

FILED UNDER SEAL

EXHIBIT 23

Deprecated -

1. **Strategy Deck:** [HYPERLINK

"https://docs.google.com/presentation/d/1BMK_Csl2WgaLLnhagfKtQsB5du1qsZWDDzgZ2T1aeh/edit" \h]

2. **Brief:** [HYPERLINK "https://docs.google.com/document/d/1FE54aPpadZF520-

cQ6dqBh1lhfb3cVplJDAZsLRsZ8E/edit" \l "heading=h.2o365c49dmwl" \h]

Today we take down CSAM + NCII content on any public threads or feeds IGD, MSGR, IG Feeds, Newsfeed through MMS

With E2EE we cant do anything

- We took an approach of generalized nudity and porn
- All nudity and porn and not just CSAM (the approach and our models are trained to detect nudity)
- We only put soft action (treatment) not take down
- Only on message threads (E2EE)
- Community chat will have server side solution
- We train our models on Ansa to take down model
- [REDACTED]
- Adults - opt in
- Teens we will likely push for default on
 - Technical perspective we need this decision sooner
 - We were always default off
 - We need this decision on this keeping E2EE in mind, So Default should carry over from open and E2EE.
- Katie to help with settings
- **Apple was defining for under 13**
- Are we rolling out to open threads then encryption
 - **IG E2EE population TLC will start in February**
 - **Open thread**
 - **2 sided sword**
 - We can get PRE signals
 - What are the users doing
 - On open thread, we can track events (are using hiding, unhiding, send friction, are they blocking, sending..)
 - We cant log any of the above in E2EE
 - Are they unblurring and viewing even after?
 - In e2ee, we can log settings changes. Only cant log message related.

Deprecated -

1. Strategy Deck: [HYPERLINK

"https://docs.google.com/presentation/d/1BMK_Csl2WgaLLnhagfKtQsB5du1qsZWDDzgZ2T1aehI/edit" \h]

2. Brief: [HYPERLINK "https://docs.google.com/document/d/1FE54aPpadZF520-

cQ6dqBh1lhfb3cVplJDAZsLRsZ8E/edit" \l "heading=h.2o365c49dmwl" \h]

■ We cant measure how many images are blurred in E2EE (we need federated analytics)

- Because we are already taking down images using our banks, there will be VERY few images that even have this treatment
-
- 50% of the images may already be taken down, so ODNC has small opportunity
- The timing is better
- Are we going to launch open first, then E2EE? (parallel)

Btw, thinking more about the open vs e2ee threads.. Whats the expected timeframe for when all IG threads will be cutover to e2ee?

Should we work backwards to figure out how long it would take to build ODNC for open threads, and see how long it would even be available to users on open?

They start in February I think with a small test and then gradually move all threads

I think it by end of 2024 till everyone on IGD is on e2ee messages

based on my experiences on how we did e2ee rollout on messenger

We can talk to IGD team and target segments which will get e2ee at the end. Like last time they had 5 high risk countries that got armadillo at the very end

Links

- [HYPERLINK "https://docs.google.com/presentation/d/1BMK_Csl2WgaLLnhagfKtQsB5du1qsZWDDzgZ2T1aehI/edit" \h]
- [HYPERLINK "https://docs.google.com/document/d/1PmuARJ9weTmLlj7RCeEI2MmjQRpXuwMjtPeaa5NI8IA/edit" \h]
- [HYPERLINK "https://app.asana.com/0/1201880975354561/list" \h]

MLT Review Outcome

1. Yes, MLT is aligned and supportive to look into this further and believe that depending on how we design this, these experiences can be consistent with our privacy stance.

Workstreams

1. Design Specs and UXR

Deprecated -

1. Strategy Deck: [HYPERLINK

"https://docs.google.com/presentation/d/1BMK_Csl2WgaLLnhagfKtQsB5du1qsZWDDzgZ2T1aehl/edit" \h]

2. Brief: [HYPERLINK "https://docs.google.com/document/d/1FE54aPpadZF520-

cQ6dqBh1lhfb3cVplJDAZsLRsZ8E/edit" \l "heading=h.2o365c49dmwl" \h]

- a. Share learnings with iterated and updated designs with MLT following UXR with teens in mid-March.
 - b. Scope
 - i. Fleshing out design specs
 - ii. Partnering with teen safety and aligning
 - iii. Going through UXR in March
 - c. Als:
 - i. Working model - [REDACTED] to follow up with [REDACTED] [REDACTED] to connect with [REDACTED]
 1. Ownership for CD,PD, PM, ENG etc.
 - ii. [REDACTED] - will kick-off, but tbd in March and will fup
 - iii. [REDACTED] to own UXR
2. Privacy and Safety Stakeholders (Internal and External)
- a. Review with external safety and privacy groups and get buy-in that this investment is compelling from a child safety perspective while maintaining a strong E2EE privacy bar.
 - b. Get PXFN approval.
 - c. Als:
 - i. [REDACTED] to connect with Nate and establish the plan for privacy groups
 - ii. [REDACTED] - kickoff the PXFN process.
 - d. Notes:
 - i. Going back to global safety groups back in April to close that group -
3. Technical Scoping and Model Improvements
- a. Begin technical scoping and identify whether this tech can work at high precision, establish model benchmarks for success, and validate assumptions with users. Share findings and evaluation with MLT.
 - b. Als:
 - i. [REDACTED] to follow up with Nidhi/Ram + partner teams and establish working model.
4. Teen safety education outside of app
- a. Als:
 - i. [REDACTED] to connect with [REDACTED] Safety Policy, Trevor and Cedric

Timeline

1. UXR week of 3/14
2. DR - April 19

Commented [1]: [REDACTED] -
Will need iterations based on PXFN.

[REDACTED] -
To update after working model discussions and alignment with partners.

[REDACTED]
Pending capacity and timeline planning

[REDACTED] -
End of march seems reasonable for a plan and updates. Going to follow up with Raul on removing Trust AH video.

Commented [2]: @ [REDACTED]@fb.com what are our tentative dates for UXR?

Commented [3]: @ [REDACTED]@fb.com week of 3/14 - will we be ready to put something in front of teens? Sorry to have missed the mtg yesterday! Needed to protect my Wednesday for research synthesis!

Deprecated -

1. **Strategy Deck:** [HYPERLINK

"https://docs.google.com/presentation/d/1BMK_Csl2WgaLLnhagfKtQsB5du1qsZWDDzgZ2T1aehI/edit" \h]

2. **Brief:** [HYPERLINK "https://docs.google.com/document/d/1FE54aPpadZF520-

cQ6dqBh1lhfb3cVplJDAZsLRsZ8E/edit" \l "heading=h.2o365c49dmwl" \h]

Deliverables

Deprecated -

1. **Strategy Deck:** [[HYPERLINK](https://docs.google.com/presentation/d/1BMK_Csl2WgaLLnhagfKtQsB5du1qsZWDDzgZ2T1aeh/edit)

"https://docs.google.com/presentation/d/1BMK_Csl2WgaLLnhagfKtQsB5du1qsZWDDzgZ2T1aeh/edit" \h]

2. **Brief:** [[HYPERLINK](https://docs.google.com/document/d/1FE54aPpadZF520-cQ6dqBh1lhfb3cVpJJDAZsLRsZ8E/edit) "https://docs.google.com/document/d/1FE54aPpadZF520-

cQ6dqBh1lhfb3cVpJJDAZsLRsZ8E/edit" \l "heading=h.2o365c49dmwl" \h]

GRAVEYARD

Notes from MLT Review

Messenger Product Review: Messenger On-Device Nudity Control Strategy

Deck: [[HYPERLINK](https://docs.google.com/presentation/d/1BMK_Csl2WgaLLnhagfKtQsB5du1qsZWDDzgZ2T1aeh/edit)

"https://docs.google.com/presentation/d/1BMK_Csl2WgaLLnhagfKtQsB5du1qsZWDDzgZ2T1aeh/edit" \l "slide=id.g73a13741e6_0_109" \h]

Date: 2/17/22

Note: Due to A/C Priv and the sensitive nature of this topic, sharing review notes via email.

Context

The Messenger Well-Being team presented a strategy for a generalized and optional on-device nudity control for MSGR and IGD that helps reduce the creation, exposure, and spread of nude images where teens (13-17) are the primary, but not only, audience.

The control offers contextual and targeted user education, protection from unwanted exposure, and reshare warnings.

Review Goals

1. Align on investing in an on-device nudity control.
2. Decide on whether the control should be opt-in or opt-out.

Decisions

1. [Decision] Are we aligned on investing in an on-device nudity control to reduce the creation, exposure, and resharing of nude images, where teens (ages 13-17) are the primary, but not only, audience?
 - a. Yes, MLT is aligned and supportive to look into this further and believe that depending on how we design this, these experiences can be consistent with our privacy stance.
2. [Decision] Should we go with a setting for teens that is opt-out to optimize for safety and adoption or opt-in to optimize for privacy?
 - a. MLT recommends that we start with an opt-in, empowering user choice, and invest in education and explore ways to integrate the feature in parental controls

Feedback

1. Ensure the opt-in control experience and subsequent notices feel private to everyone both in terms of content and implementation. Validate this with users and key external stakeholders.
2. Getting the tech to work at high precision will be critical. Establish what the benchmark for success is and the trade-offs between precision and recall. Validate with users.
3. Separate the sending/receiving experiences and work closely with UXR to understand what teens believe is happening at every step and iterate based on these learnings.
4. Break out the experience outcomes we seek for each audience (Adults, Teens, Parents).
5. Explore and specify if and how this scales for content beyond images, such as videos.

Deprecated -

1. Strategy Deck: [HYPERLINK

"https://docs.google.com/presentation/d/1BMK_Csl2WgaLLnhagfKtQsB5du1qsZWDDzgZ2T1aeh/edit" \h]

2. Brief: [HYPERLINK "https://docs.google.com/document/d/1FE54aPpadZF520-

cQ6dqBh1lhfb3cVplJDAZsLRsZ8E/edit" \l "heading=h.2o365c49dmwl" \h]

6. Success for this workstream can be seen as two milestones. First, get buy-in from regulators that this investment is compelling to assuage concerns around Child Safety as we transition to E2EE. Second, focus on effectiveness in terms of adoption using education and with a close eye on user feedback.
7. Ensure that we are not creating policy or determination on what defines nudity. Partner with the appropriate experts and leverage existing industry standard definitions.
8. Consider how to educate teens on the dangers of sending, receiving, and re-sharing nudes on social media outside of Messenger. Engage with PMM resources.

Follow Ups

1. Share learnings with iterated and updated designs with MLT following UXR with teens in mid-March.
2. Begin technical scoping and identify whether this tech can work at high precision, establish model benchmarks for success, and validate assumptions with users. Share findings and evaluation with MLT.
3. Review with external safety and privacy groups and get buy-in that this investment is compelling from a child safety perspective while maintaining a strong E2EE privacy bar.
4. MWB Teen Safety team in London, Safety Policy, and Product Marketing to update MLT on any plans for education, resources, and support for teens outside of the app on risks of sharing nude images.

Deck Outline

1. Purpose of review
 - a. Make a decision on whether we should invest in on-device solutions for explicit images related to nudity as a part of our minor and teen safety efforts.
2. People and business problem
 - a. Target audience
 - i. Teens: 13-17
 - b. People problems related to minors.
 - i. Minors are
 1. **Creating** and sharing explicit images with others, without knowing the risks associated with it.
 2. **Receiving** unwanted explicit images, that makes them feel uncomfortable and harassed.
 3. **Resharing** explicit images they've received with their friends, further victimizing and perpetuating the harm.
 - ii. ~~Both Minors and Adults are~~
 1. ~~Receiving unwanted explicit images, that makes them feel uncomfortable and harassed.~~
 - iii. Adults are
 1. Resharing explicit images of minors in outrage, not realizing they are doing more harm and victimizing the minor.

Commented [4]: explicitly out of scope - bad actors and networks

Deprecated -

1. Strategy Deck: [HYPERLINK

"https://docs.google.com/presentation/d/1BMK_Csl2WgaLLnhagfKtQsB5du1qsZWDDzgZ2T1aehl/edit" \h]

2. Brief: [HYPERLINK "https://docs.google.com/document/d/1FE54aPpadZF520-cQ6dqBh1lhfb3cVplJDAZsLRsZ8E/edit" \l "heading=h.2o365c49dmwl" \h]

iv. Parents are

1. Trying to keep their children safe from harms and risks related to sharing and receiving explicit images.

c. Business problem

- i. With E2EE, we lose the ability to identify and remove the ~600K shares of CSAM and NCII per day. Leading to a further lack of trust in our safety products, particularly for vulnerable segments of minors and women. Requiring investment in innovative measures to help address this gap.

d. Out of scope

- i. Minors U13 - will require a follow-up discussion
- ii. Bad actors and networks related to CSAM

3. Research?

a. Thorn?

4. Competitive landscape

- a. Apple iMessage has an opt-in control for minors when receiving or sending photos that contain nudity. The feature blurs the photo, warns the minor, and gives resources. ([HYPERLINK "https://www.apple.com/child-safety/" \h])
- b. Apple CSAM detection ([HYPERLINK "https://www.macworld.com/article/352875/ios-15-csam-scanning-icloud-photos-messages-siri-search-faq.html" \h])
- c. Discord options: scan direct messages from everyone, scan direct messages from non-friends, don't scan any messages. Messages scanned that contain explicit imagery will be automatically deleted. ([HYPERLINK "https://support.discord.com/hc/en-us/articles/115000068672-Discord-Safety-Safe-Messaging-" \h])

5. Existing Teen Safety Principles & Roadmap

6. Goal and North Star

- a. Goal: Leverage on-device scanning to help reduce the creation, exposure, and spread of explicit images of minors

7. Principles

- a. Choice - Should be able to opt-in or out of the on-device scanning at any time.
- b. Simple -
- c. Private
- d.

8. Product Strategy

a. Setting

- i. Control for minors and parents

Commented [5]: @ [REDACTED] @fb.com to help with additional points/reframing if needed

Commented [6]: @ [REDACTED] @fb.com is there any research we can pull out from Thorn that may be applicable here?
Assigned to [REDACTED] @fb.com

Commented [7]: Call out WAs position on on-device scanning

During family messaging privacy conversations last year discussed how to narrow the options

Commented [8]: Opt-in, transparency, and control

Commented [9]: 1. Iterative process for guardian controls

2. U13 conversation is likely a follow up

Commented [10]: Separate out what the controls are for creation side vs recipient side.

Deprecated -

1. Strategy Deck: [HYPERLINK

"https://docs.google.com/presentation/d/1BMK_Csl2WgaLLnhagfKtQsB5du1qsZWDDzgZ2T1aeh/edit" \h]

2. Brief: [HYPERLINK "https://docs.google.com/document/d/1FE54aPpadZF520-cQ6dqBh1lhfb3cVplJDAZsLRsZ8E/edit" \l "heading=h.2o365c49dmwl" \h]

ii. Options

1. Minors

a. Default on for minors with opt-out

b. Default off for minors with opt-in + upsells

2. Parents / Guardian Controls

a. Opt-in

b. Creating

i. Contextual education about risks and friction

ii. Options

1. Educational resources

2. Soft friction - are you sure

3. Hard friction - prevent send

c. Receiving / Exposure

i. Prevent unwanted exposure

ii. Options

1. Auto-delete incoming explicit images

2. Blur and tap to reveal explicit images

3. Talk to a trusted adult

4. Resources for help center content + external

5. Accessible and easy reporting

d. Resharing

i. Warn and add friction

ii. Options

1. Education - established deterrence playbook

2. Soft friction - are you sure

9. ~~Trade-Offs~~

10. Solution Explorations

Deprecated -

1. Strategy Deck: [HYPERLINK

"https://docs.google.com/presentation/d/1BMK_Csl2WgaLLnhagfKtQsB5du1qsZWDDzgZ2T1aehl/edit" \h]

2. Brief: [HYPERLINK "https://docs.google.com/document/d/1FE54aPpadZF520-cQ6dqBh1lhfb3cVplJDAZsLRsZ8E/edit" \l "heading=h.2o365c49dmwl" \h]

Trust Mission

1. Earn people's trust by making messaging private & safe, across the FoA/RL.

MWB Mission

1. Build a safe and inclusive product, by minimizing harm and improving well-being of individuals and communities who use Messenger and Instagram Direct.

Context

1. Today, in a non-E2EE world, we use proactive detection and enforcement of nudity via MMS to remove (as of 12/24/21)
 - a. CSAM:
 - i. MSGR: [HYPERLINK "https://www.internalfb.com/intern/daiquery/workspace/1040891199811416/995751091153276/" \h]
 - ii. IGD: ~15K per day
 - b. NCII:
 - i. MSGR: [HYPERLINK "https://www.internalfb.com/intern/daiquery/workspace/1040891199811416/413087547167332/" \h]
 - ii. IGD:
2. With E2EE the MMS system will no longer be able to run on content. Meaning the content that would be removed today will continue to exist in the messaging ecosystem and likely to be propagated even further.
3. Teens are sharing self-generated CSAM (SG-CSAM) at increasing rates. Based on the most [HYPERLINK "https://info.thorn.org/hubfs/Research/SGCSAM_Attitudes&Experiences_YouthMonitoring_FullReport_2021_FINAL%20(1).pdf?utm_source=website&utm_medium=blog&utm_campaign=SG_monitoring_2021" \h], ~20% of teens have shared SG-CSAM and ~35% of teens agree that sharing SG-CSAM is normal. Of those who have shared nudes, 50% reported they had shared with someone they had not met in real life and 41% reported they had shared with someone age 18 or older.
4. Non-consensually re-sharing SG-CSAM is also pervasive. ~30% of teens surveyed say they have seen non-consensually re-shared SG-CSAM.

Commented [11]: Need to reframe context to not focus so much on MMS

Commented [12]: @ [REDACTED] @fb.com
@ [REDACTED] @fb.com These are reading more as behaviors we see that cause problems for us that we need solve, rather than the [unintended] problems a user experiences because of these behaviors.

Is there a way to reframe from the perspective of the user? (there are some good examples in the user journey section) From a content perspective at least, this would help build a framework (tone, resources offered, settings, terminology etc) for features we might build and frame to the user.

E.g.

1) I shared a nude with someone I trusted, but now they won't stop harassing me for more pictures

2) I sent a personal photo to my close group of friends, but I got a message from one of them a few weeks later saying they saw it in a school group

Nudity Based Problems

CSAM

Commented [13]: Great point! I can edit these to make them from the perspective of users. Will created a section below and rename this one "problems"

Deprecated -

1. Strategy Deck: [HYPERLINK

"https://docs.google.com/presentation/d/1BMK_Csl2WgaLLnhagfKtQsB5du1qsZWDDzgZ2T1aehl/edit" \h]

2. Brief: [HYPERLINK "https://docs.google.com/document/d/1FE54aPpadZF520-

cQ6dqBh1lhfb3cVplJDAZsLRsZ8E/edit" \l "heading=h.2o365c49dmwl" \h]

1. Minors are sharing SG-CSAM without realizing the potential consequences.
2. Minors are re-sharing and further spreading CSAM.
3. Adults are non-maliciously sharing and further exposing CSAM due to outrage.
4. ~~Bad actors are sharing and trading CSAM. (Out of scope for this proposal)~~

Commented [14]: with groomers vs other minors

Commented [15]: Might remove this one if we think these people are not likely to use on-device scanning

NCII

1. People are non-maliciously sharing and propagating images and videos with nudity, not realizing it is NCII.
2. ~~People are being extorted with threats of their nude images being shared to their family and friends. (Out of scope for this proposal, part of stop NCII)~~

Commented [16]: Sync with [REDACTED] - on stop NCII

Sexual Harassment

1. People are receiving unsolicited sexual images from known and unknown contacts.

Preferences

1. People don't want to be unintentionally exposed to images and videos with nudity.

People problems

1. I shared a nude with someone I trusted; later they started harassing me for more images
2. I sent a nude to my boyfriend/girlfriend, not expecting them to show other people. They share my photo with others.
3. I see photos/posts with vulnerable children and share in outrage. I don't think about the fact that I'm doing more harm and further victimizing the child.
4. I see images or videos and share them, not realizing the damage i'm doing to the person in the imagery (NCII)
5. I get messages with nudity. It grosses me out and makes me uncomfortable.

Business Problem

1. With the launch of E2EE, we will lose our ability to enforce on a significant amount of the 600K CSAM and NCII that is removed today. Leading to a lack of trust in our safety products, particularly impacting vulnerable segments of minors and women.

Goal

1. Empower users with E2EE resilient controls that help protect them and others from nudity based harms (CSAM and NCII sharing and exposure).

Commented [17]: TBD - pick 1-2

Commented [18]: [REDACTED] - lean into empowerment and user control and narrative. Investing in controls that users have to protect their own safety.

Deprecated -

1. Strategy Deck: [HYPERLINK

"https://docs.google.com/presentation/d/1BMK_Csl2WgaLLnhagfKtQsB5du1qsZWDDzgZ2T1aehI/edit" \h]

2. Brief: [HYPERLINK "https://docs.google.com/document/d/1FE54aPpadZF520-

cQ6dqBh1lhfb3cVplJDAZsLRsZ8E/edit" \l "heading=h.2o365c49dmwl" \h]

2. Leverage on-device scanning to reduce nudity related or initiated bad experiences people have.
3. Help protect vulnerable populations from nudity related bad experiences.
4. Reduce the amount of CSAM and NCII in the post E2EE messaging ecosystem.

Competitive Landscape

1. Apple iMessage has an opt-in control for minors when receiving or sending photos that contain nudity. The feature blurs the photo, warns the minor, and gives resources. ([HYPERLINK "https://www.apple.com/child-safety/" \h])
2. Discord options: scan direct messages from everyone, scan direct messages from non-friends, don't scan any messages. Messages scanned that contain explicit imagery will be automatically deleted. ([HYPERLINK "https://support.discord.com/hc/en-us/articles/115000068672-Discord-Safety-Safe-Messaging-" \h])
3. Apple CSAM detection ([HYPERLINK "https://www.macworld.com/article/352875/ios-15-csam-scanning-icloud-photos-messages-siri-search-faq.html" \h])

Any IGWB On-Device Existing Work

Principles (From [HYPERLINK

"https://docs.google.com/document/d/1sOjZPQpozExyP6jbNZaD-wZ3eK3cFqj3fEbPVQotbg/edit" \l "bookmark=id.p9phwzpodi06" \h])

1. Comply with privacy promises. Interventions...
 - a. Benefit the device user, not Facebook
 - b. Require explicit opt-in
 - c. Will run on the device
 - d. Only apply to incoming messages
2. Transparency
 - a. Easy to understand data disclosures
 - b. Provide transparency on interventions
 - c. Classifier updates must be made available to all users
3. Simple and understandable experience
 - a. Options to control the experience
 - b. Prioritize interventions based on severity of problem
 - c. Bundle safety interventions by related problems they solve
 - d. Interventions should not degrade app or device experience
4. Protect users data and only share with FB when explicitly shared

Deprecated -

1. Strategy Deck: [HYPERLINK

"https://docs.google.com/presentation/d/1BMK_Csl2WgaLLnhagfKtQsB5du1qsZWDDzgZ2T1aehl/edit" \h]

2. Brief: [HYPERLINK "https://docs.google.com/document/d/1FE54aPpadZF520-

cQ6dqBh1lhf3cVplJDAZsLRsZ8E/edit" \l "heading=h.2o365c49dmwl" \h]

User Journey

Minor (consensual share becomes non-consensual)

1. Minor takes an explicit picture and share it with a friend/contact (minor/adult) in MSGR (self-generated)
 - a. Either prompted by someone else or self initiated
2. Recipient further shares with 2-3 friends
3. Original minor whose photo it is may be informed of it that the photo has been shared

Commented [19]: CSAM majority of shares are coming from outrage.

Non-preferential - exposed to teen content and start sharing that.

Commented [20]: [REDACTED]@fb.com added this to reflect - as we heard in today's Thorn preso - that many of these relationships, even if consensually shared, start with online-only contacts, whether they be other minors or adults

Minor (implicitly non consensual share to explicitly non consensual)

1. Minor begins speaking to an adult impersonating a minor
2. Adult impersonator shares explicit photos from online pretending to be a minor
3. Minor shares explicit photos with impersonator
4. Impersonator begins extorting minor for more photos

Solutions

1. Disappearing / vanish mode -
 - a. image screenshot notification
 - b. Increase usage of vanish mode
2. Nudity image blurring
3. Linked resources based on nudity
4. [HYPERLINK "https://sites.google.com/technologycoalition.org/technologycoalitionresources/resources/deterrence-and-intervention-messaging?authuser=1" \h]
 - a. "This may not be consensual, please reconsider"
 - b. "Sharing has criminal consequences of consensual"
5. Message level controls (can't forward/save)
6. Forwarding notification / limits?
7. Reporting upsell for nude images
 - a. Could be linked to resources
8. On-device banking and scanning
9. Automated deletion
 - a. Of all images classified with nudity
10. Messaging options
 - a. Generic for all users
 - b. Specific to teens
 - c. Specific to adults / outrage

Commented [21]: Principle: Non-judgemental

Commented [22]: Segment by audience

Commented [23]: Separate out solutions by receiver side vs sender side vs paired opt ins

Commented [24]: Inbox vs MRs - where do these solutions apply?

Deprecated -

1. Strategy Deck: [HYPERLINK

"https://docs.google.com/presentation/d/1BMK_Csl2WgaLLnhagfKtQsB5du1qsZWDDzgZ2T1aeh/edit" \h]

2. Brief: [HYPERLINK "https://docs.google.com/document/d/1FE54aPpadZF520-

cQ6dqBh1lhfb3cVplJDAZsLRsZ8E/edit" \l "heading=h.2o365c49dmwl" \h]

11. Parents and parental controls

a. Solutions where the parent is the person setting the control

b. [HYPERLINK "https://fb.workplace.com/work/knowledge/5556175567764346" \h]

12. Opt-in

a. Creative upsell for minors

Risk/Trade-offs

- Perceptions of privacy
- Acknowledging that these problems (CSAM sharing, sexual harassment) happen on our platforms // drawing attention to high-severity harms that people might not otherwise be aware of

Future Research

- Qualitative concept feedback
- Quantitative sentiment survey

Team

Timeline

1. 2/3/22 - Review first draft of strategy w/ MWB Leads
2. 2/5/22 - Revised strategy based on feedback
3. 2/9/22 - Review with Pavni and Antigone
4. TBD - MLT Review

Resources and Research

1. Sai -
 - a. On-Device Resources [HYPERLINK "https://fb.quip.com/9Zh9AkoHdp4k" \h]
 - b. Privacy AI Proof of Concept: [HYPERLINK "https://fb.quip.com/NigJAxP4XUp5" \h]
2. Nidhi
 - a. Primal FUP
3. Policy and Safety
 - a. [HYPERLINK "https://docs.google.com/document/d/1epfof4TrqSoUPuevWCRvCZDkEYFBPDR36eh3Ei5UFQQ/edit" \l "heading=h.mtzfu04w08h2" \h]
 - b. Child Safety Lit Review: [HYPERLINK "https://fb.quip.com/SlueAa596ZiZ" \h]

Commented [25]: Pavni is also interested in the discussion of over 13 vs under 13.

Deprecated -

1. Strategy Deck: [HYPERLINK

"https://docs.google.com/presentation/d/1BMK_Csl2WgaLLnhagfKtQsB5du1qsZWDDzgZ2T1aeh/edit" \h]

2. Brief: [HYPERLINK "https://docs.google.com/document/d/1FE54aPpadZF520-

cQ6dqBh1lhfb3cVplJDAZsLRsZ8E/edit" \l "heading=h.2o365c49dmwl" \h]

4. [HYPERLINK

"https://docs.google.com/document/d/1XHI2lelZhO2YJeNCxqxrQDnXubC0iPceBfi7FyLGcXY/edit" \h]

5. [HYPERLINK "https://www.theverge.com/2021/8/6/22613365/apple-icloud-csam-scanning-whatsapp-surveillance-reactions" \h]

a. Is it strange that MSGR would deviate from WA response?

6. IG Guardian (Parental Controls)

a. [HYPERLINK

"https://docs.google.com/presentation/d/1_rqreJRMeg7EWvcg94qO5KJ1OrH6BXhhG-lexAlofw4/edit" \l "slide=id.g887c1e7c60_2_0" \h]

7. Messenger Kids POC

a. [REDACTED]

8. Thorn Self-Generated CSAM Report

9. Youth Safety Design Sprint: [HYPERLINK

"https://www.figma.com/file/pcgH7xiC6vml73pwbhrcw/Teen-Safety-in-Messaging?node-id=1%3A5" \h]

10. [HYPERLINK

"https://docs.google.com/presentation/d/1wQTF4sPljMlqm6k8s3UZFuLiG-PfaxTxMnWCvU-Wts/edit" \l "slide=id.g1072b52b3a2_0_12" \h]

11. Kristy's WIP Design Concepts - [HYPERLINK

"https://www.figma.com/file/jZ1dUbYAGivzzZaEUwLBPG/On-Device-Nudity?node-id=6%3A8" \h]

12. Risks of [HYPERLINK "https://arxiv.org/pdf/2110.07450.pdf" \h]

13. SafeChat design doc - [HYPERLINK "https://fb.quip.com/NigJAXp4XUp5" \h]

14.

Discussion Topics

1. If we block sharing of nude images by minors/teens they are likely to just go to other platforms and continue to do it. -

Deprecated -

1. Strategy Deck: [HYPERLINK

"https://docs.google.com/presentation/d/1BMK_Csl2WgaLLnhagfKtQsB5du1qsZWDDzgZ2T1aeh/edit" \h]

2. Brief: [HYPERLINK "https://docs.google.com/document/d/1FE54aPpadZF520-cQ6dqBh1lhfb3cVplJDAZsLRsZ8E/edit" \l "heading=h.2o365c49dmwl" \h]

Meeting Notes

1/12/22

Sync [REDACTED] [REDACTED] [REDACTED]

1.

1/10/22

[REDACTED] [REDACTED] and [REDACTED]

1. Will teens use this?

- a. Parents still have some influence on settings of teens devices still.
- b. Self-protections framing - unsolicited sexual images from others
 - i. People could position this as maybe this shouldn't happen in the first place.

2. Will adults use this?

- a. Outraged parents would likely use this.

3. What value are we providing?

4. Add risks and trade-offs

5. Two audiences

- a. Teens/Minors
- b. Adults non-malicious
- c. Stopping non-violators from becoming violators
 - i. We do this on FB (and maybe IG) search terms

1:1 [REDACTED] and [REDACTED]

1. 1:1

- a. Where is the intervention happening?
 - i. Sender side is not ePD and Armadillo compliant
 - ii. Receiver side
- b. Frame opt-in and opt-out from the perspective of minors ability to recognize the harm and choose to turn on
- c. Discussion with WA
 - i. Privacy AI scanning was denied
 - ii. Legal obligation for reporting when non-maliciously shared
 - iii. What privacy promises are we making?

2. Talk to [REDACTED]

3. Frame this as one specific segment

4. Defensibility is a big one

- a. Of the database of the links, words, images
- b. Do they trust facebook or should it be 3rd party

Deprecated -

1. **Strategy Deck:** [HYPERLINK

"https://docs.google.com/presentation/d/1BMK_Csl2WgaLLnhagfKtQsB5du1qsZWDDzgZ2T1aehI/edit" \h]

2. **Brief:** [HYPERLINK "https://docs.google.com/document/d/1FE54aPpadZF520-

cQ6dqBh1lhfb3cVplJDAZsLRsZ8E/edit" \l "heading=h.2o365c49dmwl" \h]

5.