FTC v. Meta Platforms, Inc.

EXPERT OPINION OF PROF. DAMON MCCOY

May 6, 2025

ASSIGNMENT

- Assess whether acquisitions were **necessary** to achieve the stated integrity benefits
- Assess whether Meta **actually achieved** the stated integrity benefits

META'S CLAIMED INTEGRITY BENEFITS

- Instagram: "Meta's tools and expertise helped Instagram attack its integrity problems at scale"
- WhatsApp: "Meta's integrity systems and know-how provided WhatsApp with significant and unique benefits"

McCoy Report ¶¶ 13, 129, 164, 214; McCoy Rebuttal Report ¶ ¶ 4, 257

FRAMEWORK FOR ANALYSIS

NECESSARY

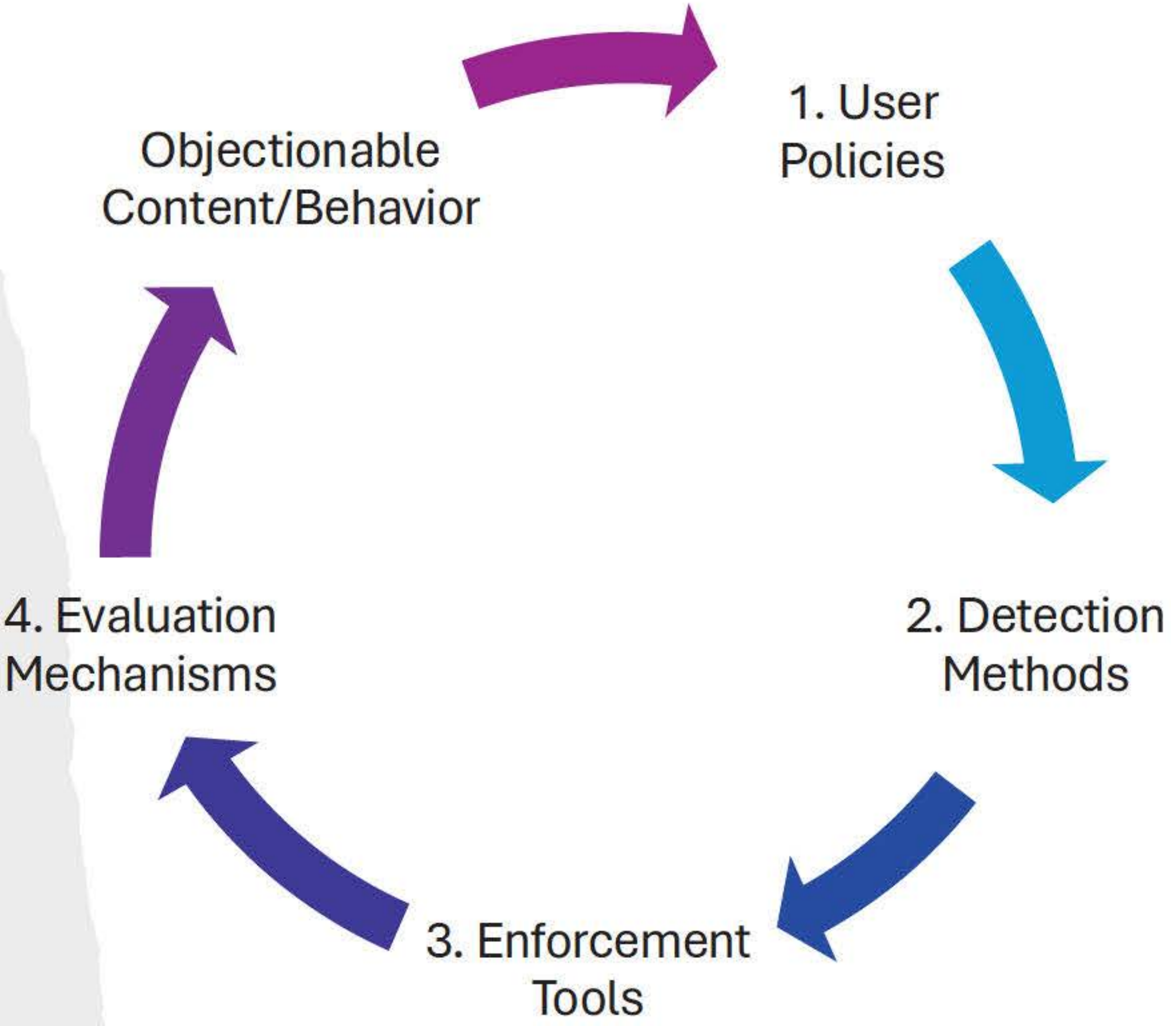
- Whether Instagram & WhatsApp were capably addressing integrity issues at the time of the acquisitions
- Whether Instagram & WhatsApp could have continued growing their capabilities without the acquisitions
- Whether the techniques and know-how Meta claims to have provided were unique

ACTUALLY ACHIEVED

- Whether quantitative metrics support Meta's stated benefits
- Whether qualitative analysis supports Meta's stated benefits

McCoy Report ¶¶ 16
McCoy Rebuttal Report ¶¶ 67, 314, 394, 395

THE CYCLE OF AN EFFECTIVE INTEGRITY SYSTEM



Objectionable Content/Behavior

1. User Policies

2. Detection Methods

3. Enforcement Tools

4. Evaluation Mechanisms

McCoy Report ¶¶ 26-28, 30;
McCoy Rebuttal Report ¶¶ 217-218

“AN ADVERSARIAL GAME”

“When thinking about something like spam, **spammers tend to change their behavior pretty frequently. We think of it as sort of an adversarial game**. And so **we need to continue to evolve our tactics as they evolve theirs**. We've also continued to update our community guidelines and policies as we **identify new categories of harm**, new things that we want to keep off the platform. And so yeah, as we evolve on the -- I guess we've evolved on the policy side, and then **we've also just continued to invest on the product side to be able to catch more and more of this content**.”

Julia Roberts (Pinterest) Trial, 150:21-151:6 (4.28.2025)

“We have adversaries that are trying to adapt to the things that you are doing. And **this is true of any service that has enough users.”**

Arturo Bejar, Former Head of Facebook Site Integrity,
Security Infrastructure, and Product Infrastructure
Arturo Bejar Dep. Tr., 81:18-21

“NOT ROCKET SCIENCE”

Q. Had any industry standards or playbooks developed for dealing with problematic content?

A. Yes. In fact, **most of our enforcement of problematic content was taken from a collection of community guidelines that existed elsewhere on other services**, things like Twitter or Facebook. And most of those community guidelines were posted publicly, and **we would take them and kind of make them our own.** So I think it was fairly straightforward, especially, you know, when we were smaller and less international. **As we became much larger, I think it became increasingly complex, but again, nothing – it wasn't rocket science.**

Kevin Systrom Trial, 55:17-56:3 (4.22.25)

“Solving spam, it's a nuanced problem, but it's not rocket science.”

Mark Zuckerberg Trial,162:12-20 (4.16.2025)

“NOT PLUG AND PLAY”

Q. So Instagram had a different backend structure, and that meant that some of Facebook’s integrity tools could not be immediately and automatically used by Instagram?

A. That's correct. They -- **because they were on a different backend, it was not -- you know, it wasn’t plug and play.**

Gregg Stefancik
Former Facebook Head of Site Integrity
Dep. Tr., 65:22-66:4

Q. ...So my question is: Was that a plug-and-play process?

A. There were -- no. **There had to be engineers dedicated to building things so that the system could be integrated on both site integrity and on Instagram's part.**

Arturo Bejar
Former Facebook Head of Site Integrity, Security Infrastructure, and Product Infrastructure
Dep. Tr., 34:13-18

Q. What type of resources, if any, were required to provide Instagram access to the [integrity] system?

A. Similar to delta, **dedicated engineers to create the interfaces. And there would also need to be work on creating the models because the features that Instagram would have would be different from the features that Facebook would have in order to make a decision.** Some were similar, but there would be some differences.

Arturo Bejar
Former Facebook Head of Site Integrity, Security Infrastructure, and Product Infrastructure
Dep. Tr., 35:9-19

McCoy Report ¶¶ 35, 61

WHATAPP'S ACQUISITION WAS NOT NECESSARY

PRE-ACQUISITION WHATSAPP CAPABLY MANAGED ITS INTEGRITY ISSUES

PRE-ACQUISITION WHATSAPP COULD HAVE CONTINUED DEVELOPING ITS INTEGRITY TOOLS WITHOUT THE ACQUISITION BY META

META'S INTEGRITY TECHNIQUES & KNOW-HOW WERE AND ARE NOT UNIQUE

McCoy Report ¶¶ 54-55, 144-145, 177-178; 217-218;
McCoy Rebuttal Report ¶¶ 264, 272-276

PRE-ACQUISITION WHATSAPP HAD INTEGRITY TOOLS

CO-FOUNDERS' TESTIMONY

Q. [] What were you doing to deal with spam?

A. [...] We would look at things like somebody joins our network and sends **the same message to 100 people** and the message goes to people who are **not in each other's address book**.

We would have, for example, a feedback mechanism where **if I send you a spam message you could report me**, and if we got enough of those reports from different users we would **block a number for being a spammer.**

We had some other ways to detect a user who would sign up a **suspicious number** and start messaging right away.

We had a way to **detect a bot**, where somebody signed up using a software instead of an actual person typing with their fingers because the **speed with which they would send a message** out. We would know that that's a computer and not an actual person and so it would trigger and potentially **block it.**

So we had a number of mechanisms in place to catch potential spammers and abusers in our system.

Jan Koum IH Tr., 188:24-189:23

Q. Prior to Facebook's acquisition of WhatsApp, **had WhatsApp built a system to deal with spam** on its application?

A. **Yes**.

Brian Acton Dep. Tr., 118:19-24

McCoy Rebuttal Report ¶ 397

WHATSAPP COULD HAVE CONTINUED DEVELOPING ITS INTEGRITY TOOLS

CO-FOUNDERS' TESTIMONY

Q. If Facebook had not acquired WhatsApp, could you have continued to improve your ability to deal with spam issues?

A. **Yes.**

Q. How would you have done so?

A. **We would hire people** to go build a team that specializes in spam abuse and spam detection and **give them enough resources** and **server capacity to go build what would be needed to solve that problem, just like we solved any problem that came up until 2014**, by hiring smart people and giving them resources to solve the problem.

Jan Koum IH Tr., 190:13-24

Q. Was **WhatsApp making improvements to that system** prior to the acquisition?

A. **Yes.**

Q. Would WhatsApp have built its own **second-generation system** to fight spam?

A. We -- **we certainly had the knowledge, expertise, and contact network of people to hire to build it. [...]**

Brian Acton Dep. Tr., 118:25-119:2

McCoy Rebuttal Report ¶ 398

UNVERIFIABLE INTEGRITY BENEFITS

NO RELIABLE QUANTITATIVE METRICS

- Only data on the number of WhatsApp accounts banned annually available
- Annual accounts banned data dates to May 2020

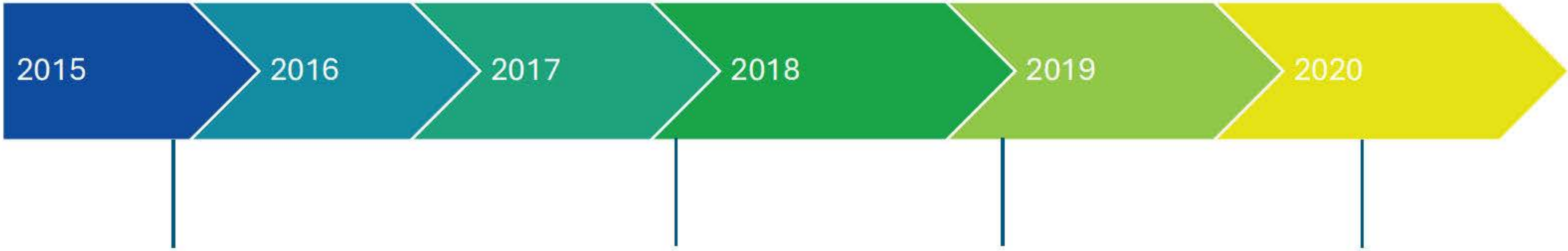
QUALITATIVE ANALYSIS INDICATES THAT META DELAYED INTEGRATING WHATSAPP INTO ITS INTEGRITY SYSTEM

McCoy Report ¶¶ 54-55, 144-145, 177-178; 217-218;
McCoy Rebuttal Report ¶¶ 264, 272-276

QUALITATIVE ANALYSIS

DELAYED INTEGRATION

QUALITATIVE ANALYSIS: META DELAYED INTEGRATING WHATSAPP INTO ITS INTEGRITY SYSTEMS



Late 2015-Early 2016
WhatsApp partial connection to some of Meta's spam-fighting tools

Jan. 2018
"we don't have a formal Integrity team..."

May 2019
Integrity 2019 Review
"the WhatsApp Integrity team owns all problems on WhatsApp...WhatsApp is currently the least centralized of all surface teams."

Nov. 2020
Guy Rosen: "Right now my team doesn't do much on WA"

McCoy Rebuttal Report ¶¶ 401-404; 410, 423

Brian Acton IH Tr., at 237:4-16; PX15206; PX3081; PX12271

INSTAGRAM'S ACQUISITION WAS NOT NECESSARY

PRE-ACQUISITION INSTAGRAM CAPABLY MANAGED ITS INTEGRITY ISSUES

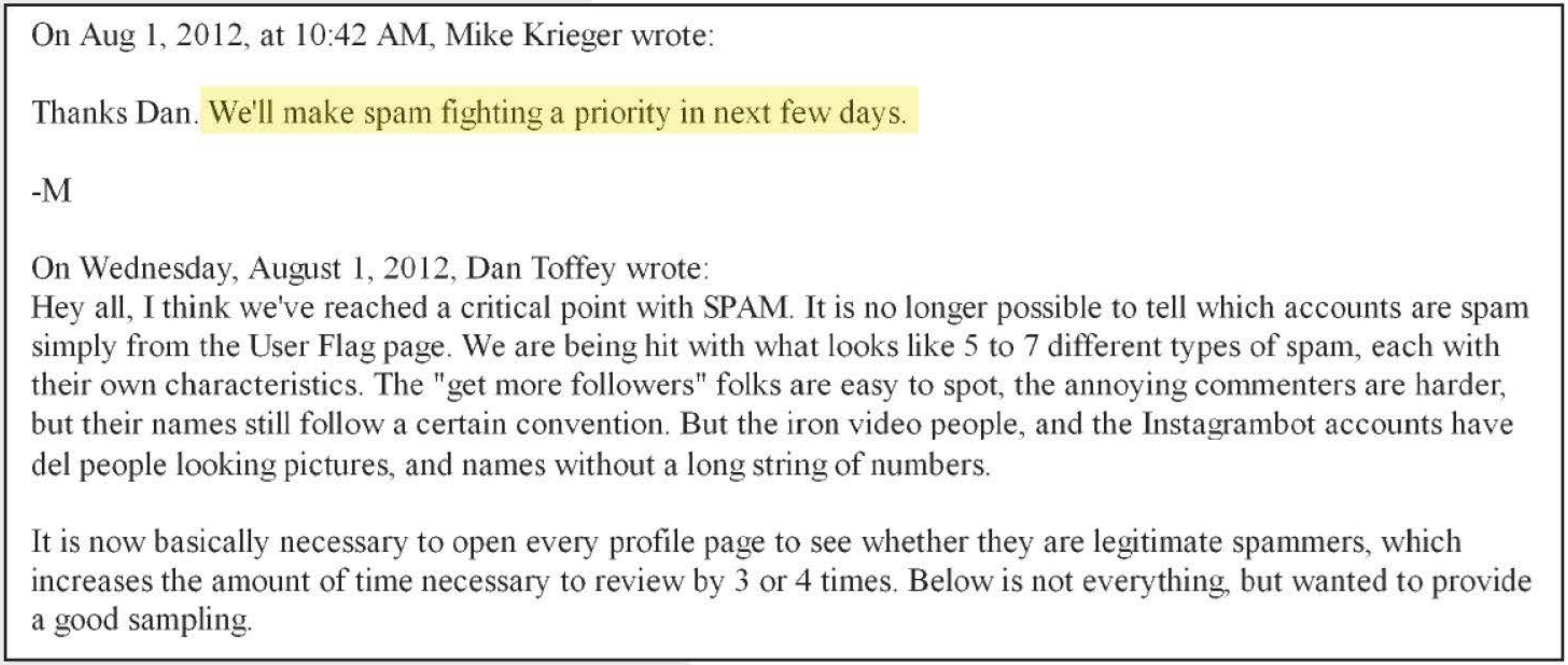
- Instagram had built a capable integrity system to address its integrity issues
- Instagram was scaling its integrity system in line with its growing user base and attack types

PRE-ACQUISITION INSTAGRAM COULD HAVE CONTINUED DEVELOPING ITS INTEGRITY TOOLS WITHOUT THE ACQUISITION

McCoy Report ¶¶ 54-55, 144-145, 177-178; 217-218;
McCoy Rebuttal Report ¶¶ 264, 272-276

INSTAGRAM'S ACQUISITION WAS NOT NECESSARY: INSTAGRAM PRE-ACQUISITION

AUGUST 1, 2012



On Aug 1, 2012, at 10:42 AM, Mike Krieger wrote:

Thanks Dan. We'll make spam fighting a priority in next few days.

-M

On Wednesday, August 1, 2012, Dan Toffey wrote:
Hey all, I think we've reached a critical point with SPAM. It is no longer possible to tell which accounts are spam simply from the User Flag page. We are being hit with what looks like 5 to 7 different types of spam, each with their own characteristics. The "get more followers" folks are easy to spot, the annoying commenters are harder, but their names still follow a certain convention. But the iron video people, and the Instagrambot accounts have del people looking pictures, and names without a long string of numbers.

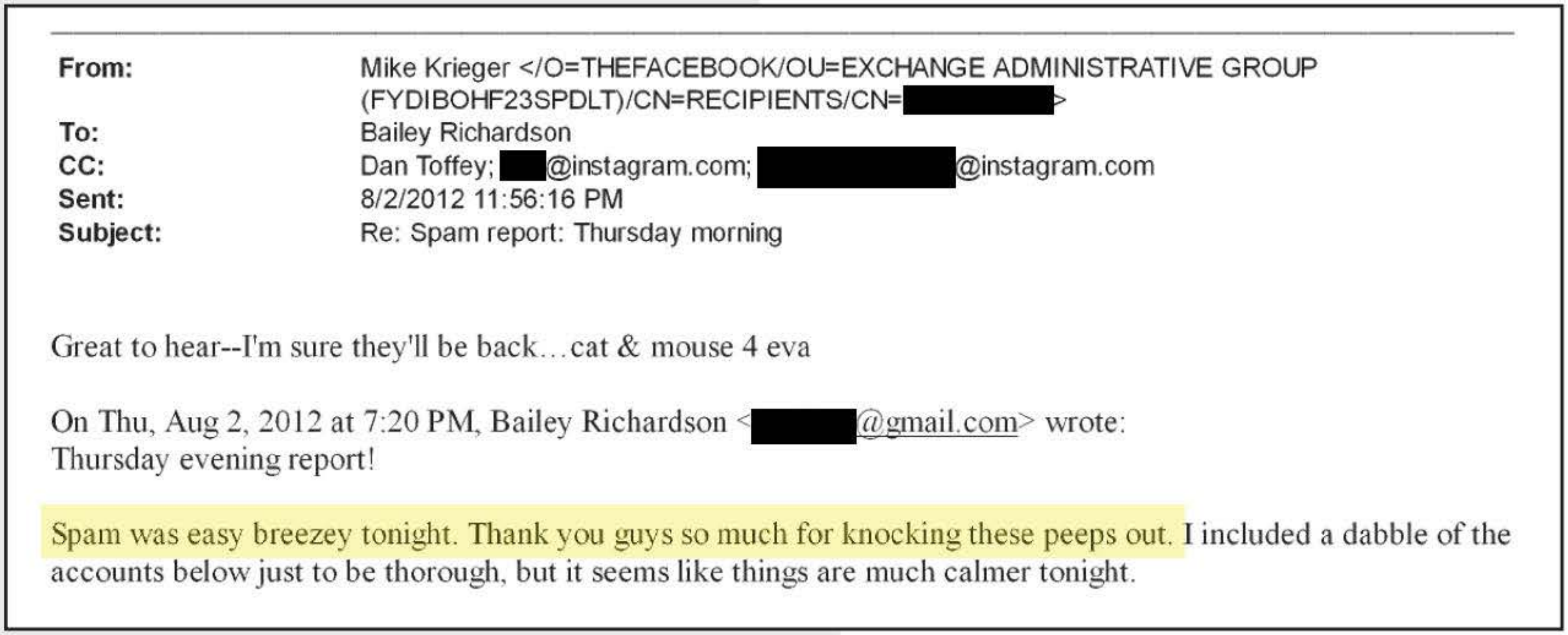
It is now basically necessary to open every profile page to see whether they are legitimate spammers, which increases the amount of time necessary to review by 3 or 4 times. Below is not everything, but wanted to provide a good sampling.

PX10020

McCoy Report ¶ 145

INSTAGRAM'S ACQUISITION WAS NOT NECESSARY: INSTAGRAM PRE-ACQUISITION

AUGUST 2, 2012

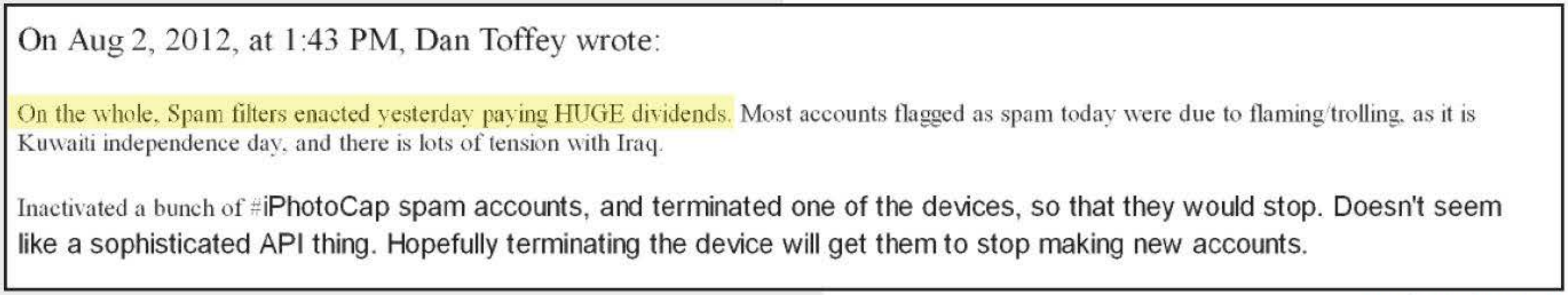


From:	Mike Krieger </O=THEFACEBOOK/OU=EXCHANGE ADMINISTRATIVE GROUP (FYDIBOHF23SPDLT)/CN=RECIPIENTS/CN=[redacted]>
To:	Bailey Richardson
CC:	Dan Toffey; [redacted]@instagram.com; [redacted]@instagram.com
Sent:	8/2/2012 11:56:16 PM
Subject:	Re: Spam report: Thursday morning

Great to hear--I'm sure they'll be back…cat & mouse 4 eva

On Thu, Aug 2, 2012 at 7:20 PM, Bailey Richardson <[redacted]@gmail.com> wrote:
Thursday evening report!

Spam was easy breezey tonight. Thank you guys so much for knocking these peeps out. I included a dabble of the accounts below just to be thorough, but it seems like things are much calmer tonight.



On Aug 2, 2012, at 1:43 PM, Dan Toffey wrote:

On the whole, Spam filters enacted yesterday paying HUGE dividends. Most accounts flagged as spam today were due to flaming/trolling, as it is Kuwaiti independence day, and there is lots of tension with Iraq.

Inactivated a bunch of #iPhotoCap spam accounts, and terminated one of the devices, so that they would stop. Doesn't seem like a sophisticated API thing. Hopefully terminating the device will get them to stop making new accounts.

PX3062

McCoy Rebuttal Report ¶¶ 200-201

INSTAGRAM'S ACQUISITION WAS NOT NECESSARY: INSTAGRAM PRE-ACQUISITION

Q. So had Instagram been tackling spam pretty well as of June 2012?

A. **It had, yes...**

Q. Before you joined Meta, had you found a need to engage third-party spam-fighting services?

A. **No.**

Q. Why is that?

A. **As I stated in the email, I thought we had it under control.**

Kevin Systrom Trial, 53:17-54:7 (4.22.2025)

Q. Did you ever see evidence that Instagram's growth was slowing down due to spam?

A. **No, not especially.**

Mike Krieger Dep. Tr., 135:12-17

Q. Did problematic content ever threaten Instagram's ability to continue growing?

A. **No.**

Kevin Systrom Trial, 55:4-6 (4.22.2025)

CO-FOUNDERS' TRIAL TESTIMONY

Q. Are you aware of any reason why Instagram wouldn't have been able to continue tackling spam on the platform if it had remained independent?

A. **I don't see any reason we wouldn't have been able to continue to fight it the way we had been fighting it.**

Kevin Systrom Trial, 54:3-7 (4.22.2025)

Q. Are you aware of any reason why Instagram wouldn't have been able to scale its response to problematic content without Facebook's help if it had stayed independent?

A. **I don't have any reason to believe that we wouldn't have been able to scale it...**

Kevin Systrom Trial, 55:7-11 (4.22.2025)

INSTAGRAM'S ACQUISITION WAS NOT NECESSARY: INSTAGRAM PRE-ACQUISITION

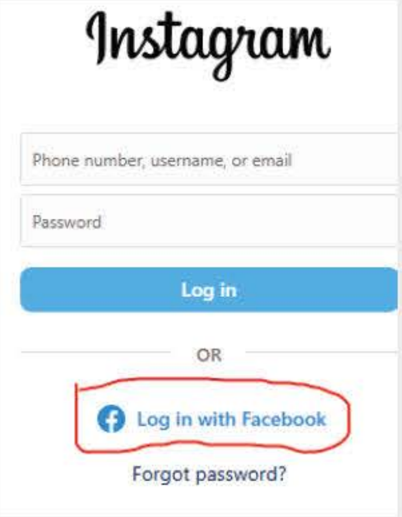
CrowdFlower

Mutual Non-Disclosure Agreement

Q. Why did Instagram use CrowdFlower?

A. **The primary motivators were being able to process the queues on an ongoing basis, so not relying on shifts from our own team** and ... **Processing the content in a more scalable way, in a more around-the-clock way.**

Mike Krieger Dep. Tr., 143:9-25



PhotoDNA enables the creation of a unique digital signature of an image. Once created, a signature can be compared to other images' signatures to find matches. This technique is known in the tech industry as "hashing"

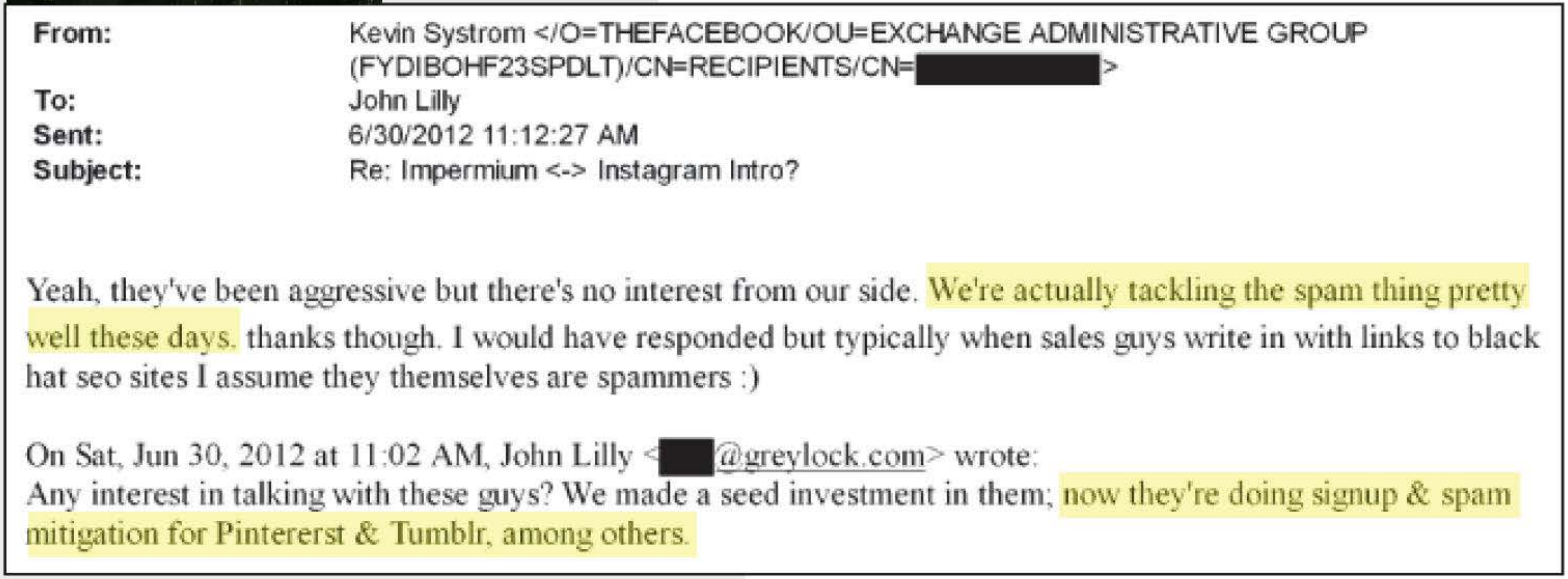
McCoy Report ¶¶ 180, 222, 234, 236

PX3057

INSTAGRAM'S ACQUISITION WAS NOT NECESSARY: INSTAGRAM PRE-ACQUISITION

JUNE 30, 2012

IMPERMIUM



From:	Kevin Systrom </O=THEFACEBOOK/OU=EXCHANGE ADMINISTRATIVE GROUP (FYDIBOHF23SPDLT)/CN=RECIPIENTS/CN=[redacted]>
To:	John Lilly
Sent:	6/30/2012 11:12:27 AM
Subject:	Re: Impermium <-> Instagram Intro?

Yeah, they've been aggressive but there's no interest from our side. We're actually tackling the spam thing pretty well these days. thanks though. I would have responded but typically when sales guys write in with links to black hat seo sites I assume they themselves are spammers :)

On Sat, Jun 30, 2012 at 11:02 AM, John Lilly <[redacted]@greylock.com> wrote:
Any interest in talking with these guys? We made a seed investment in them; now they're doing signup & spam mitigation for Pintererst & Tumblr, among others.

PX15223

McCoy Report ¶¶ 139, 149, 240
McCoy Rebuttal Report ¶ 202

INSTAGRAM'S ACQUISITION WAS NOT NECESSARY

META'S INTEGRITY TECHNIQUES & KNOW-HOW WERE AND ARE NOT UNIQUE

- Pre-acquisition Instagram was using similar techniques
- Other platforms used then and today use similar techniques to successfully tackle integrity issues at scale
- Meta hired integrity professionals from other companies
- Meta & other platforms regularly publish their integrity insights

McCoy Report ¶¶ 17a, 121, 122, 124

“KNOWLEDGE SHARE”

Former Head of Facebook’s Integrity Team

Q. And Facebook organized the Security @Scale conference to **improve security sharing best practices outside of the company**?

A. Yes, that's correct. To -- and I talked about earlier to sort of **raise all ships on – in the ocean.** I've always believed that, you know, both **in the terms of security and integrity, industry benefits from actually sharing with each other in order to make a better experience for folks. [...]**

Q. And who attended Fighting Spam @Scale? [...]

A. I mean, **a variety of folks that we invited for -- from across industry.** Folks from **Google, Netflix**. I don't know why **Airbnb** stands out for some reason, but basically the whole spectrum of companies that would be -- would have integrity issues. Even **LinkedIn. [...]**

Q. Turning back to the Fighting Spam @Scale conference, **companies other than Facebook also presented** at that conference; is that right?

A. Yes. That was the intent was to have **joint presentations and learn from each other.**

Gregg Stefancik Dep. Tr., 74:18-75:18, 78:2-6

Pinterest

Q. Are you familiar with the term "knowledge share" in the context of trust and safety?

A. Yes.

Q. What does "**knowledge share"** mean in that context?

A. It tends to mean sharing -- whether **best practices for identifying this type of content, for combating spam**, yeah identifying spam users, that sort of thing. We tend to do these with **different companies that might have similar problems to us.**

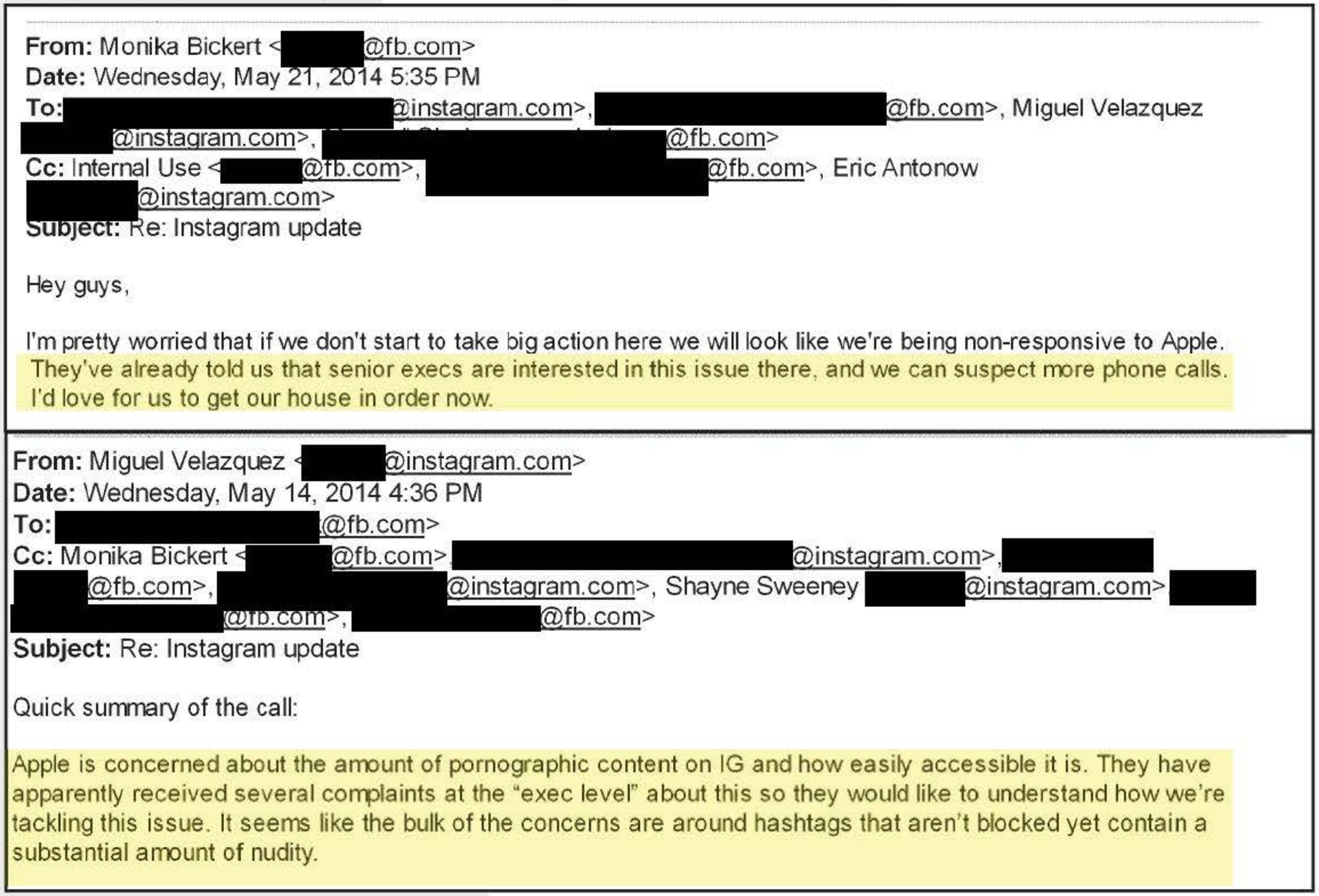
Q. Does Pinterest engage in knowledge share in the trust and safety context?

A. We do.

Julia Roberts Trial (Pinterest), 153:20-7 (4.28.2025)

INSTAGRAM'S ACQUISITION WAS NOT NECESSARY: META'S INTEGRITY TECHNIQUES & KNOW-HOW NOT UNIQUE

MAY 14-21, 2014



From: Monika Bickert <[redacted]@fb.com>
Date: Wednesday, May 21, 2014 5:35 PM
To: [redacted]@instagram.com>, [redacted]@fb.com>, Miguel Velazquez
[redacted]@instagram.com>, [redacted]@fb.com>
Cc: Internal Use <[redacted]@fb.com>, [redacted]@fb.com>, Eric Antonow
[redacted]@instagram.com>
Subject: Re: Instagram update

Hey guys,

I'm pretty worried that if we don't start to take big action here we will look like we're being non-responsive to Apple. They've already told us that senior execs are interested in this issue there, and we can suspect more phone calls. I'd love for us to get our house in order now.

From: Miguel Velazquez <[redacted]@instagram.com>
Date: Wednesday, May 14, 2014 4:36 PM
To: [redacted]@fb.com>
Cc: Monika Bickert <[redacted]@fb.com>, [redacted]@instagram.com>, [redacted]
[redacted]@fb.com>, [redacted]@instagram.com>, Shayne Sweeney [redacted]@instagram.com>, [redacted]
[redacted]@fb.com>, [redacted]@fb.com>
Subject: Re: Instagram update

Quick summary of the call:

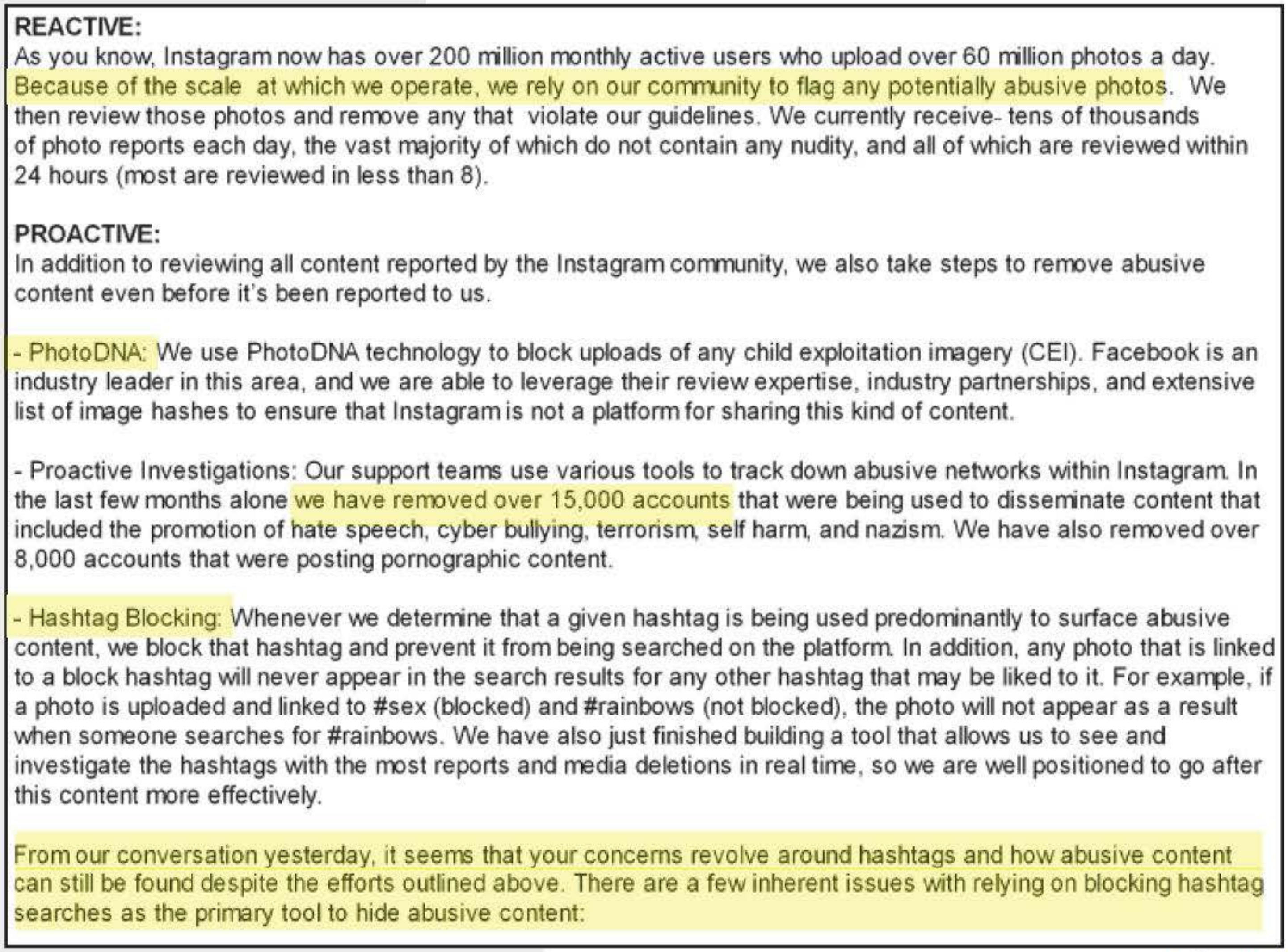
Apple is concerned about the amount of pornographic content on IG and how easily accessible it is. They have apparently received several complaints at the "exec level" about this so they would like to understand how we're tackling this issue. It seems like the bulk of the concerns are around hashtags that aren't blocked yet contain a substantial amount of nudity.

PX3105 / PX10160

McCoy Report ¶¶ 256-58

INSTAGRAM'S ACQUISITION WAS NOT NECESSARY: META'S INTEGRITY TECHNIQUES & KNOW-HOW NOT UNIQUE

MAY 2014 CONTINUED



REACTIVE:

As you know, Instagram now has over 200 million monthly active users who upload over 60 million photos a day. Because of the scale at which we operate, we rely on our community to flag any potentially abusive photos. We then review those photos and remove any that violate our guidelines. We currently receive- tens of thousands of photo reports each day, the vast majority of which do not contain any nudity, and all of which are reviewed within 24 hours (most are reviewed in less than 8).

PROACTIVE:

In addition to reviewing all content reported by the Instagram community, we also take steps to remove abusive content even before it's been reported to us.

- PhotoDNA: We use PhotoDNA technology to block uploads of any child exploitation imagery (CEI). Facebook is an industry leader in this area, and we are able to leverage their review expertise, industry partnerships, and extensive list of image hashes to ensure that Instagram is not a platform for sharing this kind of content.

- Proactive Investigations: Our support teams use various tools to track down abusive networks within Instagram. In the last few months alone we have removed over 15,000 accounts that were being used to disseminate content that included the promotion of hate speech, cyber bullying, terrorism, self harm, and nazism. We have also removed over 8,000 accounts that were posting pornographic content.

- Hashtag Blocking: Whenever we determine that a given hashtag is being used predominantly to surface abusive content, we block that hashtag and prevent it from being searched on the platform. In addition, any photo that is linked to a block hashtag will never appear in the search results for any other hashtag that may be liked to it. For example, if a photo is uploaded and linked to #sex (blocked) and #rainbows (not blocked), the photo will not appear as a result when someone searches for #rainbows. We have also just finished building a tool that allows us to see and investigate the hashtags with the most reports and media deletions in real time, so we are well positioned to go after this content more effectively.

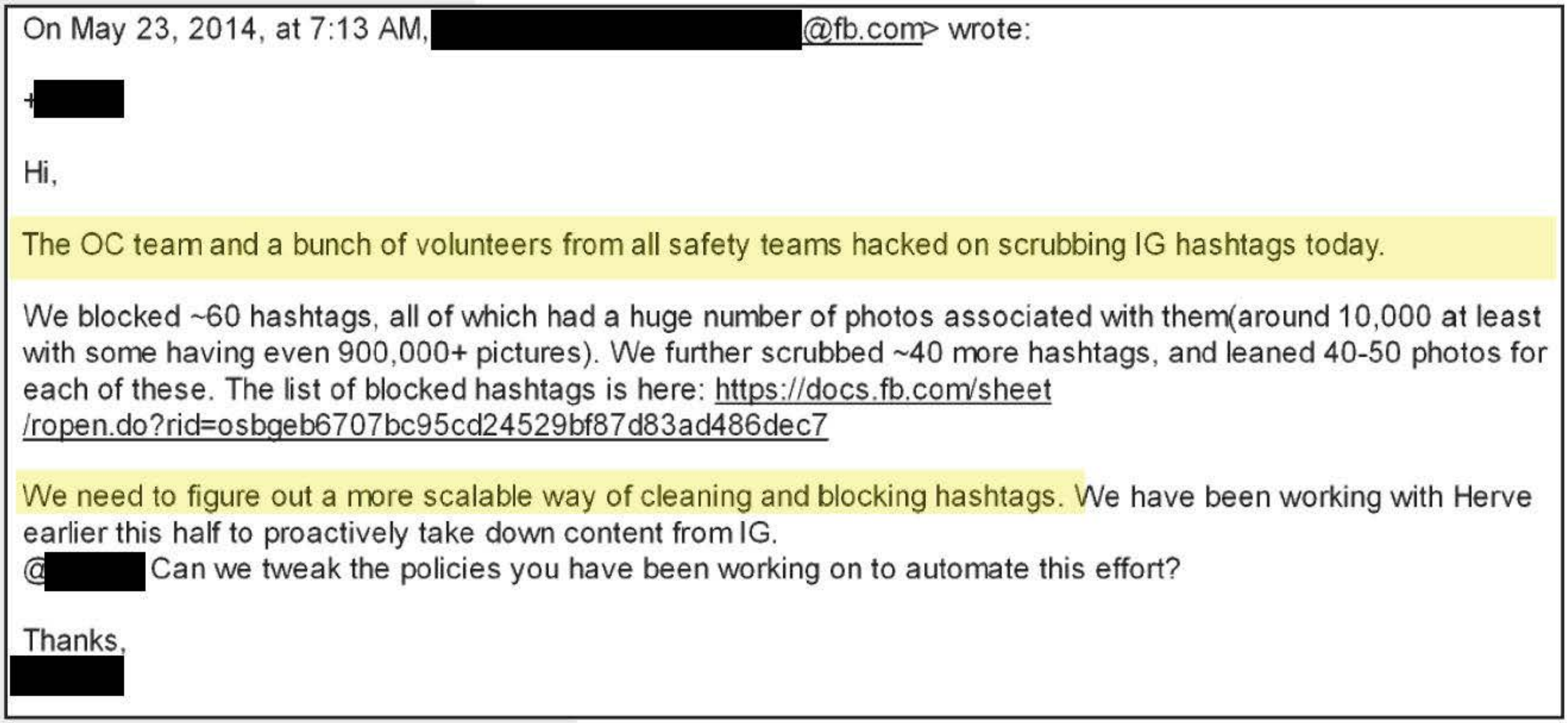
From our conversation yesterday, it seems that your concerns revolve around hashtags and how abusive content can still be found despite the efforts outlined above. There are a few inherent issues with relying on blocking hashtag searches as the primary tool to hide abusive content:

PX3105 / PX10160

McCoy Report ¶¶ 256-58

INSTAGRAM'S ACQUISITION WAS NOT NECESSARY: META'S INTEGRITY TECHNIQUES & KNOW-HOW NOT UNIQUE

MAY 2014 CONTINUED



On May 23, 2014, at 7:13 AM, [redacted]@fb.com> wrote:

+[redacted]

Hi,

The OC team and a bunch of volunteers from all safety teams hacked on scrubbing IG hashtags today.

We blocked ~60 hashtags, all of which had a huge number of photos associated with them(around 10,000 at least with some having even 900,000+ pictures). We further scrubbed ~40 more hashtags, and leaned 40-50 photos for each of these. The list of blocked hashtags is here: https://docs.fb.com/sheet/ropen.do?rid=osbgeb6707bc95cd24529bf87d83ad486dec7

We need to figure out a more scalable way of cleaning and blocking hashtags. We have been working with Herve earlier this half to proactively take down content from IG.
@[redacted] Can we tweak the policies you have been working on to automate this effort?

Thanks,
[redacted]

PX3105

McCoy Report ¶¶ 256-58
McCoy Rebuttal Report ¶¶ 22, 104, 226

INSTAGRAM'S ACQUISITION WAS NOT NECESSARY:
META'S INTEGRITY TECHNIQUES & KNOW-HOW NOT UNIQUE

GOOGLE, SNAPCHAT, TWITTER

Q. Was **Google's delta equivalent comparable to Facebook's?**

A. **Yes. [...]**

Q. [...] have any companies built integrity systems that are comparable in scope and capability?

[...]

A. I think that the **larger social media services that face similar pressures have built equivalent systems**...But I know that **Snapchat** built their set of systems, and I had some conversations with them about that. I know that **Twitter** built certain systems. And I definitely spoke with **Google** about the systems that they built. [...]

Arturo Bejar Dep. Tr., 179:5-7, 181:12-182:17

TRIAL TESTIMONY X/TWITTER & PINTEREST

Q. And when you joined X in 2016, did X use tools to identify child sexual abuse material?

A. Yes. [...]

Q. And was one tool that X used PhotoDNA?

A. I believe so, yes.

Q. So is it fair to say that when you joined in 2016, X was able to employ **blacklists, rules engines, PhotoDNA, and machine learning** to trust in safety without relying on Meta?

A. Yes.

Keith Coleman (Twitter) Trial, 42:5-19, 43:1-4 (4.28.2025)

Q. How does Pinterest determine if a user has violated its community guidelines?

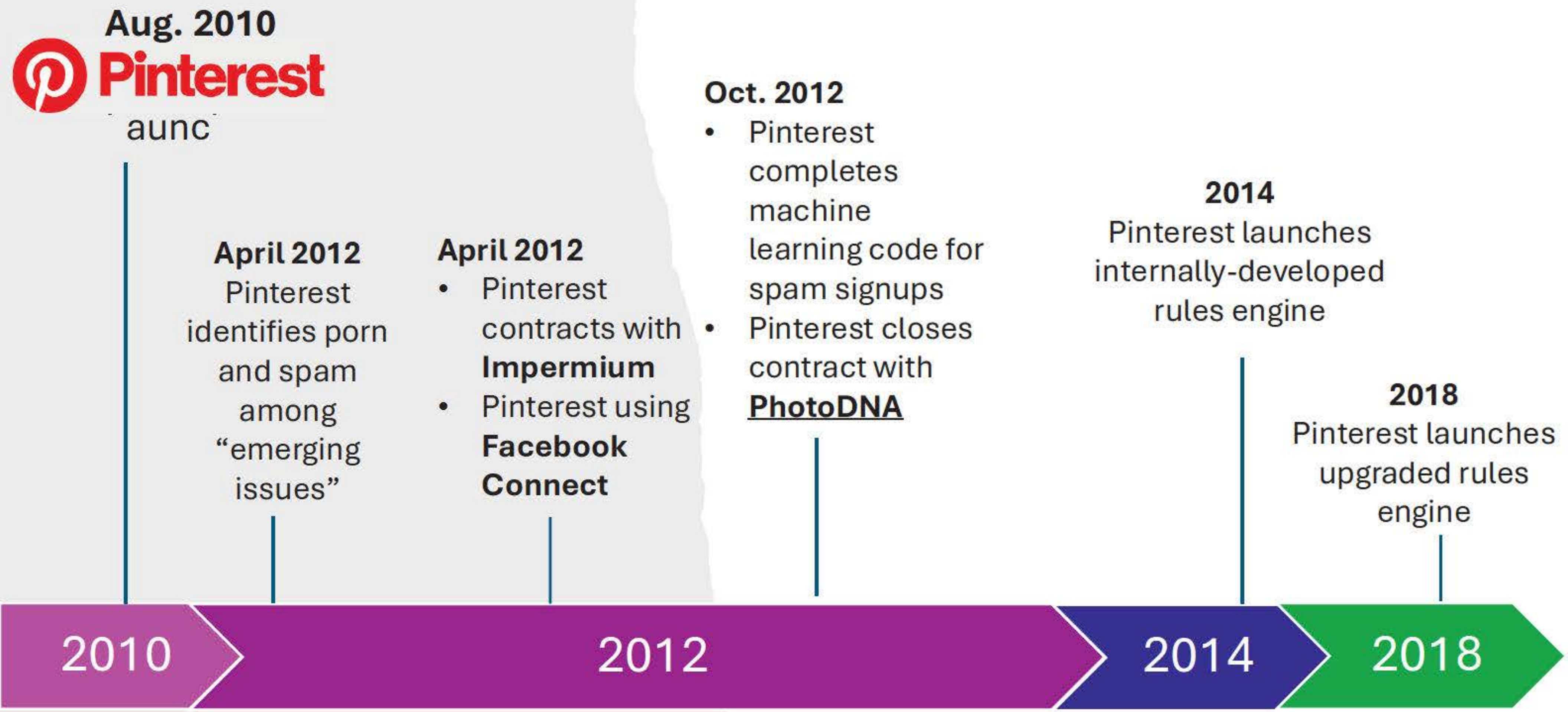
A. We have a combination of tactics. So the first would be sort of **automated enforcement**. So we **build signals and rules** that are aimed to identify content or behavior that violate our policies. [...] We also have a way to **report users, pins, boards, content on the platform** that then would be sent to a human reviewer [...] sometimes if we **identify content that is unsafe or against our guidelines, we will fan out to find other similar content.**

Q. [...] when did Pinterest begin developing those tools?

A. **I believe the first version, which was Stingray, would have been 2013. We probably had something before that**, but yeah, the first real one that I know of was 2013.

Julia Roberts (Pinterest) Trial,152:10-153:3 (4.28.2025)

INSTAGRAM'S ACQUISITION WAS NOT NECESSARY: META'S INTEGRITY TECHNIQUES & KNOW-HOW NOT UNIQUE



Aug. 2010

Pinterest

aunc

April 2012
Pinterest identifies porn and spam among "emerging issues"

April 2012

- Pinterest contracts with **Impermium**
- Pinterest using **Facebook Connect**

Oct. 2012

- Pinterest completes machine learning code for spam signups
- Pinterest closes contract with **PhotoDNA**

2014
Pinterest launches internally-developed rules engine

2018
Pinterest launches upgraded rules engine

PX7083; PX7016; PX0569; PX0568; PX12632; PX15223

McCoy Report ¶¶ 138b, 140; McCoy Rebuttal Report ¶¶ 202, 248

UNABLE TO VERIFY STATED INTEGRITY BENEFITS

NO RELIABLE QUANTITATIVE METRICS

- No reliable metrics prior to 2021 for Instagram

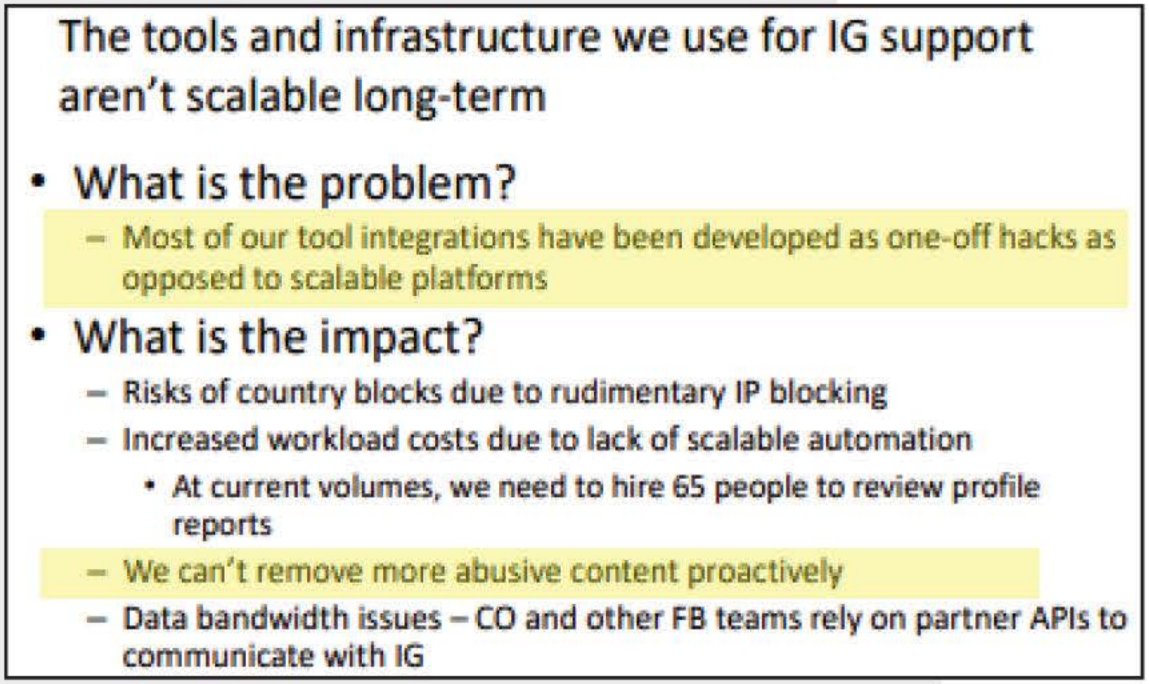
QUALITATIVE ANALYSIS INDICATES:

- Inadequate customization
- Under-resourcing
- Facebook Blue Prioritization

McCoy Report ¶¶ 17b, 52a-c

QUALITATIVE ANALYSIS: INADEQUATE CUSTOMIZATION

APRIL 2014



The tools and infrastructure we use for IG support aren't scalable long-term

- What is the problem?
  - Most of our tool integrations have been developed as one-off hacks as opposed to scalable platforms
- What is the impact?
  - Risks of country blocks due to rudimentary IP blocking
  - Increased workload costs due to lack of scalable automation
    - At current volumes, we need to hire 65 people to review profile reports
  - We can't remove more abusive content proactively
  - Data bandwidth issues – CO and other FB teams rely on partner APIs to communicate with IG

Q. ...Can you explain why your scalable or lack of scalable solutions is hindering your ability to remove content proactively? [...]

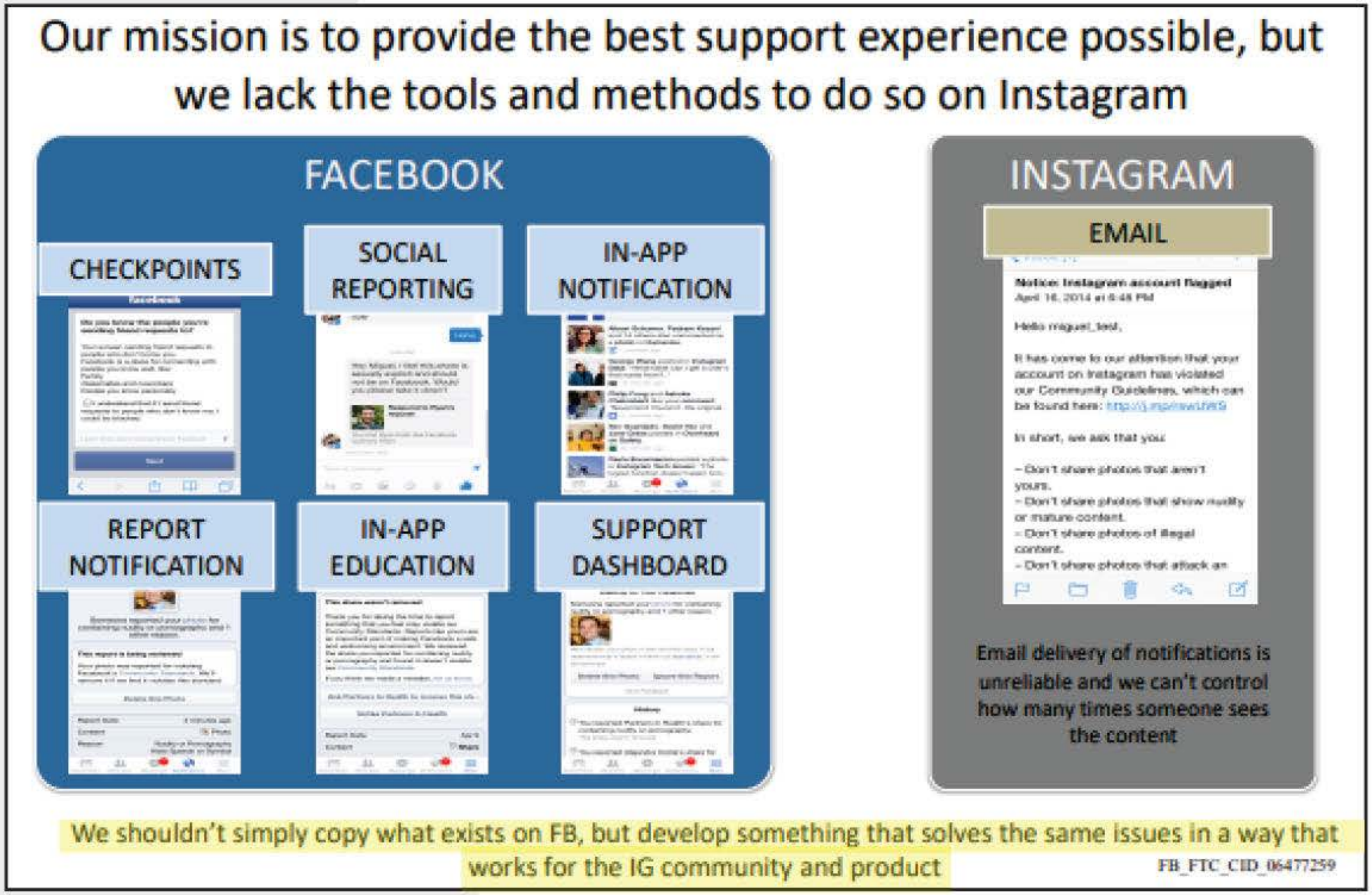
A. I think within, within the context of **there were tools that we weren't like fully plugged into on the Facebook side yet that, just because they required, you know, engineering investment to, to plug into them...**

Miguel Velazquez,
Community Support Experience Specialist,
Dep. Tr., 155:22-156:12

PX10156

McCoy Report ¶¶ 63-64

APRIL 2014



Our mission is to provide the best support experience possible, but we lack the tools and methods to do so on Instagram

FACEBOOK

CHECKPOINTS

SOCIAL REPORTING

IN-APP NOTIFICATION

REPORT NOTIFICATION

IN-APP EDUCATION

SUPPORT DASHBOARD

INSTAGRAM

EMAIL

Notice: Instagram account flagged
April 16, 2014 at 6:48 PM

Hello miguel_test,

It has come to our attention that your account on Instagram has violated our Community Guidelines, which can be found here: http://j.mp/newUHR9

In short, we ask that you:

- Don't share photos that aren't yours.
- Don't share photos that show nudity or mature content.
- Don't share photos of illegal content.
- Don't share photos that attack an

Email delivery of notifications is unreliable and we can't control how many times someone sees the content

We shouldn't simply copy what exists on FB, but develop something that solves the same issues in a way that works for the IG community and product

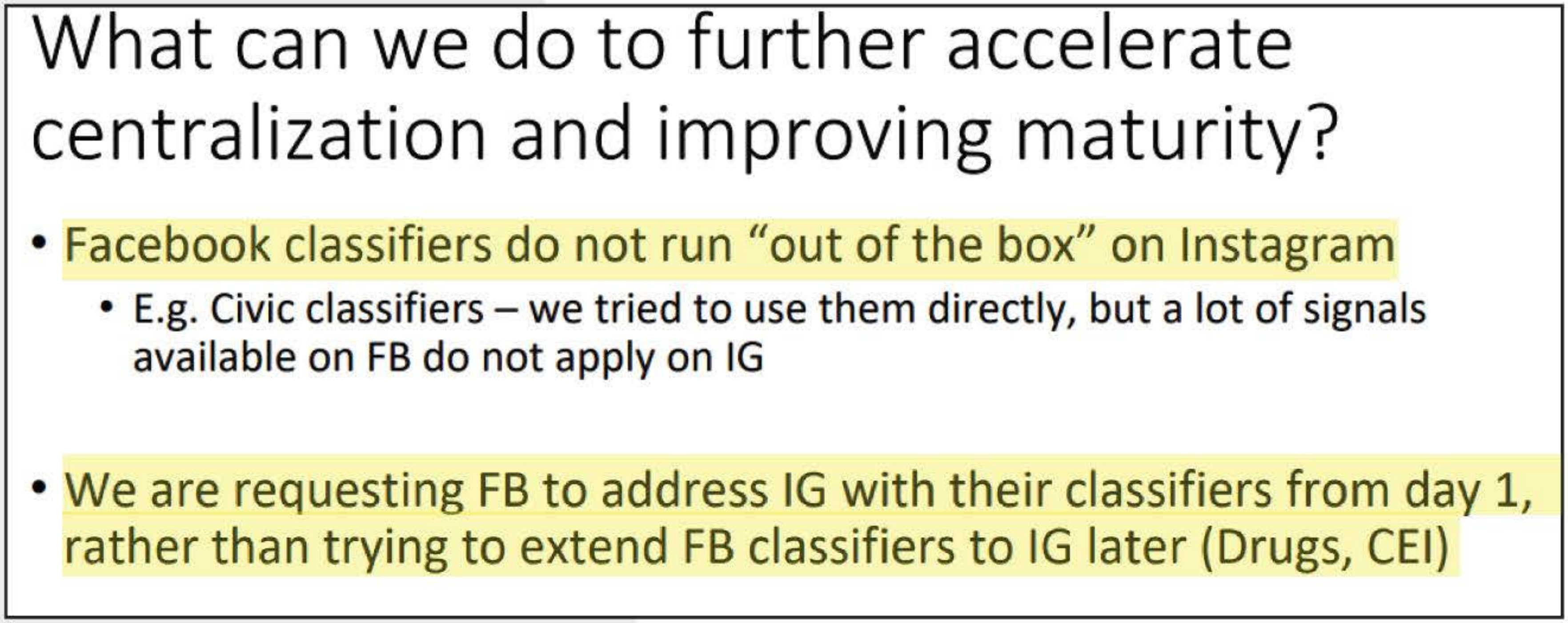
FB_FTC_CID_06477259

PX10156

McCoy Report ¶¶ 63-64

QUALITATIVE ANALYSIS: INADEQUATE CUSTOMIZATION

OCTOBER 2018



What can we do to further accelerate centralization and improving maturity?

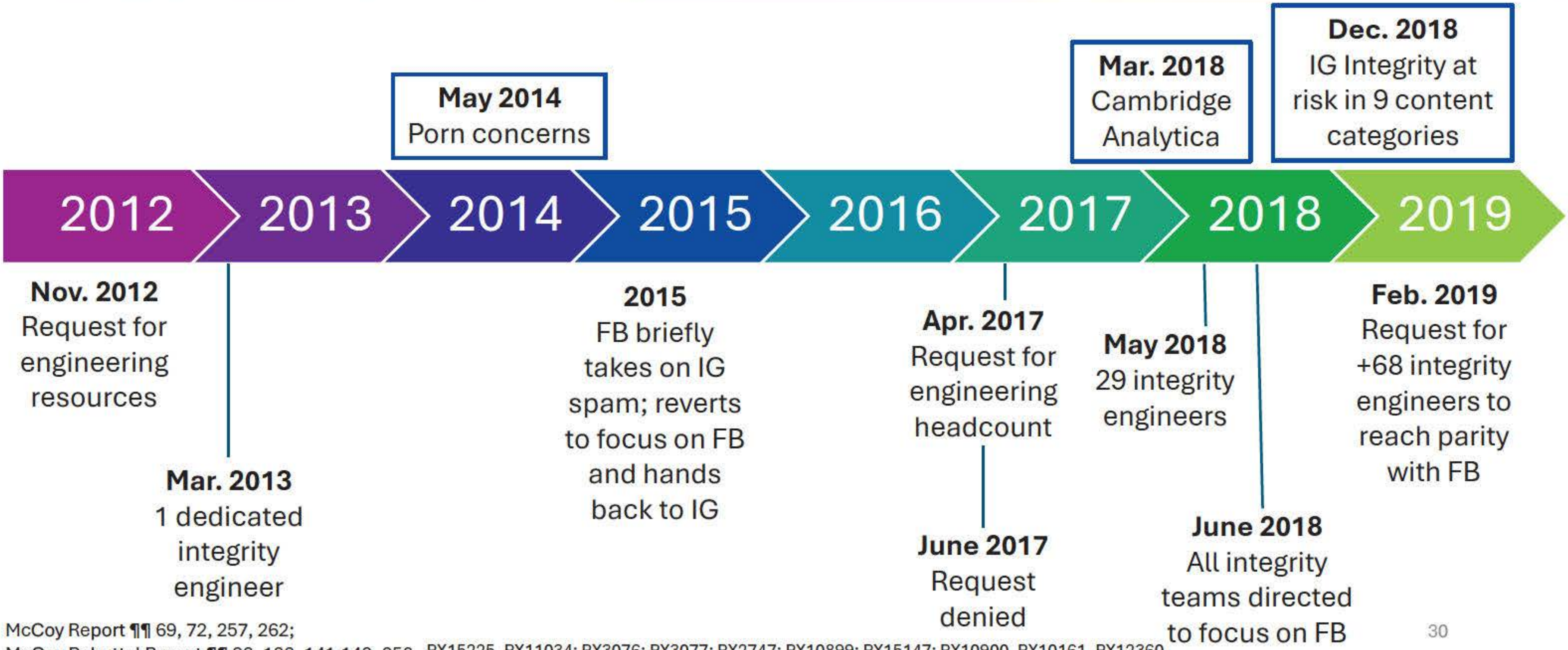
- Facebook classifiers do not run "out of the box" on Instagram
  - E.g. Civic classifiers – we tried to use them directly, but a lot of signals available on FB do not apply on IG
- We are requesting FB to address IG with their classifiers from day 1, rather than trying to extend FB classifiers to IG later (Drugs, CEI)

PX3605

See McCoy Report ¶¶ 35, 61

QUALITATIVE ANALYSIS

UNDER-RESOURCING TIMELINE



2012 2013 2014 2015 2016 2017 2018 2019

Nov. 2012
Request for engineering resources

Mar. 2013
1 dedicated integrity engineer

May 2014
Porn concerns

2015
FB briefly takes on IG spam; reverts to focus on FB and hands back to IG

Apr. 2017
Request for engineering headcount

June 2017
Request denied

Mar. 2018
Cambridge Analytica

May 2018
29 integrity engineers

June 2018
All integrity teams directed to focus on FB

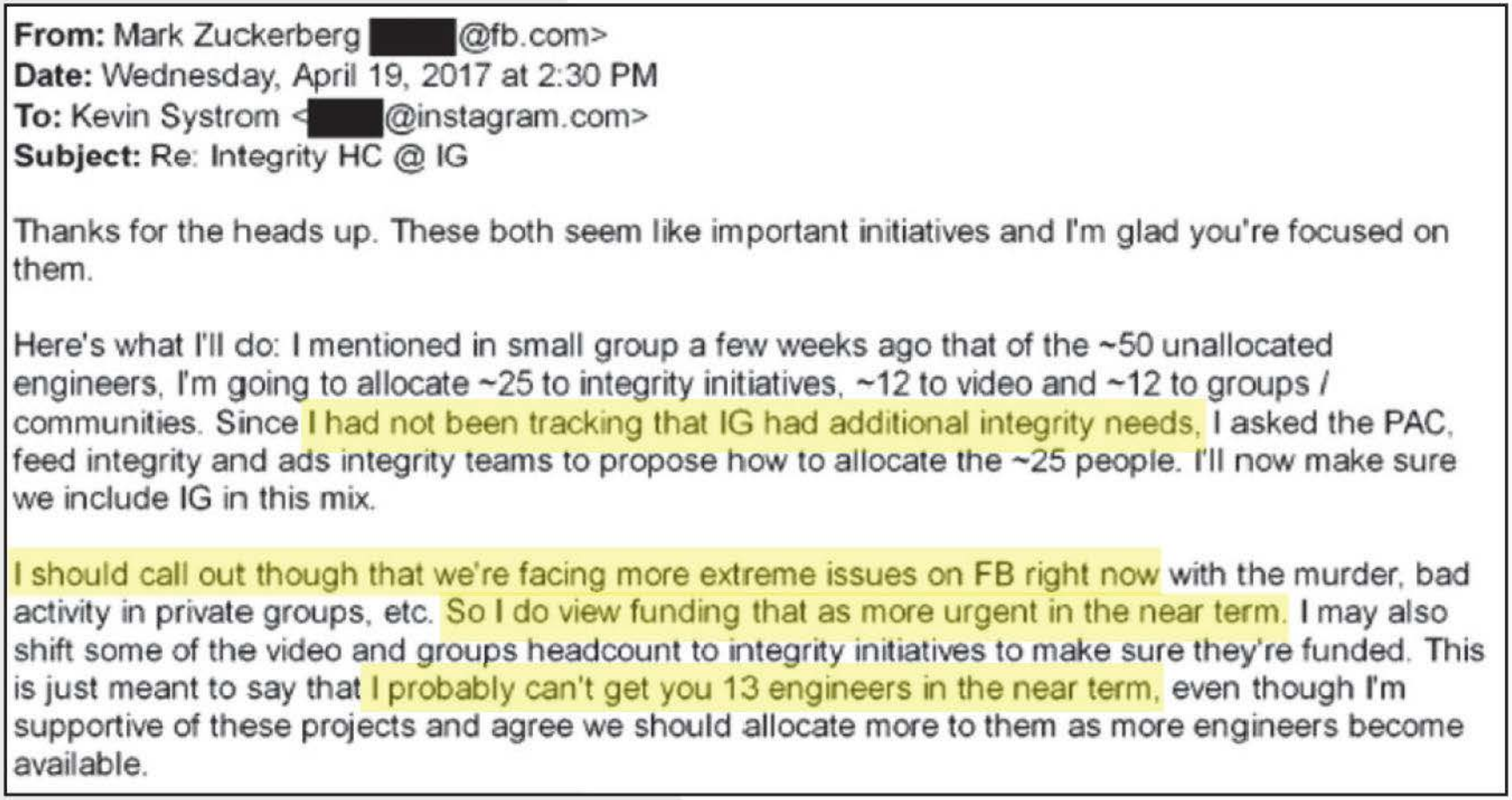
Dec. 2018
IG Integrity at risk in 9 content categories

Feb. 2019
Request for +68 integrity engineers to reach parity with FB

McCoy Report ¶¶ 69, 72, 257, 262;
McCoy Rebuttal Report ¶¶ 92, 126, 141-142, 258 PX15225, PX11034; PX3076; PX3077; PX2747; PX10899; PX15147; PX10900, PX10161, PX12360

QUALITATIVE ANALYSIS: UNDER-RESOURCING

2017



From: Mark Zuckerberg [redacted]@fb.com>
Date: Wednesday, April 19, 2017 at 2:30 PM
To: Kevin Systrom <[redacted]@instagram.com>
Subject: Re: Integrity HC @ IG

Thanks for the heads up. These both seem like important initiatives and I'm glad you're focused on them.

Here's what I'll do: I mentioned in small group a few weeks ago that of the ~50 unallocated engineers, I'm going to allocate ~25 to integrity initiatives, ~12 to video and ~12 to groups / communities. Since I had not been tracking that IG had additional integrity needs, I asked the PAC, feed integrity and ads integrity teams to propose how to allocate the ~25 people. I'll now make sure we include IG in this mix.

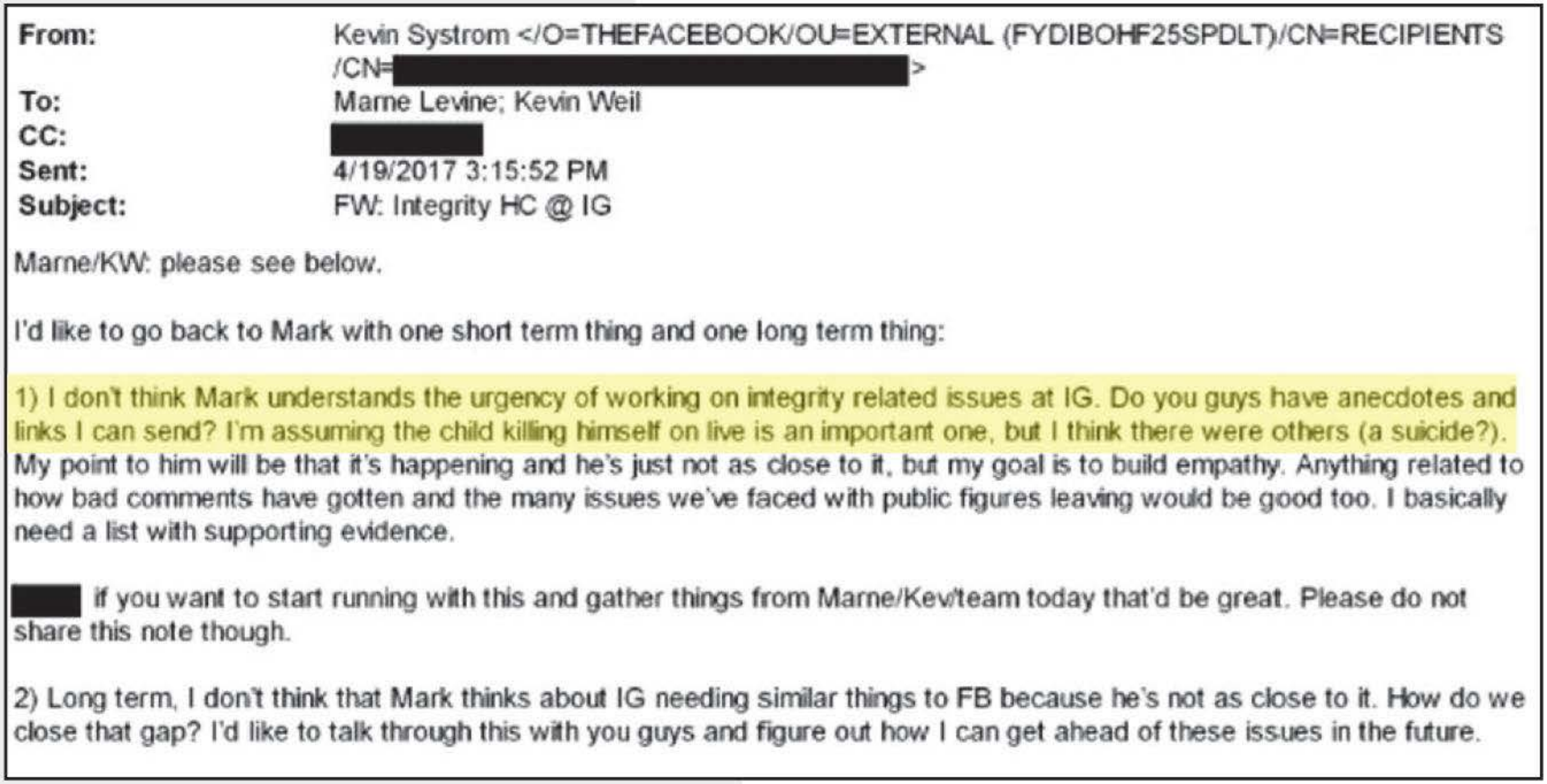
I should call out though that we're facing more extreme issues on FB right now with the murder, bad activity in private groups, etc. So I do view funding that as more urgent in the near term. I may also shift some of the video and groups headcount to integrity initiatives to make sure they're funded. This is just meant to say that I probably can't get you 13 engineers in the near term, even though I'm supportive of these projects and agree we should allocate more to them as more engineers become available.

PX15225

McCoy Rebuttal Report ¶¶ 141, 154

QUALITATIVE ANALYSIS: UNDER-RESOURCING

2017



From: Kevin Systrom </O=THEFACEBOOK/OU=EXTERNAL (FYDIBOHF25SPDLT)/CN=RECIPIENTS /CN=[redacted]>
To: Marne Levine; Kevin Weil
CC: [redacted]
Sent: 4/19/2017 3:15:52 PM
Subject: FW: Integrity HC @ IG

Marne/KW: please see below.

I'd like to go back to Mark with one short term thing and one long term thing:

1) I don't think Mark understands the urgency of working on integrity related issues at IG. Do you guys have anecdotes and links I can send? I'm assuming the child killing himself on live is an important one, but I think there were others (a suicide?). My point to him will be that it's happening and he's just not as close to it, but my goal is to build empathy. Anything related to how bad comments have gotten and the many issues we've faced with public figures leaving would be good too. I basically need a list with supporting evidence.

[redacted] if you want to start running with this and gather things from Marne/Kev/team today that'd be great. Please do not share this note though.

2) Long term, I don't think that Mark thinks about IG needing similar things to FB because he's not as close to it. How do we close that gap? I'd like to talk through this with you guys and figure out how I can get ahead of these issues in the future.

PX15225

McCoy Report ¶ 60;
McCoy Rebuttal Report ¶¶ 141, 154

QUALITATIVE ANALYSIS: UNDER-RESOURCING

2017

April 2017: Systrom writes to Parikh to request headcount (PX11034)

Systrom: “At SG two weeks ago Mark mentioned unlocking HC around integrity projects. Is this something you're working on? For context, **we have three areas that are unstaffed right now and a small amount of add'l headcount to staff them would unlock great work…”**

June 2017: Patel writes to Mosseri to request headcount (PX12330)

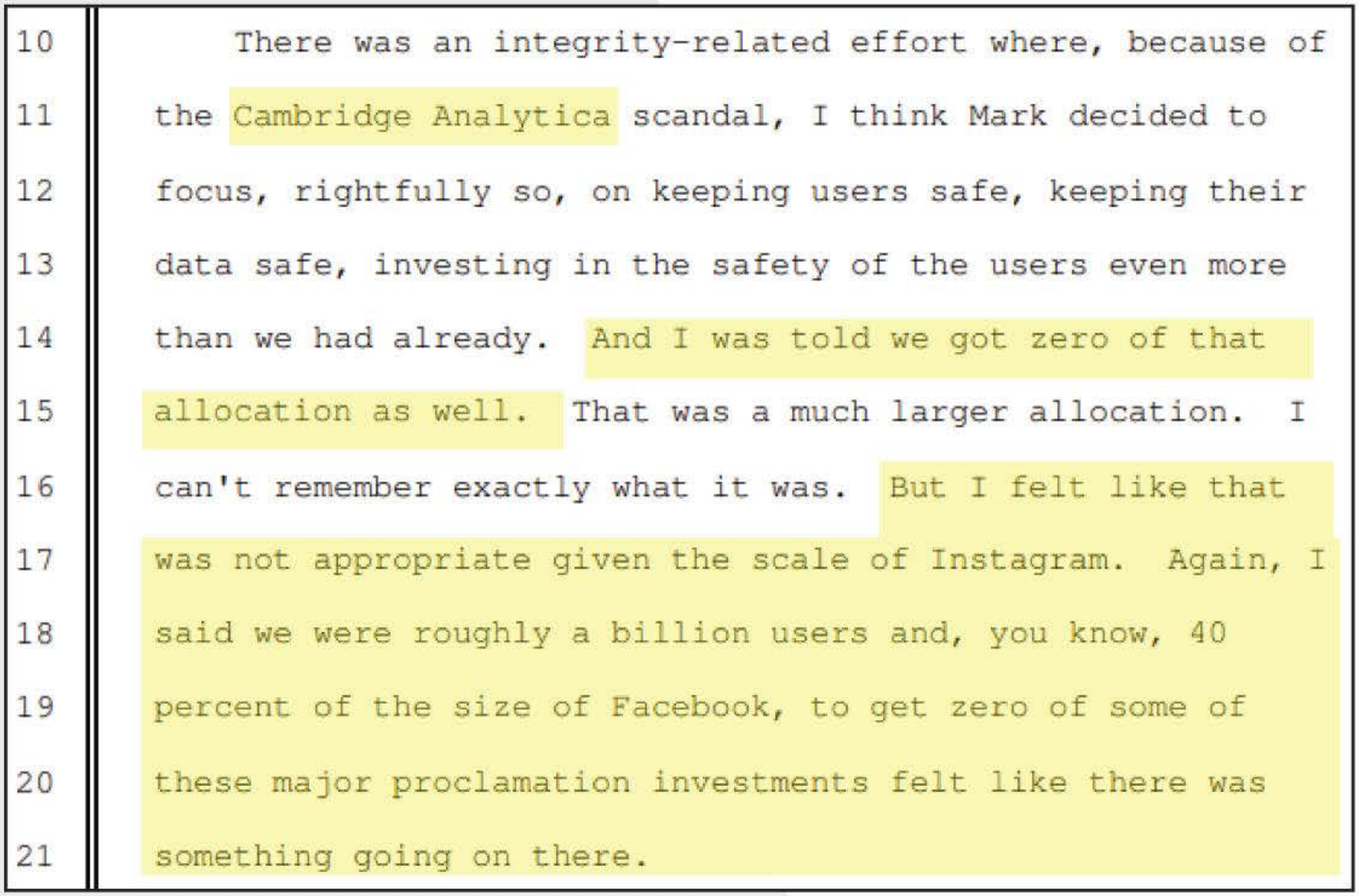
Mosseri: “High level I think you're doing the right[] thing, I think **IG is underinvested in integrity relative to its scale and importance** to the business. **I'm not sure I can help you pull headcount out of a hat…**”

June 2017: Systrom complains that Central Integrity does not focus on IG (PX3076)

Systrom: Re: Central Integrity “they need to stop being fb specific…**they have all these systems and really smart people yet don’t focus on ig…** Seems silly to invest in protecting 1.9b but ignore 750m… Esp when we have more live”

McCoy Report ¶¶ 68-70, 83; McCoy Rebuttal Report ¶ 140

Kevin Systrom's Trial Testimony



There was an integrity-related effort where, because of the Cambridge Analytica scandal, I think Mark decided to focus, rightfully so, on keeping users safe, keeping their data safe, investing in the safety of the users even more than we had already. And I was told we got zero of that allocation as well. That was a much larger allocation. I can't remember exactly what it was. But I felt like that was not appropriate given the scale of Instagram. Again, I said we were roughly a billion users and, you know, 40 percent of the size of Facebook, to get zero of some of these major proclamation investments felt like there was something going on there.

Kevin Systrom Trial, 91:10-21 (4.22.2025)

QUALITATIVE ANALYSIS: FACEBOOK BLUE PRIORITIZATION

MAY-JUNE 2018

May 2018: Systrom complains about "Blue Bias" (PX10899)

Mosseri: "**KevinS is still worried that CI has a strong Blue bias…** My sense is the teams are working to collaborate more and more, but **it does seem like our default…**"
Rosen: "**I think this is right and it's a muscle we need to change.**"

May 2018: IG's Well Being team had 29 engineers (PX3077)

Krieger: "**Our 29 engineer Well-being team is a small fraction of the size of any of the major Facebook integrity teams…** [centralization] is going to require technical work on both sides, as well as clear top-down direction for those teams that **they should care about the Instagram app surfaces in addition to the Facebook ones.**"

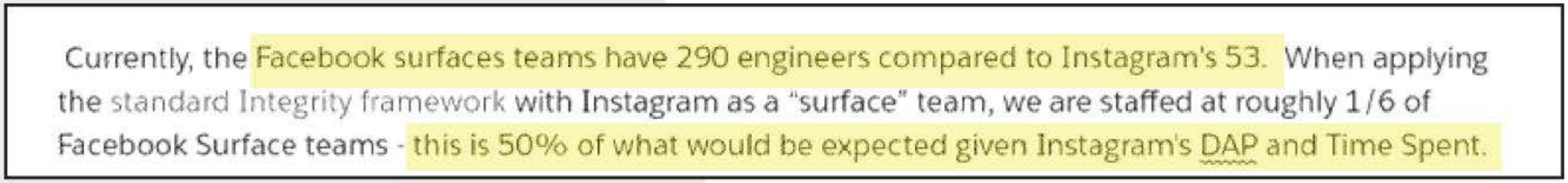
June 2018: Mark Zuckerberg's Directive (PX15147)

"The general direction will likely be that **teams focused on** growth, business platform, **integrity**, video infra, social good, and AI **should continue to focus on FB as their primary target, even if there are low-hanging fruit in other apps.**"

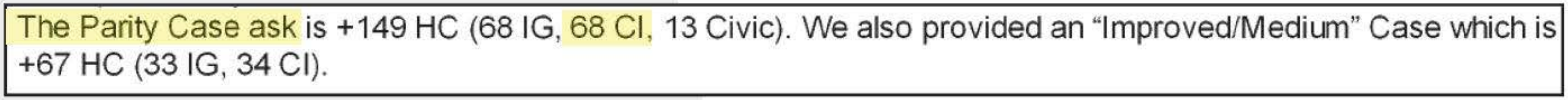
McCoy Report ¶¶ 47, 70-71; McCoy Rebuttal Report ¶¶ 83, 147, 150, 154

QUALITATIVE ANALYSIS: UNDER-RESOURCING & FACEBOOK BLUE PRIORITIZATION

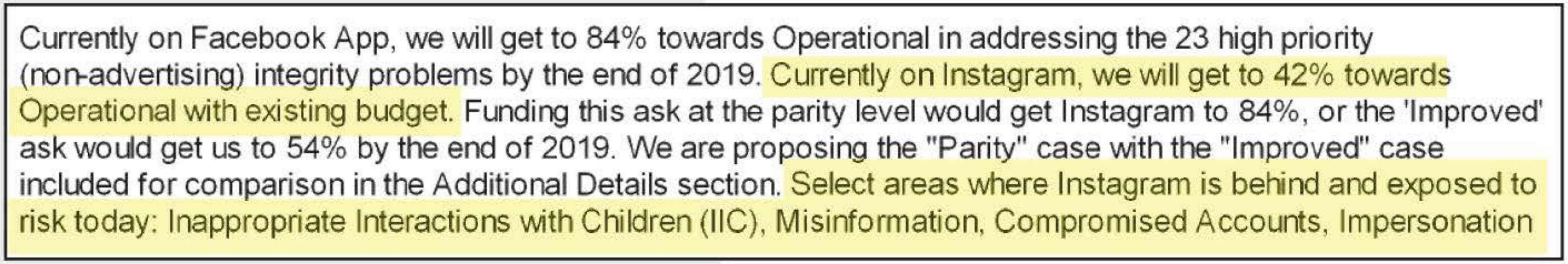
FEBRUARY 2019



Currently, the Facebook surfaces teams have 290 engineers compared to Instagram's 53. When applying the standard Integrity framework with Instagram as a "surface" team, we are staffed at roughly 1/6 of Facebook Surface teams - this is 50% of what would be expected given Instagram's DAP and Time Spent.



The Parity Case ask is +149 HC (68 IG, 68 CI, 13 Civic). We also provided an "Improved/Medium" Case which is +67 HC (33 IG, 34 CI).

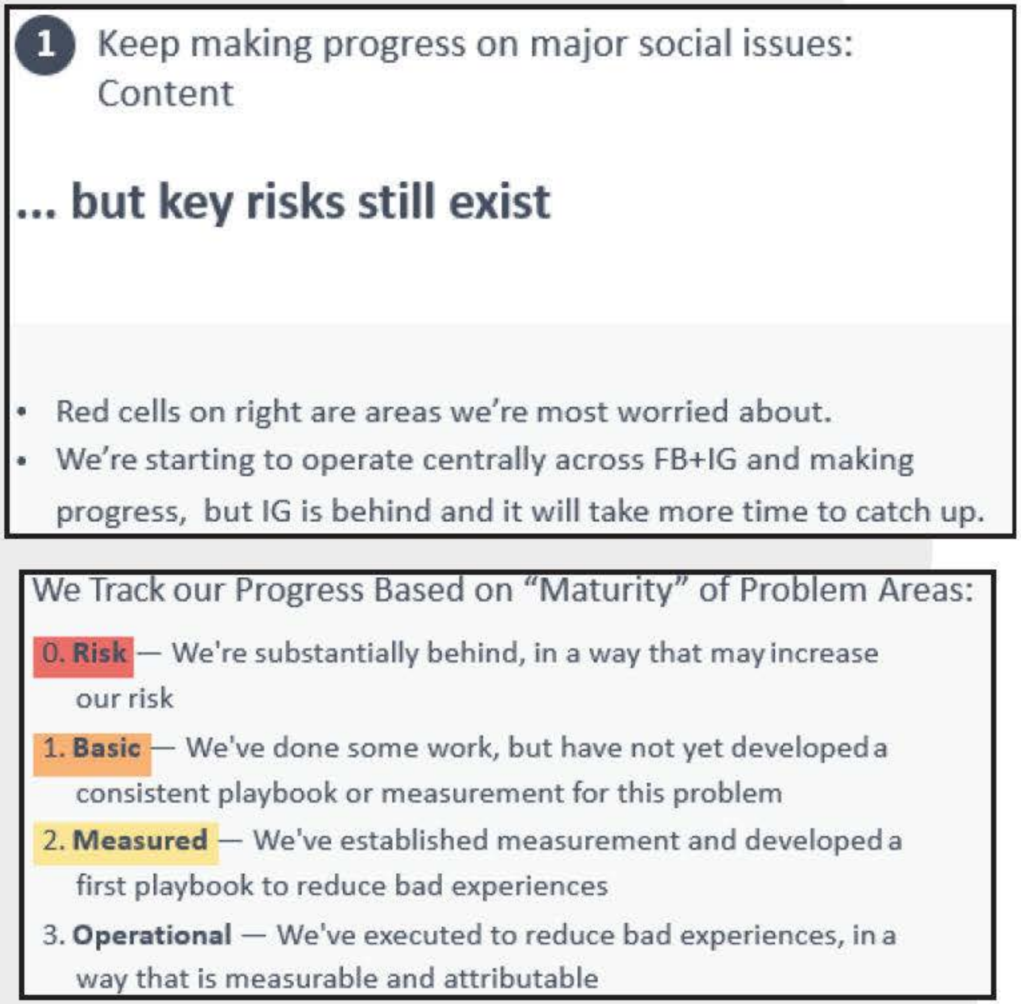


Currently on Facebook App, we will get to 84% towards Operational in addressing the 23 high priority (non-advertising) integrity problems by the end of 2019. Currently on Instagram, we will get to 42% towards Operational with existing budget. Funding this ask at the parity level would get Instagram to 84%, or the 'Improved' ask would get us to 54% by the end of 2019. We are proposing the "Parity" case with the "Improved" case included for comparison in the Additional Details section. Select areas where Instagram is behind and exposed to risk today: Inappropriate Interactions with Children (IIC), Misinformation, Compromised Accounts, Impersonation

PX3811; PX10900

McCoy Report ¶ 72;
McCoy Rebuttal Report ¶¶ 142-143

QUALITATIVE ANALYSIS: UNDER-RESOURCING & FACEBOOK BLUE PRIORITIZATION



1 Keep making progress on major social issues: Content

... but key risks still exist

- Red cells on right are areas we're most worried about.
- We're starting to operate centrally across FB+IG and making progress, but IG is behind and it will take more time to catch up.

We Track our Progress Based on "Maturity" of Problem Areas:

0. **Risk** — We're substantially behind, in a way that may increase our risk

1. **Basic** — We've done some work, but have not yet developed a consistent playbook or measurement for this problem

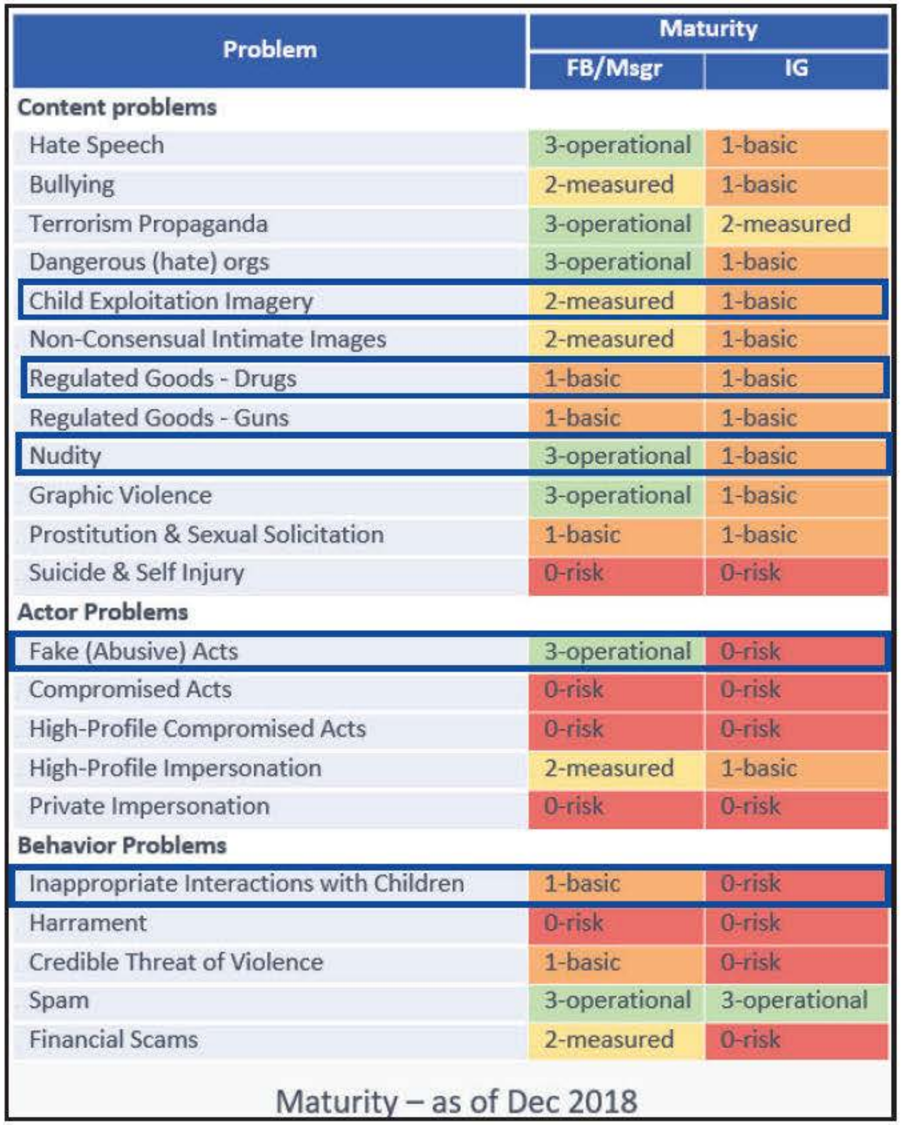
2. **Measured** — We've established measurement and developed a first playbook to reduce bad experiences

3. **Operational** — We've executed to reduce bad experiences, in a way that is measurable and attributable

PX12360

McCoy Report ¶¶ 77-78

INSTAGRAM'S MATURITY AS OF DECEMBER 2018

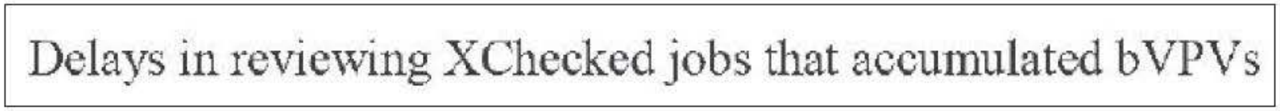


Problem	Maturity	
	FB/Msgr	IG
Content problems		
Hate Speech	3-operational	1-basic
Bullying	2-measured	1-basic
Terrorism Propaganda	3-operational	2-measured
Dangerous (hate) orgs	3-operational	1-basic
Child Exploitation Imagery	2-measured	1-basic
Non-Consensual Intimate Images	2-measured	1-basic
Regulated Goods - Drugs	1-basic	1-basic
Regulated Goods - Guns	1-basic	1-basic
Nudity	3-operational	1-basic
Graphic Violence	3-operational	1-basic
Prostitution & Sexual Solicitation	1-basic	1-basic
Suicide & Self Injury	0-risk	0-risk
Actor Problems		
Fake (Abusive) Acts	3-operational	0-risk
Compromised Acts	0-risk	0-risk
High-Profile Compromised Acts	0-risk	0-risk
High-Profile Impersonation	2-measured	1-basic
Private Impersonation	0-risk	0-risk
Behavior Problems		
Inappropriate Interactions with Children	1-basic	0-risk
Harrament	0-risk	0-risk
Credible Threat of Violence	1-basic	0-risk
Spam	3-operational	3-operational
Financial Scams	2-measured	0-risk

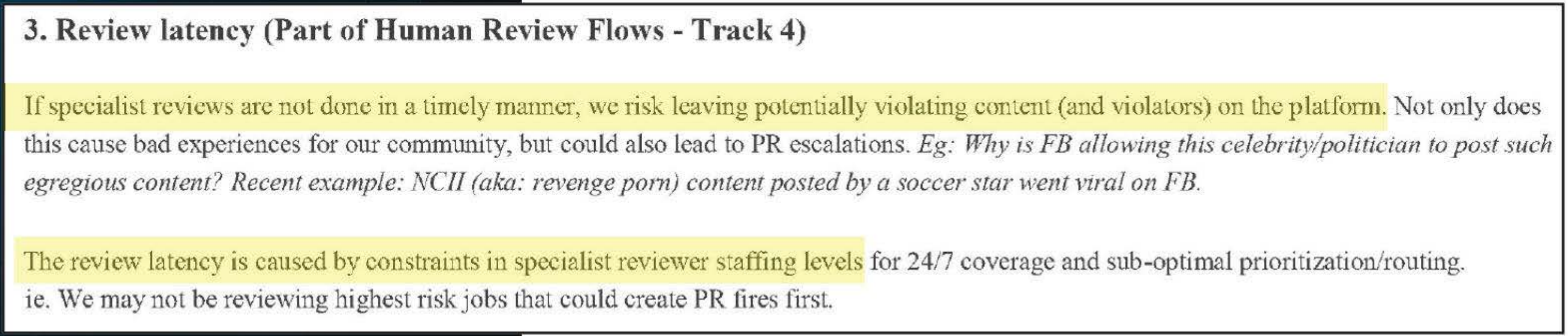
Maturity – as of Dec 2018

UNDER-RESOURCING: XCheck

2019



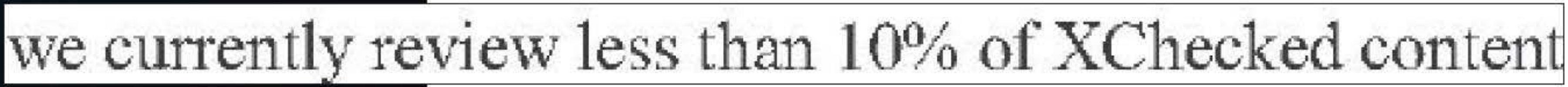
Delays in reviewing XChecked jobs that accumulated bVPVs



3. Review latency (Part of Human Review Flows - Track 4)

If specialist reviews are not done in a timely manner, we risk leaving potentially violating content (and violators) on the platform. Not only does this cause bad experiences for our community, but could also lead to PR escalations. *Eg: Why is FB allowing this celebrity/politician to post such egregious content? Recent example: NCII (aka: revenge porn) content posted by a soccer star went viral on FB.*

The review latency is caused by constraints in specialist reviewer staffing levels for 24/7 coverage and sub-optimal prioritization/routing. ie. We may not be reviewing highest risk jobs that could create PR fires first.

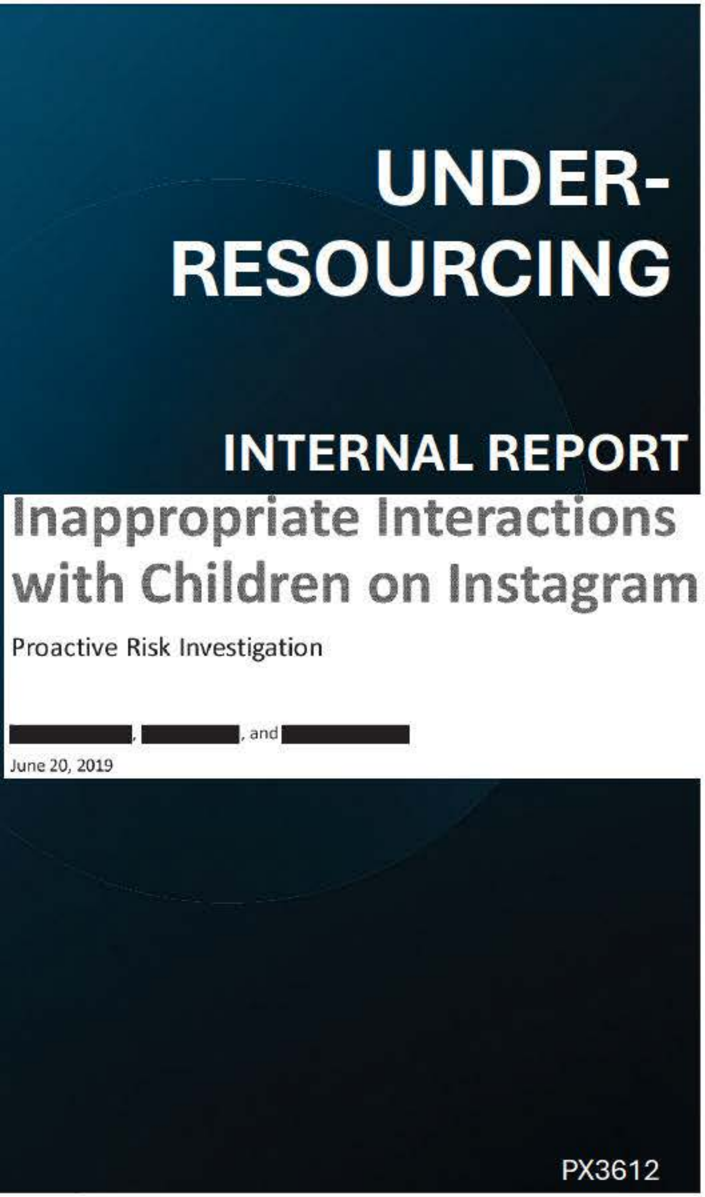


we currently review less than 10% of XChecked content

PX10939

McCoy Report ¶¶ 84, 89;
McCoy Rebuttal Report ¶ 163

UNDER-RESOURCING



INTERNAL REPORT

Inappropriate Interactions with Children on Instagram

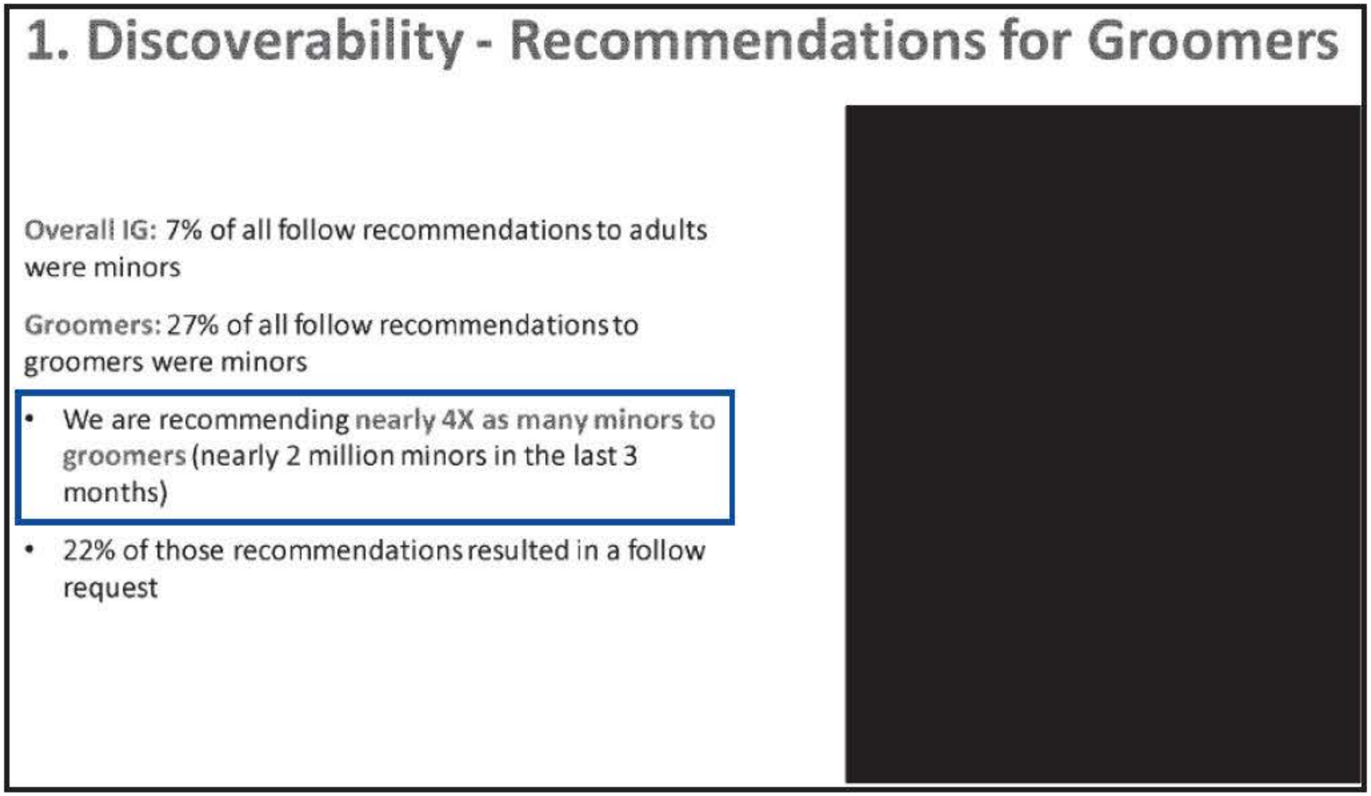
Proactive Risk Investigation

[redacted], [redacted], and [redacted]

June 20, 2019

PX3612

JUNE 2019



1. Discoverability - Recommendations for Groomers

Overall IG: 7% of all follow recommendations to adults were minors

Groomers: 27% of all follow recommendations to groomers were minors

- We are recommending nearly 4X as many minors to groomers (nearly 2 million minors in the last 3 months)
- 22% of those recommendations resulted in a follow request

[redacted]

See McCoy Report ¶ 248;
McCoy Rebuttal Report ¶¶ 53, 56, 363

UNDER-RESOURCING

CONTINUED

Inappropriate Interactions with Children on Instagram

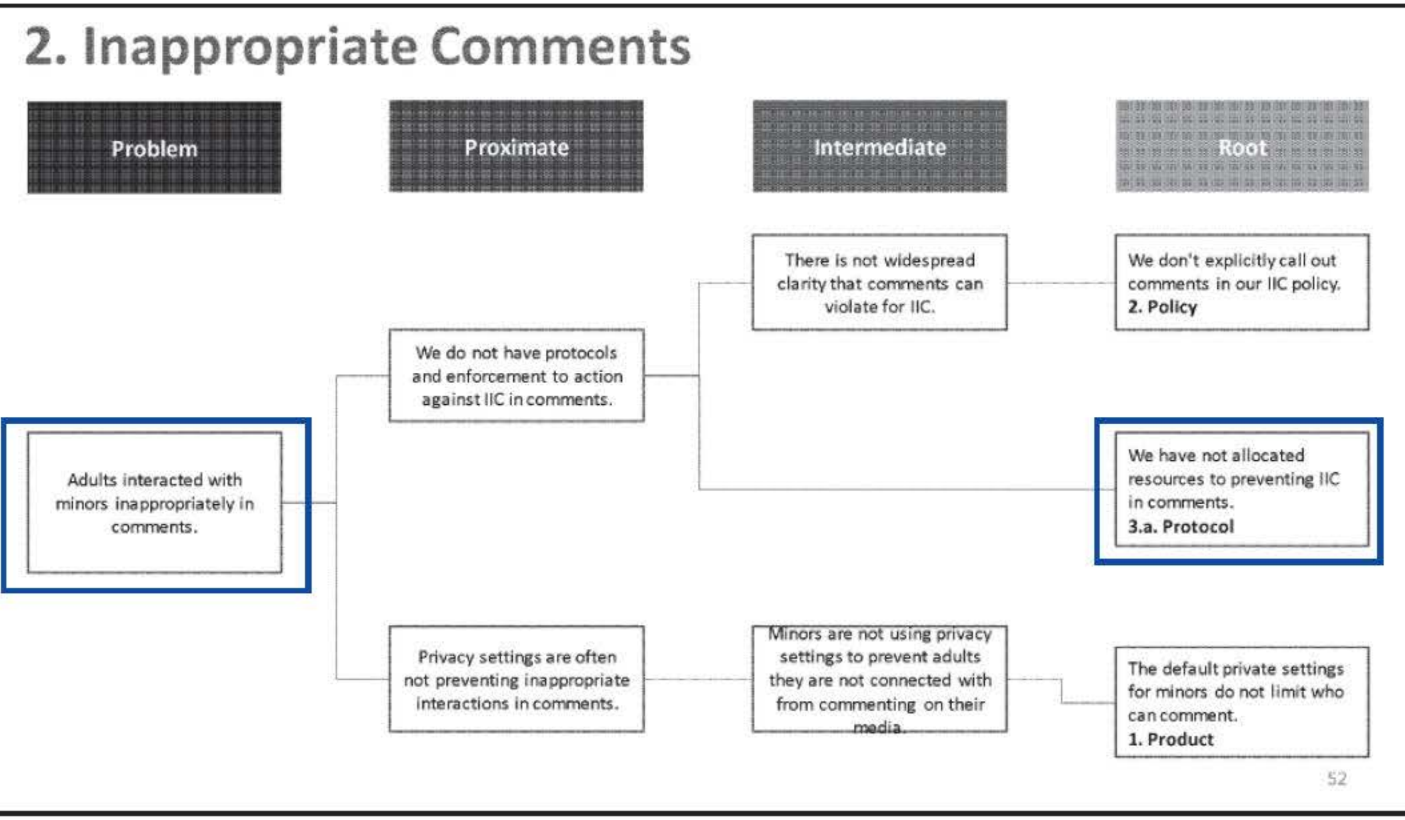
Proactive Risk Investigation

[redacted], [redacted], and [redacted]

June 20, 2019

PX3612

JUNE 2019



2. Inappropriate Comments

Problem | Proximate | Intermediate | Root

Adults interacted with minors inappropriately in comments.

We do not have protocols and enforcement to action against IIC in comments.

There is not widespread clarity that comments can violate for IIC.

We don't explicitly call out comments in our IIC policy. **2. Policy**

We have not allocated resources to preventing IIC in comments. **3.a. Protocol**

Privacy settings are often not preventing inappropriate interactions in comments.

Minors are not using privacy settings to prevent adults they are not connected with from commenting on their media.

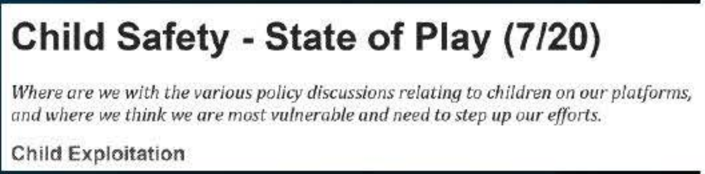
The default private settings for minors do not limit who can comment. **1. Product**

52

See McCoy Report ¶ 79;
McCoy Rebuttal Report ¶¶ 53-54, 56

UNDER-RESOURCING

INTERNAL REPORT



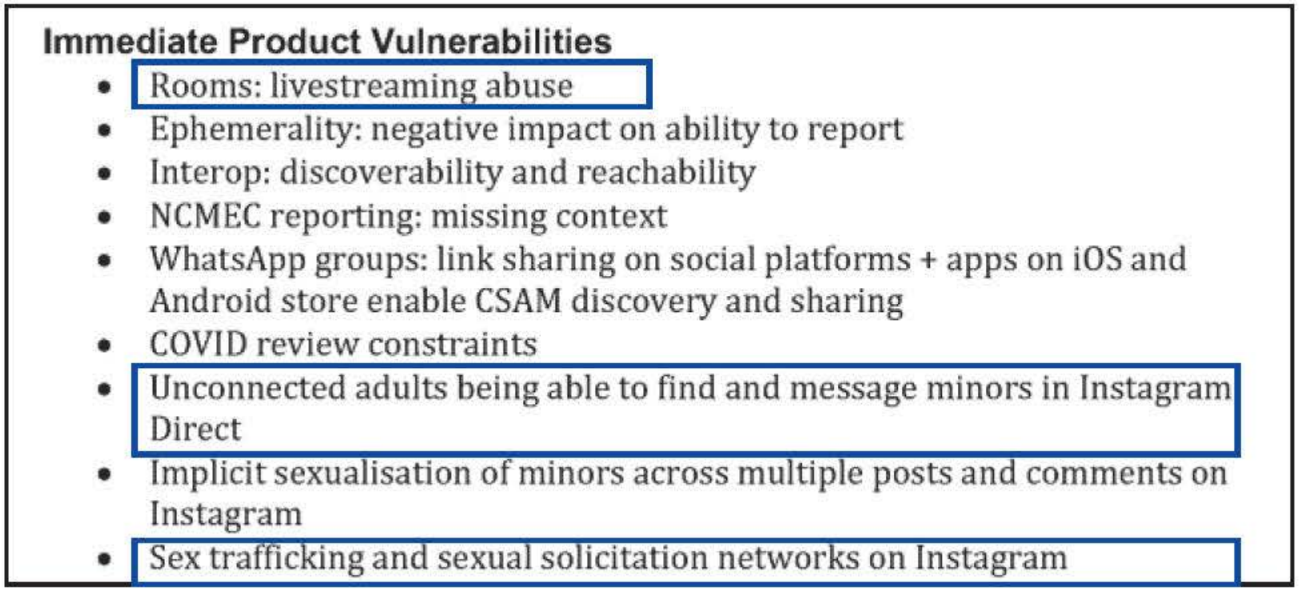
Child Safety - State of Play (7/20)

Where are we with the various policy discussions relating to children on our platforms, and where we think we are most vulnerable and need to step up our efforts.

Child Exploitation

PX3094; PX3618

JULY 2020



Immediate Product Vulnerabilities

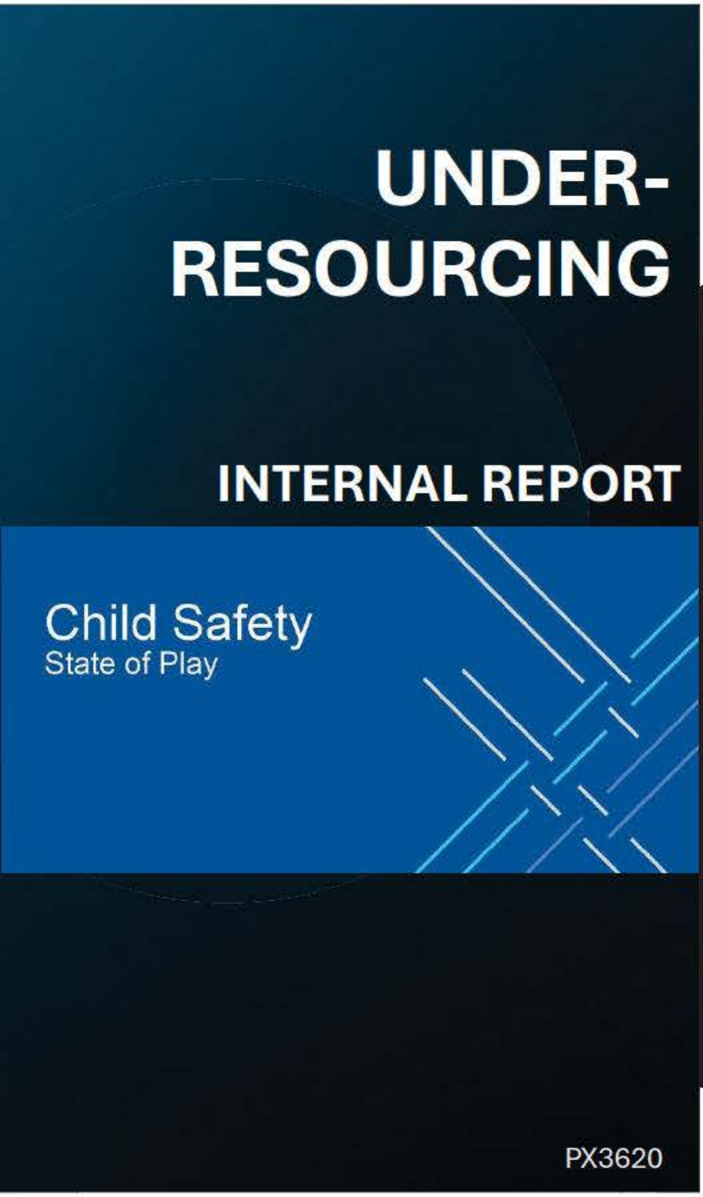
- Rooms: livestreaming abuse
- Ephemerality: negative impact on ability to report
- Interop: discoverability and reachability
- NCMEC reporting: missing context
- WhatsApp groups: link sharing on social platforms + apps on iOS and Android store enable CSAM discovery and sharing
- COVID review constraints
- Unconnected adults being able to find and message minors in Instagram Direct
- Implicit sexualisation of minors across multiple posts and comments on Instagram
- Sex trafficking and sexual solicitation networks on Instagram

DECEMBER 2021

"~1 in 100 IG teen users in the US are exposed to IIC convos (child grooming) with adults everyday."

See McCoy Report ¶¶ 21, 248;
McCoy Rebuttal Report ¶¶ 54, 57, 96

UNDER-RESOURCING

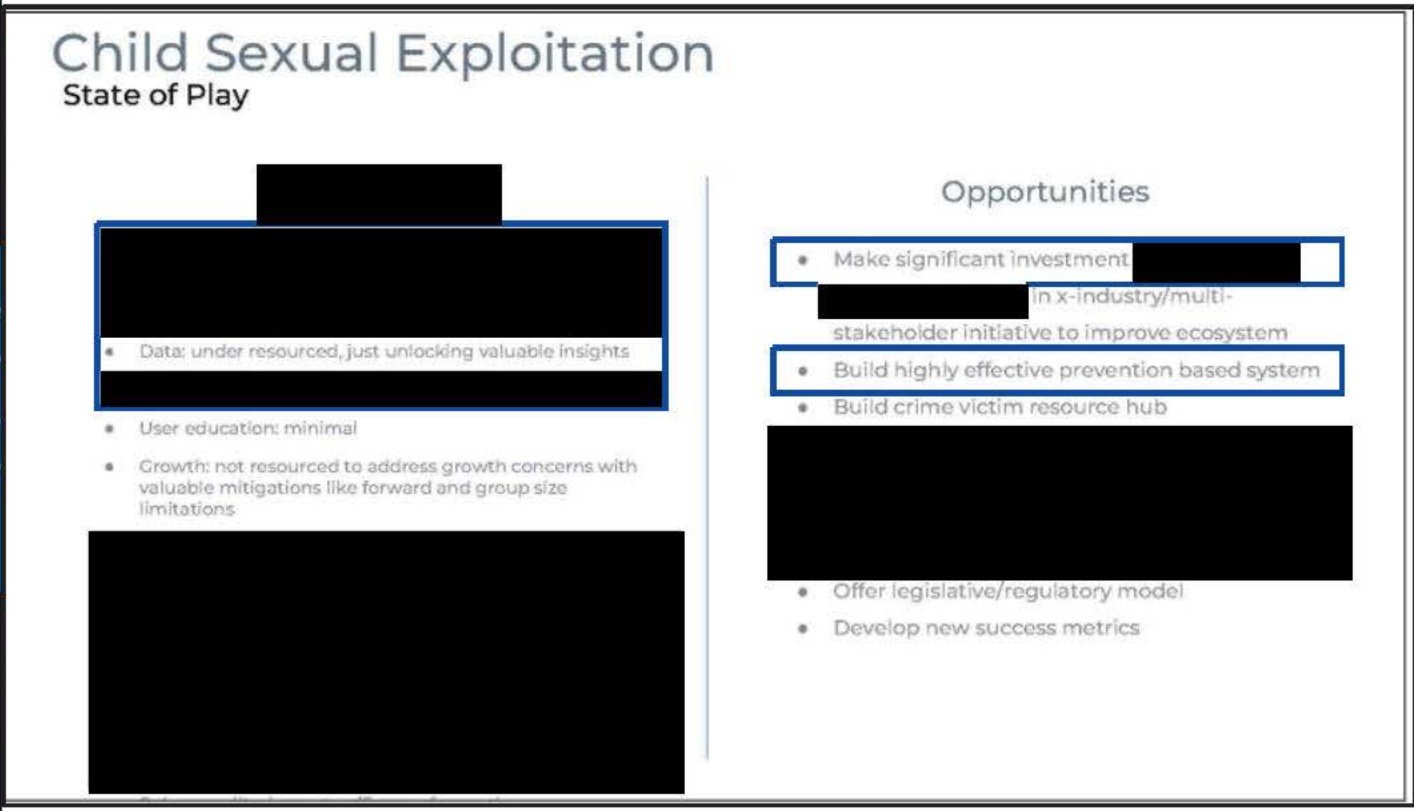


INTERNAL REPORT

Child Safety
State of Play

PX3620

MAY 2022



Child Sexual Exploitation
State of Play

[redacted]

[redacted]

- Data: under resourced, just unlocking valuable insights

[redacted]

- User education: minimal
- Growth: not resourced to address growth concerns with valuable mitigations like forward and group size limitations

[redacted]

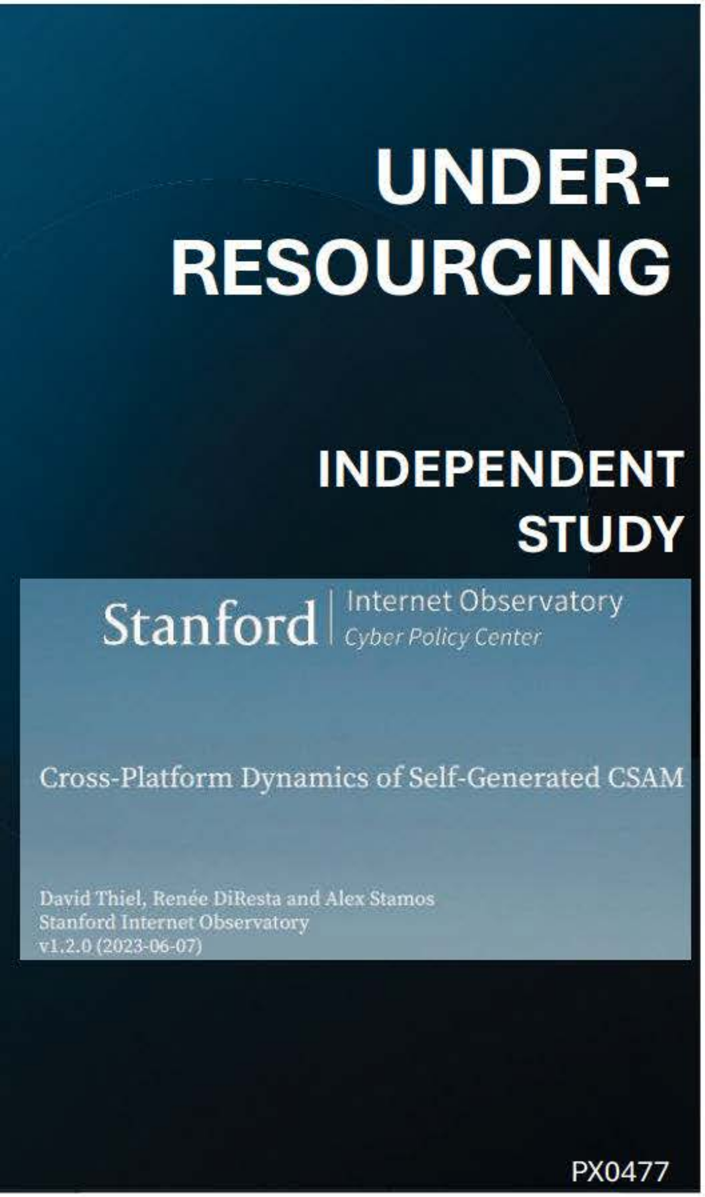
Opportunities

- Make significant investment [redacted] [redacted] in x-industry/multi-stakeholder initiative to improve ecosystem
- Build highly effective prevention based system
- Build crime victim resource hub

[redacted]

- Offer legislative/regulatory model
- Develop new success metrics

See McCoy Report ¶¶ 40, 78, 79, 248;
McCoy Rebuttal Report ¶¶ 56-57



UNDER-RESOURCING

INDEPENDENT STUDY

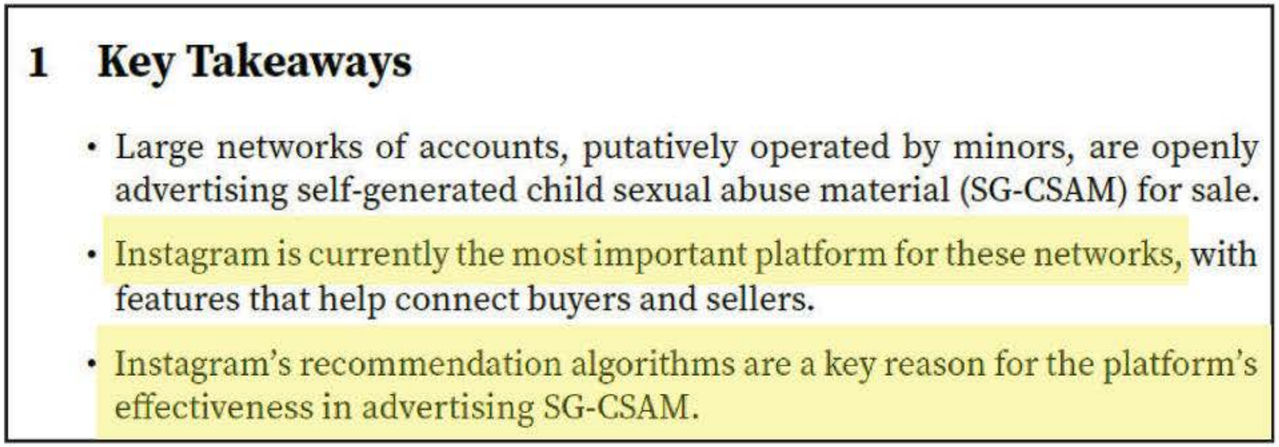
Stanford | Internet Observatory
Cyber Policy Center

Cross-Platform Dynamics of Self-Generated CSAM

David Thiel, Renée DiResta and Alex Stamos
Stanford Internet Observatory
v1.2.0 (2023-06-07)

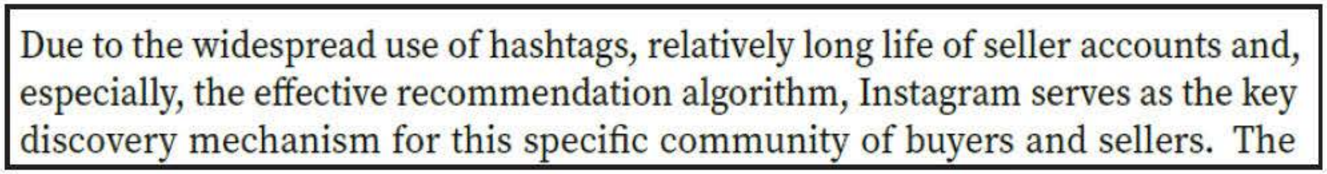
PX0477

JUNE 2023

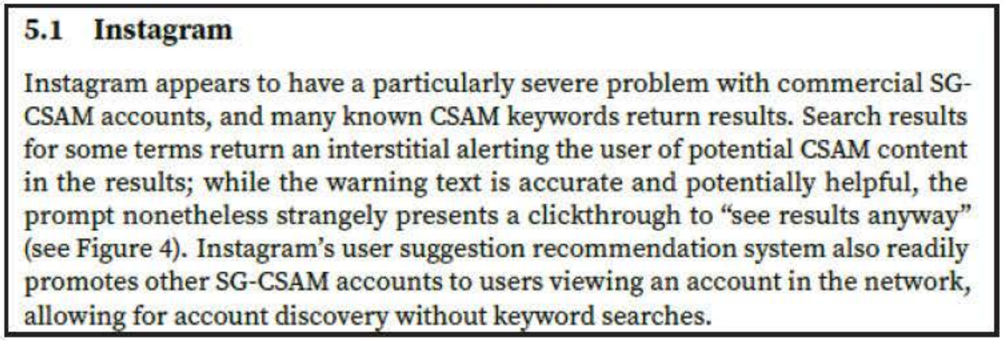


1 Key Takeaways

- Large networks of accounts, putatively operated by minors, are openly advertising self-generated child sexual abuse material (SG-CSAM) for sale.
- Instagram is currently the most important platform for these networks, with features that help connect buyers and sellers.
- Instagram's recommendation algorithms are a key reason for the platform's effectiveness in advertising SG-CSAM.



Due to the widespread use of hashtags, relatively long life of seller accounts and, especially, the effective recommendation algorithm, Instagram serves as the key discovery mechanism for this specific community of buyers and sellers. The



5.1 Instagram

Instagram appears to have a particularly severe problem with commercial SG-CSAM accounts, and many known CSAM keywords return results. Search results for some terms return an interstitial alerting the user of potential CSAM content in the results; while the warning text is accurate and potentially helpful, the prompt nonetheless strangely presents a clickthrough to "see results anyway" (see Figure 4). Instagram's user suggestion recommendation system also readily promotes other SG-CSAM accounts to users viewing an account in the network, allowing for account discovery without keyword searches.

McCoy Report ¶¶ 21, 248;
McCoy Rebuttal Report ¶ 351

CONCLUSIONS

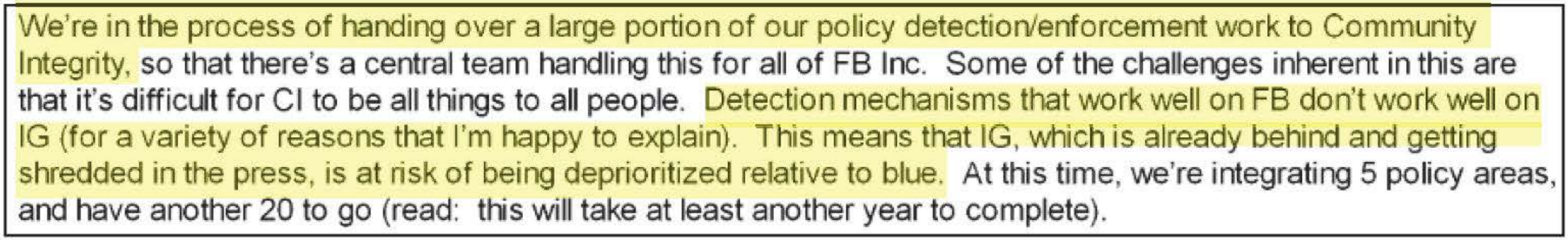
- Meta's acquisition of Instagram was not necessary for Instagram to address its integrity issues at scale because Meta's integrity solutions are not unique.
- Moreover, pre-acquisition Instagram was capably addressing its integrity issues and had access to the third-party integrity solutions and know-how it would need to continue improving its tools and scaling as other online platforms have done.
- Meta's governance of Instagram has been associated with meaningful integrity failures likely caused by Meta's under-resourcing of Instagram's integrity.

McCoy Report ¶¶ 17, 93, 116, 130; McCoy Rebuttal Report ¶¶ 4, 7-8, 367, 385, 391

APPENDIX

DEVELOPING EFFECTIVE INTEGRITY SYSTEMS

OCTOBER 2018



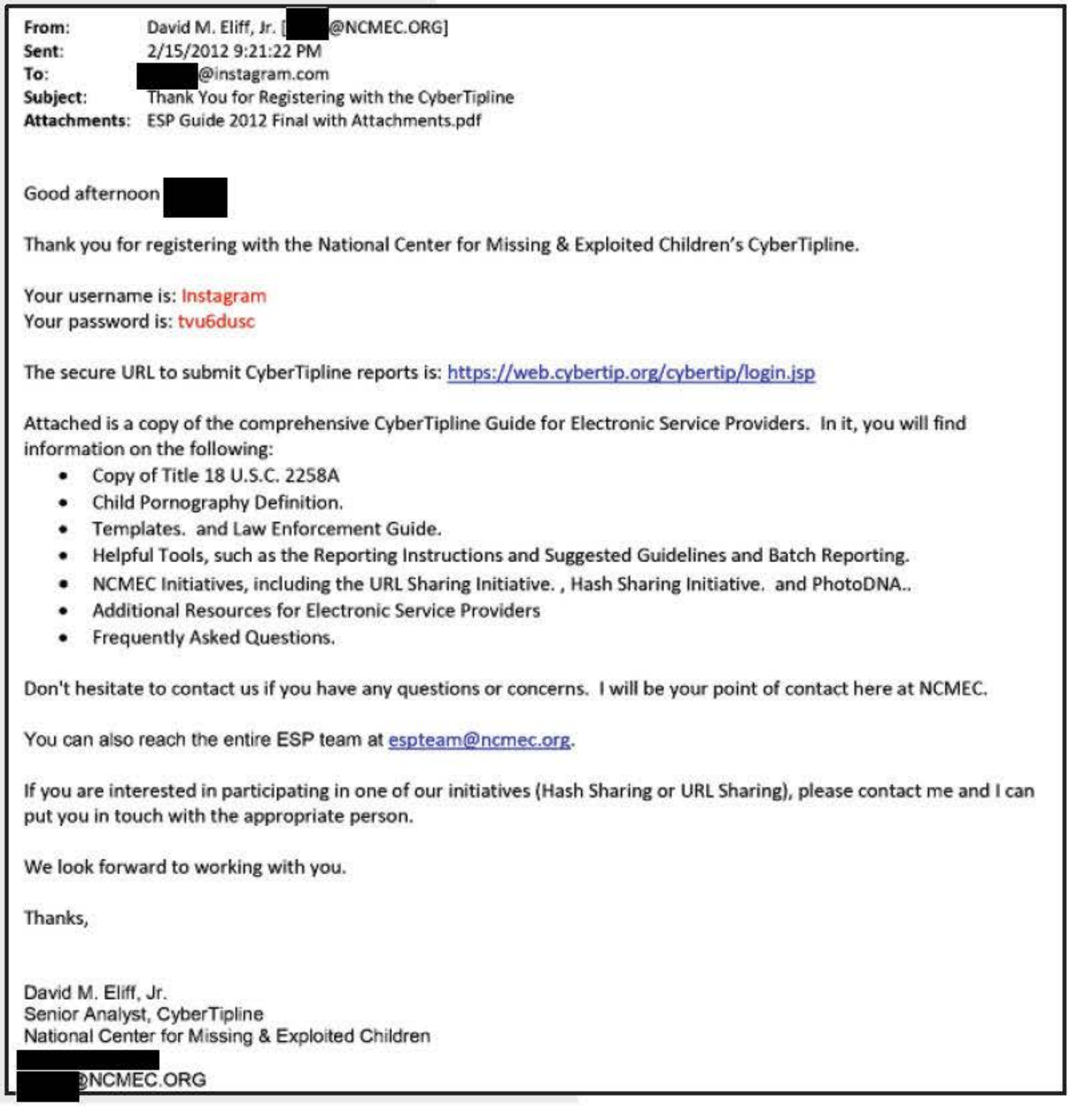
We're in the process of handing over a large portion of our policy detection/enforcement work to Community Integrity, so that there's a central team handling this for all of FB Inc. Some of the challenges inherent in this are that it's difficult for CI to be all things to all people. Detection mechanisms that work well on FB don't work well on IG (for a variety of reasons that I'm happy to explain). This means that IG, which is already behind and getting shredded in the press, is at risk of being deprioritized relative to blue. At this time, we're integrating 5 policy areas, and have another 20 to go (read: this will take at least another year to complete).

PX3079

McCoy Report ¶ 65

INSTAGRAM'S ACQUISITION NOT NECESSARY: INSTAGRAM PRE-ACQUISITION

FEBRUARY 15, 2012



From: David M. Eliff, Jr. [[redacted]@NCMEC.ORG]
Sent: 2/15/2012 9:21:22 PM
To: [redacted]@instagram.com
Subject: Thank You for Registering with the CyberTipline
Attachments: ESP Guide 2012 Final with Attachments.pdf

Good afternoon [redacted]

Thank you for registering with the National Center for Missing & Exploited Children's CyberTipline.

Your username is: Instagram
Your password is: tvu6dusc

The secure URL to submit CyberTipline reports is: https://web.cybertip.org/cybertip/login.jsp

Attached is a copy of the comprehensive CyberTipline Guide for Electronic Service Providers. In it, you will find information on the following:

- Copy of Title 18 U.S.C. 2258A
- Child Pornography Definition.
- Templates. and Law Enforcement Guide.
- Helpful Tools, such as the Reporting Instructions and Suggested Guidelines and Batch Reporting.
- NCMEC Initiatives, including the URL Sharing Initiative. , Hash Sharing Initiative. and PhotoDNA..
- Additional Resources for Electronic Service Providers
- Frequently Asked Questions.

Don't hesitate to contact us if you have any questions or concerns. I will be your point of contact here at NCMEC.

You can also reach the entire ESP team at espteam@ncmec.org.

If you are interested in participating in one of our initiatives (Hash Sharing or URL Sharing), please contact me and I can put you in touch with the appropriate person.

We look forward to working with you.

Thanks,

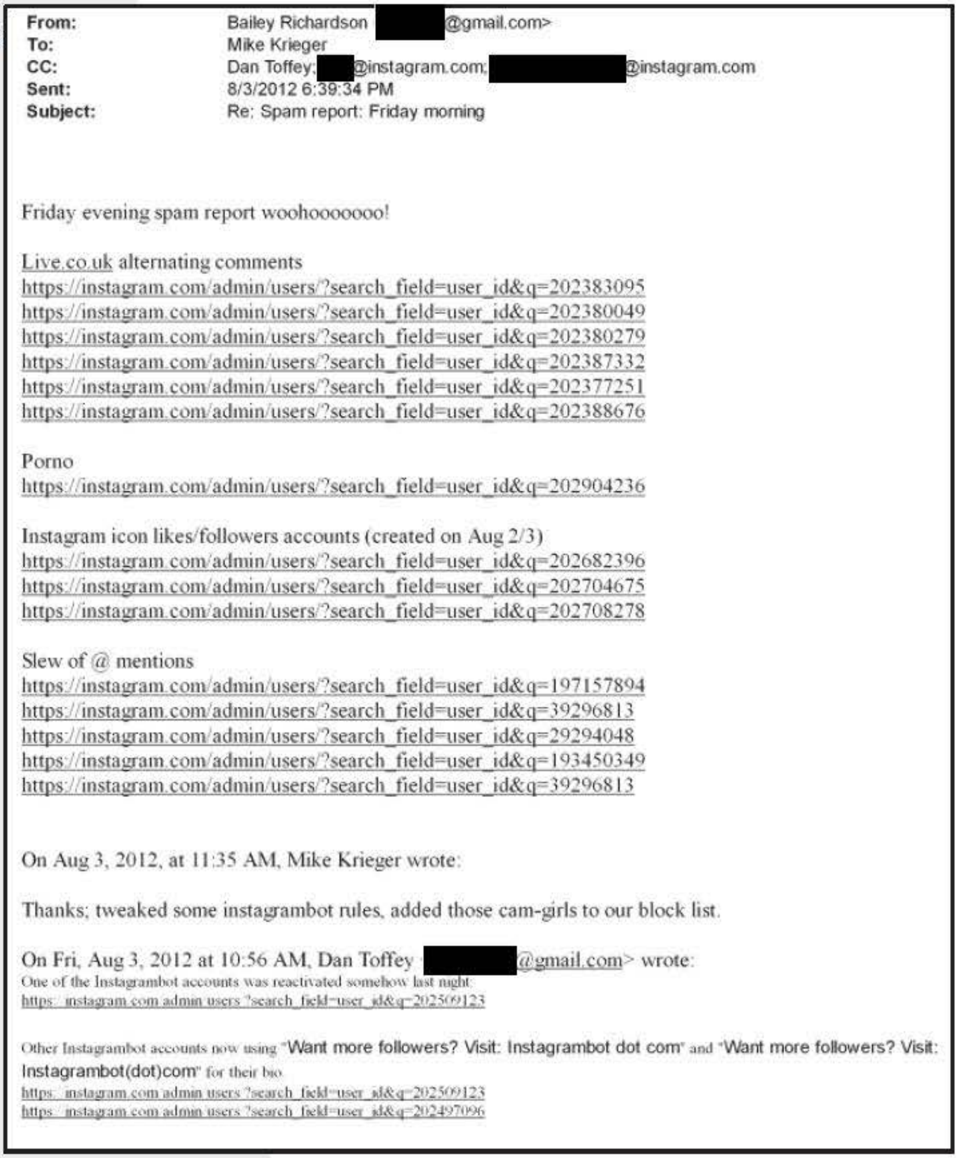
David M. Eliff, Jr.
Senior Analyst, CyberTipline
National Center for Missing & Exploited Children
[redacted]NCMEC.ORG

PX3803

McCoy Rebuttal Report ¶ 355

INSTAGRAM'S ACQUISITION NOT NECESSARY: INSTAGRAM PRE-ACQUISITION

AUGUST 3, 2012



From: Bailey Richardson [redacted]@gmail.com>
To: Mike Krieger
CC: Dan Toffey; [redacted]@instagram.com; [redacted]@instagram.com
Sent: 8/3/2012 6:39:34 PM
Subject: Re: Spam report: Friday morning

Friday evening spam report woohoooooooo!

Live.co.uk alternating comments
https://instagram.com/admin/users/?search_field=user_id&q=202383095
https://instagram.com/admin/users/?search_field=user_id&q=202380049
https://instagram.com/admin/users/?search_field=user_id&q=202380279
https://instagram.com/admin/users/?search_field=user_id&q=202387332
https://instagram.com/admin/users/?search_field=user_id&q=202377251
https://instagram.com/admin/users/?search_field=user_id&q=202388676

Porno
https://instagram.com/admin/users/?search_field=user_id&q=202904236

Instagram icon likes/followers accounts (created on Aug 2/3)
https://instagram.com/admin/users/?search_field=user_id&q=202682396
https://instagram.com/admin/users/?search_field=user_id&q=202704675
https://instagram.com/admin/users/?search_field=user_id&q=202708278

Slew of @ mentions
https://instagram.com/admin/users/?search_field=user_id&q=197157894
https://instagram.com/admin/users/?search_field=user_id&q=39296813
https://instagram.com/admin/users/?search_field=user_id&q=29294048
https://instagram.com/admin/users/?search_field=user_id&q=193450349
https://instagram.com/admin/users/?search_field=user_id&q=39296813

On Aug 3, 2012, at 11:35 AM, Mike Krieger wrote:

Thanks; tweaked some instagrambot rules, added those cam-girls to our block list.

On Fri, Aug 3, 2012 at 10:56 AM, Dan Toffey [redacted]@gmail.com> wrote:
One of the Instagrambot accounts was reactivated somehow last night:
https://instagram.com/admin/users/?search_field=user_id&q=202509123

Other Instagrambot accounts now using "Want more followers? Visit: Instagrambot dot com" and "Want more followers? Visit: Instagrambot(dot)com" for their bio
https://instagram.com/admin/users/?search_field=user_id&q=202509123
https://instagram.com/admin/users/?search_field=user_id&q=202497096

PX3066

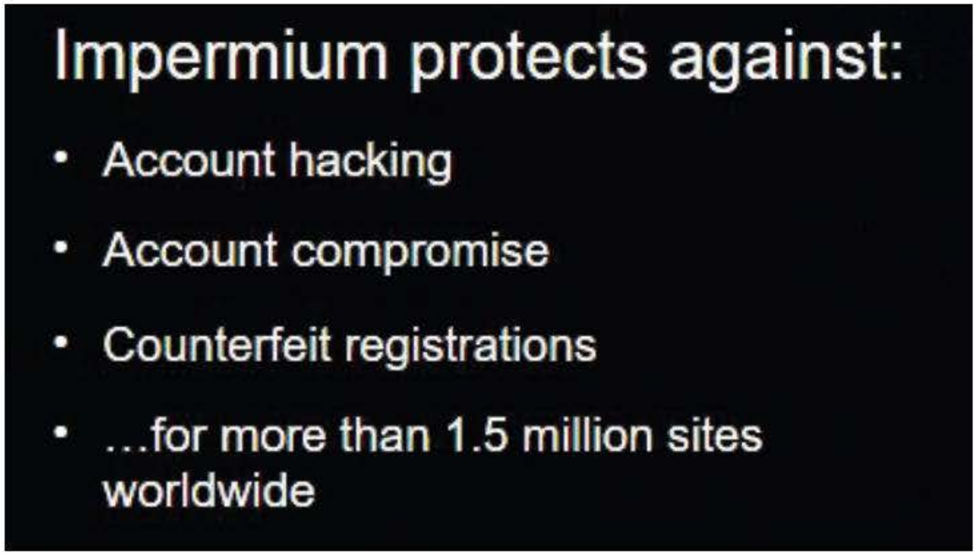
McCoy Rebuttal Report ¶ 201

INSTAGRAM'S ACQUISITION NOT NECESSARY: INSTAGRAM PRE-ACQUISITION

IMPERMIUM

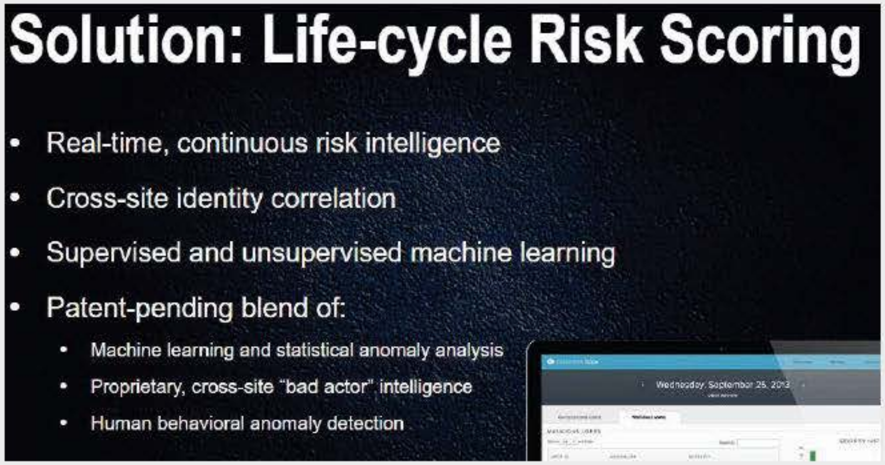
Account Compromise

Oct 11, 2013



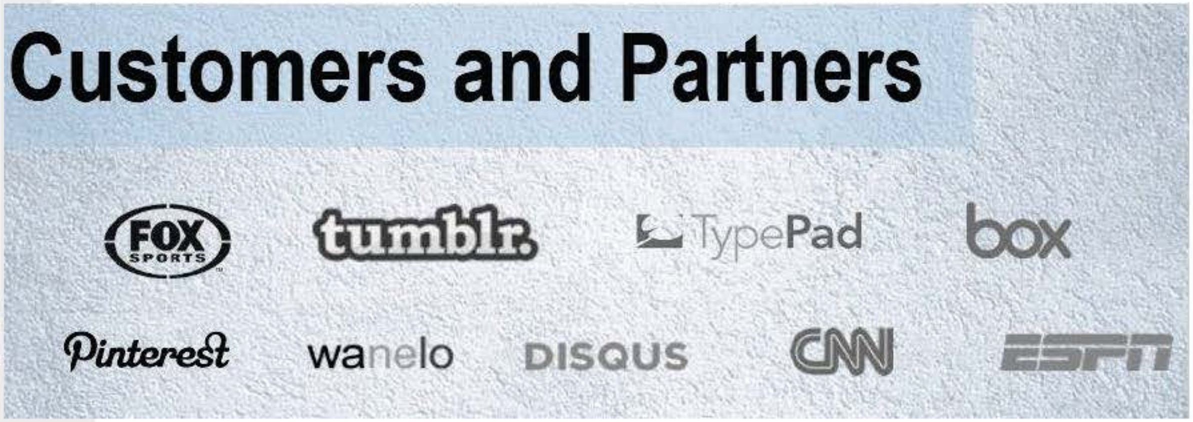
Impermium protects against:

- Account hacking
- Account compromise
- Counterfeit registrations
- ...for more than 1.5 million sites worldwide



Solution: Life-cycle Risk Scoring

- Real-time, continuous risk intelligence
- Cross-site identity correlation
- Supervised and unsupervised machine learning
- Patent-pending blend of:
  - Machine learning and statistical anomaly analysis
  - Proprietary, cross-site "bad actor" intelligence
  - Human behavioral anomaly detection



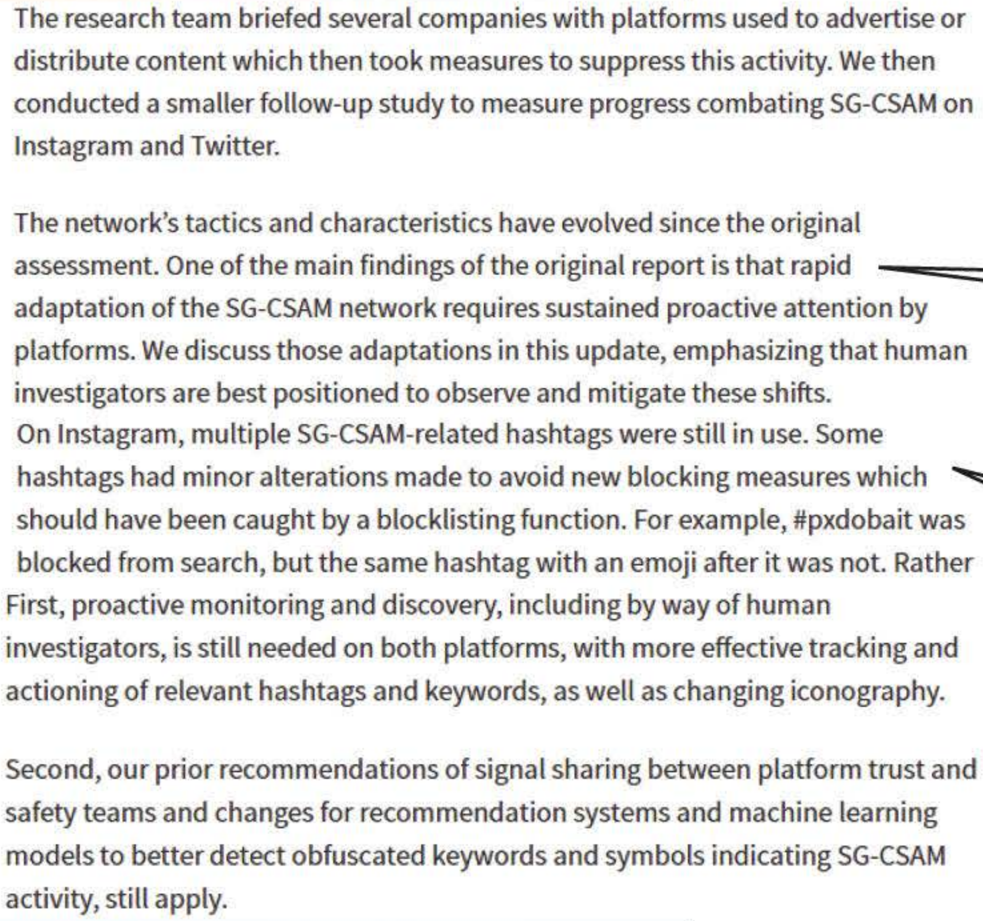
Customers and Partners

FOX SPORTS
tumblr.
TypePad
box
Pinterest
wanelo
DISQUS
CNN
ESPN

TWTR-META00109310 at 2, 4, 6

McCoy Report ¶138

UNDER-RESOURCING



The research team briefed several companies with platforms used to advertise or distribute content which then took measures to suppress this activity. We then conducted a smaller follow-up study to measure progress combating SG-CSAM on Instagram and Twitter.

The network's tactics and characteristics have evolved since the original assessment. One of the main findings of the original report is that rapid adaptation of the SG-CSAM network requires sustained proactive attention by platforms. We discuss those adaptations in this update, emphasizing that human investigators are best positioned to observe and mitigate these shifts.

On Instagram, multiple SG-CSAM-related hashtags were still in use. Some hashtags had minor alterations made to avoid new blocking measures which should have been caught by a blocklisting function. For example, #pxdobait was blocked from search, but the same hashtag with an emoji after it was not. Rather

First, proactive monitoring and discovery, including by way of human investigators, is still needed on both platforms, with more effective tracking and actioning of relevant hashtags and keywords, as well as changing iconography.

Second, our prior recommendations of signal sharing between platform trust and safety teams and changes for recommendation systems and machine learning models to better detect obfuscated keywords and symbols indicating SG-CSAM activity, still apply.

Stan. Internet Observatory: Cyber Policy Center, *An update on the SG-CSAM ecosystem* (Sept. 21, 2023))

All Cyber News / Blogs / September 21, 2023

An update on the SG-CSAM ecosystem

Stanford Internet Observatory

"One of the main findings of the original report is that rapid **adaptation of the SG-CSAM network requires sustained proactive attention by platforms**."

Findings: Weak hashtag enforcement

"On Instagram, **multiple SG-CSAM-related hashtags were still in use. Some hashtags had minor alterations** made to avoid new blocking measures which should have been caught by a blocklisting function."

McCoy Rebuttal Report ¶ 55