliation, caste, sexual orientation, sex, gender identity and serious disease. We know hate speech constantly evolves. That's why we partner with experts and organizations to stay ahead of trends and keep our policies current. We also ban the use of harmful stereotypes —

¹ See page 75 of our [2021 Annual Human Rights Report](#).

² See page 40 of our 2022 [Annual Human Rights Report](#).

defined by us as dehumanizing comparisons that have historically been used to attack, intimidate or exclude specific groups.

- Under our [Bullying and Harassment Policy](#), we do not tolerate bullying behavior. Bullying and harassment take varying forms, including threatening messages, unwanted malicious contact and the release of personal information. Meta views public figures and private individuals differently to allow discussion. For public figures, Meta removes posts that use derogatory terms, call for sexual assault or exploitation, call for mass harassment or threaten to release private information. For human rights defenders, journalists, and private individuals, Meta's protection goes further. We remove content that's meant to degrade or shame someone for their sexual orientation or gender identity, among other protections. You can find out more details and relevant tools in the relevant section of our multilingual [Safety Center](#).
- The sharing of [non-consensual intimate images](#) violates our policies, as do threats to share those images. We remove images shared on Facebook and Instagram in revenge or without permission, as well as photos or videos depicting incidents of sexual violence. We also remove content that threatens or promotes sexual violence or exploitation. We encourage people to report when someone shares your intimate images without your consent or is threatening to do so. We encourage people to report sextortion and are participating in an important relevant multi-stakeholder initiative, [StopNCII.org](#).

The development of these policies, tools and resources has benefited from our teams' extensive stakeholder engagement with over 850 safety partners around the world, including in the MENA region, who are experts in online safety issues, including LGBTQIA+ advocacy and crisis support organizations. The inclusion of specific LGBTQIA+ risks in due diligence, for example from our [Philippines](#) human rights due diligence, has also contributed to the development of these policies and tools.

Our stakeholder engagement in the MENA region

We have specific commitments to help protect human rights defenders and vulnerable groups in Meta's [Corporate Human Rights Policy](#). We also seek to meaningfully engage with potentially affected groups and other stakeholders through our Stakeholder Engagement program, centered around [three core principles](#): inclusiveness, expertise and transparency.

In addition, our policy teams in MENA locally conduct regular engagements with women and LGBTQIA+ human rights defenders to understand their experiences on our platforms and listen to their feedback on products. Our concern for the wellbeing of vulnerable communities in MENA led to the creation of a team focused solely on engaging with high risk communities within our Africa, Middle East and Türkiye (AMET) public policy team called the Community Engagement and Advocacy (CEA) team.

For example, this year during the South African Pride Month, the CEA team launched a campaign centered on LGBTQIA+ activism aimed at elevating the visibility of the safety resources and casting a spotlight on Meta's newly relaunched [LGBTQ+ Safety Center](#), the [Stop NCII project](#), and the team's [online digital security training for activists](#). They collaborated with four prominent LGBTQIA+ activists in Sub-Saharan Africa (SSA) who have been using their social media platforms as a tool for movement-building, social justice campaigning, and spotlighting the injustices faced by their communities.

The [Defenders Safe](#) Website launched in December of 2023 is a comprehensive resource for human rights defenders in AMET including a [digital security toolkit](#) developed in partnership with the Jordan Open Source Association, the [Africa Human Rights Defenders Fund](#) with Africandefenders providing support to targeted activists, a [social advocacy campaign](#) with Devex showcasing the power of the global south to foster dialogue on digital rights.

They also developed a [lexicon](#) of homophobic terms and slurs that were being used colloquially in MENA to abuse the LGBTQIA+ community online. This lexicon was developed with civil society partners Helem and shared with our market experts to increase awareness and improve contextual understanding. We aim to build further on this work. We also have proactive detection of harmful content broadly, as described in our [Transparency Center](#), and specifically initiatives relevant to vulnerable MENA defenders and periods of heightened risk, such as Pride month.

In 2023, we initiated a series of roundtable discussions to help guide mitigation strategies related to harmful content at the regional level, including focused engagements in partnership with global human rights organizations among others. We also ran user experience engagements designed to feed into product design, specifically with the LGBTQIA+ community in Iran and the diaspora. This included roundtable product policy consultations with LGBTQIA+ individuals and organizations across Türkiye, Iran and Arabic speaking MENA countries that helped shape the language provided for non-binary pronouns available on Instagram in those languages when they launched in 2023.

Escalating harmful content

We've built the trusted partner programme to foster partnerships with civil society, strengthen the social media monitoring capacity of local organizations and improve our policies, enforcement processes and products to help keep users safe on our platforms. Our network of trusted partners includes over 400 non-governmental organizations, humanitarian agencies, human rights defenders and researchers from 113 countries around the globe which you can read about in our [Annual Human Rights Report](#). In MENA, we work with LGBTQIA+ focused Trusted Partners, among others, and hope to add more that can flag content to us and act as a bridge between us and local communities. In addition, our [StopNCII](#) program focuses on partnerships and enforcement mechanisms addressing non-consensual sharing of private images of a sexual nature together with more than 50 non-governmental organization partners around the world. This platform empowers people around the globe to proactively thwart the non-consensual sharing of their intimate images on participating tech companies' platforms, giving victims more control and security over their images. Our team and the AMET Community Engagement and Advocacy team ensure that we have an individualized and survivor centered approach to managing reports and escalations. We have regular contact with a number of high risk communities, including from the LGBTQIA+ community, with whom we seek regular feedback on how reporting systems and provide feedback on escalation where applicable.

We also offer human rights organizations direct channels to our teams, so they can flag issues to us quickly. We work with local and regional human rights organizations in this way to ensure local level coverage is provided. We offer additional specific measures to protect their safety and well-being, as HRW is aware, and we are happy to provide a confidential briefing if useful.

We also note that the report and this letter allude to content examples that, as of yet, we have not been able to directly assess, as they have not been forwarded to us by HRW. Without seeing the examples, it is not possible to analyze such content, or ascertain whether and how it might infringe upon our [Community Standards](#) or [Community Guidelines](#). We encourage you to share those examples with us and/or coordinate with the Access Now Digital Security Helpline in urgent cases.

Our content enforcement

In your letter, you request among other things specific data about our user safety investments, staffing levels, content moderation processes and automation in the Arabic language. While we understand and appreciate your interest, we do not provide breakdowns on the number of people 'working' on content from a particular country or in a particular language because that wouldn't accurately reflect the complex [enforcement](#) system we have in place. This system is made up of thousands of engineers building [the technology we use](#) to find and detect content proactively; thousands of content reviewers - some of which requires native language skills and

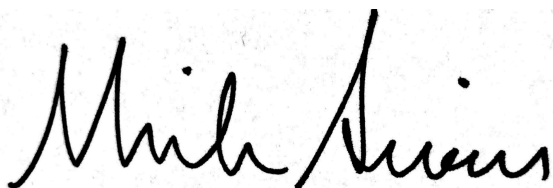
local cultural understanding, some of which (e.g. adult nudity violations, graphic violence violations) doesn't.

That being said, we have large and diverse teams to review content, some with native language skills and/or an understanding of the local cultural context. This includes teams of native Arabic speakers with the skills and the cultural context for a range of countries across the Middle East and North Africa. We strive to ensure coverage of our most spoken languages in the region such as Arabic, Farsi and Kurdish, and we strive to accommodate different dialect groups such as North African, Kurmanji and Sorani. We also have systems in place which use technology to help determine what language content is in, to help with this review process.

We regularly explore ways to improve our processes. For example, in Meta's [September 2023 Update](#) on our Israel Palestine human rights due diligence, we also explained that we had conducted analysis on building a dialect- specific Arabic classifier for detection of any content in that language and that we would add expanded language identification functionality to our systems that will be able to recognise content in different Arabic dialects. We are also reviewing the development of mechanisms to efficiently route content by Arabic dialect to improve the accuracy of Arabic content review and better prevent under- and over-enforcement issues.

Our continued engagement with HRW and other global and regional organizations on this topic has been fruitful; we hope to expand this work and deepen our collaboration with you on the issues outlined.

Yours sincerely,

A handwritten signature in black ink that reads "Miranda Sissons". The signature is written in a cursive, flowing style.

Miranda Sissons
Director, Human Rights Policy Team