FILED UNDER SEAL

EXHIBIT 9

Message

From: [redacted] [/O=THEFACEBOOK/OU=EXCHANGE ADMINISTRATIVE GROUP (FYDIBOHF23SPDLT)/CN=RECIPIENTS/CN[redacted]]
Sent: 2/25/2019 8:36:43 PM
To: Nick Clegg [[redacted]@fb.com]; Guy Rosen [[redacted]@fb.com]; Chris Sonderby [[redacted]@fb.com]
Subject: Note ahead of MZ E2EE meeting - for discussion this afternoon

Nick, Chris, Guy

Here is a note that policy have pulled together analyzing the impact of E2EE in Messenger on safety.

It's long and detailed, so feel free to focus on the points suggested for your discussion today and the TL;DR.

Some points are still moving around - there will be a final version of this later tonight.

[redacted]

Suggested goals for Guy/ Nick/ Chris pre- meeting on 2/25

Align on integrity risks, process and priorities for mitigations:

1. Agree major concerns:
    1. NCMEC reporting (and other detection that involves image detection, such as terrorist propaganda and non-consensual intimate imagery (revenge porn);
    2. Inability to proactively and reactively assist law enforcement in preventing harm;
    3. Inability to identify other policy violations and take action (including information operations, fake accounts, spam, malware)
2. Discuss how we ensure we are still reporting useful content to NCMEC:
    1. question to ask: how technically/ legally difficult is device-side hash analysis?
3. Discuss how we should respond to harm we would have previously detected after other signals would have pushed us to investigate Messenger contact? The options:
    1. Do nothing - this harm exists separate from Facebook and is not our problem.
    2. Try to prevent it at other pinch points (i.e. preventing victim identification for child grooming)?
    3. Push the burden onto law enforcement (i.e. give weaker signals to law enforcement, sharing more unverified information on potential school shooters, terrorists or suicide attempts with law enforcement).
    4. Push the burden onto users (ie developing very easy reporting mechanisms; pushing reporting mechanisms when there are off-Messenger signals).
    5. NB. we should be careful about believing metadata behavioral analysis will work across all harm types. Spam is far easier to spot, for example, than child groomers, sexploiters or terrorists plotters, due to the small sample sizes and overlap with typical friends behavior. Good behavioral analysis through encrypted metadata is only in its infancy.
4. Agree our work on mitigations should be data driven and focus on prevention, detection and helping users seek assistance for each different type of harm.
    1. We should have a data driven, systematic process for testing mitigations.
    2. Harm against children should be the starting point.

5. Agree we should have a comprehensive engagement process with external stakeholders to understand different perspectives and build legitimacy.
6. Discuss how and if we re-calibrate our relationship with law enforcement.
    1. Focus on joint work we can do to build safer communities, if we cannot provide evidence of crimes. This could include better community engagement, prevention campaigns, working with them on how metadata can be useful (including changing relevant laws). This is long term, and will not immediately solve the problem of evidence provision.
7. Discuss how much effort we should put into future efforts to detect bad behavior in Messenger.
    1. this is tough: behavioral pattern analysis is difficult for many threats that mirror normal friendship behavior, in particular grooming or conversations between child sex offenders or terrorists. This is even more difficult with only metadata. Should we be investing in additional AI? We are already doing some work on deep neural network analysis.

General

Tl;dr

- *Without robust mitigations, E2EE on Messenger will mean we are **significantly less able to prevent harm against children.** Whilst our other Community Standards Enforcement figures are unlikely to change, our community standard enforcement reporting on Child Nudity and Sexual Exploitation and public National Centre for Missing and Exploited Children (NCMEC) numbers will indicate we are keeping children less safe.*
    - ***Without mitigations on e2ee that enable us to credibly identify images of child sexual exploitation within Messenger, total reporting to NCMEC will drop (at least) from 18.4 million to 6 million (a 65% drop). Our community standard enforcement reporting on Child Nudity and Sexual Exploitation would fall by a similar percentage from circa 8.8 million to 3 million a quarter.***
    - ***This could trigger media reports of Facebook endangering children, a potential Congressional review in the U.S., and scrutiny in other countries where children's safety is high profile.***
- We will also be less able to spot harms related to the private sharing of other *known* images (i.e. where people tell us about them), including non consensual intimate imagery and terrorist propaganda:
    - our reporting on terrorist propaganda could fall by circa 1 million a quarter to 2 million violations identified . (This is an informed guestimate, as our technology and practice is constantly improving).
    - our ability to spot non consensual intimate imagery (revenge porn) would be reduced to 100,000s from 2,400,000.
- ***E2EE will severely impact our proactive and reactive work with law enforcement, including terrorism, child sexual exploitation, suicide and school shootings emergencies.** Depending on mitigations/exemptions, Facebook may not be able to identify or verify current or future harm suggested by other factors, including user or trusted third party reports, our automated detection systems or law enforcement requests.*
    - *Law enforcement, politicians, family members and the media will report harm cases that may not have occurred if we had not moved to e2ee.*
- ***Because we are able to assist less, Governments (particularly US, UK, Australia, Brazil and India) will increase existing pressure on us to provide technical assistance to work around encryption** (there is less pressure break E2EE per se, although that is being discussed by some).*

- *E2EE will severely impact our own ability to carry out further research in Messenger to confirm violations of our policies, after it is suggested by other factors, including automated signals on Facebook and Instagram, press reporting, user or trusted third party.*
    - *Identifying fake accounts, information operations and espionage violations relies heavily on reactive investigation of Messenger for additional signals.*
- *Implementing E2EE will also potentially create additional harms given the connection between FBM and Facebook. Given that perpetrators can identify and befriend victims on Facebook, before taking them easily into the secret world of E2EE, these additional risks are likely;*
    - *We are more likely to attract bad actors, who will know that Facebook will not be able identify or verify current or future harm. We know that users of Secret Conversation users are more likely to have violated policy than Messenger users, two times as likely for terrorism, dangerous org, or adult sexual exploitation violations.*
    - *We will not be able to use any improvement in our techniques on machine learning, behavioural analysis to get better at identifying sexploitation, bullying, hate speech etc in Messenger.*
- ***We should focus on mitigations to ensure that:***
    - ***we are still able to report to NCMEC;***
    - ***respond to harm we would have previously detected after other signals would have pushed us to investigate Messenger content;***
    - ***support law enforcement in its work.***
- ***Our work on mitigations should be data driven** and focus on prevention, detection and helping users seek assistance for each different type of harm. We should have a **systematic process for testing mitigations***
    - *Harm against children should be the starting point.*
- *We should have a**n engagement process with external stakeholders to understand different perspectives and build legitimacy.***
- *Although other messaging systems are undoubtedly used for harm, **we will be held to a higher standard**, due to the link with Facebook Blue app and our championing of our safety efforts.*
- *It is likely that the **negative press related to other Facebook issues** - including the pending FTC settlement - **could increase attention** on any product that shares data across services or that is perceived as consolidating our market position or increasing harm.*

Overarching principles that Mark is working towards

- **Mark's Company Priorities for 2019**
    - Continue making progress on the major social issues facing the internet and our company.
    - **Build new experiences that meaningfully improve people's lives and set the stage for even bigger improvements in the future.**
        - (UME2EE falls under this priority)
    - Keep building our business by supporting the millions of businesses - mostly small businesses - that rely on our services to grow and create jobs.
    - Communicate more transparently to make the case for the role our services play in the world.
- **UME2EE Mission**
    - **UM Mission:** To let people talk to anyone, anywhere, without needing to think about where to reach them.
    - **e2ee Mission:** The messaging world has been moving toward default e2e encryption and we're at risk of falling behind.

- **Other company principles that are relevant:** "Principles are what we stand for, what we will fight to provide for people, and what kind of community we want to build. They are beliefs we hold deeply and that we already make real tradeoffs to pursue."
    - **Keep people safe and protect privacy** (Internal Company Mission, Values and Principles)
        - *People try to use our services for good and bad, and we have a responsibility to promote the good and prevent harm. That's why we have the initiatives on counter-terrorism and self-harm – and 20,000 people working on security. And it's why, even though we care about giving people a voice, we take down a lot of content that is bullying, harassing, and attacking people.*
    - **Facebook's privacy principle**s (External source)
        - We give you control of your privacy
        - We help people understand how their data is used
        - We design privacy into our products from the outset
        - We work hard to keep your information secure
        - You own and can delete your information
        - Improvement is constant
        - We are accountable
    - **Facebook's community standard principles** (external https://www.facebook.com/communitystandards/)
        - **Safety:** People need to feel safe in order to build community. We are committed to removing content that encourages real-world harm, including (but not limited to) physical, financial, and emotional injury.
        - **Voice:** Our mission is all about embracing diverse views. We err on the side of allowing content, even when some find it objectionable, unless removing that content can prevent a specific harm. Moreover, at times we will allow content that might otherwise violate our standards if we feel that it is newsworthy, significant, or important to the public interest. We do this only after weighing the public interest value of the content against the risk of real-world harm.
        - **Equity:** Our community is global and diverse. Our policies may seem broad, but that is because we apply them consistently and fairly to a community that transcends regions, cultures, and languages. As a result, our Community Standards can sometimes appear less nuanced than we would like, leading to an outcome that is at odds with their underlying purpose. For that reason, in some cases, and when we are provided with additional context, we make a decision based on the spirit, rather than the letter, of the policy.

Detail

End to End Encryption

- **E2EE is likely to have a major impact on our integrity work, user safety and law enforcement relationship**
    - **Public numbers indicate we are keeping less people safe**:
        - The Department of Justice Office of Juvenile Justice and Delinquency, which funds the National Centre for Missing and Exploited Children (NCMEC), is required under US law to

make public regular (every three years) studies on missing and sexually exploited children.

- We are mandated by US law to report information related to child sexual exploitation (CSE) cases, including online enticement or "sextortion," child sex trafficking, child sexual molestation, and CSE images to NCMEC. Where relevant, we should include the CSE image.
    - NCMEC then has a legal responsibility to forward this information to the relevant law enforcement agency, nationally or internationally.
- We currently scan Messenger content for images of child sexual exploitation, using a hashing system is used by all major companies (this compiles a digital signature or fingerprint of images (known as a "hash"), which can be matched against a database of *known* child pornography images).
    - **In annual year 2018, we reported 12.4 million cases of *Messenge*r users exchanging CSE imagery to NCMEC. 12 million of these were detected by us using hashes, the rest were reported by users. Total NCMEC reporting for 2018 was 18.4 million.**
        - Without our ability to scan for hashes, NCMEC reporting would have been 6.4 million.
        - NCMEC would also be unable to report 12 million viable investigative leads on national and law enforcement, leading to the possible disruption or arrest of those involved in the exchange of CSE.
- We publicly report every quarter violations against our community standards in our enforcement report. ***Without our ability to proactively use machine learning and photo hash technology to scan Messenger content in Q3 2018 we would have missed*** the harm below. The list is in order of magnitude and there is a significant gap between identifying harms related to known images (non consensual intimate imagery, child sexual exploitation and terrorist propaganda) and identifying other harms automatically:
    - (not yet reported publicly: **2.39 million of 2.4m violations of our non consensual intimate imagery (revenge porn) (99%)** [Precise numbers to follow]**)**
        - We are very good at spotting violations using hash technology to spot images that have been shared with us by victims or potential victims. In Q3 2018, we found and flagged around 99% of all app content we subsequently removed after users reported it to us but before it was subsequently re-shared.
    - **5,850,388 of 8,815,671 violations of our child nudity and sexual exploitation policy (66.36%)**
        - We are very good at spotting violations using hash technology. In Q3 2018, we found and flagged around 99.2% of all app content we subsequently removed before users reported it to us.
    - **9625 of 2,997,261 violations of our policy on terrorist propaganda (0.32% BUT more accurate figure is closer to 41%)**
        - The Messenger deletions for GTO have not successfully been automated. Therefore, these are taken in bulk actions currently, and thus the Q3 "deletes" do not represent the "true volume" of GTO in Q3 Messenger (some deletes on Q3 detections were taken in Q4). A more accurate number would be the total of Q3 and Q4 Messenger deletes as a proportion of total deletes. **The combined metric for H2 2018 is 41% of deletes coming from Messenger images and videos.**

        - We have identified a number of large GTO trading networks in Messenger, and these actions (in Q3 and Q4) have included ~4000 account disables and at least 15k pieces of previously unbanked content.
    - **At least 540,031 cases of 1,231,119,671 violations of our spam policy (more than 0.04%)**
        - We take down bad URLs in all of our apps and in Q3 almost 100% of this was done through automated systems spotting the url.
        - The number of violations in Messenger is likely however to be somewhat higher than the current number as we missed counting some Messenger violations due to logging issue (i.e. if a url shared in a message without content, the automated logging system did not count it).
    - **6852 of 2,903,905 violations of our hate speech (0.24%)**
        - we largely take reactive actions on Messenger, and will also only take proactive actions if we see further validation.
        - Overall on hate speech, in Q3 2018, we only found around 51.6% of the content across the apps that we took action on before users reported it to us.
            - Technology has limitations.Our software is built with machine learning to recognize patterns, based on the violation type and local language. In some cases, our software hasn't been sufficiently trained to automatically detect violations at scale. Other
              violation types, such as hate speech or graphic violence, require us to
              understand context when we review reports and therefore require review
              by our trained teams. As emphasized above, it's important that people
              continue to report violations to help our enforcement efforts.
    - **5929 of 15,322,521 violations of our policy on violence and graphic content (0.04%);**
        - we largely take reactive actions on Messenger, and will also only take proactive actions if we see further validation.
    - **1196 of 30,902,886 violations of our adult nudity policy (0.0039%)**
        - we largely take reactive actions on Messenger, and will also only take proactive actions if we see further validation.
    - **less than 312,900 of 2,100,000 violations of our bullying and harassment policy (less than 14.9%)**
        - In Q3 2018 across all apps, we found and flagged 312,900 violations of our policy, *while searching for other types of violations*. We do not have data on how many of the 312,900 involved Messenger.
            - Bullying and harassment has a relatively low proactive rate due to the personal nature of the category. (This is reflected in our policies, which often require someone to report content before we can consider it a violation and take action.)We believe there is room for improvement here to optimize proactive detection when we can see the content.
        - We acted on the other 85.1% because users reported it first.

- **Law Enforcement, politicians, family members and the media report harm cases where we have not helped** (suicides, terrorism, child sexual exploitation, sexploitation and information operations) and we have no comeback on what we are doing to keep people safe, or to help others keep people safe:
  - The Facebook graph makes things worse: perpetrators can identify and befriend victims on Facebook, before taking them easily into the secret world of E2EE;
  - Facebook can't verify actual harm happening when other factors indicate we should do further research. Other factors would include user or trusted third party reports, our automated detection systems or law enforcement requests;
    - Without Messenger content that we could access, it is highly likely we could not have proactively alerted US and global law enforcement (via the FBI) to imminent risks to life and children:
      - 600 child exploitation cases referred directly to law enforcement in 2018, with 36 arrests;
        - in total in 2018, 1,496 victims were escalated to NCMEC; 1,454 of these escalation were for Sextortion.
      - 6 active terrorist cases involving 13 individuals for H2 2018
        - 98% of proactive terrorism referrals to law enforcement depended on MSGR content;
        - ***example of attack in Turkey***: FB referred a case to LE based on an individual's Messenger content; attacker was arrested in Istanbul; LE found a stockpile of explosives and believed user was days away from an attack; LE had no awareness of user before our report.
        - **example of Rotterdam arrest**: machine translation of Messenger content uncovered a planned purchase of guns and explosives and delivery to a suicide bomber; FBI responded within 15 minutes and Dutch authorities made an arrest that day.
      - [X] school shootings [will update when we have this number]
      - [X] potential suicides [will update when we have this number]
        - 60% of 2018 threatened suicide cases relied on Messenger data to verify that there was a threat;
    - E2EE will severely impacts on our own ability to carry out further research in Messenger to confirm violations of our policies, after it is suggested by other factors, including automated signals on Facebook and Instagram, press reporting, user or trusted third party.
      - Identifying fake accounts, information operations and espionage violations relies heavily on reactive investigation of Messenger for additional signals.
        - All of our 2018 20 Info Ops disruptions relied on Messenger content as a high confidence signal of coordination and a clear signal of campaign command and control. It also helps us quickly identified recidivists ("This is X, my other account is in FB jail so message here."). We may be able to learn from WhatsApp:
          - ***Example of Myanmar military content***: messenger content confirmed military personnel had command and control over distribution of violent content

                - ***Example of Myanmar inciting content:*** 9/11 jihad warning messages targeted Buddhist users to incite violence against Muslims were spread only via messenger
                - ***Example of attempt to influence 2019 Indonesian*** election by selling Fake IDs for authentication and account passwords for hacked pages through Messenger.
            - We will not be able to carry out additional checks for violations, meaning that if we were to rely on behavioral signals on Facebook and on E2EE we are more likely to miss harm and/ or take action against false positives.
    - Facebook also give enough data to law enforcement to enable them to get a warrant for further investigations/ arrests - we can no longer help them meet the 'necessity and proportionality' level in Europe or 'probable cause' in the US;
        - Due to US law restrictions, we currently provide stored and live time content data within the US (or through US partners) and non-content data outside of the US.(With the exception of Ireland, which has minimal requests.) We can also - exceptionally - provide content to non-US law enforcement where life or the welfare of a child is at risk.
            - In H1 2018, we responded to 23801 US Search Warrant requests for stored content, including Messenger content, providing data in 88% of cases; we responded to 261 Title III wiretap requests involving Messenger content, providing live time monitoring in 95% of cases.
                - This compares to 9396 subpoena requests for stored non-content and 4591 Pen Trap and Trace requests for live time non-content, including Messenger metadata.
            - [X] [will update when we have this number]of the 23801 requests originated from Mutual Legal Assistance, meaning this is content data that would have gone to foreign law enforcement to assist their investigations.
            - In H1 2018, we also responded to 3173 emergency cases in the US, where life or the welfare of a child was at risk (including 122 school shooting threats), providing some data in 85% of cases. These cases are highly likely to have included provision of Messenger content.
                - **Example of California school shooting**: FBI submitted an Emergency Request based on threats made on IG though proxy accounts; by reviewing user's Messenger content, we identified the person, his location, and his school.
            - In H1 2018, we also responded to 6000 emergency cases outside of the US, where life or the welfare of a child was at risk, providing some data in 79.8% of cases. Many of these included provision of Messenger content which we are legally permitted to do in emergencies.
        - We have promised we will share Messenger data with other countries when they have CLOUD Act agreements. Currently, no countries have CLOUD Act agreements. But when they do, and if we are unable to provide MSGR data, we will be breaking our promise. That broken promise will hit hard.
- **Governments** (particularly U.S., UK, E.U., India, Brazil, and Turkey) **increase pressure on us to provide technical assistance to work around encryption (not to break E2EE per se)**:
    - Think adding a secret person to a WhatsApp Group, using our platforms to push malware, asking us to store and analyze more metadata;

    - removing our ability to provide data to the US government is likely to increase pressure building up to our own 'Apple vs the FBI' moment
        - The **US Government** is currently seeking a court order mandating that encrypted services either decrypt or provide access to voice calls made through MSGR for suspected MS-13 members.
    - International technical assistance pressures is also mounting:
        - For example, the governments of the Five Eyes (US UK AUS NZ CA) issued a joint statement in 2018 that "should governments continue to encounter impediments to lawful access to information necessary to aid the protection of the citizens of our countries, we may pursue technological, enforcement, legislative or other measures to achieve lawful access solutions."
        - The **UK** and **Australia** already have technical assistance legislation, although we currently think it is unlikely they would use it against us. This could change that calculation.
    - In **India**, we have heard recently from the regulator that national security is a government mandate, and that while e2ee has benefits, tech companies must find a solution to support the government against security threats.
    - **Brazil** has already proposed bills that aim at forcing companies to hand in encryption keys to law enforcement agents in the process of criminal investigations and otherwise allow LE access to communications apps. Because of our inability to provide the content of WA messages, WA has been blocked in Brazil multiple times.
    - **Turkey** may react to our 'doubling down' on e2ee by banning WA, MSGR and IG Direct because of frustration that we are unable to access and police messaging on those services; the government has already threatened WA with a ban and announced its intention to restrict services that are non-compliant with Turkish law, and Turkey's 2018 ban of Uber would suggest that this is not an empty threat.

What Other Ways Could E2EE make things worse?

- *As well potentially preventing us from identifying harm, and impacting on our relationship with law enforcement, implementing E2EE will also potentially create additional harms. Given that perpetrators can identify and befriend victims on Facebook, before taking them easily into the secret world of E2EE, these additional risks are likely;*
    - *We are more likely to attract bad actors, who will know that Facebook will not be able identify or verify current or future harm. We know that users of Secret Conversation users are more likely to have violated policy than Messenger users, two times as likely for terrorism, dangerous org, or adult sexual exploitation violations.*
    - *We will not be able to use any improvement in our techniques on machine learning, behavioural analysis to get better at identifying sexploitation, bullying, hate speech etc in Messenger.*

Thinking about mitigations:

<u>Top-Line Message: We need 12-18 months to explore effective mitigations that can keep our users safe.</u>

When it comes to mitigations we will equally have to be data and evidence driven.

- We should focus on mitigations to ensure that:

    - we are still able to report to NCMEC;
    - respond to harm we would have previously detected after other signals would have pushed us to investigate Messenger content;
        - Do we try to prevent it at other pinch points? (ie preventing victim identification for child grooming; sharing more unverified information on potential school shooters, terrorists or suicide attempts with law enforcement; developing very easy reporting mechanisms; pushing reporting mechanisms when there are off-Messenger signals?)
    - support law enforcement in its work.
- Our work on mitigations should be evidence and data driven and focus on prevention, detection and helping users seek assistance for each different type of harm:
    - Where are the pain points in each harm cycle and where can we prove mitigations do or don't work?
        - For example on spam our research shows we **can** successfully use metadata to identify spam in 95% of the time; my understanding from talking to the safety teams today is that behavioral analysis of metadata for child grooming is only 1% successful and throws up thousands of false positives.
        - Equally we should not assume that a Facebook friendship means there is no harm:
            - 61% of 2018 NCMEC reports from Messenger are between friends.
            - 86% of 2018 revenge porn cases is between friends
    - We'll also need to dig into WhatsApp's experience on things like terrorism, and their struggle to tackle it through anything other than reporting;
- We should have a data driven, systematic process for testing mitigations, and an engagement process with external stakeholders to understand different perspectives and build legitimacy.
    - Harm against children should be the starting point.

- Whilst we are skeptical that many mitigations will not work, there are some 'big bet' opportunities for developing:
    - world class sexploitation and bullying reporting systems, that use behavioral analysis to push individuals to report content to us and law enforcement;
    - a new way to work collaboratively with LE, on joint ways to understand metadata and also on crime prevention education initiatives using our platform.

What specific external engagement efforts will blunt risk?

- **Educate stakeholders surrounding perceptions of safety/privacy issues, and distinguish them from actual safety/privacy issues.**
- Identify actual safety risks and develop detailed prevention, detection and alert strategies for these risks, involving technical and policy mitigations, and then **work with external partners for some mitigations** (for example sharing behavioral analysis with safety groups).
- **Closely collaborate and partner with a vetted group of external experts from a variety of stakeholder communities with the goals of building a world leading product and using it publicly for validation in the press.**
- **Develop testing systems for our strategies and tactics, involving partners and user research.**
    - Some prevention, detection and investigation will take place off our platform (i.e. by law enforcement and vendors) and in order to have the safest product, working with these groups is essential.
    - Build on relationships with law enforcement, focusing on joint work we can do to build safer communities, if we cannot provide evidence of crimes. This could include better community

engagement, prevention campaigns, working with them on how metadata can be useful (including changing relevant laws).

Lessons learned from Prior external engagement

- We have models for and learnings from past external engagements - namely the External Review Board - that can inform our third party engagement for UM and e2ee. From past experience, we think successful external engagement involves three principal steps:

1. **Building sustained relationships with influencers,** people who disproportionately impact sentiment of Facebook.
    1. Identify people who are influencers across all products and relevant stakeholders for specific products, based on their subject-matter expertise.
    2. And we have criteria for creating an influencer list and a playbook for how to engage with them.
2. **Involving stakeholders in the design process for products and programs.**
    1. There isn't one form of collaborative engagement - 1x1 conversations, proposals or other pilots or forums can all be suitable depending on the initiative.
    2. The outcome is critical, however, and Facebook must actively seek input on outstanding questions, respond and incorporate feedback into product decisions.
    3. And we have strategies for gathering feedback and scripts for these conversations on hand.
3. Through the above process, we **validate that we are taking the right approach** and, consequently, **receive external validation from our target influencers.**
    1. Our methodology or product is externally defensible, and we have developed a list of influencers who, at least, appreciate we are listening so will not criticize our approach and, at best, will publicly endorse our approach.

Which top MSGR markets would be upset if we encrypted MSGR?

Note: Top 15 MSGR markets (ranked) are: US, India, Brazil, Indonesia, Mexico, Philippines, Vietnam, Thailand, UK, Egypt, Turkey, France, Germany, Italy, Pakistan (MAP at bottom).

Of MSGR's top markets, we can expect the expansion of e2ee to prompt a significant response from regulators and law enforcement in the US, EU, UK, India, Brazil, and Turkey:

- **US:** The government is currently seeking a court order mandating that encrypted services either decrypt or provide access to voice calls made through MSGR for suspected MS-13 members.
- **UK and Australia:** Already have technical assistance legislation, although we currently think it is unlikely that they would use it against us. Expanding e2ee could change that calculation.
    - The US, UK and Australia are also (alongside New Zealand and Canada) members of the Five Eyes intelligence community, which vowed in a 2018 statement to "pursue technological, enforcement, legislative or other measures to achieve lawful access solutions" if governments continue to encounter impediments to lawful access.
- **EU:** We have seen attempts to force US-based technology companies to register as telecommunications services, this way facilitating law enforcement access obligations on those companies.
- **India:** We have heard recently from the regulator that national security is a government mandate, and that while e2ee has benefits, tech companies must find a solution to support the government against security threats.

- **Brazil:** Has proposed bills that aim at forcing companies to hand in encryption keys to LE agents in criminal investigations and allow LE access to communications apps. WA has been blocked in Brazil multiple times because of our inability to provide the content of WA messages.
- **Turkey:** The government has threatened WA with a ban and announced its intention to restrict services that are non-compliant with Turkish law; may react to our expanding encryption by banning our messaging services.

We should also expect pressure in Vietnam, Thailand, the Philippines, and Indonesia:

- **Vietnam:** Vietnam has a large security apparatus that monitors online activity and just passed a new law on cybersecurity that will create additional pressure to provide access to data. Expanding e2ee is likely to be of concern to law enforcement and there could be pressure to provide backdoors.
- **Thailand:** FB has been under considerable pressure in the country to remove undesirable content, and expanding e2ee will increase this pressure from lawmakers, particularly if the military returns to power after the March elections.
- **Philippines:** Will likely increase pressure from LE - particularly intelligence and military agencies working on terrorism - and from safety organizations that believe that MSGR is one of the most used platforms for child exploitation and harassment.
- **Indonesia:** Telegram was reportedly blocked for purported failure to moderate content.

What is actual versus perceived harm?

- **Actual**
  - Moving to e2ee without effective mitigations will reduce our ability to keep vulnerable individuals safe from bad actors and predatory behavior.
  - Because of our recent commitment to (and request that the public judge us on) transparency on our enforcement numbers, our diminished efforts and success in keeping people safe will be visible through data we have already committed to releasing.
  - In short, we can expect the move to e2ee to result in numbers that will be interpreted as FB keeping fewer people safe.
  - Further, in explorations to date, we have not identified a mitigation measure that allows us to materially retain our safety abilities or our ability to hand actionable information over to e2ee.
  - And we currently assess that even with mitigations, moving to e2ee will have an impact on safety and the perception of how much we care about people.
  - It is well known that WhatsApp offers e2ee, and any step that would weaken encryption on WA would significantly undermine the overall unification effort.
- **Perceived**
  - The carveouts in the extent to which we offer e2ee on other services (even if for safety/integrity purposes) will be viewed by some critics as detrimental to privacy.
    - To the extent these exceptions make it harder to understand when a message/service is e2ee or create ways we can force people out of e2ee based on government requests, this can create actual harm.

Note on Data Security Positives of Encryption and E2EE

- End to end encryption is viewed by most privacy stakeholders as an essential component of any privacy protective messaging service. both because it protects traffic from interception and because it prevents the service provider from viewing or using message content.
- The data security benefits of encryption - including e2ee “where appropriate” - are noted in the forthcoming EU telecoms code:
    - *In order to safeguard security of networks and services, and without prejudice to the Member States’ powers to ensure the protection of their essential security interests and public security, and to permit the investigation, detection and prosecution of criminal offences,* ***the use of encryption for example, end-to-end where appropriate****,* ***should be promoted and, where necessary, encryption should be mandatory*** *in accordance with the principles of security and privacy by default and by design.*