Tennessee v. Meta
Case No. 23-1364-IV
**PX-00109**

# Document Provided in Native Format

# Child Safety on Messenger



Messenger Academy

001



1 My role
Sarah [redacted]
2 How long I've been at Meta
Safety Policy Director
3 Favorite tv show

Messenger Academy

**Introduce** yourself

002



1 My role
Ravi Sinha
2 How long I've been at Meta
Safety Policy Manager
3 Favorite tv show

Messenger Academy

**Introduce** yourself

003

# Our Agenda

- Safety & Wellbeing Overview
- Child Safety Foundations
- How we're taking action and what you can do


Warning
This program contains descriptions
of abuse that may be disturbing.

R

005



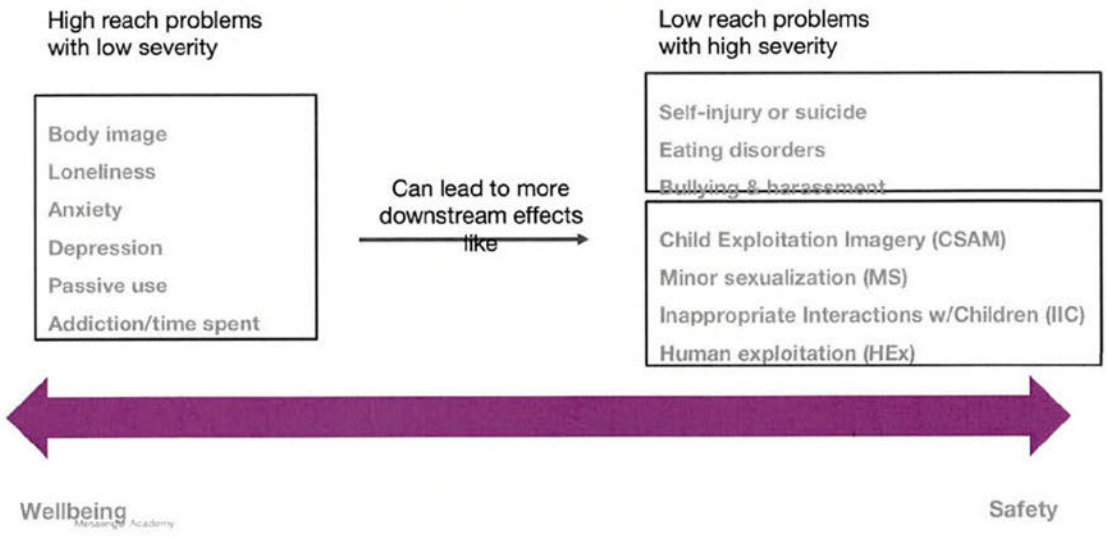
# Safety & Wellbeing Overview

Messenger Academy

S

006

# Physical and mental safety & wellbeing behaviors




S

Fills a spectrum of behaviors

007

# External impressions of youth safety & wellbeing

On our family of apps

Children are being "surveilled and manipulated" and "apps are pushing kids to spend unhealthy amounts of time online." - **US Senate**

"Facebook Messenger and Instagram could become "superplatforms" for pedophiles if end-to-end encryption is added, expert warns" - **Daily Mail, UK**

"Facebook contributes to significant harassment and cyberbullying that results in anxiety, depression, and loneliness among minors and young adults, especially women [who] are routinely extorted on Facebook" - **Bangladesh**

"Instagram most recorded platform used in child grooming crimes during lockdown" - **NSPCC**

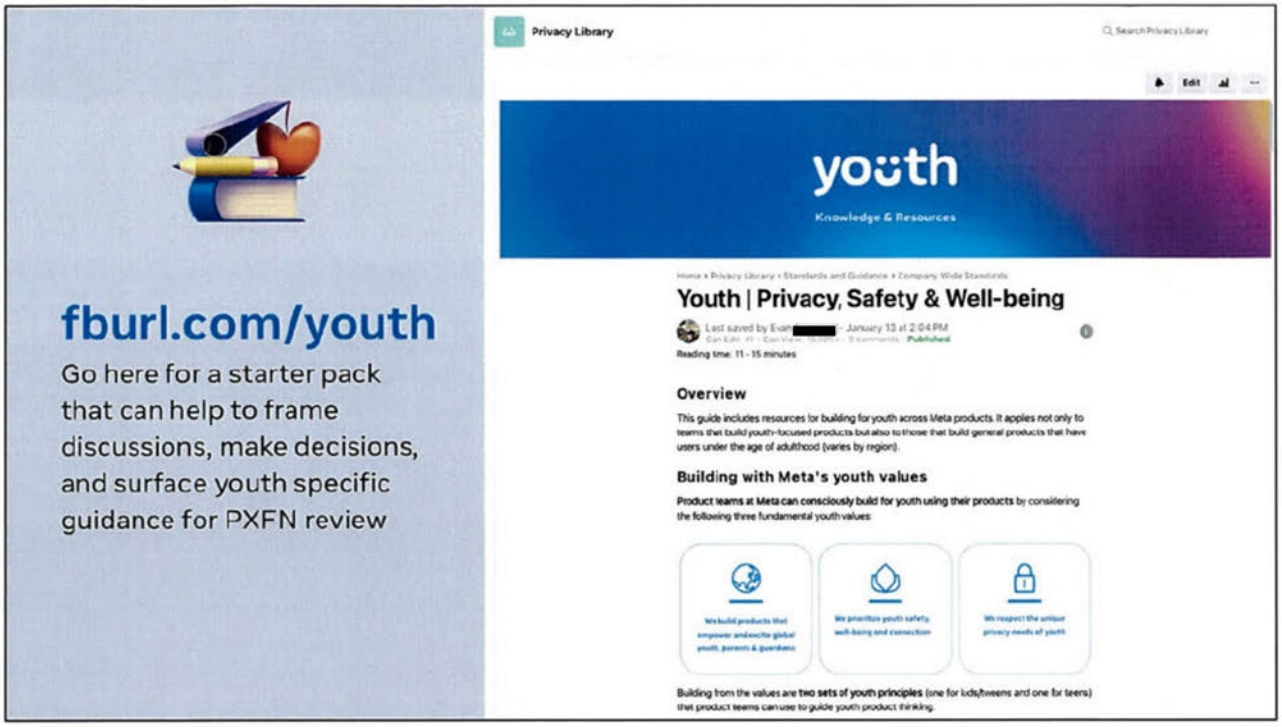
Messenger Academy

S

In addition to the endless headlines that look very much like these, regulatory risk is heating up, especially in the EU and India where legislators are proposing intermediary laws that hold tech companies and executives responsible for not taking down CSAM and NCII immediately and pressuring us to do more proactively to keep children and young people safe online.

On e2ee: World governments and child safety organizations continue to use child safety as the wedge issue to run coordinated campaigns against our move to e2ee. We're already seeing increased legislative and regulatory activity globally seeking to undermine e2ee protections.

NSPCC = National Society for the Prevention of Cruelty to Children, UK children's charity


fburl.com/youth
Go here for a starter pack that can help to frame discussions, make decisions, and surface youth specific guidance for PXFN review
Privacy Library
youth
Knowledge & Resources
Youth | Privacy, Safety & Well-being
Reading time: 11 - 15 minutes
Overview
This guide includes resources for building for youth across Meta products. It applies not only to teams that build youth-focused products but also to those that build general products that have users under the age of adulthood (varies by region).
Building with Meta's youth values
Product teams at Meta can consciously build for youth using their products by considering the following three fundamental youth values:
We prioritize youth safety, well-being and connection
We respect the unique privacy needs of youth
Building from the values are two sets of youth principles (one for kids/tweens and one for teens) that product teams can use to guide youth product thinking.

Sarah

009

## Different apps in our family define safety & wellbeing in distinctive ways



We envision an Instagram where everyone feels safe and supported to express themselves and to push culture forward



Build a safe and inclusive product, by minimizing harm and improving well-being of individuals and communities.



Helping people connect with friends and family, build community, and find and offer social support.



Minimize interactions that incite, inflict, or deepen physical harm while remaining world class on user privacy.

Messenger Academy

S

1. We envision an Instagram where everyone, including teens, creators, and underserved communities, feel safe and supported to express themselves and to push culture forward
2. Build a safe and inclusive product, by minimizing harm and improving well-being of individuals and communities. Our strategy to achieve this is through Prevent, Control, and Care
3. Helping people connect with friends and family, build community, and find and offer social support. Our strategy focuses on connections, bullying and harassment, emotional impact, & problematic use
4. Minimize interactions that incite, inflict, or deepen physical harm while remaining world class on user privacy

010

# Different apps in our family define safety & wellbeing in distinctive ways









Safety notices warning minors of suspicious actors

Direct messaging reachability controls

Muted words on IGD

Hiding Likes

SSI and ED interstitials and resources in-app

Teen privacy settings

Messenger Academy

MSGR crisis support

Forward limits

Unwanted contact protections

Sexual Solicitation policy

Comment filters

Remove minors from PYMK

Remove friend button for potential victims

CEI search intervention

Forward limits

Naming of groups on public surfaces

S

011

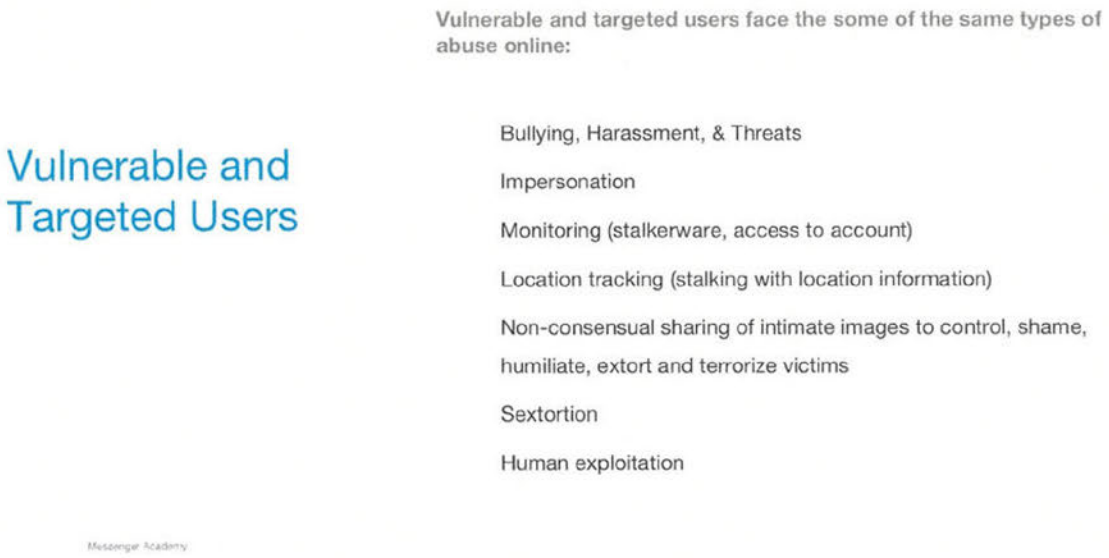


# Child Safety 101

Messenger Academy

R

012




R

Vulnerable users, such as children, and historically targeted users, such as women and members of the LGBTQ community, often face the some of the same types of abuse online. Though the focus of our presentation is on child safety--i.e. safety of vulnerable users-- it's important to call out that a lot of the safety concerns that targeted adult user face also apply to children. They include: (go through each one and give a quick definition).

Insert notes

# Bullying

For child safety

Messenger Academy

**S**

**Child Safety** includes the following areas of abuse:

- **CEI** (Child Exploitation Imagery) or CSAM (Child Sexual Abuse Material)
  - Images or videos that memorialize the sexual abuse of a child
  - Using our platform to further the production, distribution, advertisement or solicitation of CEI
- **IIC** (Inappropriate Interactions with Children), including (but not limited to):
  - Sexualizing minors
  - Engaging in sexual conversations with minors
  - Grooming a child for eventual solicitation, enticement, or coercion of self-produced CEI
  - Grooming a child for in-real-life sexual contact and abuse
  - Sex tourism to engage in sex with a minor
  - Grooming or engaging with an adult with access to a child to produce CEI or CSAM or engage in sex with the minor for purposes of producing CEI/CSAM.
- **Human Exploitation**

- Sexual solicitation and prostitution involving a minor
- Human trafficking and smuggling

- **Bullying and Harassment**
- Suicide & Self Injury (SSI)
- Physical abuse or other harmful/dangerous acts

**NCMEC** is the National Center for Missing & Exploited Children, which acts a clearinghouse for CEI reports

## Some important definitions

For child safety

**CSE** (Child Sexual Exploitation)

- Direct CSE (child is present)
  - **IIC** (Inappropriate Interactions with Children)
    - **IIC Tier 1** includes many abuse types, e.g. child sextortion, plans to meet a minor IRL for sex, online sex with a child, enticing a child to produce CSAM. **Please do not refer to this a 'real world harm'.**
    - **IIC Tier 2** sexually-motivated harassment of a child, creepy and sex talk with a child. **Please do not refer to this as 'online harm'.**
- Indirect CSE (child is not present)
  - **CSAM** (Child Sexual Abuse Material)
  - **CSx** (Child Sexualization)

**NCMEC** is the National Center for Missing & Exploited Children, which acts a clearinghouse for CSE reports

Messenger Academy

S

- Youth: "Youth" refers to anyone who is under the age of 18 or has not reached adulthood as defined by law in their jurisdiction. It is important to note that generally speaking, 13 years old is the minimum age for account registration across our Family of Apps with the exception of Messenger Kids, which is an app specifically designed for under 13 year olds with parental supervision. Certain jurisdictions have laws that set different age requirements. For example, in South Korea and Spain, people must be 14 years old to use our services.

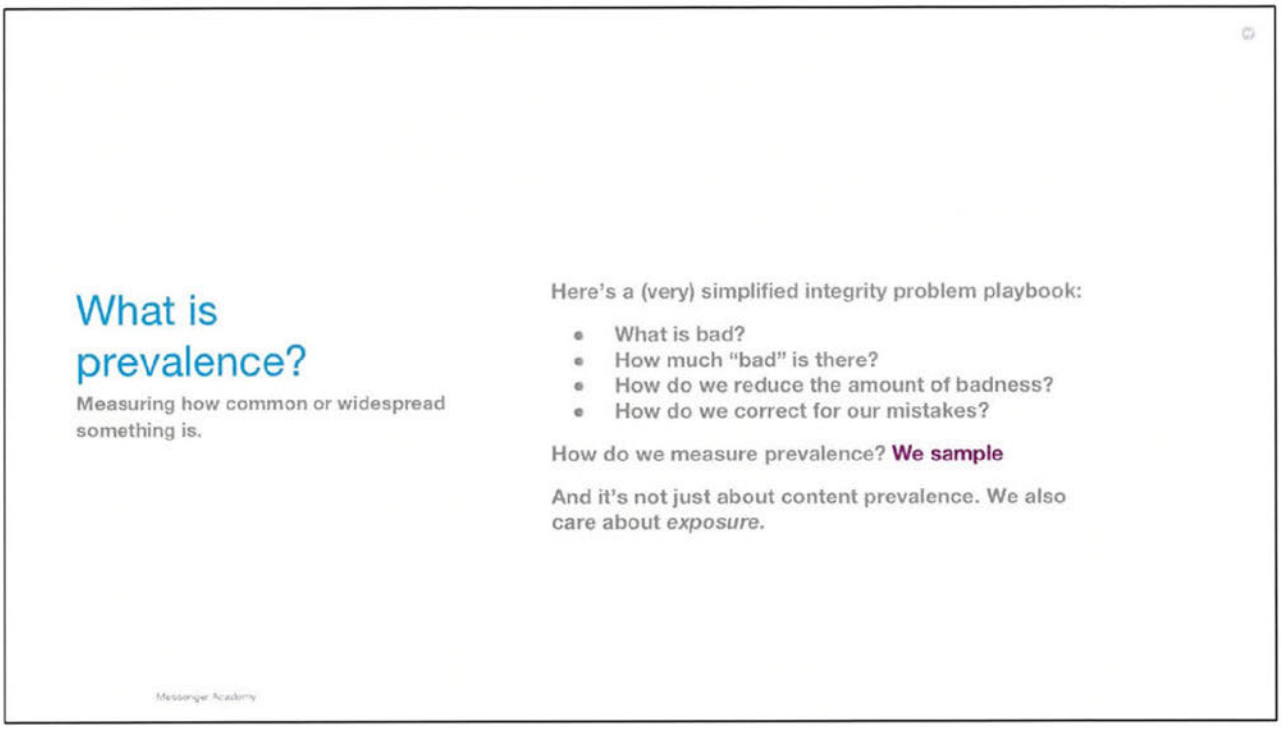
- Child Sexual Exploitation (CSE): Child Sexual Exploitation is a broad term (also used by our Community Standards) that encompasses all forms of child sexual exploitation and may include all other terms defined below, such as child sexual abuse material, child sex trafficking, and sextortion. We use this term to talk about the issue as a whole. Please note that Global Operations, Central Integrity, and Policy teams who focus on all Youth concerns included under CSE are referred to as the "Child Safety" team.

- CSAM (Child Sexual Abuse Material): Images, videos, live streaming, and digital or computer generated images indistinguishable from an

actual minor that depict sexually explicit activities involving a child.

**In the past, we described IIC T1 as 'real world harm' that required in-person sexual contact with a child (as opposed to 'online harm' which did not). However, this both does not reflect the threshold for illegal offences in many countries, but especially this excludes online interactions that external and internal research increasingly recognise as extremely harmful to the victim. We will instead describe these abuses as 'imminent IIC'.**

**NCMEC** is the National Center for Missing & Exploited Children, which acts a clearinghouse for CEI reports




**S**

Prevalence is a measure of how common or widespread something is.
It's an essential metric in integrity work because our fundamental goal is to reduce the amount of bad stuff on our platforms.

1. What is bad? (i.e. write a policy)
2. How much "bad" is there? <<<Prevalence goes here
3. How do we reduce the amount of badness? (eg. delete, down-rank etc)
4. How do we correct for our mistakes? (user reports, appeals, HERO, HIPO)

You might ask Why can't you just use your detection capabilities to measure how much badness there is? Why do you have to sample and human review? B/c a classifier might be GREAT at finding tabby cats but not great at finding photos of other breeds. If you're using that classifier to measure the overall prevalence of cat photos, you're incorrectly thinking that there are mainly tabby cat photos on FB and you won't be properly measuring how commonplace photos of ALL cats are. This is why a sampling methodology that is *independent* of your detection capabilities is so important.

One last note on prevalence in the integrity space -- we don't just care about content prevalence. You should aim to measure prevalence in a way that best captures the *harm* you're trying to prevent. For example, in the Child safety space, it matters that this memorialized instance of a child's sexual exploitation is being *seen* by hundreds of thousands of people. We call this VPV prevalence (the numbers of impressions of violating content).

# The impact on victims is devastating and lifelong

"It was the worst moment of my life," Nicole said of seeing the pictures of herself. "In a way, I didn't remember it being that bad with my father — and then I saw that it was. **Knowing that other people, all over, had seen me like that, I just froze. I could hear my mother crying, but I couldn't cry."**

"That's in my head all the time — knowing those pictures are out there," said E. **"Because of the way the internet works, that's not something that's going to go away."**

Messenger Academy

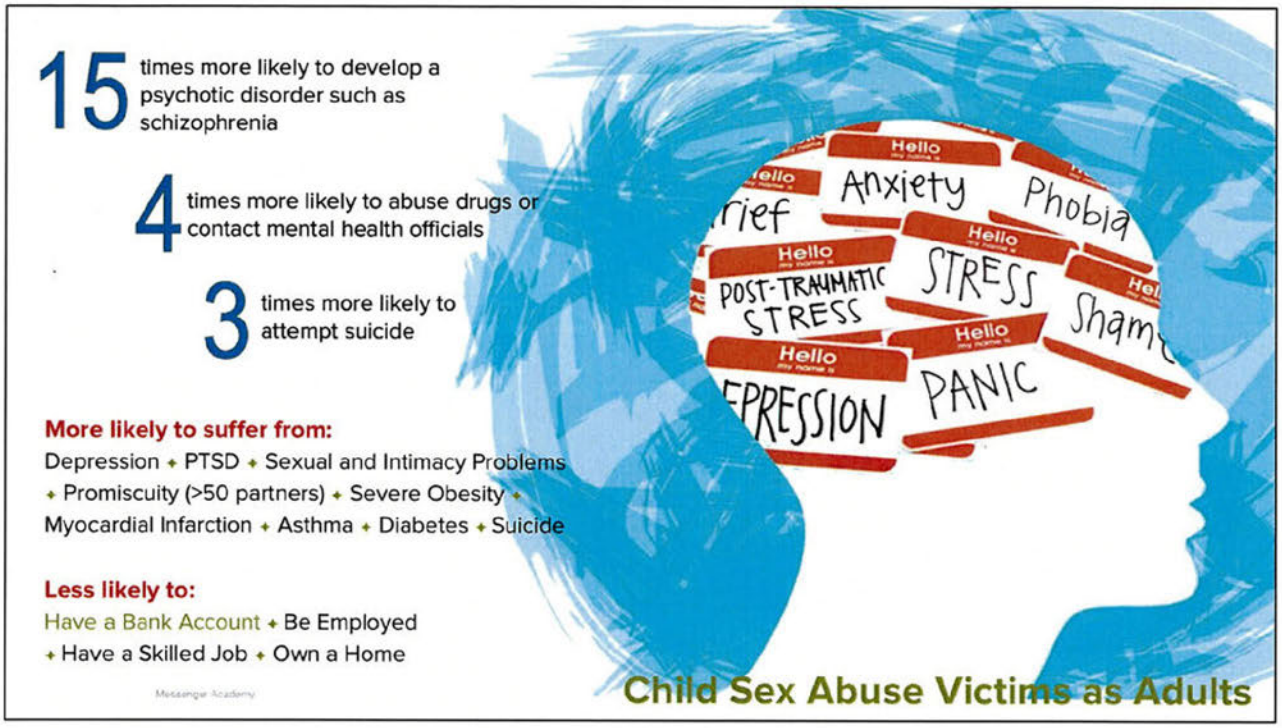
R

**NYTimes, 11/9/2019, "Child Abusers Run Rampant as Tech Companies Look the Other Way"**

**NYTimes, 1/27/13, "How Much Restitution Can Help Victims of Child Pornography"**

Photo: **https://americansecuritytoday.com/ncmec-john-f-clark-honored-in-2019-astors-awards-multi-video/**

NCMEC White Paper on why CEI is uniquely harmful to its victims




R

See generally: Janet Currie & Cathy Spatz Widom, Long-Term Consequences of Child Abuse and Neglect on Adult Economic Well-Being, 15 Child Maltreatment 2, 111-120 (2010); David. W. Brown et. al.; Adverse Childhood Experiences and the Risk of Premature Mortality, 37 Am. J. of Preventative Med. 5, 389, 396 (2009)(stressful childhood experiences have negative neurodevelopmental impacts leading to increases in behavioral, health, and social problems and associated with premature death); Stephen A. Wonderlich et. al., Eating Disturbance and Sexual Trauma in Childhood and Adulthood, 30 Int'l J. of Eating Disorders 4, 401, 401 (2001)(survivors of childhood sexual abuse exhibit high levels of eating disorder behavior and other self-destructive behaviors); Frank W. Putnam, Ten-Year Research Update Review: Child Sexual Abuse, 42 J. of the Am. Academy of Child & Adolesc. Psych. 3, 1, 5 (2003)(listing psychiatric conditions associated with childhood sexual abuse); Lex L. Merrill et. al., Predicting the Impact of Child Sexual Abuse on Women: The Role of Abuse Severity, Parental Support, and Coping Strategies, 69 J. of Consulting and Clin. Psych 6, 992, 1001(childhood sexual abuse is a significant predictor of long-term psychological difficulties); Patricia Cohen et. al., Child Abuse and Neglect and the Development of Mental Disorders in the General Population, 13 Dev. & Psychopath., 981-99 (2001)(reporting higher rates of personality disorders, anxiety, and substance abuse in childhood sexual abuse victims compared to other victims); Ingo Shafer et. al., Childhood Trauma and Psychosis – What is the Evidence?, 13 Dialogues in Clin. Neurosci. 3, 360-365 (2011)(biological mechanisms like reduced cortical thickness and dysregulated Cortisol, following exposure to childhood trauma, contribute to the development of psychosis); Abigail Powers et. al., Childhood Trauma, PTSD, and Psychosis: Findings from a Highly

Traumatized, Minority Sample, 58 Child Abuse Negl., 111-118 (2016)(early life trauma is connected to PTSD and psychotic disorders); Yael Dvir et. al., Childhood Trauma and Psychosis, 22 Child Adolesc. Psych. Clin., 629-641 (2013).
https://www.ajpmonline.org/article/S0749-3797(98)00017-8/fulltext;
https://www.frontiersin.org/articles/10.3389/fpsyg.2018.02535/full#B11
https://openriver.winona.edu/cgi/viewcontent.cgi?article=1072&context=counseloreducationcapstones;




**R**

**More likely to suffer from:**
Depression ✦ PTSD ✦ Sexual and Intimacy Problems ✦ Promiscuity (>50 partners) ✦ Severe Obesity ✦ Myocardial Infarction ✦ Asthma ✦ Diabetes ✦ Suicide

**Less likely to:**
Have a Bank Account ✦ Be Employed ✦ Have a Skilled Job ✦ Own a Home

See generally: Janet Currie & Cathy Spatz Widom, Long-Term Consequences of Child Abuse and Neglect on Adult Economic Well-Being, 15 Child Maltreatment 2, 111-120 (2010); David. W. Brown et. al.; Adverse Childhood Experiences and the Risk of Premature Mortality, 37 Am. J. of Preventative Med. 5, 389, 396 (2009)(stressful childhood experiences have negative neurodevelopmental impacts leading to increases in behavioral, health, and social problems and associated with premature death); Stephen A. Wonderlich et. al., Eating Disturbance and Sexual Trauma in Childhood and Adulthood, 30 Int'l J. of Eating Disorders 4, 401, 401 (2001)(survivors of childhood sexual abuse exhibit high levels of eating disorder behavior and other self-destructive behaviors); Frank W. Putnam, Ten-Year Research Update Review: Child Sexual Abuse, 42 J. of the Am. Academy of Child &

Adolesc. Psych. 3, 1, 5 (2003)(listing psychiatric conditions associated with childhood sexual abuse); Lex L. Merrill et. al., Predicting the Impact of Child Sexual Abuse on Women: The Role of Abuse Severity, Parental Support, and Coping Strategies, 69 J. of Consulting and Clin. Psych 6, 992, 1001(childhood sexual abuse is a significant predictor of long-term psychological difficulties); Patricia Cohen et. al., Child Abuse and Neglect and the Development of Mental Disorders in the General Population, 13 Dev. & Psychopath., 981-99 (2001)(reporting higher rates of personality disorders, anxiety, and substance abuse in childhood sexual abuse victims compared to other victims); Ingo Shafer et. al., Childhood Trauma and Psychosis – What is the Evidence?, 13 Dialogues in Clin. Neurosci. 3, 360-365 (2011)(biological mechanisms like reduced cortical thickness and dysregulated Cortisol, following exposure to childhood trauma, contribute to the development of psychosis); Abigail Powers et. al., Childhood Trauma, PTSD, and Psychosis: Findings from a Highly Traumatized, Minority Sample, 58 Child Abuse Negl., 111-118 (2016)(early life trauma is connected to PTSD and psychotic disorders); Yael Dvir et. al., Childhood Trauma and Psychosis, 22 Child Adolesc. Psych. Clin., 629-641 (2013).
https://www.ajpmonline.org/article/S0749-3797(98)00017-8/fulltext;
https://www.frontiersin.org/articles/10.3389/fpsyg.2018.02535/full#B11
https://openriver.winona.edu/cgi/viewcontent.cgi?article=1072&context=counseloreducationcapstones;

## High political and reputational risk

We continue to be at the center of threats to e2ee.

Child safety continues to be used as a wedge issue by world governments

FB contributes over 90% of the total reports received by NCMEC

We are feeling increasing pressure to demonstrate end-to-end encryption resilient safety mitigations

Encrypted mega chats like Signal are already receiving public scrutiny for lack of safety mitigations

Messenger Academy

S

World governments and child safety organizations continue to use child safety as the wedge issue to run coordinated campaigns against our move to e2ee. We're already seeing increased legislative and regulatory activity globally seeking to undermine e2ee protections.

In 2021, Messenger was responsible for over 80% of all reports of child sexual exploitation made by Facebook to Law Enforcement. Out of 29.3M total reports, Meta provided 25.5M.

Child safety continues to be used as a wedge issue by world governments to feed the narrative that we cannot be trusted to moderate ourselves or get e2ee right, fueling content moderation and liability laws.

Our NCMEC numbers continue to be used against us. FB contributes >90% of the total reports received by NCMEC and >80% of CEI content reported by FB is from MSGR. The public is aware and concerned we will lose the ability to enforce post-e2ee.

024

We are feeling increasing pressure to demonstrate end-to-end encryption resilient safety mitigations that maintain our privacy promises and measurably improve safety (particularly in areas of key government concern).

Encrypted mega chats like Signal are already receiving public scrutiny for lack of safety mitigations.



## Messenger

Where we detect most of our CSAM reports made to authorities

Over 80% of CSAM reported by FB is from MSGR.
At least 50% of CSAM deletes are a result of *retroactive* proactive detection (not caught at time of upload).

***Most of what's shared in MSGR are images not yet banked.***

***And with e2ee coming, an enforcement model which relies on finding bad content and bad actors after the harm is done is not enough, which is why we are pivoting to prevention.***

Messenger Academy

S

- Important to note that this doesn't necessarily mean a 1:1 ratio of image to victim. There are a lot of duplicates and similar photos that circulate of a single image. We're doing some important contextualization work here which you can read about in Antigone Davis' NRP, Preventing Child Exploitation on our Apps:

- We found that more than 90% of this content was the same as or visually similar to previously reported content. And copies of just six videos were responsible for more than half of the child exploitative content we reported in that time period. While this data indicates that the number of pieces of content does not equal the number of victims, and that the same content, potentially slightly altered, is being shared repeatedly, one victim of this horrible crime is one too many.
  - At least 50% of CEI deletes are a result of *retroactive* proactive detection (not caught at time of upload).
  - ***Most of what's shared in MSGR are new images*** for which we have no hash set and it's only after our CEI proactive detection catches it or it's posted somewhere on our public platform and someone reports it that it's possible for us to bank the image, run detection across threads and find that hundreds or thousands of copies have already been distributed via MSGR.

## Why Prevention?

### Enforcement requires that children are harmed

Focusing on finding actors after they abuse by definition requires a children to be harmed. We know that safety isn't top of mind for teens and they aren't likely to report. [1]

### Prevention is possible for potential violators

Emerging research highlights that it is possible for potential violators to moderate their behaviour. [2] Many people who violate are opportunistic offenders or just beginning to act out on desires—there is help available and measures we can take to make our surfaces less suitable for abuse.

### This is the only feasible way post-E2EE

Without the ability to find inbox-based abuse, we have to stop abuse before it has the chance to reach encrypted surfaces.

## Meta Teen Safety Strategy

**Prevent:** Pivot resources upstream to prevent harm. Stop possible violators (identified with e2ee-resilient behavioral means) from interacting with children, with one another, and with child sexual abuse material (CSAM).

**Control:** Empower minors with options to control their experience, so they can better recognize harm, and remove friction for minors to report harm they experience to us.

**Care:** Increase recall and volume of reports, and provide support to minors who are victimized on our platform.

Attorney Client Privilege - Confidential, Internal Use Only



**Sarah**



## Questions?

Messenger Academy

S

- In 2019, Messenger was responsible for over 80% of all reports of child sexual exploitation made by Facebook to Law Enforcement.
- 94% of CEI Content Deletions as a result of proactive detection happen on MSGR.
- At least 50% of CEI deletes are a result of *retroactive* proactive detection (not caught at time of upload).
- ***Most of what's shared in MSGR are new images*** for which we have no hash set and it's only after our CEI proactive detection catches it or it's posted somewhere on our public platform and someone reports it that it's possible for us to bank the image, run detection across threads and find that hundreds or thousands of copies have already been distributed via MSGR.

Important to note that this doesn't necessarily mean a 1:1 ratio of image to victim. There are a lot of duplicates and similar photos that circulate of a single image. We're doing some important contextualization work here which you can read about in Antigone Davis' NRP, Preventing Child Exploitation on our Apps:

We found that more than 90% of this content was the same as or visually similar to previously reported content. And copies of just six videos were responsible for more than half of the child exploitative content we reported in that time period. While this data indicates that the number of pieces of content does not equal the number of victims, and that the same content, potentially slightly altered, is being shared repeatedly, one victim of this horrible crime is one too many.

## Additional risks

### Child Safety Risks during COVID

In 2020, FBI issued a warning that school closings due to COVID presented increased risk for child exploitation on the Internet.

Increasing efforts to impose ESP intermediary liability for child safety in US, India, PH, Korea, etc.

Interpol reported seeing increased online activity by pedophiles seeking CSAM in 2020

The age management landscape is growing increasingly hostile while our efforts are largely hamstrung by our inability to verify age and predict age globally

Coordinated campaigns to build back doors or stop e2ee altogether
Groomers are viewing quarantine as opportunity when kids are more vulnerable to the psychosocial manipulation tactics

### Regulatory risk

### Brand and Comms Risk

"These virtual spaces are essentially hunting grounds", Mary Anne Franks, president of Cyber Civil Rights Initiative.

Multiple exposes in the New York Times detailing child exploitation on Facebook and sexual predators in gaming communities.

NCMEC says reports of child sexual exploitation on the Internet have surged since March 2020

Messenger Academy

S

Brand and Comms Risk

- "These virtual spaces are essentially hunting grounds", Mary Anne Franks, president of Cyber Civil Rights Initiative.
- *Child Abusers Run Rampant as Tech Companies Look the Other Way* (New York Times, November 2019) - detailing child exploitation on Facebook and other online platforms.
- *Video Games and Online Chats are 'Hunting Grounds' for Sexual Predators* (New York Times, December 2019) - detailing how sexual predators use online gaming platforms and chat apps to dupe children into sharing sexually explicit photos and videos of themselves, used to blackmail them for more imagery.

Regulatory Risk

- Given regulatory environment linked to our e2ee announcement (EARN IT, UK technical notice, and DOJ's June 17, 2020 announcement of its official recommendations for CDA reform) launching products without safety mitigations plays into accusations that we are abandoning children to black boxes of abuse and could refuel this discussion, which lost momentum due to focus on COVID pandemic but is back and we expect continued and increasing

030

attention on e2ee and its connection to child protection overall in 2021 fueled by a spike in cybertips in NCMEC during this period, by upcoming product advancements towards interoperable e2ee messaging and changes in the policy landscape.

- The age management landscape is growing increasingly hostile, especially in Europe, the US, and India. We are facing heightened scrutiny from clients, regulators, and advocates around all aspects of our handling of age — collecting age, checking whether minors are lying about age, and proactively applying protections for minors in areas such as safety and marketing.

Child Safety Risks during COVID

- On March 23, 2020, FBI issued a warning that school closings due to COVID presented increased risk for child exploitation on the Internet.
- In April, 2020, Interpol reported seeing increased online activity by pedophiles seeking CSAM.
- NCMEC says reports of child sexual exploitation on the internet have surged since March 2020 and there is Dark Web chatter amongst CSAM offenders coaching one another to take advantage of quarantine to target more children.
- On May 14, 2020, A21 (NGO), NCMEC, and DOJ teamed up to present a FB Live event highlighting this issue, stating that groomers are viewing quarantine as opportunity when kids are more vulnerable to the psychosocial manipulation tactics used by groomers.



The offenders

Messenger Academy

S

## Taxonomic groupings of CSAM actors

**Malicious Users**

- Preferential offenders
- Commercial offenders
- Situational offenders

**Non-malicious Users**

- Unintentional offenders
- Minor non-exploitative users
- Situational "risky" offenders

Messenger Academy

S

In the "malicious" group are people we believe intended to harm children with their behavior, and in the "nonmalicious" group are people whose behavior is problematic and potentially harmful, but who we believe based on contextual clues and other behaviours likely did not intend to cause harm to children. For example, they shared the content with an expression of abhorrence.

The category "minor nonexploitative" came from feedback from our partners about the consensual, developmentally appropriate behavior that older teens may engage in. Throughout our consultations, a number of global experts, organizations, and academic researchers highlighted that children are, at times, a distinct grouping in some ways. Children can and do sexually offend against other children, but children also can engage in consensual, developmentally appropriate sexual behavior with one another. While the content produced is technically illegal and the behavior risky – as this imagery can be later exploited by others – it was important for us to separate out the nonexploitative sharing of sexual imagery between teens.

Lastly, we observed users who were sharing a large amount of adult sexual content and amongst that content was child exploitative imagery, to which the user was potentially

unaware that the imagery represented a child (which may depict children in their late teens, whose primary and secondary sexual characters appear developed to that of adulthood). We report these images to NCMEC as they depict the abuse of children that we are aware of, but the users in these situations may not be aware that the image depicts a child. We are still concerned about their behavior and may want to offer interventions to prevent any escalation in behavior.

To read the full research paper published by FB, see here.

## New risks, qualitatively different




Messenger Academy

R

Groups are fundamentally different than single actors:

- 'Wouldn't have done it but for the group'
- Normalize behavior
- Use encouragement and peer pressure to get additional material/abuse
- Direct new--usually more painful and demeaning---types of abuse or torture
- Lionize prolific abusers- people who might not get positive feedback elsewhere
- Use sophisticated and organized methods to produce new CEI at scale
- Prolific producers

## Why are online networks so valuable?

**Community**
Offenders can now reach across the globe to find others

**Trade Best Practices**
Offenders can swap expert knowledge

**Remote Presence**
They don't need to be physically present to abuse

**Novelty**
Presents opportunities to acquire or produce exclusive content and gain notoriety

**Collection and Storage**
Repositories of "high value" content

Messenger Academy

R

- **Community:** Instead of being alone with their stigmatized sexual desires for children, they can now reach across the globe and create virtual communities and sub-cultures where sex with children is normalized and encouraged. This then turns into organized crime.
- **Trade Best Practices:** Swap expert knowledge on how to best target and groom children for sexual abuse and evade detection by LE and tech companies.
- **Remote Presence:** They don't have to be physically present in a room in order to participate in a child's abuse (by watching or giving live direction).
- **Novelty:** It presents opportunity to help one another produce exclusive content with "signatures" (hand-written notes to one another featured in videos or victims marked with an offender's handle or site name) or "proof" that the video stream is in fact live and "fresh."
- **Collection and Storage:** Community found through the internet becomes source for "high-value" content and storage of treasured collections. They collect "series" and "legendary" CEI to boast about.

## Social Media + Messaging + Video Calling + Payments:

**Can become tools for bad actors**

- Target-rich environment with low entry barriers
- Research targets and tailor approach
- Use friend graphs to locate next victims
- Work at volume with minimal consequences
- Use messaging for grooming
  - Limited friction and risk:
    - Private video chat reduces risk to offenders and makes victims more willing to participate (versus recording)
    - Group video calls allow for private strip or sex shows where all attendees are willing participants (and unlikely to report)
    - On-platform video and payment facilitates real-time quid pro quo

Messenger Academy

Ravi - LEGAL OBLIGATIONS, ASK EINAT

Intersection of these 3 = complex risk

- Target-rich environment with low entry barriers
- Research targets and tailor approach
- Use friend graphs to locate next victims
- Work at volume with minimal consequences
- Use messaging for grooming
  - Limited friction and risk:
    - Private video chat reduces risk to offenders and makes victims more willing to participate (versus recording)

- Group video calls allow for private strip or sex shows where all attendees are willing participants (an unlikely to report)
- On-platform video and payment facilitates real-time quid pro quo



## Anthony Lozowski

19-year old accused of sexually exploiting girls on Instagram

| | |
|---|---|
| Used IG and IGD to contact minors | Groomed minors, victims as young as 11, obtained feet photos, moved to extorting CEI and sex acts.<br><br>Lozowski claimed to have exploited ~100 minors. |
| Victims as young as 11 | Minors consistently begged him to stop and told him they were suicidal.<br><br>Told 12-year-old that he would distribute her images unless she travelled to have sex with him. |
| Post Miranda Interview | Said **it was so easy** |

Messenger Academy

R

More info.

# Paul Stabile case study




R

Messenger and RTC are used by groomers to befriend a child, gain their trust, and eventually get them to self-produce CEI and transmit them via chat thread or RTC.

1. October 2017 - Paul J. Stabile, 39 year old Rhode Island man creates online persona "Cindi Mennillo" to pretend to be a 16-year-old girl from Newport, RI.
2. December 2017 - Paul J. Stabile posing as "Cindi" befriends numerous young boys between the ages of 13-15 and persuades them to engage in sexually explicit activity over RTC and Skype.
3. 2018 - Investigation begins when mother of a 14 year old Nevada boy reports to the FBI that her son was a victim of online enticement.
    a. MSGR Chat threads between Stabile, pretending to be "Cindi", and the Nevada boy shows that Stabile had directed the minor victim to send live video streams of himself masturbating. Stabile falsely claimed to the boy that technical difficulties prevented "Cindi" from reciprocating with live video. Instead, Stabile would send the minor victim sexually explicit CEI images of a minor purported to be "Cindi."
4. November 2018 - Federal Search Warrant executed. Computers and other media devices seized reveal CEI images and videos, including videos of dozens of young teenage boys performing sex acts in a similar manner to the Nevada-based minor

victim. Images of "Cindi" sent to the minor victims are also found.



Questions?

Messenger Academy



NCMEC

Messenger Academy

R




R

Rescuing children and preventing abuse by:

- Allowing law enforcement to identify and prioritize repeat and high-volume child exploitation offenders
- Providing probable cause for residential search warrants and arrest warrants for:
  - Charges based directly on the NCMEC-reported activity
  - Other online child exploitation offenses by showing lack of mistaken identity
  - Related charges such as child rape by showing an abuser's sexual interest in children
- Supporting search warrants for digital devices, such as phones, where both hands-on and online child sex offenders commonly store evidence of their abuse
- Supporting felony indictments for:
  - Charges based directly on the NCMEC-reported activity
  - Other online child exploitation offenses by showing lack of mistaken identity
  - Related charges, such as child rape and incest
- Supporting pretrial detention of repeat child predators
- Corroborating child victims' disclosures to authorities by showing an

044

abuser's:

- Charges based directly on the NCMEC- reported activity
- sexual interest in children
- specific age or gender affinity
- interest in specific acts or abuse types such as S&M, scat, or infants

- Supporting and corroborating child rape and online child exploitation victims' trial testimony
- Preventing child victims from having to testify by facilitating plea agreements to avoid mandatory minimums
- Enhancing sentences of child predators
- Supporting Child Protective Services efforts to remove child victims from the homes of sexual predators

## The value of NCMEC and user reports

Reported: 1/13/20
Investigated: 1/14/20
Child Rescued: 1/15/20

**Facts:** A 34-year-old female and a 41 year-old male used MSGR and MK to entice a 7-year-old female to produce and send CEI and engage in sexual activity.

Conversations indicate that the adult female may have direct access and proximity to the child as her mother and may have engaged in sexual activity with the child.

41 y.o. male: I want a picture of you in my favorite pose please 😈🌹
7 y.o. girl: *Sends an image depicting a prepubescent female in underwear on a bed.*
41 y.o. male: That's beautiful!! But your legs are suppose to be spread open for me babygirl
7 y.o. girl: *Sends an image depicting a prepubescent female in underwear on a bed, spreading her legs open*
41 y.o. male: Omg😍😍😍😍. PERFECT 😍😍👅🍑. *Sends an image depicting adult male genitalia.*
7 y.o. girl: Arevyou going to ido call me or whar
41 y.o. male: Love you
7 y.o. girl: I love you more

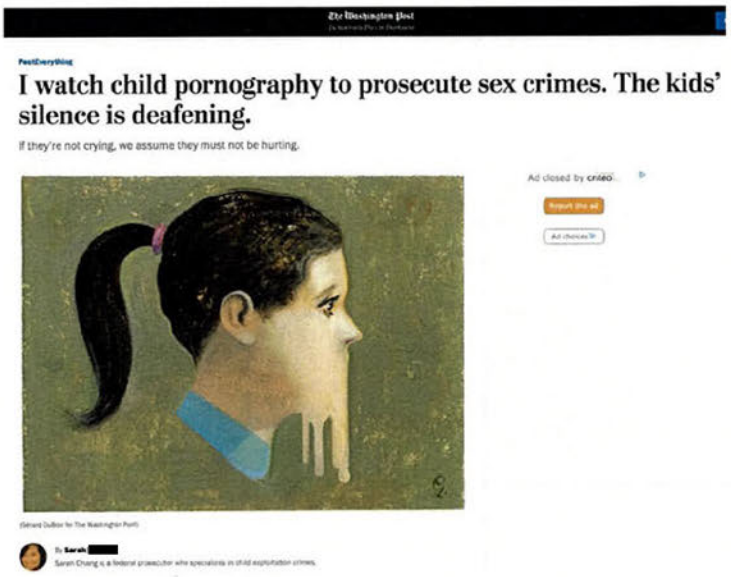
Messenger Academy

R

In addition to that case, we had two other cases in the same 6 month period where a child was rescued within days of the user reports:
Reported: 4/16/20
Investigated: 4/16/20
Child Rescued: 4/20/20
Two additional children rescued: 4/22/20

Reported: 5/7/20
Investigated: 5/7/20
Child Rescued: 5/8/20

# User Reporting is important but it's not enough



The Washington Post

PostEverything

**I watch child pornography to prosecute sex crimes. The kids' silence is deafening.**

If they're not crying, we assume they must not be hurting.

Ad closed by criteo.

- Minors rarely report their own abuse.
- Tech savvy offenders engage in extensive vetting of one another when sharing CEI or exchanging tips.
- Bystanders will have limited access or visibility into what's really happening.
- For more insight, read Sarah Chang's op-ed.

R

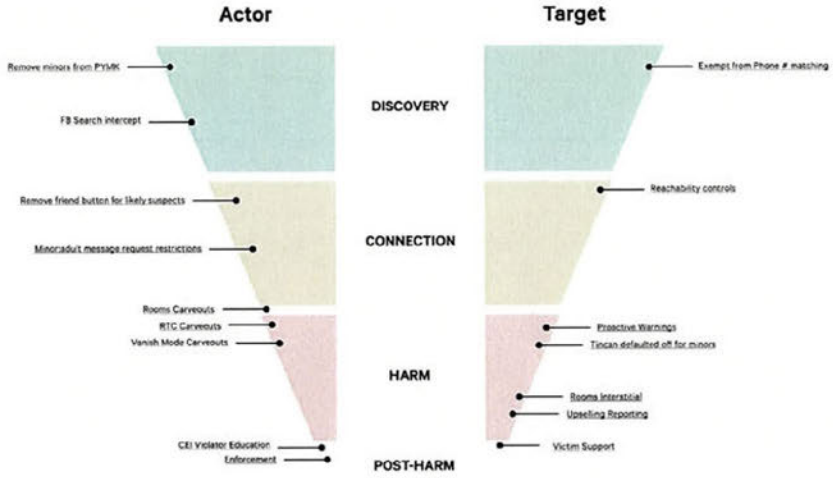


# Taking action

Messenger Academy

## Interventions

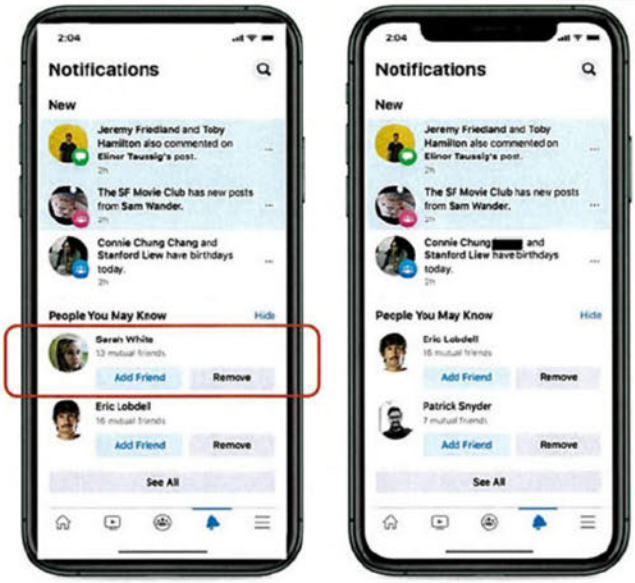
We have made some progress in indentifying and intervening to stop and enforce on abuse.



Internal use only: DO NOT DISTRIBUTE

## Prevent victim discovery at scale

We remove teens from the suggested friends of potentially suspicious adults.

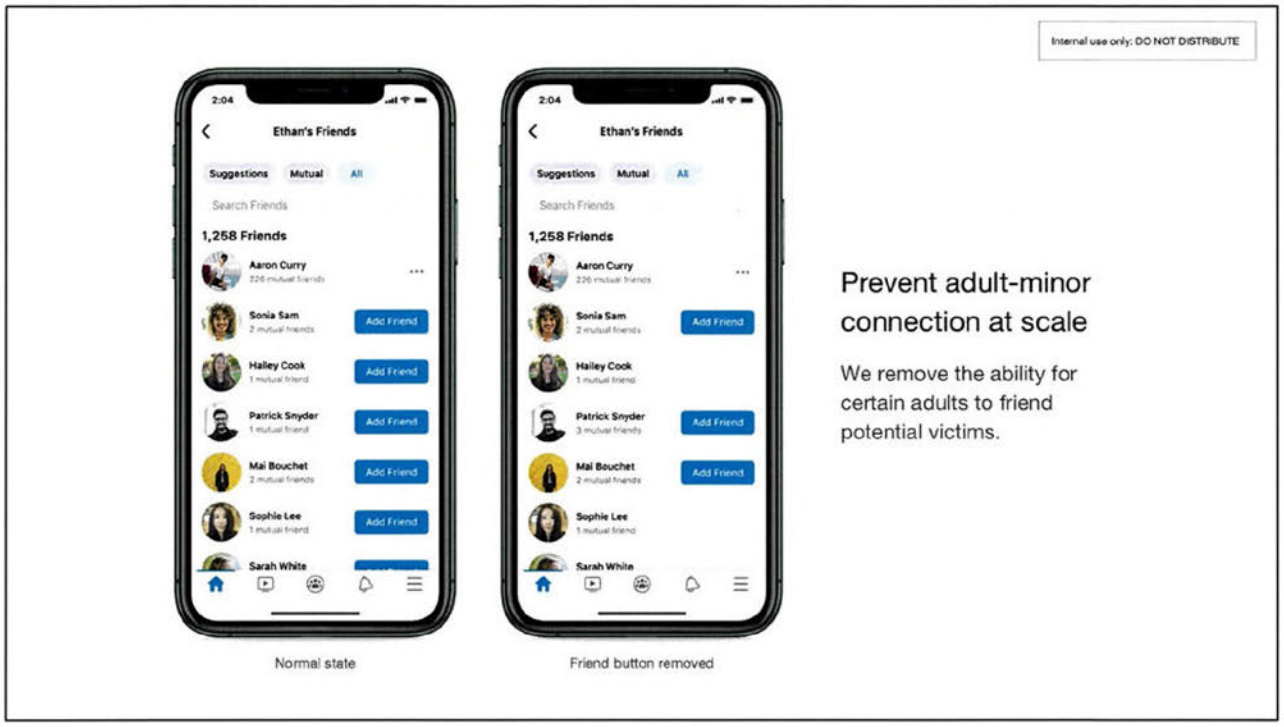



Minor in suggestions

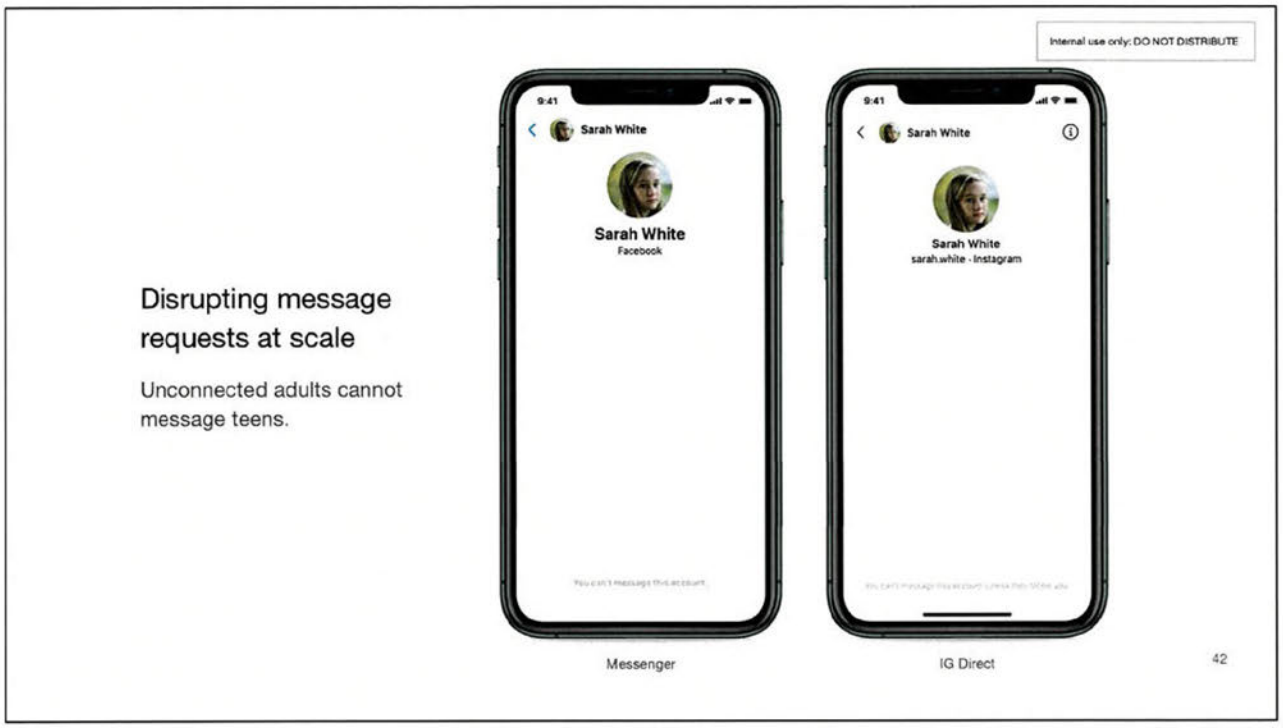
Minor removed

SARAH:
In an end-to-end encrypted environment, we will use artificial intelligence to proactively detect accounts engaged in malicious patterns of behavior instead of scanning your private messages. Our machine learning technology will look across non-encrypted parts of our platforms — like account information and photos uploaded to public spaces — to detect suspicious activity and abuse.

For example, if an adult repeatedly sets up new profiles and tries to connect with minors they don't know or messages a large number of strangers, we can intervene to take action, such as preventing them from interacting with minors, and one specific step we can take that you see illustrated here is to remove minors from suggested friends.



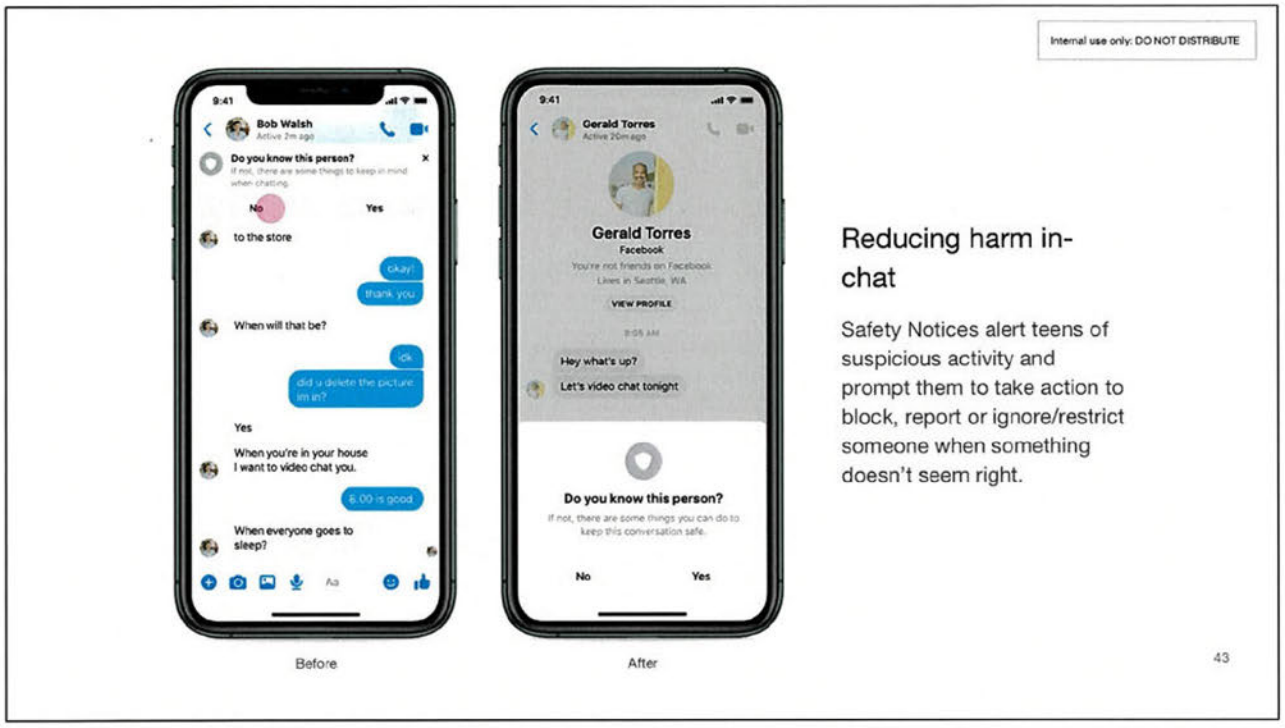

SARAH
We're also taking steps to remove the ability for certain adults to friend potential victims by removing the Friend button altogether, for example.




SARAH
To further protect teens from unwanted contact from adults, we're preventing adults from sending messages to people under 18 who don't Follow them or are connected to them.

[further detail if asked: For example, when an adult tries to message a teen who doesn't follow them, they receive a notification that DM'ing them isn't an option. This feature relies on our work to predict peoples' ages using machine learning technology, and the age people give us when they sign up. As we move to end-to-end encryption, we're investing in features that protect privacy and keep people safe without accessing the content of DMs.]

[If pressed on further detail for age prediction: We train the technology using multiple signals. We look at things like people wishing you a happy birthday and the age written in those messages, for example, "Happy 21st Bday!" or "Happy Quinceañera." We also look at the age you shared with us on Facebook and apply it to our other apps where you have linked your accounts and vice versa — so if you share your birthday with us on Facebook, we'll use the same for your linked account on Instagram. This technology isn't perfect, and we're always working to improve it, but that's why it's important we use it alongside many other signals to understand people's ages.]



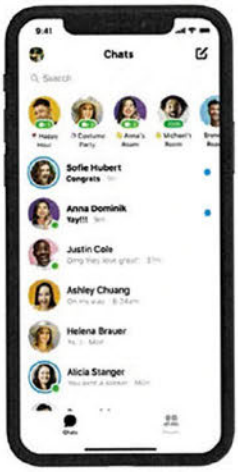

SARAH:

We also educate young people with in-app advice on avoiding unwanted interactions. Safety Notices alert minors of suspicious activity and prompt them to take action to block, report or ignore/restrict someone when something doesn't seem right.

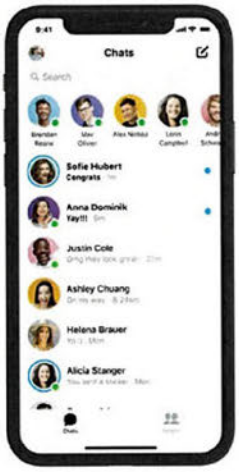
We've recently improved the design of this product to be more prominent and harder to dismiss/ignore b/c you have to choose yes/no to return to the chat.

## Carveouts

We also gate high risk features from potential suspects.




New Accounts can't create Rooms




IIC Suspects > 0.6 can't use Rooms

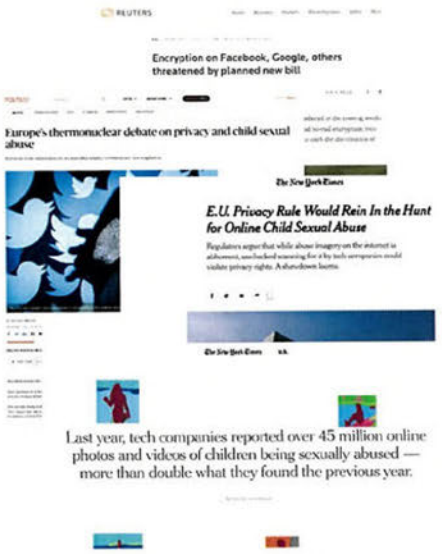



IIC Suspects > 0.6 can't use VM
(Sends fail silently)

Facebook
Messenger

High Level Takeaway

**Our current strategy focused on detection and enforcement is not working well today, and will slowly stop working as we move towards E2EE.**



## Prevent harm to children in our community

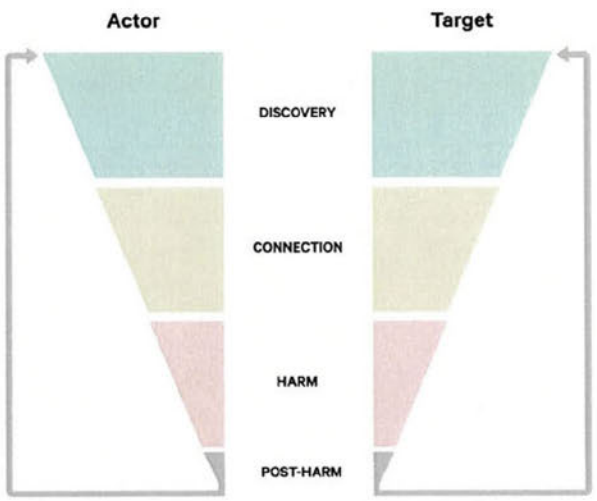
**Prevent** possible violators from:

1. Interacting with children.
2. Interacting with one another.
3. Interacting with child sexual exploitation material (CSEM) or child nudity.

Facebook
Messenger

## Cycle of Abuse

Today, those who offend continue to, and those who are victims may continue to be victimised, forming a vicious cycle.



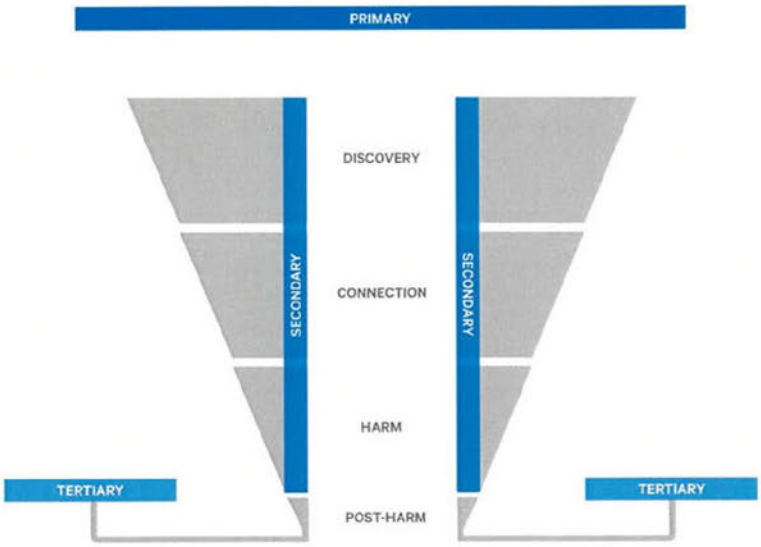
Facebook
Messenger

## How we apply Prevention Strategy

**Primary prevention** helps reduce the total pool of potential offenders and victims.

**Secondary prevention** shrinks the width of each stage of the funnel.

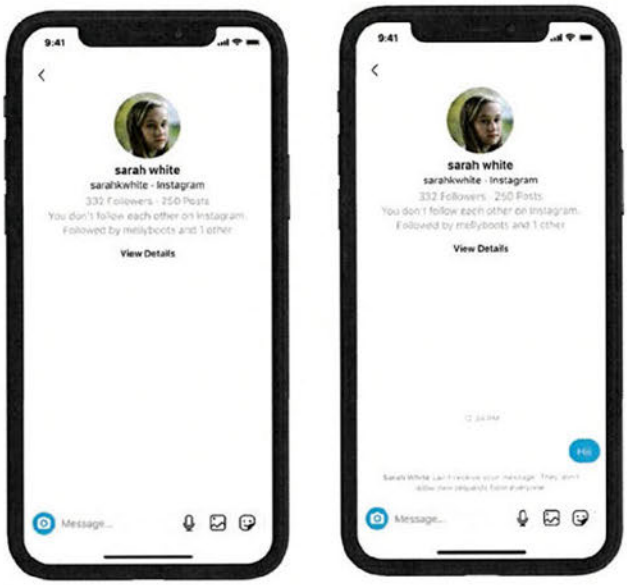
**Tertiary prevention** breaks the cycle.



⏳ **Things in Progress**

## IGD Message Request Restrictions

Adults will not be able to message minors who don't follow them back.




Thread Null

Send Fails

Facebook
Messenger

## Groups Restrictions

Work is underway to develop policies and techniques for how to enforce on groups, given we're losing MMS and content generally to prevent CEI trading.

There are a lot of encryption-resilient signals we can use to detect questionable groups and limit their ability to grow and distribute material.




### Potential Signals

Large group size
CEI shared in group
Many members IIC score
High FAI scores

### Potential Restrictions

Disable adding new members
Disable forwarding
De-link
No media sharing
Upsell reporting

# How you can get involved

Additional resources

Youth XFN FYI

Safety Center

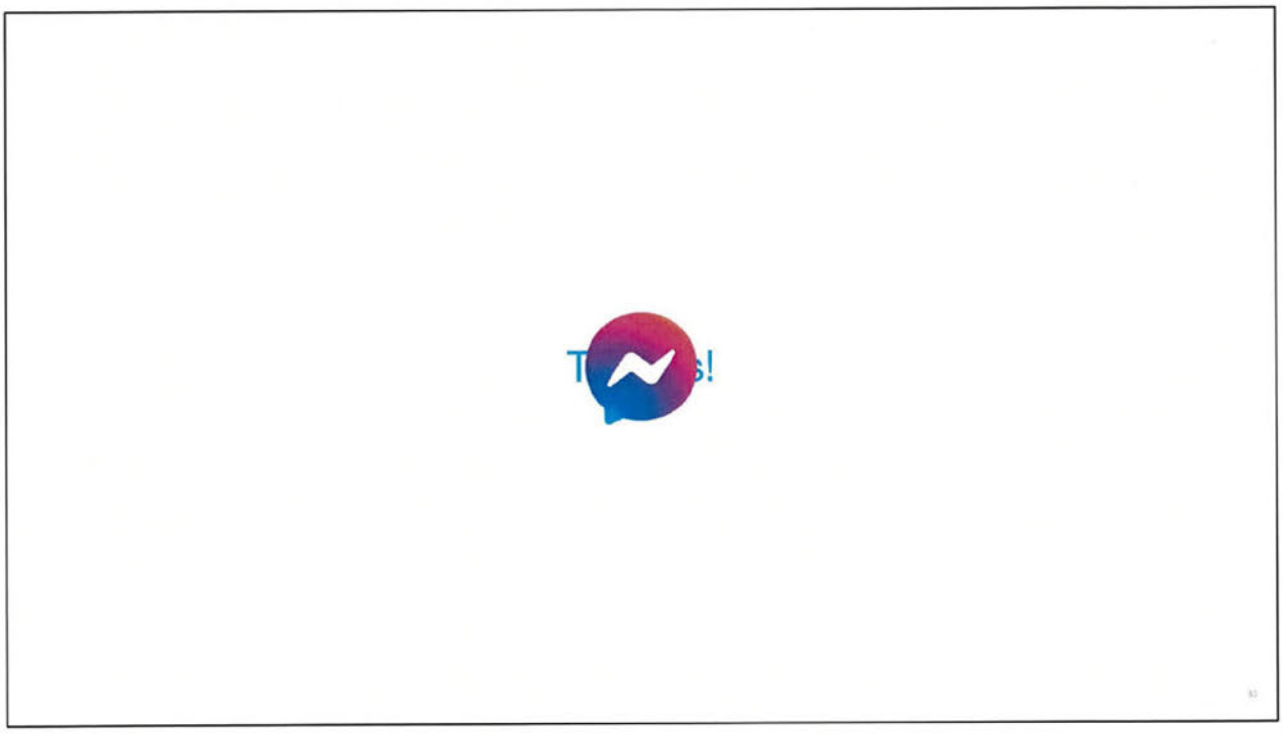
Messenger Well being FYI

Child Safety FYI

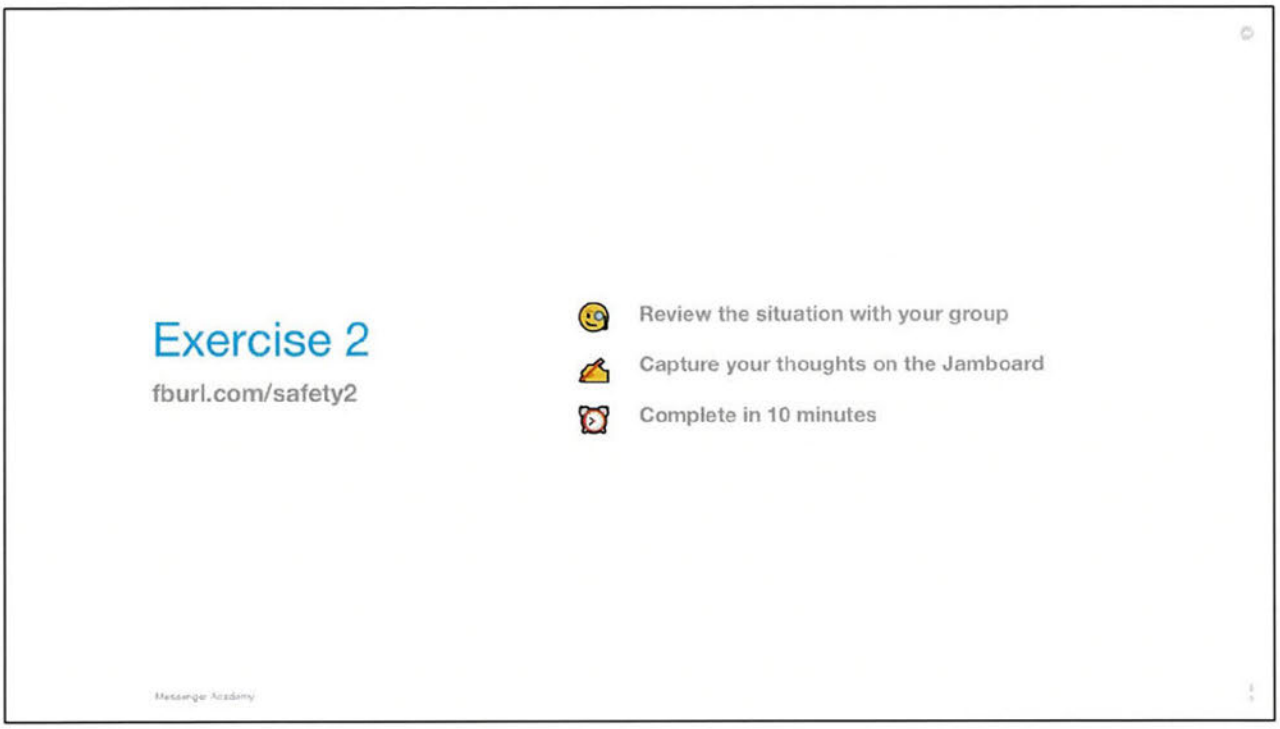
THORN.org

Youth Safety & Wellbeing Guide: Child Sexual Exploitation (CSE)

Youth Knowledge Library



# Appendix


Exercise 2
fburl.com/safety2
Review the situation with your group
Capture your thoughts on the Jamboard
Complete in 10 minutes
Messenger Academy

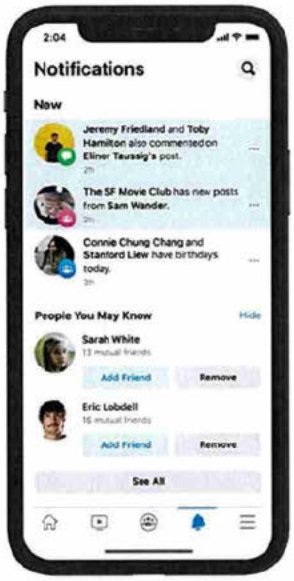
R

Graveyard

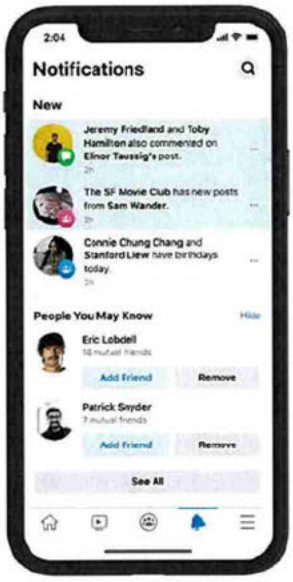
## Remove Minors from PYMK

Before this intervention, **59%** of Tier 2 IIC conversations were initiated by friends that originally connected via suggestions provided via PYMK.

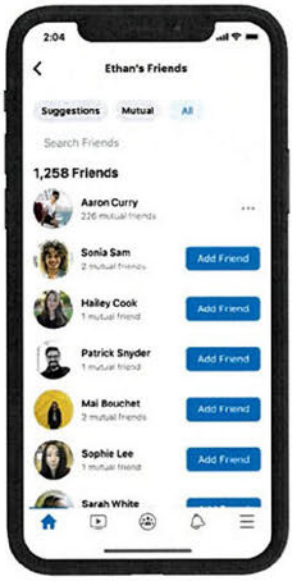
This intervention removes minors from PYMK for users who have an IIC classifier score [redacted]



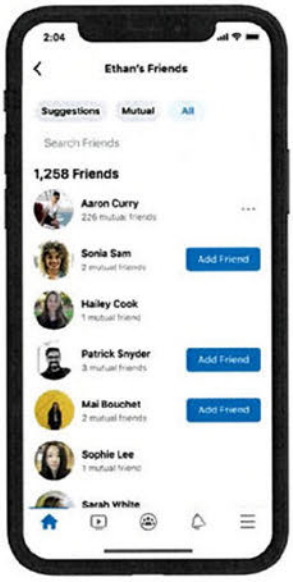

Normal View




Suspect's View: Minor Removed




Normal View




Suspect's View: Minor Removed

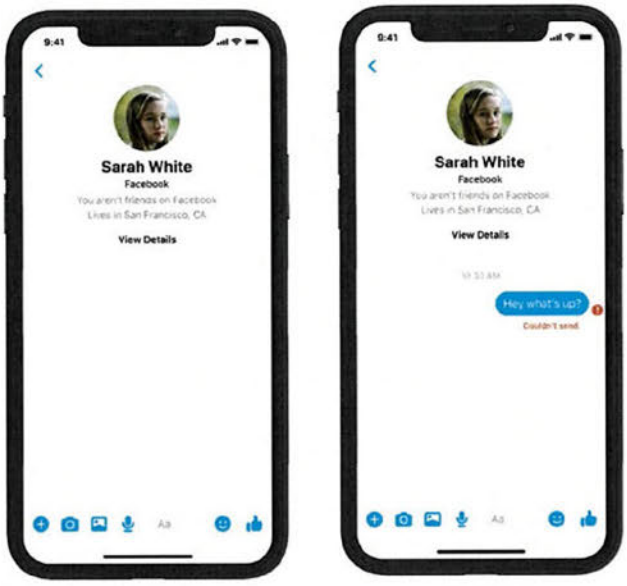
## Remove Suspect's Ability to Friend Minors

This intervention removes the ability of adults to friend unconnected minors if they have been blocked, reported or had a message deleted by a minor in the last 10 days.

**Impact:** ~23% reduction in Tier 2 IIC prevelance

## Message Request Restrictions

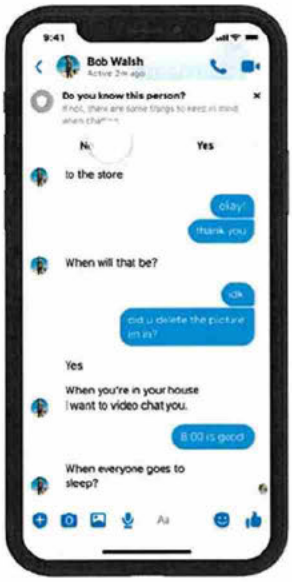
Adults with no friends in common with a minor are unable to send a message request to them.



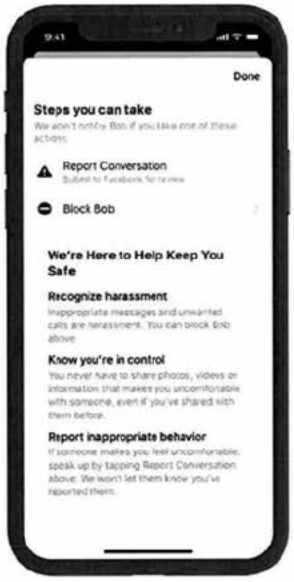

Thread Null

Send Fails

Now also available on IGD<>IGD




In-Thread QP




Detail View

## Proactive Warnings

When a minor is chatting to an adult with [redacted] we show a warning banner in thread, with a link to more info and remediation tools.

**Impact:** 37k blocks/day


Metadata
Doc ID 1641098
Reference Id
Pages 1
Custodian [redacted]Sarah
Bates No. META3047MDL-046-00319359
Tags None
Folders META3047MDL VOL046; zMETAProductionDocuments
Filenames Copy of Messenger Child Safety Foundations_1O84UZ_qdTrD5u9Fco72Cn5Llu44aWPe-Bcjdo8YC_bQ.pptx
Author [redacted]
Title Copy of Messenger Child Safety Foundations
Created May 16, 2022 12:00AM EDT
Modified May 16, 2022 12:00AM EDT
Path /[redacted]Sarah
BEGATTACH META3047MDL-046-00319359
ENDATTACH META3047MDL-046-00319359
PRODVOL META3047MDL_VOL046
Confidentialit... HIGHLY CONFIDENTIAL (COMPETITOR)
HASHVALUE 683131f25506e61dd8e4e88c45c53a7d
DOCUMENT ... E-document
TRACK CHAN... N
COMMENTS N
HASHIDDEN... N
DOCEXT pptx
REDACTIONS N
DuplicateStat... Primary