
— PROTECTED & SENSITIVE WHISTLEBLOWER DISCLOSURE —

July 6, 2022

U.S. Securities and Exchange Commission
Office of the Whistleblower (c/o ENF-CPU)
14420 Albemarle Point Place, Suite 102
Chantilly, VA 20151-1750

Reenah L. Kim
Bureau of Consumer Protection – Division of Enforcement
U.S. Federal Trade Commission
600 Pennsylvania Avenue NW, CC-9528
Washington DC 20580

Brian M. Boynton
Principal Deputy Assistant Attorney General
Civil Division

Jonathan Kanter
Assistant Attorney General
Antitrust Division

U.S. Department of Justice
950 Pennsylvania Avenue, N.W.
Washington, DC 20530-0001
Via FedEx

**Re: Protected Disclosures of Federal Trade Commission Act
Violations, Material Misrepresentations and Omissions, and Fraud by
Twitter, Inc. (NASDAQ: TWTR) and CEO Parag Agrawal, SEC TCR#
16571-665-930-019**

Whistleblower Aid is a U.S. tax-exempt, 501(c)(3) organization, EIN 26-4716045.

<https://WhistleblowerAid.org> — Anonymously via Tor Browser:

<http://p6ufg73qskew53cglxt6hktyt35rbl46yultzyuytq3tvicywa3plid.onion>

Contact via **SecureDrop** over Tor: <http://whistlebloweraid.securedrop.tor.onion> — via **Signal App**: +1 201-773-1371

To the Securities and Exchange Commission (“SEC”), Federal Trade Commission (“FTC”), and Department of Justice (“DOJ”):

I. Legal Violations and Deceit

1. We are lawyers representing **Peiter “Mudge”¹ Zatk**, who was employed as “Security Lead”, a member of the senior executive team responsible for Information Security, Privacy, Physical Security, Information Technology, and “Twitter Service” (the corporate division responsible for global content moderation enforcement) at **Twitter, Inc.** from November 16, 2020,² until the morning of January 19, 2022, when CEO Parag Agrawal terminated Mudge. During Mudge’s employment, he uncovered extreme, egregious deficiencies by Twitter in every area of his mandate including (as described in detail below) user privacy, digital and physical security, and platform integrity / content moderation. As described below, these deficiencies are the basis for our client’s reasonable belief in extensive legal violations by Twitter, Inc.
2. In this submission, Mudge makes protected, lawful disclosures of original evidence³ showing that the corporation, CEO Parag Agrawal, particular senior executives, and particular members of its Board of Directors, **since 2011 and on an ongoing basis**, have engaged in:
 - a. Extensive, repeated, uninterrupted violations of the Federal Trade Commission Act by making **false and misleading statements** to users and

¹ Decades ago, anonymous articles signed with the pseudonym “Mudge” began to appear, exposing security vulnerabilities. These articles helped people to understand and correct the problems, but they were contentious. Companies called out by Mudge did everything except fix the problems: They denied everything, threatened litigation, even sought to get Mudge fired from jobs elsewhere. Over the course of years of pioneering research on buffer overflow, code injection and other fundamental vulnerabilities, Mudge became recognized as a luminary in the information security field. Mudge’s real identity remained obscured until the late 1990s when Mudge was invited to meet the President, and White House staff accidentally disclosed that Mudge was in fact Peiter Zatk.

² Before joining Twitter, Mudge held senior positions at Google and Stripe, and within the Department of Defense, where he was authorized to access Top Secret / Special Compartmented Information for work on programs at the bleeding edge of both offensive and defensive cyber operations. Mudge has been formally recognized by the CIA, White House, U.S. Army, and The Office of the Secretary of Defense bestowed upon Mudge the Exceptional Public Service Award, the highest medal honor available to civilian, non-career officials. See https://en.wikipedia.org/wiki/Peiter_Zatko. Former Twitter CEO Jack Dorsey cited this track record of speaking truth to power as a primary reason for recruiting Mudge.

³ As described in more detail below, we have worked carefully to ensure that no Attorney-Client Privileged materials or communications are included in this disclosure or exhibits.

-
- the FTC about, *inter alia*, the Twitter platform’s **security, privacy, and integrity**;
- b. Violations of **SEC rules governing public companies** including, *inter alia*, auditing requirements;
 - c. **Fraudulent and material misrepresentations** in communications with the Board of Directors and investors, constituting securities law violations;
 - d. Negligence and even complicity with respect to efforts by **foreign governments to infiltrate, control, exploit, surveil and/or censor** the company’s platform, staff, and operations.
3. **Particular episodes of fraud** and deliberate efforts to mislead include, among other examples:
- a. In or around February 2021, after Mudge had prepared comprehensive written materials to educate the Board on his findings about the company’s extensive security, privacy and integrity problems, Mudge was instructed **not to send them** to the Board of Directors.
 - b. On multiple occasions during 2021, described in greater detail below, Mudge witnessed senior executives engaging in deceitful and/or misleading communications affecting Board members, users and shareholders. In contrast, Mudge spent 2021 designing and implementing a long-term strategy to reform and address Twitter’s privacy, security and integrity vulnerabilities. On December 14, 2021, against Mudge’s recommendation, CEO Agrawal explicitly **instructed Mudge to provide documents which both of them knew to be false and misleading**, regarding vital information security matters, to the Risk Committee of Twitter’s Board of Directors.⁴
 - c. In January 2022, Mudge began working to document evidence of fraud. Twitter’s Chief Compliance Officer opened a fraud investigation based on

⁴ Before Agrawal was appointed CEO on November 29, 2021, he had served over four years as Twitter’s Chief Technology Officer. Agrawal’s hiring as CEO had been contentious, with some Board Directors opposed. Our client reasonably believes that Agrawal became defensive about many of the problems that our client identified, because Agrawal had caused them, or allowed them to fester, in his role as CTO.

Mudge's allegations. On January 18, **CEO Agrawal** lied about Mudge's efforts to rectify the previous month's fraud.⁵

- d. Agrawal terminated Mudge the next day, January 19.
4. **Astonishingly**, hours after Twitter terminated Mudge's employment, including immediately denying him access to corporate systems, Twitter's Chief Compliance Officer began emailing Mudge at his personal gmail account, seeking to obtain his latest disclosures of fraud. The Compliance Officer's reference to "your conversation this morning" was the video call in which Mudge had been terminated, and the "matters already under investigation" was Agrawal's instructions to knowingly present inaccurate materials to the Board:⁶

On Wed, Jan 19, 2022 at 11:59 PM Marianne Fogarty <[REDACTED]@twitter.com> wrote:
Mudge,

[REDACTED] [REDACTED] advised me that during your conversation this morning you mentioned the concerns you raised about information shared with the Risk Committee in December. If you were referring to matters that are not already under investigation, please let me know so we can schedule time to talk right away.

We appreciate that you raised your concerns and want to be sure they are fully and appropriately addressed. While my investigation is not complete, we intend to bring all of the concerns you raised with me and Omid, as well as the document you have been preparing in response to my email of January 11, to the Audit Committee and the full Risk Committee in the coming days. If there is anything else you want to include or recommend we do with respect to this issue so it concludes to your satisfaction, please let me know.

I'm available to hear any further concerns you may have as well as any additional thoughts you may have to resolve this matter.

Regards,

Marianne

5. **Apparently, Twitter's own compliance officers understood the gravity** of a situation in which the CEO had deliberately misled the Board. (Twitter's compliance team could face personal liability for letting fraud allegations go unaddressed.) Between Mudge's termination on January 19 and January 27, Twitter's Chief Compliance Officer emailed Mudge at least five times to obtain his corrected disclosures concerning information security.
6. Mudge ultimately **worked at least 150 hours—after he was terminated, without pay, and without access to his Twitter accounts or laptop**—to do his best to document the underlying facts about information security, and the fraud he had identified. Details of these events, including Mudge's emails with Twitter's Chief

⁵ When Mudge tried to correct the record, Board Member Omid Kordestani interrupted, and refused to let Mudge speak or provide facts.

⁶ See Exhibit 16; These post-termination communications, made while the parties were adverse and without any expectation of confidentiality, are not subject to Attorney-Client Privilege, as explained below.

Compliance Officer and his final report to the Board to articulate specific fraud he was identifying, are all included with this disclosure.⁷

7. **No privileged materials included:** Mudge has carefully limited disclosure of internal corporate documents to those relevant and “reasonably necessary” to demonstrate Twitter’s legal violations.⁸ In order to identify any materials subject to a claim of attorney-client privilege,⁹ with assistance of independent filter counsel, we conducted a review of every Exhibit to this disclosure. We determined that none of the exhibits included in this disclosure are protected by attorney-client privilege:
- a. Some Exhibits include the words “**Privileged and Confidential**” or a similar designation. These labels, which Twitter staff often applied indiscriminately and without legal guidance, do not determine whether a document is in fact subject to a valid claim of privilege. Therefore, after a careful review of the documents and the applicable law, we included some documents that contain such a label but nevertheless do not contain privileged communications.
 - b. Similarly, the **mere presence of a lawyer** in a communication does not mean the communication is covered by the attorney-client privilege. A lawyer may be operating in a non-legal capacity, or may have a dual role that encompasses legal as well as business or operational functions. Even when operating in a legal capacity, a lawyer’s communication may not be related to the request for, or provision of, legal advice. And emails between lawyers and their clients are not necessarily privileged if they are not made “in confidence.” After a review of the applicable law and the documents

⁷ See Exhibits 1 and 16

⁸ Cf. *Cafasso v. General Dynamics C4 Systems, Inc.*, 637 F.3d 1047, 1062 (9th Cir. 2011) (dicta suggesting that relators under the False Claims Act should limit disclosure of internal corporate documents to those documents “reasonably necessary” to pursue their whistleblower claim).

⁹ The attorney-client privilege “protects a confidential communication between attorney and client if that communication was made for the purpose of obtaining or providing legal advice to the client.” *In re Kellogg Brown & Root, Inc.*, 756 F.3d 754, 757 (D.C. Cir. 2014). See also *United States v. Mejia*, 655 F.3d 126, 132 (2d Cir.2011) (“The attorney-client privilege protects communications (1) between a client and his or her attorney (2) that are intended to be, and in fact were, kept confidential (3) for the purpose of obtaining or providing legal advice.”). See also *Restatement Third, The Law Governing Lawyers*, § 68, 2000.

-
- themselves, we included some email communications in which a lawyer is present on the chain.
- c. Further, “[t]he protection of the privilege extends only to communications, and **not to facts**. A fact is one thing and a communication concerning that fact is an entirely different thing.”¹⁰ For example, facts about a privileged communication, including its existence, are not themselves privileged.
 - d. No attorney-client privilege attaches to the set of **post-termination communications** between Mudge and Twitter counsel. The attorney-client privilege can, depending on the circumstances, cover some communications between in-house counsel and a former employee. After a review of the documents and the case law, we determined that the privilege does not apply to the post-termination communications included here. After Twitter abruptly terminated Mudge, their interests were not aligned, but rather adverse (and in fact, during the termination call, Mudge explicitly raised the possibility that the circumstances of his termination could create a legal risk for Twitter).¹¹ Under these circumstances, Twitter could not reasonably expect that its interaction with Mudge was privileged.¹²
 - e. Finally, we made redactions on a number of Exhibits to obscure some portions over which Twitter might claim privilege. (The fact that we redacted a portion of an Exhibit is not an admission that the redacted portion is in fact privileged.) Many times, we redacted material out of an abundance of caution even when we determined that the privilege would not properly apply.
- 8. The Work Product Doctrine does not affect** this disclosure. The Work Product Doctrine applies in the context of “rule[s] dealing with discovery” requests in civil litigation and other adversarial proceedings. Hickman v. Taylor, 329 U.S. 495, 509 (1947). The rule states that “[o]rdinarily, a party may not discover documents and

¹⁰ Upjohn Co. v. United States, 449 U.S. 383 at 395-96 (1981) (internal citations omitted).

¹¹ In fact, Mudge had been so concerned with Agrawal’s conduct since December 2020 that he had already retained lawyers to advise him on how to follow whistleblower laws and, if necessary, pursue claims of unlawful retaliation. (Please note that undersigned counsel did not become involved until later.)

¹² The majority of these post-termination communications were with Twitter’s Chief Compliance Officer. Although this individual has a law license, for the purposes of these communications, this person was acting in an operational, non-legal capacity. The Chief Compliance Officer’s non-legal role in these communications is another reason why the communications are not protected by the attorney-client privilege.

tangible things that are prepared in anticipation of litigation or for trial by or for another party or its representative (including the other party's attorney, consultant, ... or agent).” Fed R. Civ. P. 26(b)(3)(A). So, “[t]he purpose of the work–product rule is not to protect the evidence from disclosure to the outside world but rather to protect it only from the knowledge of opposing counsel and his client, thereby preventing its use against the lawyer gathering the materials.” *The Work–Product Rule—Matters Protected by the Work–Product Rule*, 8 Fed. Prac. & Proc. Civ. § 2024 (3d ed.).¹³

9. As an example, if our client were to sue Twitter, the Work Product Doctrine might allow Twitter to refuse to give certain documents (that it could show were actually prepared in anticipation of the hypothetical lawsuit) to our client during the discovery process. But there is no such lawsuit, and nobody, including our client, is making any production demands on Twitter. We are not aware of any authority suggesting that the Doctrine affects voluntary, protected disclosures under the Dodd-Frank Act. (Even if the Work Product Doctrine applied here, we have not identified any documents that Twitter, under the applicable case law, prepared in “anticipation of litigation.”)
10. **Ethical Disclosure Dilemma:** Mudge is proceeding with these disclosures quite reluctantly. Mudge comes out of, even helped to create, the modern information security community of responsible security disclosures. While criminal hackers break and steal, independent security researchers (also known as “ethical hackers”) use their skills to inform people about specific vulnerabilities, strengthen security and advance human rights and democracy. When ethical researchers find a vulnerability that bad actors can exploit, they first make a quiet “responsible disclosure” so that the affected company or government can fix it. But sometimes, the vulnerable institution doesn’t want to hear the truth, or fix the problem. In those cases, ethical researchers are forced to weigh the risks of wider disclosure: Exposing vulnerabilities tips off bad actors, but it also allows users of a service to

¹³ Thus, the work-product doctrine, unlike the attorney-client privilege, “does not exist to protect a confidential relationship, but rather to promote the adversary system by safeguarding the fruits of an attorney's trial preparations from the discovery attempts of the opponent.” *United States v Am. Tel. and Tel. Co.*, 642 F2d 1285, 1299 (D.C. Cir. 1980).

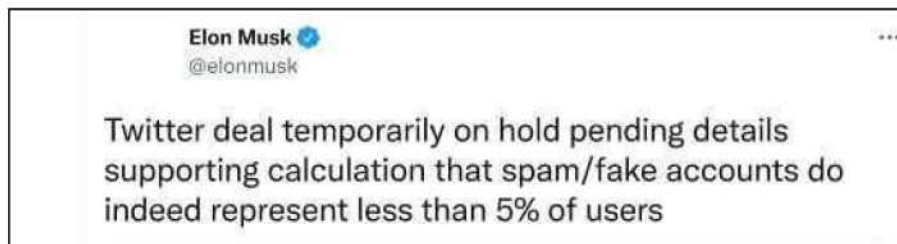
make more informed decisions, and can push the service to improve.¹⁴ Mudge made a personal commitment to Dorsey, the Twitter Board, the greater public, and to himself, that he would do his best to help fix Twitter. Mudge spent about 14 months pushing improvements from the inside, and was terminated for his efforts. With a heavy heart, Mudge has concluded that these lawful disclosures are his ethical obligation.

[Disclosure continues next page]

¹⁴ "In computer security, **coordinated vulnerability disclosure**, or "CVD" (formerly known as responsible disclosure) is a vulnerability disclosure model in which a vulnerability or an issue is disclosed to the public only after the responsible parties have been allowed sufficient time to patch or remedy the vulnerability or issue. This coordination distinguishes the CVD model from the "full disclosure" model." "Coordinated vulnerability disclosure - Wikipedia." https://en.wikipedia.org/wiki/Coordinated_vulnerability_disclosure.

II. Lying about Bots to Elon Musk

11. A recent example of misrepresentations by Twitter concerns Elon Musk's high-profile takeover attempt since April 2022.¹⁵ On May 13, Mr. Musk expressed doubts about the accuracy of Twitter's claim in legal filings that <5% of accounts are "bots," or automated spam accounts that spread propaganda and hurt the experience of real users:¹⁶



12. In response, on May 16, 2022, **CEO Agrawal tweeted false and misleading statements about Twitter's handling of bots** on the platform, starting with this:¹⁷



13. Agrawal's tweet was a lie. In fact, Agrawal knows very well that Twitter executives are **not incentivized to accurately "detect"** or report total spam bots on the platform. Here's why:

¹⁵ Please note that Mudge began preparing these disclosures in early March 2022, well before Mr. Musk expressed any interest in acquiring Twitter, and has not communicated these disclosures to anyone with a financial interest in Twitter. As a senior executive, Mudge was awarded Twitter stock, for which he previously created (and has followed) an Automatic Securities Disposition Plan pursuant to SEC rules codified at 17 C.F.R. § 240.10b5-1(c).

¹⁶ Elon Musk's Personal Twitter Page: https://twitter.com/elonmusk/with_replies?lang=en

¹⁷ Parag Agrawal's Personal Twitter Page: <https://twitter.com/paraga>

-
14. Until 2019, Twitter reported total monthly users, but stopped because the number was subject to negative swings for a variety of reasons, including situations such as the removal of large numbers of inappropriate accounts and botnets.¹⁸ Instead, Twitter announced a new, proprietary, opaque metric they called “**mDAU**” or “**Monetizable Daily Active Users**,” defined as valid user accounts that *might* click through ads and actually buy a product.¹⁹ From Twitter’s perspective, “mDAU” was an improvement because it could internally define the mDAU formula, and thereby report numbers that would reassure shareholders and advertisers. Executives’ bonuses (which can exceed \$10 million) are tied to growing mDAU.
15. Executives are incentivized to avoid counting spam bots as **mDAU**, because mDAU is reported to advertisers, and advertisers use it to calculate the effectiveness of ads. If mDAU includes spam bots that do not click through ads to buy products, then advertisers conclude the ads are less effective, and might shift their ad spending away from Twitter to other platforms with higher perceived effectiveness.
16. However there are many **millions of active accounts** that are *not* considered “mDAU,” either because they are spam bots, or because Twitter does not believe it can monetize them. These millions of non-mDAU accounts are part of the median user’s experience on the platform. And for this vast set of non-mDAU active accounts, Musk is correct: Twitter executives have little or no personal incentive to accurately “detect” or measure the prevalence of spam bots.
17. In fact, Mudge learned deliberate ignorance was the norm amongst the executive leadership team. In early 2021, as a new executive, Mudge asked the Head of Site Integrity (responsible for addressing platform manipulation including spam and botnets), what the underlying spam bot numbers were. Their response was “**we**

¹⁸ “Twitter...said it would stop reporting monthly active users (MAUs) after Q1 2019 as it would switch to a new metric called monetizable daily active users (mDAUs)...”

https://www.business-standard.com/article/news-ians/twitter-says-will-stop-reporting-monthly-active-users-119020701161_1.html But even after the switch, Twitter overcounted mDAU users, see <https://techcrunch.com/2022/04/28/twitter-says-it-overcounted-its-users-over-the-past-3-years-by-as-much-as-1-9m/>.

¹⁹ “We define mDAU as people, organizations, or other accounts who logged in or were otherwise authenticated and accessed Twitter on any given day through twitter.com or Twitter applications that are able to show ads.” See <https://www.sec.gov/Archives/edgar/data/0001418091/000141809120000037/twtr-20191231.htm>. Twitter has stated that mDAU is “not comparable to current disclosures from other companies.” *Digging Into Twitter’s First Daily User Disclosure*, 7 Feb. 2019, <https://www.fool.com/investing/2019/02/07/digging-into-twitters-first-daily-user-disclosure.aspx>.

don't really know.” The company could not even provide an accurate upper bound on the total number of spam bots on the platform. The site integrity team gave three reasons for this failure: (1) they did not know how to measure; (2) they were buried under constant firefighting and could not keep up with reacting to bots and other platform abuse; and, most troubling, (3) **senior management had no appetite to properly measure the prevalence of bot accounts**—because as Mudge later learned from a different sensitive source, they were concerned that if accurate measurements ever became public, it would harm the image and valuation of the company.

18. Even the Board of Directors understood the counterproductive incentives in place: In or about the Q3 2021 Board Risk Committee meeting, a Director asked why more progress has not been made around bots and related harmful content on the platform. Our client remembers an executive of the company **admitting to Board members that the company had “intentionally and knowingly deprioritized” platform health** to focus on growing mDAU. Afterwards, a different Twitter leader who had witnessed the exchange commented to Mudge, in reference to this admission, “it is very strange what this company does not share with board members, and then some of the statements that they do.”

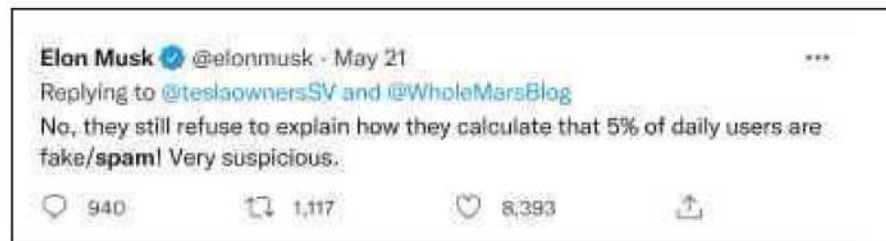
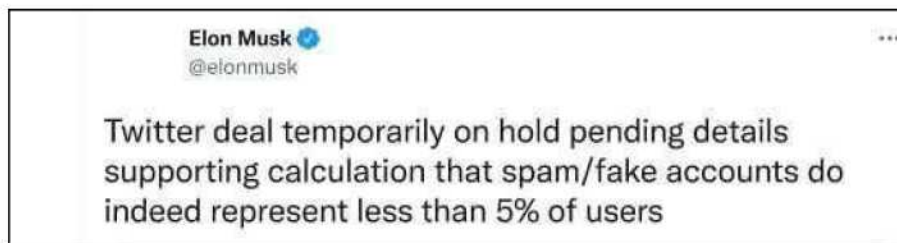
19. **Repeated Efforts to Disable ROPO:** “ROPO,” which stands for “Read-Only Phone Only,” is probably Twitter’s most volumetrically-effective mechanism for identifying and blocking spam bots. If a script identifies an account as possibly spam and triggers ROPO, the account is placed into a “Read Only” mode and is unable to post content to the platform. Twitter sends a text message to the associated phone number, with a one-time code that the recipient needs to manually enter to regain account access. Shortly into Mudge’s time at Twitter, **a senior executive (with primary responsibility for growing mDAU) proposed disabling ROPO worldwide**, based on an anecdote of a small number of unsolicited DMs (text messages) he had personally received in which users claimed they were incorrectly denied access by ROPO.²⁰ The Lead of Site Integrity told Mudge that executives responsible for growing mDAU had proposed disabling ROPO several times before. The Site Integrity Lead pleaded with Mudge, as a

²⁰ Executives at the company receive a near continuous stream of messages directed to them, complaining about the service and other requests like demanding malicious accounts be reinstated. Some percentage of the time the complaints were valid, but more often not.

senior executive, to prevent the other executives from disabling ROPO. Research later performed at Mudge’s direction showed ROPO was effectively blocking more than 10-12 million bots each month with a surprisingly low rate (<1%) of false positives.²¹

20. Therefore Musk’s suspicions are on target: senior executives earn bonuses not for cutting spam, but for growing mDAU. In fact, Twitter **created the mDAU metric precisely to avoid** having to honestly answer the very questions Mr. Musk raised.

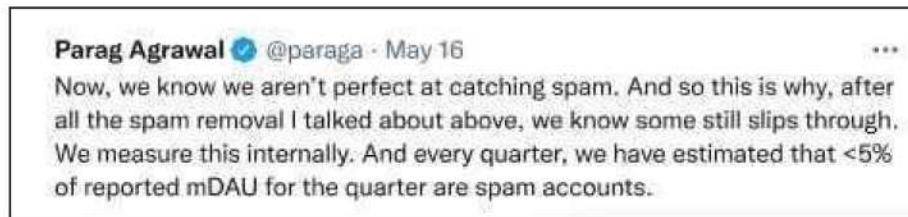
21. The rest of Agrawal’s May 16 tweets aren’t out-and-out lies but they rely on wordplay to distract and mislead Mr. Musk, and everyone else. Musk appears to be asking a valid and intuitive question, *what percent of accounts encountered by the median user are actually bots?*



²¹ Mudge does not recall whether the 10-12 million number was per month or per week as Twitter revoked his access to the notes and data on this topic when he was terminated. Here we provide the benefit of the doubt and present the smaller of the two numbers.



22. While pretending he is answering Musk’s question, in fact Agrawal is answering a very different one, namely, *Are there fewer than 5% bots in the set of mDAU accounts, as defined in secret by Twitter?* Agrawal’s reasoning might appear a bit circular since, by definition, mDAU is more or less Twitter’s best approximation of the set of accounts that aren’t bots. And Agrawal is not exactly trying to help readers understand the bait-and-switch nature of his answer:





23. Unless you're a Twitter engineer responsible for calculating mDAU, you probably wouldn't know what Agrawal is talking about. He is *not* saying that fewer than 5% of all accounts on the platform are spam. He's saying, more or less, that Twitter starts with all the accounts on the platform, tries to automatically put all the human accounts that could be convinced by advertisers to buy products (but no spam accounts) into mDAU, and then uses humans to estimate the error rate of spam accounts that nevertheless slip through into mDAU. And naturally, Twitter "can't share" its special sauce for determining mDAU.

24. In mathematical terms, Mr. Musk is asking whether the following proposition holds:

$$\frac{\text{spam bot accounts}}{\text{Total active}^{22} \text{ accounts}} < 5\%$$

25. To which Agrawal responds by affirming a rather different proposition:

$$\frac{\text{\{human estimate of spam bots that slip through into (mDAU, i.e. Twitter's secret automated estimate of Total active accounts minus spam bot and other worthless accounts)\}}}{\text{(mDAU, i.e. Twitter's secret automated estimate of Total active accounts minus spam bot and other worthless accounts)}} < 5\%$$

²² The qualifier "active" is meant to exclude accounts of users who have died or no longer use the service, etc..

26. A more meaningful and honest answer to Mr. Musk’s question would be trivial for Twitter to calculate, given that Twitter is already doing a decent job excluding spam bots and other worthless accounts from its calculation of mDAU. But this number is likely to be meaningfully higher than 5%:²³

$$\frac{\text{Total active accounts minus mDAU accounts}}{\text{Total active accounts}} \cong \% \text{ Spam and other worthless accounts}$$

27. Agrawal goes on to provide raw numbers of takedowns - again without context:



28. Is half a million a day a lot or a little, for a platform as vast as Twitter? No one knows, because there is no denominator provided for context. Is Twitter *missing* a hundred thousand new spam bots every day? No one knows. Mudge attended every Board of Directors and relevant Board subcommittee meeting in 2021, where he saw this strategy regularly deployed: executives reported items, such as bot takedowns and other metrics, as **raw numbers, without context**—never in a more useful format (e.g. percentages with well-defined numerators and denominators) that would permit Board members to understand the overall prevalence of fake accounts.

29. More broadly, Agrawal’s tweets and Twitter’s previous blog posts misleadingly imply that Twitter employs proactive, sophisticated systems to measure and block spam bots. Mudge discovered the reality: mostly outdated, unmonitored, simple

²³ To be fair, this formula doesn’t precisely measure spam bots. Rather it lumps bots in with human accounts that Twitter, for whatever reason, believes it can’t monetize, perhaps because, e.g., they are not selling ads for their region, or the company has no capacity in that user’s language. Relatedly, our client notes that lack of language capacity is a significant shortcoming at Twitter. Because Twitter lacks language capacity the platform is disturbingly deficient in regards to integrity efforts in dozens of countries worldwide, permitting well-established harms like disinformation and adverse electoral effects to fester largely unaddressed.

scripts plus overworked, inefficient, understaffed, and reactive human teams. The scripts were largely un-owned by any person or team, and their results were not tracked. Furthermore no effort was made to compare costs to benefits of the scripts, nor approaches, nor their veracity.²⁴

30. **#Protect Initiative:** Mudge was so concerned about this situation, and Twitter’s overall cybersecurity state, that during the 2021 calendar year, he developed and presented to the Board of Directors a sweeping, 3-year Board-supervised objective called “#Protect Initiative.”²⁵ Elements of the initiative would have assigned responsibility for properly measuring spam bot prevalence. The entire senior leadership team and Board of Directors received and approved Mudge’s #Protect Initiative plan. If Twitter was already accurately measuring and estimating spam bot prevalence on the platform, this issue would not have reached the Board and been a specific part of Mudge’s 2022 plans. This excerpt from the 3-year plan shows that Mudge intended to lead Twitter Services (a corporate division abbreviated as “TwS”) to inventory, obtain measurements, and improve anti-spam efforts:²⁶

4. TwS - inventory of bots, measurements, continuity of action across bot bounces, improve automation and accuracy by >25%

²⁴ In a September 2021 Twitter blog post, Twitter stated “it’s not the number of bots, (around 5%, a number Twitter reports quarterly) but the impact they have on the conversation.” Five percent of what? This was apparently an attempt to distract and mislead users about the bot problem. The post goes on to state: “First Truth: Don’t assume an account with a peculiar name must be a bot.” <https://blog.twitter.com/common-thread/en/topics/stories/2021/four-truths-about-bots> . Twitter is arguing that account names consisting of random, auto-generated sets of letters and numbers aren’t always bots. But this is a straw man. The blog post does not cite any data on what percentage of random-character handles were in fact bots, because the Site Integrity team did not have a dedicated data scientist, and this data, though available within the company, was poorly maintained and largely un-measured.

²⁵ Mudge lost access to the detailed #Protect Initiative documentation when he was terminated, but investigators should be able to acquire it easily. *But see* exhibit 8, Mudge’s #Protect board presentation.

²⁶ Confusingly, because one of Twitter’s main automation tools was called “Botmaker”, Twitter staff also used the term “bots” to describe the company’s automated scripts to identify spam bots. Whether a document’s reference to “bots” means a spam account or a Twitter script depends on the context.

31. All this is conveyed in a damning independent report²⁷ on platform integrity, produced in or about May or June of 2021, from Alethea Group (a disinformation consultancy retained for this purpose by our client in his role as Head of Security).²⁸ Here are a few of its findings (“SI”= Site Integrity, whose mandate included hunting spam bots):

Tools available to Site Integrity to work on these issues are often outdated, “hacked together,” or difficult to use, limiting Twitter’s ability to effectively enforce policies at scale. A lack of automation and sophisticated tooling means that Twitter relies on human capabilities, which are not adequately staffed or resourced, to address the misinformation and disinformation problem.

3.1.1.2 -- There are components of Twitter that are part of the disinformation and misinformation detection or response that are outside of Site Integrity / Security, and Site Integrity / Security have no access or authority to use these tools absent the good will of other teams.

3.1.1.3 -- Twitter does not have aligned incentives across the organization, and, as a result, priorities with regards to Product Safety.

3.1.1.4 -- SI relies on functions that have no accountability to SI in order to piece together solutions.

3.2.3 -- SI does not have dedicated engineering support for their tools, so even minor upgrades or changes to existing tools can take months or years to complete.

3.2.4 -- SI lacks sufficient dedicated data science support and staff with technical skills.

²⁷ Independent “filter counsel” advised us that the Alethea Group report is not subject to Attorney-Client Privilege: (1) The report does not contain or discuss legal advice or exposure, nor does it discuss legal or regulatory options or contain legal citations. (2) It was written by non-lawyers at Alethea Group for the non-lawyer executive tasked with security, privacy, and content moderation at scale, Mudge. Cf. Guo Wengui v. Clark Hill, PLC, 338 F.R.D. 7 (D.D.C. 2021) (a cybersecurity report created by a non-legal consulting firm is not privileged, even when consultant was hired by outside counsel, because the claimant’s goal was “gleaning [consultant’s] expertise in cybersecurity, not in obtaining legal advice from [its] lawyer,” *id.* at 13, internal citations omitted).

²⁸ See Exhibit 2

3.3.3 -- There are existing internal tools in other parts of Twitter that would be useful for the misinformation and disinformation use case, but SI analysts do not have access to them. Analysts also lack access to externally available tools or datastreams that would allow them to do more proactive cross-platform analysis.

3.4.1 -- SI does not have a knowledge management system to track and store findings and data. As a result, SI does not have the ability to monitor threat actors or identify changes in their tactics, techniques, and procedures (TTPs) over time, or to measure the impact of SI's work.

3.7.3 -- Policies to address misinformation/disinformation often do not address repeat offenders and are applied on a case-by-case basis, leading to a lack of scalability.

3.8.2 -- The process for labelling disinformation and misinformation content is largely manual, requires the use of multiple tools, and usually needs to be done on a case-by-case basis.

32. Unfortunately, as detailed in the rest of this disclosure, Agrawal's misrepresentations about spam bots are just the tip of the iceberg.

[Disclosure continues next page]

III. 2011 FTC Consent Order and 2020's "Largest Social Media Hack in History": Dorsey Recruits Mudge

33. Since Twitter's 2006 launch, the platform has earned a **reputation for problems**²⁹ with security, privacy and integrity (a broad term that includes disinformation, spam bots, election interference and other content-related abuses).
34. **2011 FTC Complaint:** In 2011, the FTC had filed a complaint against Twitter for its failure to properly protect nonpublic consumer information, which included users' email addresses, Internet Protocol ("IP") addresses, telephone numbers, and nonpublic information exchanged on the platform.³⁰ The complaint alleged that, from 2006 to 2009, far too many Twitter employees exercised administrative ("God mode") control within Twitter's internal systems and user data, thereby allowing any attacker with access to an employee account to easily compromise Twitter systems. And Twitter's systems were, and are, full of highly sensitive personal user data that enable a hostile government to find precise geo-location(s) for a specific user or group, and target them for arrest or violence.
35. **Consent Order:** As a result of the complaint, the FTC and Twitter entered into a consent decree in March 2011, which has the force of law for future violations.³¹ The FTC ordered Twitter to: "establish and implement, and thereafter maintain a comprehensive information security program that is reasonably designed to protect the security, privacy, confidentiality, and integrity of nonpublic consumer information." Components of this comprehensive information security program included identifying security risks and preventing, detecting, and effectively responding to cyberattacks.³² The order imposed various reporting requirements

²⁹ Eric Geller, *Twitter's security holes are now the nation's problem*,

<https://www.politico.com/news/2020/07/16/twitter-security-hack-congress-366771>

³⁰ U.S. Fed. Trade Comm'n, *In the Matter of Twitter, Inc.*, (No. C-4316), Compl. (Mar. 2, 2011), available at: <https://www.ftc.gov/sites/default/files/documents/cases/2011/03/110311twittercmpt.pdf>.

³¹ U.S. Fed. Trade Comm'n, *In the Matter of Twitter, Inc.*, (No. C-4316), Decision & Order (Mar. 2, 2011), available at: <https://www.ftc.gov/sites/default/files/documents/cases/2011/03/110311twitterdo.pdf>.

³² Per the FTC order, Twitter was required to include various components of this program, including, among other things:

- 1) designating employee(s) to be accountable for the information security program;
- 2) identifying reasonably foreseeable security risks that could expose or compromise nonpublic consumer information, taking into account various considerations, such as employee training and management, information systems, and prevention, detection and responses to various system failures, such as attacks and account takeovers;

upon Twitter to ensure it was keeping the FTC informed of its progress on its security system for ten years following issuance of the order, which had a final termination date of March 2, 2031.

36. **Hacked by Teenagers:** In July 2020—following nine years of supposed fixes, investments, compliance policies, and reports to the FTC by Twitter—the company was hacked by a 17-year old, then-recent high school graduate from Florida and his friends. The hackers managed to take over the accounts of former **President Barack Obama, then-Presidential candidate Joseph Biden, and high-profile business leaders including, but not limited to, Jeff Bezos, Bill Gates, and Elon Musk.** As part of the account takeovers, the hackers urged their tens of millions of followers to send Bitcoin cryptocurrency to an account they created.³³

37. The 2020 hack was then the **largest hack of a social media platform in history,**³⁴ and triggered a global security incident.³⁵ Moreover, the hack did not involve malware, zero-day exploits, supercomputers brute-forcing their way past encryption, or any other sophisticated approach. In fact, it was pretty simple:³⁶ Pretending to be Twitter IT support, the teenage hackers simply called some Twitter employees and asked them for their passwords. A few employees were duped and complied and—given systemic flaws in Twitter’s access controls—those credentials were enough to achieve “God Mode,” where the teenagers could imposter-tweet from any account they wanted. Twitter’s solution was to impose a **system-wide shutdown of system access to all of its employees, lasting days.** For about a

-
- 3) designing and implementing reasonable safeguards to control the risks identified through the risk assessment, and carrying out regular testing and monitoring of these safeguards; and,
 - 4) evaluating and adjusting its information security program and circumstances that may have a material impact on the effectiveness of its information security program.

Id.

³³ Kif Leswing, *Hackers targeted Twitter employees to hijack accounts of Elon Musk, Joe Biden and others in digital currency scam*, CNBC, July 15, 2020, <https://www.cnbc.com/2020/07/31/twitter-bitcoin-scam-masterminded-by-17-year-old.html>.

³⁴ "Major US Twitter accounts hacked in Bitcoin scam - BBC News." July 16, 2020, <https://www.bbc.com/news/technology-53425822>.

³⁵ Brain Fung, *Twitter's massive hack could be even worse than it seems*, CNN, July 17, 2020, <https://www.cnn.com/2020/07/16/tech/twitter-hack-security-analysis/index.html>.

³⁶ See New York State Dep't of Financial Servs., *Twitter Investigation Report* (Oct. 14, 2020), https://www.dfs.ny.gov/Twitter_Report ("In the hands of a dangerous adversary, the same access obtained by the Hackers—the ability to take control of any Twitter users' account—could cause even greater harm.").

month, hiring was paused and the company essentially shut down many basic operations to diagnose the symptoms, not the causes, of the hack.

38. Security experts agreed this extreme response demonstrated that Twitter did not have proper systems in place to understand what had happened, let alone remediate and reconstitute to a safe state.³⁷ These failures in Twitter's security raised alarms about more serious breaches that could occur in the future, especially because 2020 was a presidential election year. Bad actors with more sophisticated tools than what was used by a recent high school graduate could easily take advantage of Twitter's poor security, creating detrimental consequences for the country. As aptly framed in the *Wired* article about this incident,³⁸

**But if a teenager with access to an admin panel can
bring the company to its knees, just imagine what Vladimir Putin could do.**

39. Soon, Twitter's situation with the FTC got even worse. On July 28, 2020, the FTC filed a draft complaint alleging Twitter engaged in violations of the 2011 order.³⁹ Specifically the draft FTC complaint charged that from 2013 to 2019, Twitter misused users' phone number and/or email address data for targeted advertising when users had provided this information for safety and security purposes only. This implied Twitter still lacked basic understandings about how, what, and where its data lived, and how to responsibly protect and handle it. On May 25, 2022, the FTC announced a \$150 million fine against Twitter.⁴⁰

40. At the time of the hack and the new FTC draft complaint, Twitter had neither an executive versed in information security and privacy engineering (the executive-level Security Lead role Mudge would fill in November 2020), nor even a Chief Information Security Officer.⁴¹ As a result, **Parag Agrawal, then Twitter's Chief**

³⁷ See *id.* (former Facebook CISO explained his surprise that a phishing scheme led to a total shutdown at Twitter and a former Twitter employee stated that the company did not have the right systems in place to address such an attack, which led to this extreme response).

³⁸ "Inside the Twitter Hack—and What Happened Next | WIRED." 24 Sep. 2020, <https://www.wired.com/story/inside-twitter-hack-election-plan/>.

³⁹ https://www.ftc.gov/system/files/ftc_gov/pdf/2023062TwitterComplaint.pdf

⁴⁰ "Twitter Fined in Privacy Settlement, as Musk Commits More Equity" 25 May. 2022, <https://www.nytimes.com/2022/05/25/technology/twitter-fined-ftc-doj-privacy.html>.

⁴¹ Twitter's failure to employ a Chief Information Security Officer constitutes an independent violation of the 2011 Consent Order, which required "the designation of an employee or employees to coordinate and

Technology Officer (CTO), was the ultimate decision-maker for correcting the security vulnerabilities exposed by the hack.⁴² Mr. Agrawal made statements that acknowledged the problem that the FTC had precisely identified nine years earlier: too many Twitter staff and contractor accounts had too much access to too much user data. Here are quotes from Mr. Agrawal in *Wired* magazine⁴³ meant to assure the public—as we explain later, these were false and misleading (and therefore constitute legal violations):

But one of the first things Twitter realized in the immediate aftermath was that too many people had too much access to too many things. “It’s more about how much trust you’re putting in each individual, and in how many people do you have broad-based trust,” Agrawal says. “The amount of access, the amount of trust granted to individuals with access to these tools, is substantially lower today.”

41. Twitter’s CEO at the time, Jack Dorsey, realized the company had serious problems. To demonstrate he was serious about fixing things, Dorsey began recruiting Mudge. Mudge had other very desirable, well-paid, high-profile job opportunities. But Mudge also understood that Twitter is a critical global public resource that can build bridges between different communities and parts of the world, and serve the public conversation. Mudge recognized that Twitter’s platform could also cause real harm, and understood that it would take a lot of work to get Twitter on track. He was up for the challenge. At the request of Dorsey, and with the promise of Dorsey’s support, Mudge accepted the offer and expected to spend the rest of his career at Twitter. Mudge never expected or wanted to become a whistleblower. He was convinced that the executives and board were ready to deal with long overdue security and privacy challenges.

be accountable for the information security program.” See

<https://www.ftc.gov/sites/default/files/documents/cases/2011/03/110311twitterdo.pdf>

⁴² See Nicholas Thompson & Brian Barrett, *How Twitter Survives Its Biggest Hack—and Plans to Stop the Next One*, *WIRED* (Sept. 24, 2020), <https://www.wired.com/story/inside-twitter-hack-election-plan/> (detailing Mr. Agrawal’s role in responding to the hack).

⁴³ *Id.*; see also Parag Agrawal & Damien Kieran, Twitter Blog, *Our continued work to keep Twitter secure* (Sept. 24, 2020),

https://blog.twitter.com/en_us/topics/company/2020/our-continued-work-to-keep-twitter-secure (“We have strict principles around who is allowed access to which tools and at what time, and require specific justifications for customer data to be accessed.”).

-
42. On **November 16, 2020, four months after the summer’s publicly humiliating incident, Mudge began his new job as Security / Integrity Lead** with Twitter. **Mudge’s hire was applauded** by cybersecurity industry executives and experts, and allowed Twitter to claim credit for challenging cybersecurity problems aggressively. Kevin O’Brien, co-founder and CEO of GreatHorn, a cybersecurity firm, stated, “My take is that [Mudge] remains one of the best security minds on the planet today – Dorsey bringing him [in] speaks well for their focus on security.”⁴⁴ Dan Kaufman, Mudge’s former supervisor at DARPA, stated Mudge would “be at the top of [his] list” of names of people who could fix Twitter’s current abysmal security state.⁴⁵ Similarly, Alex Stamos, former Facebook Chief Security Officer, stated that Mudge was a great fit for the Company because of his ability to find creative solutions to security problems.⁴⁶
43. **Portfolio:** CEO Jack Dorsey assigned Mudge a vast portfolio, responsible for some of the hardest problems, with hundreds of staff and thousands of contractors in chains that reported up to him:
- a. **Information Security** - the integrity and security of all Twitter systems and data;
 - b. **Privacy** - creating the privacy policies and processes, plus engineering and executing them across all Twitter systems and data, to avoid liability with the FTC and build systems and processes that respect people’s data;
 - c. **Corporate Security** - responsible for the physical security and safety of employees, offices, and data centers;
 - d. **Information Technology** - running the internal systems for finance, HR, and internal corporate technologies and communications;

⁴⁴ David Jones, *Famed hack Mudge to lead Twitter security after summer of attacks*, Cybersecurity Dive (Nov. 17, 2020), <https://www.cybersecuritydive.com/news/twitter-mudge-security/589177/>; see also *id.* (Doug Britton, CTO of RunSafe Security, commented that Mudge was “a good hire for Twitter, because the security issues are macro and micro.”).

⁴⁵ Joseph Menn, *Twitter names famed hacker “Mudge” as head of security*, CNBC (Nov. 16, 2020), <https://www.cnbc.com/2020/11/16/twitter-names-famed-hacker-mudge-as-head-of-security.html>.

⁴⁶ *Id.*

e. **“Twitter Service”** - the company’s internal name for the division tasked with operational enforcement of global content moderation at scale, including processing and the removal of various spam and spam bots.

44. After arriving, Mudge spent two months performing an in-depth evaluation to understand how things worked, or didn’t work, at Twitter. Mudge conducted in-depth interviews of about 40 employees – from members of the executive team to engineers to salespeople – to gain a better understanding of the Company’s current status with regards to security, perceived security needs, and what employees understood about Twitter’s security.⁴⁷ He attended engineering meetings, reviewed internal technical documents, and directly evaluated some of Twitter’s key computer systems and servers. Even though the company had been under the FTC consent decree since 2011, requiring by law that Twitter address fundamental security and privacy issues, Mudge remembers early in his tenure hearing Mr. Agrawal stating to the executive team that “Twitter has 10 years of unpaid security bills.”

[Disclosure continues next page]

⁴⁷ The results of Mudge’s inquiries were reflected comprehensively in a Google Sheet that he lost access to when he was terminated. Mudge can assist investigators in identifying and finding this important document.

IV. Mudge Discovers Egregious Deficiencies, Negligence, Willful Ignorance, and Threats to National Security & Democracy

45. **Mudge’s findings were dire.** Nearly a decade after the FTC Consent Order, with total users growing to almost 400 million and daily users totaling 206 million,⁴⁸ Twitter had made little meaningful progress on basic security, integrity, and privacy systems. Years of regulatory filings in multiple countries were misleading, at best. In many ways, the situation was even worse than Dorsey feared, as the company haphazardly expanded into contentious international areas without even following existing (albeit deficient) corporate policies.
46. Mudge’s reports, all highly-experienced experts and intimately familiar with Twitter’s problems with the FTC, told Mudge unequivocally that **Twitter had never been in compliance** with the 2011 FTC Consent Order, and was **not on track to ever achieve full compliance**. Twitter’s deficiencies are described in greater detail later, and in the exhibits.⁴⁹ But at a high level, Mudge found **serious deficiencies** in:
- a. **Privacy**, including
 - i. **Ignorance and misuse of vast internal data sets**, with only about 20% of Twitter’s huge data sets registered and managed,⁵⁰
 - ii. **Mishandling Personally Identifiable Information (PII)**, including repeated marketing campaigns improperly based on user email addresses and phone numbers designated for security purposes only;⁵¹

⁴⁸ Brian Dean, *How Many People Use Twitter in 2022? [New Twitter Stats]*, Backlinko, Jan. 5, 2022, <https://backlinko.com/twitter-users>.

⁴⁹ See Exhibits 1, 2, 3, 4, 20, 23, and 26

⁵⁰ See Exhibits 1 and 4

⁵¹ This was the problem identified in the FTC’s July 2020 Draft Complaint. In mid-2021, in the midst of negotiations with the FTC, Twitter *did it again*: the product sales team saw a data set, and (in the absence of any data tracking) just started using it for ad targeting. When one Twitter executive learned that, even after the 2011 Consent Decree and 2020 Draft Complaint, this was happening again, he said: “So we only started to address the problem, and then got side tracked and forgot about it? We do that for everything.” (This may have been Twitter SIM 144.) Around the same time, the CFO complained to Mudge that his request to send a large collection of user emails to an advertiser was being blocked by a few engineers. Mudge explained that the engineers were right to be blocking it, because Twitter did not have any understanding of data-lineage and there was no indication whether Twitter sending this data to a customer would be violating the FTC consent decree. In a further irony, better data lineage and enforced handling would not only have made the company compliant, but would have enabled the company to better monetize data, a double win.

iii. **Misusing security cookies** for functionality and marketing,⁵²

iv. **Misrepresentations to the FTC** on these matters;⁵³

b. **Information Security (InfoSec)**, including

i. **Server vulnerabilities**, with over 50% of Twitter's 500,000 data center servers with non-compliant kernels or operating systems, and many unable to support encryption at rest,

ii. **Employee computers exposed**, with over 30% of devices reporting they had disabled software and security updates,⁵⁴

⁵² In December 2021, the French CNIL (*Commission Nationale de l'Informatique et des Libertés*) demanded Twitter comply with their regulations. Up until Q2/Q3 2021 Twitter did not have sufficient understanding of how, and what, cookies were used for. Cookies were used for multiple functions, such as ad tracking and session security. It was apparent Twitter was in violation of international data requirements across many regions of the world. The new Twitter Privacy Engineer team had worked tirelessly with product to disentangle cookies and permit some form of user choice and control in regards to cookies and tracking performed by Twitter. On December 31, 2021, the fix was rolled out exclusively to France and then, because Twitter lacked separate testing environments, encountered a problem and it was almost immediately rolled back and disabled. The bug was fixed in a matter of hours, but product and legal blocked rolling out the fix for another month, until January 31, 2021, in order to extract maximum profit from French users before rolling out the fix. Mudge challenged executives to claim this as anything other than an effort to prioritize incremental profits over user privacy and legal data privacy requirements. The senior leaders in that meeting confessed that Mudge was correct. Twitter even launched a proactive court case attempting to claim that all cookies were by definition critical and required, because the platform is powered by advertisements. During internal conversations, Mudge heard Twitter product staff admitted that their argument was false and made in bad faith.

⁵³ In years past, the FTC had asked Twitter whether the data of users who canceled their accounts was properly "deleted." Twitter had determined that not only had the data not been properly deleted, but that data couldn't even be accounted for. Instead of answering the question that was asked, Twitter assured the FTC that the accounts were "deactivated," hoping FTC officials wouldn't notice the difference. Mudge learned about this historical practice in 2021, and was told that fines could be \$3 million each month plus 2% of revenue

⁵⁴ Twitter did not actively monitor what employees were doing on their computers. Although against policy, it was commonplace for people to install whatever software they wanted on their work systems. Twitter employees were repeatedly found to be intentionally installing spyware on their work computers at the request of external organizations. Twitter learned of this several times only by accident, or because of employee self-reporting. In other words, in addition to a large portion of the employee computers having software updates disabled, system firewalls turned off, and remote desktop enabled for non-approved purposes, it was repeatedly demonstrated that until Twitter leadership would stumble across end-point (employee computer) problems, external people or organizations had more awareness of activity on some Twitter employee computers than Twitter itself had.

-
- iii. **No Mobile Device Management (MDM)** for employee phones, leaving the company with no visibility or control over thousands of devices used to access core company systems;⁵⁵
 - iv. **Insider Threats** were virtually unmonitored, and when found the company did not take corrective actions;⁵⁶

c. **Fundamental architecture** including

- i. **lack of development and testing environments** for all software development and testing (highly anomalous for a large tech company),⁵⁷ where engineers use live production data and test directly on the commercial service, leading to regular service disruptions,
- ii. **serious access control problems**, with far too many staff (about half of Twitter's 10,000 employees, and growing) given access to sensitive live production systems and user data in order to do their jobs, the subject of

⁵⁵ It was well known at the time that governments were targeting the cell phones of activists, journalists, and executives, yet Twitter lacked basic abilities to identify or defend against this. See "Pegasus: Spyware sold to governments 'targets activists'" 19 July, 2021. <https://www.bbc.com/news/technology-57881364>.

⁵⁶ In 2019 two Twitter employees were accused of being Saudi government agents. Ellen Nakashima & Greg Bensinger, *Former Twitter employees charged with spying for Saudi Arabia*, Wash. Post, Nov. 6, 2019, https://www.washingtonpost.com/national-security/former-twitter-employees-charged-with-spying-for-saudi-arabia-by-digging-into-the-accounts-of-kingdom-critics/2019/11/06/2e9593da-00a0-11ea-8bab-0fc209e065a8_story.html

⁵⁷ A fundamental engineering and security principle is that access to live production environments should be limited as much as possible. Engineers should mostly work in separate development, test, and/or staging environments, using test data (not live customer data). Over a decade prior, companies like Google moved development to segregated test systems. But at Twitter, engineers built, tested, and developed new software directly in production with access to live customer data and other sensitive information in Twitter's system. This ongoing arrangement, almost unheard of at modern tech companies, causes repeated problems for Twitter in bad software deployments and significantly reduces the work an attacker needs to do to acquire credentials with extremely sensitive access. Twitter's practice was a huge red flag for job candidates, who universally expressed disbelief. One particular candidate for Vice President of Information Technology considered withdrawing his application on the (accurate) rationale that Twitter's lack of basic engineering hygiene in their arrangement presaged major headaches.

specific misrepresentations in 2020⁵⁸ by then-Chief Technology Officer Parag Agrawal;

- iii. **Insufficient data center redundancy**,⁵⁹ without a plan to cold-boot or recover from even minor overlapping data center failure, raising the risk of a brief outage to that of a catastrophic and existential risk for Twitter's survival.

47. Unsurprisingly, given these and other deficiencies, Twitter suffered from an **anomalously high rate of security incidents**⁶⁰—approximately one security incident each week serious enough that Twitter was required to report it to government agencies like the FTC and SEC, or foreign agencies like Ireland's Data Protection Commission.⁶¹ In 2020 alone, Twitter had more than 40 security incidents, 70% of which were access control-related. These included 20 incidents defined as breaches; all but two of which were access control related.⁶² Mudge identified there were several exposures and vulnerabilities at the scale of the 2020 incident waiting to be discovered,⁶³ and reasonably feared Twitter could suffer an Equifax-level hack.⁶⁴

⁵⁸ In particular, Agrawal had misrepresented the truth when he told Wired magazine that “[t]he amount of access, the amount of trust granted to individuals with access to these tools, is substantially lower today.” Nicholas Thompson & Brian Barrett, *How Twitter Survives Its Biggest Hack—and Plans to Stop the Next One*, Wired (Sept. 24, 2020), <https://www.wired.com/story/inside-twitter-hack-election-plan/>; see also Parag Agrawal & Damien Kieran, Twitter Blog, *Our continued work to keep Twitter secure* (Sept. 24, 2020), https://blog.twitter.com/en_us/topics/company/2020/our-continued-work-to-keep-twitter-secure (“We have strict principles around who is allowed access to which tools and at what time, and require specific justifications for customer data to be accessed.”).

⁵⁹ Although Twitter “is relatively coy about its current data center footprint” (because that footprint is vulnerable), it has been publicly reported that during the 2020-21 timeframe, Twitter had “[a] facility in Sacramento, and ... [a] data center in Atlanta.” *Twitter plans to build out new data center as platform grows*, Sebastian Moss, February 7, 2020 <https://www.datacenterdynamics.com/en/news/twitter-plans-build-new-data-center-platform-grows/>. When Mudge joined, Twitter had recently begun some amount of load shifting between data centers, but the process was both manual, and buggy.

⁶⁰ A security incident is an incident significant enough to trigger interruptions to work and redirect teams to track down the incident, determine the scope of the incident, and, if required, report it to the government.

⁶¹ A security incident that may need to be reported would include exposure of sensitive user information like emails, passwords, phone numbers or users' credit card data. An incident that might not need reporting might be a code bug.

⁶² See Exhibit 3; Mudge noted that internal reports stated more than 200 million customers and more than 20,000 employees (current and past) were impacted or involved in such breaches.

⁶³ See Exhibit 17.

⁶⁴ A 2017 hack of Equifax exposed the data of 147 million U.S. persons (fewer than those affected by Twitter's deficiencies), and led to a \$575 million fine. See U.S. Federal Trade Comm'n, *Equifax to Pay \$575 Million as Part of Settlement with FTC, CFPB, and States Related to 2017 Data Breach* (July 22,

-
- 48. January 6 Capitol Attack:** When a violent mob attacked and invaded the U.S. Capitol Building in an attempt to prevent Congress from certifying the election, Mudge quickly went to the executive in charge of engineering and asked “how do we seal the production environment?” Not knowing if there would be acts of internal protest aligned with the rioters, Mudge did not want any employees accessing, or potentially damaging the production environment. It was at this point when he learned that it was impossible to protect the production environment. All engineers had access. There was no logging of who went into the environment or what they did. When Mudge asked what could be done to protect the integrity and stability of the service from a rogue or disgruntled engineer during this heightened period of risk he learned it was basically nothing. There were no logs, nobody knew where data lived or whether it was critical, and all engineers had some form of critical access to the production environment. (Later on January 6 after the Capitol attack, the incoming administration offered Mudge a day-one appointed position as Chief Information Security Officer for the United States; Mudge turned the position down on the grounds that he thought he could have more positive impact fixing Twitter.)
- 49. Initial report:** Mudge presented his initial findings to the senior executive team in February 2021, about one week before the Q1 Board meeting.⁶⁵ Jack Dorsey had specifically recruited Mudge for his reputation of speaking truth to power, and told Mudge to not hold back. And Twitter’s other senior leaders knew they had security problems. But even so, the rest of the executives were stunned to hear Mudge tell them just how bad things were. While Mudge highlighted some positive aspects of Twitter’s security processes, such as the Company’s well exercised (but understaffed) team tasked with scrambling to react to crises, the overall picture was dire.
- 50. Defensiveness and denial from Agrawal:** Even at the first executive team meeting where Mudge shared his initial findings, Mudge got stiff pushback. In particular, Twitter’s CTO Parag Agrawal vehemently challenged Mudge’s assessment that Twitter faced a non-negligible existential risk of even brief simultaneous, catastrophic data center failure, and had no workable disaster recovery plan.

2019),
<https://www.ftc.gov/news-events/news/press-releases/2019/07/equifax-pay-575-million-part-settlement-ftc-cfpb-states-related-2017-data-breach>.

⁶⁵See Exhibit 3

Twitter’s most senior engineers had told Mudge they did not know whether, or on what time frame, Twitter could recover from such an outage. Perhaps Agrawal’s defensiveness should not have been surprising—as a senior engineer later promoted to then Chief Technology officer for years, Twitter’s problems had developed under Agrawal’s watch.

51. Below are excerpts from Mudge’s notes prepared for the February meeting, on the particular issue of simultaneous catastrophic data center failure. While catastrophic loss of data centers would understandably be problematic, take note of the last item. Mudge was shocked to learn that even a temporary but overlapping outage of a small number of datacenters would likely result in the service going offline for weeks, months, or permanently. This was even more disturbing as small outages were not uncommon due to bad software pushes from the engineers. On top of this all engineers had some form of access to the data centers, the majority of the systems in the data centers were running out of date software no longer supported by vendors, and there was minimal visibility due to extremely poor logging. This meant that of the four threats cited below what would normally be viewed as the least surprising, and was the statistically most likely, issue carried the greatest damage to the company; an existential company ending event:

With all of the above helping to provide context around our environment, and some of what is slowing us down or making it difficult to execute on our strategy and operations, let me share the existential threat that surprised me.

Threat matrix of effect:

- [REDACTED] data centers physically destroyed
 - Twitter unable to do business - full stop (not surprising)
- [REDACTED] goes down (hard or soft)
 - Twitter continues to run out of [REDACTED]
- [REDACTED] goes down (hard or soft)
 - Twitter operates, but impaired - and more impaired as time goes on
- [REDACTED] data centers gracefully go down and come back up
 - We don't know - best guess is weeks to months to bring the service back online
 - We can't boot(?)
 - Known unknown we really should know

52. Instructions to withhold information from Board: After the executive team meeting, Mudge was instructed not to send a detailed written report to the Board of Directors, but instead convey his findings orally, at a high level only.⁶⁶ Mudge found the request unusual, but as a new team member, complied. With the benefit of hindsight, Mudge now interprets this instruction as an overt act in furtherance of an ongoing effort to restrict critical information and defraud the Board of Directors and Twitter shareholders.⁶⁷

53. Executive action: As Twitter's Security Lead, Mudge was responsible not just for identifying the problems, but also for fixing them. And over the course of 2021, he designed and implemented a long-term strategy for reform. Among other things, Mudge:

- a. Stood up a world-class **Privacy Engineering team**, recruited some of the best leadership talent in the world, quantified the problem for Twitter for the

⁶⁶ On multiple occasions, executive team members shared that they believed that the best type of Board was one that was uninformed so as to keep them very hands off and mostly out of Twitter's business. Note that Dorsey always encouraged Mudge to be direct, unfiltered, honest, and transparent with the Board of Directors.

⁶⁷ On information and belief, with respect to this episode, the particular Twitter staff lawyer was acting on instructions from Twitter General Counsel Vijaya Gadde. Mudge prefers not to provide the name of the staff lawyer that conveyed the instructions, on the grounds that he (Mudge) has no reason to suspect that particular staff lawyer of harboring fraudulent intent.

-
- first time,⁶⁸ and achieved more progress in only 8 months than had been made over several years prior⁶⁹;
- b. Solicited independent report by the Alethea Group to formally identify **platform integrity (manipulation, disinformation, and spam) capabilities and gaps** for the first time;
 - c. Procured resources and head count to enable significant growth of the **InfoSec team** for the purpose of enabling reform and accountability;
 - d. Created the **#Protect Initiative**, a formal 3-year Board-reported objective to address critical privacy, security, and platform manipulation issues;
 - e. Oversaw **improvements on user safety cases** that drove down a backlog⁷⁰ of more than 1M cases to approximately 200K, and placed Twitter on its way to running within internal support service level agreements (“SLAs,” in which Twitter defines the level of service that it needs to meet in order to responsibly serve its customers) in 2022 for the first time ever;⁷¹
 - f. Demanded **data-driven metrics, and accountable ownership** of every process;
 - g. Began **aggressively recruiting diverse top talent** from across the industry.

54. Disengaged CEO: CEO Jack Dorsey had recruited Mudge personally. They got along well, and Mudge has never suspected Dorsey of harboring bad intent. But Dorsey, the high-profile CEO of one of the most prominent companies on earth, was experiencing a drastic loss of focus in 2021. Dorsey attended meetings

⁶⁸ General Counsel Vijaya Gadde did not only provide legal advice, but supervised operational (non-legal) staff on non-legal matters. With respect to one operational (non-legal) privacy matter, after Gadde was shown quantified data for the first time, she stated approximately “...so this proves that we [Twitter] haven’t made any progress over the past 4 years.”

⁶⁹ Mudge notes that Twitter engineers worked very hard in the prior years in good faith, but without leadership having domain knowledge and expertise to direct them to measure the problem and correctly direct the effort and solutions the underlying problem grew larger, not smaller.

⁷⁰ Backlogs included items such as harassment, violations of various rules, and reported accounts and tweets, problems with accounts, etc. It was historically the norm that cases in backlogs would eventually become so old that they would be silently closed, which most would agree is inappropriate support.

⁷¹ In or around October 2021, Mudge learned that even the @TwitterSupport account was historically unmanned. Through new leadership, brought on by Mudge, other overlooked fundamental issues such as language support and staffing safety and abuse agents to match timezones when issues were being reported were identified and improved.

sporadically, and when he did, he was extremely disengaged.⁷² In some meetings—even after he was briefed on complex corporate issues—Dorsey *did not speak a word*. Mudge heard from his colleagues that Dorsey would remain silent for days or weeks. Worried about Dorsey’s health, the senior team mostly tried to cover up for him,⁷³ but even mid- and lower-level staff could tell that the ship was rudderless.

55. Lack of Support: Whatever the cause, Dorsey’s absent behavior was anomalous and unhelpful in summoning the herculean effort needed to fix Twitter’s problems. In theory, Dorsey supported Mudge, and delegated him a huge amount of responsibility. But in deed, Mudge was getting little to no actual support for his task of fundamentally changing the risky behaviors of over 8,000 employees, and the entire corporate culture. Other senior executives took advantage of Dorsey’s absence to stay in their separate silos, pursuing their separate interests without interference. Unsurprisingly, this dynamic had negative consequences.

56. Cascading data center problems: In or around the spring of 2021, Twitter’s primary data center began to experience problems from a runaway engineering process, requiring the company to move operations to other systems outside of this datacenter. But, the other systems could not handle these rapid changes and also began experiencing problems. Engineers flagged the catastrophic danger that all the data centers might go offline simultaneously. A couple months earlier in February, Mudge had flagged this precise risk to the Board because Twitter data centers were fragile, and Twitter lacked plans and processes to “cold boot.” That meant that if all the centers went offline simultaneously, even briefly, Twitter was unsure if they could bring the service back up. Downtime estimates ranged from weeks of round-the-clock work, to permanent irreparable failure.

57. “Black Swan” existential threat: In fact, in or about Spring of 2021, just such an event was underway, and shutdown looked imminent. Hundreds of engineers nervously watched the data centers struggle to stay running. The senior executive

⁷² Over the course of 12 months, Mudge had no more than 6 one-on-one telephone calls with Dorsey, each lasting less than 30 minutes and almost all at the request of Mudge. During these calls, Dorsey cumulatively spoke perhaps fifty words. The total set of their electronic communications, again predominantly initiated by Mudge, came to no more than a couple dozen text messages.

⁷³ One executive team member bragged to Mudge about trying to get Dorsey to break his silence by prodding and aggravating him (Dorsey).

who supervised the Head of Engineering, aware that the incident was on the verge of taking Twitter offline for weeks, months or permanently, insisted the Board of Directors be informed of an impending catastrophic “Black Swan” event. Board Member Robert Zoellick responded with words to the effect of “Isn’t this exactly what Mudge warned us about?” Mudge told Mr. Zoellick that he was correct. In the end, Twitter engineers working around the clock were narrowly able to stabilize the problem before the whole platform shut down.⁷⁴

- 58. Software Development Life Cycle (SDLC):** An SDLC is a uniform process to develop and test software, and a basic best practice for engineering development at commercial companies. Twitter’s need to implement an SDLC was more than a best practice, it had been required since the 2011 FTC Consent Order⁷⁵ and reported regularly to the Board of Directors.⁷⁶ In or around May 2021, Mudge instructed that the Board Risk Committee receive accurate data showing that the company only had a *template* for the SDLC, not even a functioning process, and by Q2 2021 that template had only been rolled out for roughly 8 to 12% of projects.
- 59. Board Misled on SDLC:** Board Chair Patrick Pichette became incensed and noted that for years the board had been hearing “the (SDLC) effort was getting closer to being complete.” Mr. Pichette realized he and the Board had been misled, and was not happy. After the meeting, an executive called Mudge to state that he and Agrawal were upset with Mudge for providing accurate information to the Risk Committee, and that he and Agrawal deserved credit for their efforts. The call was a turning point for Mudge. He realized that for years, Agrawal and other executives had been misleading the board by **reporting their efforts, not actual results.**

⁷⁴ This is one of the specific items on which Agrawal had challenged Mudge earlier that year, which Mudge interpreted as defensiveness over problems that had developed on Agrawal’s watch as Chief Technology Officer, see discussion above.

⁷⁵ See FTC 2011 Decision & Order (requiring Twitter to implement technical safeguards appropriate for its size and complexity, the nature and scope of its activities, and the sensitivity of the nonpublic consumer information). U.S. Fed. Trade Comm’n, *In the Matter of Twitter, Inc.*, (No. C-4316), Decision & Order (Mar. 2, 2011), available at:

<https://www.ftc.gov/sites/default/files/documents/cases/2011/03/110311twitterdo.pdf>.

⁷⁶ See FTC 2011 Decision & Order requiring Twitter to implement technical safeguards appropriate for its size and complexity, the nature and scope of its activities, and the sensitivity of the nonpublic consumer information. U.S. Fed. Trade Comm’n, *In the Matter of Twitter, Inc.*, (No. C-4316), Decision & Order (Mar. 2, 2011), available at:

<https://www.ftc.gov/sites/default/files/documents/cases/2011/03/110311twitterdo.pdf>.

-
- 60. Attempts to engage Agrawal:** On many of these deficiencies, the Chief Technology Officer behaved defensively. Beginning mid-2021, Mudge initiated bi-weekly one-on-one meetings with Agrawal to flag issues with him first, and tried to get Agrawal's buy-in for reforms. For a period of time, these regular meetings seemed to improve their working relationship.
- 61. Anomalous Handling of Report on Platform Integrity:** With authority to engage consultants, Mudge hired Alethea Group to produce a report on Twitter's capacity to combat mis- and dis-information, fight spam and hostile actors, and promote overall platform integrity. In or around May or June of 2021, Alethea Group reported its initial findings, which were devastating. Word got out to other senior executives, who became concerned about the impact on Twitter's reputation were the findings to become publicly known. Without notifying Mudge, others on the executive team approached Alethea Group and ordered them to open a separate contract with an outside law firm. Alethea Group was told to send their report there first; the external law firm was responsible for removing factual information that would be especially embarrassing for Twitter, and return to Alethea Group a "clean" version to present to Mudge. When Mudge learned of this, he wondered whether this process was illegal or unethical. Further, despite the fact that the report did not touch on legal or compliance issues in any way, lawyers applied the erroneous label "Privileged and Confidential / Attorney Work Product" to the report.⁷⁷ Twitter counsel explicitly told Mudge that **this was intended to hide the findings and prevent them from becoming known internally or externally.**⁷⁸
- 62. Perverse bonus structure:** In or around July 2021, Twitter announced the "Value Creation Award,"⁷⁹ a new bonus structure in which top executives could individually earn over \$10 million for generating short-term growth of mDAU ("monetizable daily active users," see description above Section II). No bonus was provided for

⁷⁷ See Exhibit 2; After consulting with independent "filter counsel," we have concluded that this document is not in fact subject to attorney-client privilege. (1) The report does not contain or discuss legal advice or exposure, nor does it discuss legal or regulatory options or contain legal citations. (2) It was written by non-lawyers at Alethea Group for Mudge, a non-lawyer executive tasked with security, privacy, and content moderation at scale.

⁷⁸ *Id.*; As described above, we have obtained an independent legal opinion that this document is not in fact subject to Attorney-Client Privilege, and also that the Work Product Doctrine has no application in the context of voluntary, protected disclosures under the Dodd-Frank Act.

⁷⁹ See page 62, SEC Schedule 14A, available at https://www.sec.gov/Archives/edgar/data/1418091/000114036122012589/ny20001921x1_pre14a.htm#tDC

improving platform privacy, security or integrity. Mudge came to believe that short-sighted incentives like this were an important cause of Twitter's egregious ongoing deficiencies.

63. **Failed “stemming” for hateful ad targeting:** Twitter maintains a list of hateful terms and slurs that cannot be used for ad targeting. But Mudge learned that the list was not “stemming” properly, meaning that even minor variations on slurs were able to be used for targeting for an unknown period (Twitter SIM 154).
64. **Failed logins:** In or around August 2021, Mudge notified then-CTO Agrawal and others that the login system for Twitter's engineers was registering, on average, between 1500 and 3000 failed logins every day, a huge red flag. Agrawal acknowledged that no one knew that, and never assigned anyone to diagnose why this was happening or how to fix it.
65. **No employee computer backups:** In or around Q3 or Q4 2021, Mudge learned that no Twitter employee computers were being backed up at all. Supposedly, Twitter's IT department had managed a backup system for years, but it had never been tested and Mudge learned that it was not functioning correctly. Obviously this raised fundamental risks for corporate data integrity, including financial data, for any information needing to be recovered that was located exclusively on employee laptops. (To the extent that financial staff's data was at risk, it could constitute material weaknesses in internal financial controls required under SEC regulations.) Other Twitter executives—aware that the company was chronically out of compliance with most government requests for information—tried to look on the bright side, noting that going forward Twitter would have a valid excuse for not responding to regulator queries about which data particular employees had access to on which days. They explicitly decided not to replace or fix the employee backup system, but instead discontinued the service entirely.
66. Knowing that employees often had actual data from production systems on their laptops several executives and leaders commented to the effect of “this is actually a good thing because it means we [Twitter] cannot comply with [legal requests] and have less exposure”.

67. At the end of 2021, the Head of Privacy Engineering and the Chief Privacy Officer reported accurately to the Board that:⁸⁰

Every new employee has access to data they do not need to have access to for the purpose of their role. Until we have implemented a mature centrally owned and operated system to manage access to data (e.g., entitlements and review, Role Based Access Controls, audits, etc) we are at risk of inappropriate access or use of data. Our inability to delete data compounds that risk, as we retain data that we should not have and which is therefore accessible by people who do not need to have access to this data.

68. **Deficient moderation for “Spaces”:** In December 2021, an executive incorrectly told staff and Board members that Twitter’s “Spaces” product was being appropriately moderated. But Mudge researched and discovered that about half of “Spaces” content flagged for review was in a language that the moderators did not speak, and that there was little to no moderation happening.

69. **Log4j:** In December 2021, the world discovered that “Log4j,” a very common piece of software deployed in hundreds of applications across hundreds of millions (or billions) of computers worldwide, contained a previously-unrecognized (“zero-day”) security vulnerability. Overnight, a huge number of computers around the globe needed patching, or else they would be easy for adversaries to exploit. Left unaddressed, Log4j lets hackers break into systems, steal passwords and login information, extract data, and infect networks with malicious software. Log4j was *already* actively being exploited to compromise computers worldwide by criminals and governments alike. As “the most severe computer vulnerability in years,”⁸¹ the FTC instructed companies to pursue remediation, and that they could request detailed explanation and data on a company’s Log4j remediation efforts.⁸² In January 2022, Mudge determined and reported to the executive team that (because of poor engineering architecture decisions that preceded Mudge’s employment) Twitter had over 300 corporate systems and upwards of 10,000 services that might still be affected, but Twitter was unable to thoroughly assess its exposure to Log4j,

⁸⁰ See Exhibit 1

⁸¹ “What the Log4j vulnerability is, who is affected - NCSC.GOV.UK.”
<https://www.ncsc.gov.uk/information/log4j-vulnerability-what-everyone-needs-to-know>.

⁸² “FTC warns companies to remediate Log4j security vulnerability.” 4 Jan. 2022,
<https://www.ftc.gov/policy/advocacy-research/tech-at-ftc/2022/01/ftc-warns-companies-remediate-log4j-security-vulnerability>.

and did not have capacity, if pressed in a formal investigation, to show to the FTC that the company had properly remediated the problem.⁸³

- 70. Unlicensed machine learning materials for core algorithms:** In January 2022, in the days before he was terminated, Mudge learned that Twitter had never acquired proper legal rights to training material used to build Twitter’s key Machine Learning models.⁸⁴ The Machine Learning models at issue were some of the core models running the company’s most basic products, like which Tweets to show each user. Days before Mudge was fired in January 2022, Mudge learned that Twitter executives had been informed of this glaring deficiency several times over the past years, yet they never took remedial action.
- 71. Misleading regulators in multiple countries:** When, years earlier, the FTC had asked questions about the training material used to build Twitter’s machine learning models, Twitter realized that truthful answers would implicate the company in extensive copyright / intellectual property rights violations. Twitter’s strategy, which executives explicitly acknowledged was deceptive, was to decline to provide the FTC with the requested training material, and instead pointed the FTC towards particular models that would not expose Twitter’s failure to acquire appropriate IP rights. In early 2022, the Irish-DPC and French-CNIL were expected to ask similar questions, and a senior privacy employee told Mudge that Twitter was going to attempt the same deception. Unless circumstances have changed since Mudge was fired in January, then Twitter’s continued operation of many of its basic products is most likely unlawful and could be subject to an injunction, which could take down most or all of the Twitter platform. Before Mudge could dig deeper into this issue he was terminated.
- 72. Penetration by Foreign Intelligence & Threats to Democracy:** Over the course of 2021, Mudge became aware of multiple episodes suggesting that Twitter had been

⁸³ The head of the Detection and Remediation Team had told Mudge that the lack of visibility across the Twitter system meant that there was no way to determine whether log4j was successfully remediated and that upwards of 10,000 instances of the vulnerability could still be running and would not be able to be identified. Similarly, multiple teams were reporting conflicting internal numbers of systems needing remediation or evaluation or that had been fixed.

⁸⁴ See Exhibit 39

penetrated by foreign intelligence agencies and/or was complicit in threats to democratic governance, including:⁸⁵

- a. **The Indian government** forced Twitter to hire specific individual(s) who were government agents, who (because of Twitter's basic architectural flaws) would have access to vast amounts of Twitter sensitive data. Twitter's transparency reports purported to quantify the number of government data requests from the Indian government, but the company did not in fact disclose to users that it was believed by the executive team that the Indian government had succeeded in placing agents on the company payroll. By knowingly permitting an Indian government agent direct unsupervised access to the company's systems and user data, Twitter executives violated the company's articulated commitments to its users.⁸⁶
- b. Twitter executives opted to allow Twitter to become more dependent upon revenue coming from **Chinese entities** even though the Twitter service is blocked in China. After Chinese entities paid money to Twitter, there were concerns within Twitter that the information the Chinese entities could receive would allow them to identify and learn sensitive information about Chinese users who successfully circumvented the block, and other users around the world. Twitter executives knew that accepting Chinese money risked endangering users in China (where employing VPNs or other circumvention technologies to access the platform is prohibited) and elsewhere. Twitter executives understood this constituted a major ethical compromise. Mr. Zatzko was told that Twitter was too dependent upon the revenue stream at this point to do anything other than attempt to increase it.
- c. The **Nigerian government** blocked Twitter in June 2021, then falsely claimed to be in negotiations with Twitter executives; Twitter's failure to correct the false record on many reported non-existent discussions with the Nigerian government permitted Nigeria to negotiate unilaterally through media and

⁸⁵ Mudge has submitted a separate disclosure including details and documentation of these incidents to the Counterintelligence and Export Controls Section within the National Security Division of the U.S. Department of Justice, and to the Senate Select Committee on Intelligence.

⁸⁶ "India - Twitter Transparency Center." <https://transparency.twitter.com/en/reports/countries/in.html>.

-
- dictate unfavorable terms for final resolution.⁸⁷ Twitter's deliberate decision to refrain from correcting misinformation about Twitter's proposed negotiations with the Nigerian government directly harmed Twitter shareholders. Permitting the Nigerian government to impose various conditions on the platform harmed free expression rights and democratic accountability for Nigerian citizens;⁸⁸
- d. A few months before CTO Parag Agrawal was promoted to CEO, Agrawal suggested to Mudge that Twitter should **consider ceding to the Russian Federation's** censorship and surveillance demands as a way to grow users in Russia. Although Mr. Agrawal's suggestion was never pursued or implemented, the fact that Twitter's current CEO even suggested Twitter become complicit with the Putin regime is cause for concern about Twitter's effects on U.S. national security. This was a strong departure from the message Mr. Dorsey had conveyed to Mr. Zlatos. This interaction was notable because Mr. Zlatos was already directing teams to prepare for possible Russian incursions into Ukraine;
 - e. Shortly before Mudge was unlawfully terminated, Twitter received specific information from a U.S. government source that **one or more particular company employees were working on behalf of another particular foreign intelligence agency.**
73. In none of these cases did Twitter choose to focus on the long term health of the platform and company. Mudge's inference from these and other episodes was that some senior executives were intent on hiding bad news, or even misrepresenting it, instead of trying to fix it. This was possibly because (a) executives had personal financial incentives to grow mDAU / active users; or (b) they didn't know any better; or (c) some of them had built the broken system in the first place.
74. **Squeezing Local Staff:** Countries where Twitter had a physical presence, including actual full time employees (FTEs), and particularly where Twitter had official offices, represented heightened risk to Twitter and the Twitter platform. In addition to the risk exposed by Twitter's fundamental lack of information security and privacy

⁸⁷Ruth Maclean, *Nigeria Lifts 7-Month Ban on Twitter*, New York Times, Jan. 13, 2022, <https://www.nytimes.com/2022/01/13/world/africa/nigeria-lifts-twitter-ban.html>

⁸⁸ See Exhibit 40

control, described in other disclosures, there was the physical safety of the employees to consider. The threat of harm to Twitter employees was sufficient to cause Twitter to seriously consider complying with foreign government requests that Twitter would otherwise fundamentally oppose. The governments of **India, Nigeria, Turkey and Russia** sought, with varying success, to force Twitter to hire local FTEs that could be used as leverage.

75. As of the date of Mudge's termination, January 19, 2022, Twitter remained out of compliance in multiple respects with the 2011 FTC Consent Order, which has the force of law following Twitter's consent to its terms. While the company had made progress on privacy because of Mudge's leadership, it was not on track to ever achieve compliance for other important items, especially in the area of information security (in which Mudge's reform efforts had been repeatedly and unreasonably blocked). Twitter's non-compliance constitutes violations of the Federal Trade Commission Act, 15 U.S. Code §§ 41-58.

[Disclosure continues next page]

V. New CEO Enables Fraud

76. **Dorsey Out, Agrawal In:** In November 2021, Twitter announced that Dorsey was stepping down, and would be replaced by Agrawal, effective November 29.
77. **Agrawal's vote of confidence:** Agrawal called Mudge shortly after the announcement and they spoke at length, one-on-one. Agrawal said that he (Agrawal) appreciated Mudge's efforts, acknowledged that Mudge had been put in a challenging spot for many reasons, and requested that Mudge stay on. Based on that call, Mudge believed that they shared the goal of leading with integrity and transparency. In contrast, Agrawal immediately terminated two other senior executives, the Head of Engineering and the Head of Design.
78. Meanwhile, Agrawal's first Board meeting as CEO was fast approaching. The full Board of Directors was set to meet on December 9, 2021, followed by a Board Risk Committee meeting on December 16. And Mudge was becoming increasingly concerned about the accuracy of the information that Board members would receive.
79. **Inaccurate Board Materials:** In or about the first week of December 2021, a senior employee of the company sent Mudge via email a draft PowerPoint presentation and other materials that were to be distributed to the Board of Directors before the December 9 meeting, which would then be used as a basis for the presentation to the Board. Upon review of the materials, Mudge determined that the materials contained false and misleading statements about the state of Twitter's information security and privacy.⁸⁹
80. Mudge's concerns are described in greater detail below, and in the exhibits to this disclosure. At a high level, Mudge flagged four issues which were improperly omitted or presented in a misleading way:

⁸⁹ Some of the emails evidencing deliberate misrepresentations to the Board in December 2021 constitute privileged attorney-client communications. Therefore, although Mudge is unable to share those particular communications, he can nevertheless describe the underlying facts. "[T]he protection of the [attorney-client] privilege extends only to communications, and not to facts. A fact is one thing and a communication concerning that fact is an entirely different thing." Upjohn Co. v. United States, 449 U.S. 383 at 395-96 (1981) (internal citations omitted).

-
- a. **Basic security protections on software (software and system patching):** The materials reported a misleading statistic that 92% of employee computers had security software installed, implying those computers were secure. In fact, that software's most basic function was to determine whether that particular computer's software and settings met basic security standards—and the software was reporting that one-third of the systems were critically insecure. A full 30% of employee systems were reporting that they had disabled critical safety settings such as software updates. Other critical flaws reported by the software brought this number closer to 50%. This crucial context was not included for the Board.
- b. **Access control to systems and data:** On the issue that led to the FTC complaint in 2011 and the July 2020 hack,⁹⁰ the materials did not contain the overall numbers, which were getting worse. Instead, the materials took a small, cherry-picked subset of data that could be made to look like a positive trend, and turned that into a graph with a downward trajectory. The graph misleadingly suggesting that Twitter was making significant progress in reducing access to production systems. Mudge knew that the actual underlying data showed that at the end of 2021, 51% of the ~11 thousand full-time employees had privileged access to Twitter's production systems, a 5% increase from the 46% of total employees in February of 2021 that Mudge had shared in his initial findings delivered to the Board in early 2021.
- c. **Volume and frequency of security incidents:** A graphic in the document showed only a subset of security incidents, presented as if to encompass all security incidents.⁹¹ Twitter's actual total number of security incidents in 2021 was closer to 60, a marked difference from what was being implied. The misleading graphic also attributed only 7% of incidents to access control,

⁹⁰ As Mudge noted in his February 2021 oral presentation to the Board, Twitter engineers have privileged access to Twitter's production systems, which meant 46% of full-time employees had such access to Twitter's live data. While the deck inaccurately reported improvements in this area, the problem was actually getting worse. Data showed that at the end of 2021, 51% of full-time employees had privileged access to Twitter's production systems, a 5% increase from Mudge's February 2021 report to the Board.

⁹¹ Twitter's actual total number of security incidents in 2021 was closer to 60, an alarmingly high number and particularly concerning, given that Twitter faced scrutiny from the FTC and similar international regulatory agencies. Because the document miscategorized various incidents, the graphic misattributed only 7% of incidents to access control, when in reality the root causes of 60% of incidents were actually access control issues, a problem that plagued Twitter's security and had not been properly reported to the Board historically.

when in reality access control was the root cause of 60% of security incidents.

- d. **Lack of Software Development Life Cycle (or related processes and compliance)**, was presented as largely completed, instead of still in the initial phases of planning, as discussed above in Section IV.

81. Mudge immediately contacted CEO Agrawal and other senior staff, seeking to correct the misinformation in the draft materials and/or stop their delivery to the Board. While those discussions were in process, and to help provide more focus on privacy and security to Board members, Mudge substituted corrected summaries for the package sent to the full Board before December 9.
82. **December 9:** At the December 9 full Board meeting, the inaccurate materials were not shared only because of Mudge's strenuous efforts. Mudge replaced the materials with a short, factual summary of the work on Information Security and Privacy teams. Mudge indicated details would be provided in the Board Risk Committee meeting on December 16th, a week later. Mudge then briefed the Board on his 3-year #Protect Initiative that would provide Board-level quantified visibility into the areas of privacy, information security, and scope and remediation efforts around malicious actors and activity on the platform.

83. **December 12:** For months, Mudge had been blocked from fixing a particular, sensitive problem that culminated with the misleading PowerPoint deck. Now the issue had metastasized and required direct CEO involvement. Mudge sent this email noting his inability to meet his obligations to the imminent Board Risk Committee meeting:⁹²

On Sun, Dec 12, 2021 at 7:55 AM Peiter "Mudge" Zatko <[REDACTED]@twitter.com> wrote:
Privileged and Confidential

Parag and Dalana ([REDACTED]),

I apologize for the need to bring this to your attention and for the time critical nature of the ask. We all have enough on our plates and this should have been able to be handled without more executive support than myself.

I need your support in resolving the matter of [REDACTED] Monday (Dec 13).

Twitter [REDACTED] is stuck in regards to actioning this item. The item was supposed to have been resolved weeks ago. Multiple promised deadlines have come and gone. [REDACTED]

[REDACTED] I am blocked from being able to correctly perform key portions of my duties and obligations for Twitter, including for the board and Risk Committee, and we are causing Twitter increasing harm through these continued delays.

[REDACTED] have context on this item.

Thank you for your attention to this (now) urgent matter.

Kindest,

Mudge Zatko

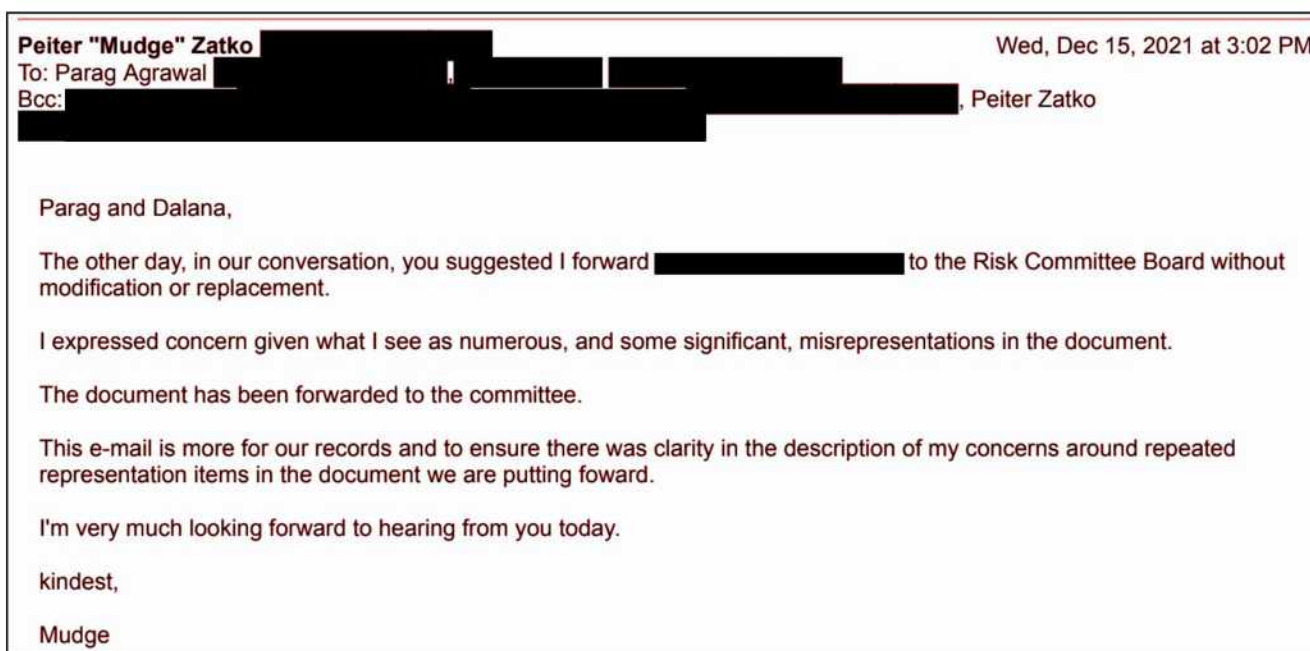
⁹² See Exhibit 41. This email is primarily from a non-lawyer to non-lawyers concerning non-legal matters. We have redacted parts of the email that Twitter might consider privileged, and/or to protect the privacy of third parties.

-
84. **December 12:** In a call on or around December 12 with Agrawal and at least one other (non-lawyer) witness, Mudge expressed his deep concerns about the upcoming Risk Committee meeting because a senior employee in the company intended to present the misleading material to board members at Risk Committee meeting on December 16, 2021. On the call, Mudge offered to rewrite the problematic report intended for the Risk Committee meeting to ensure it was accurate and complete. Agrawal told him not to correct the information and instead asked for time to look into this issue further.
85. **December 14:** Agrawal called two days later⁹³ He said that the materials as they stood would be presented to the Board Risk Committee on December 16, despite Mudge’s informing Agrawal that the information in the materials were materially misleading.
86. Mudge again offered to create corrected materials, but Agrawal told Mudge not to do so. Agrawal gave Mudge a different “solution”: Mudge would be forced to send the inaccurate materials, unchanged, to the Risk Committee, and then correct the inaccuracies presented to the Board in real time—and without accurate written materials. Also Agrawal told Mudge that at the end of the meeting in the closed session with Board members, Mudge should do his best to “walk back” and correct the misinformation presented to the Risk Committee verbally and electronically. Agrawal promised that he (Agrawal) would personally call the members of the Committee after the fact to help ensure they were not misled, if Mudge so requested.

[Disclosure continues next page]

⁹³ On information and belief, Agrawal called from a landline telephone. A call detail record may have captured the call time and duration.

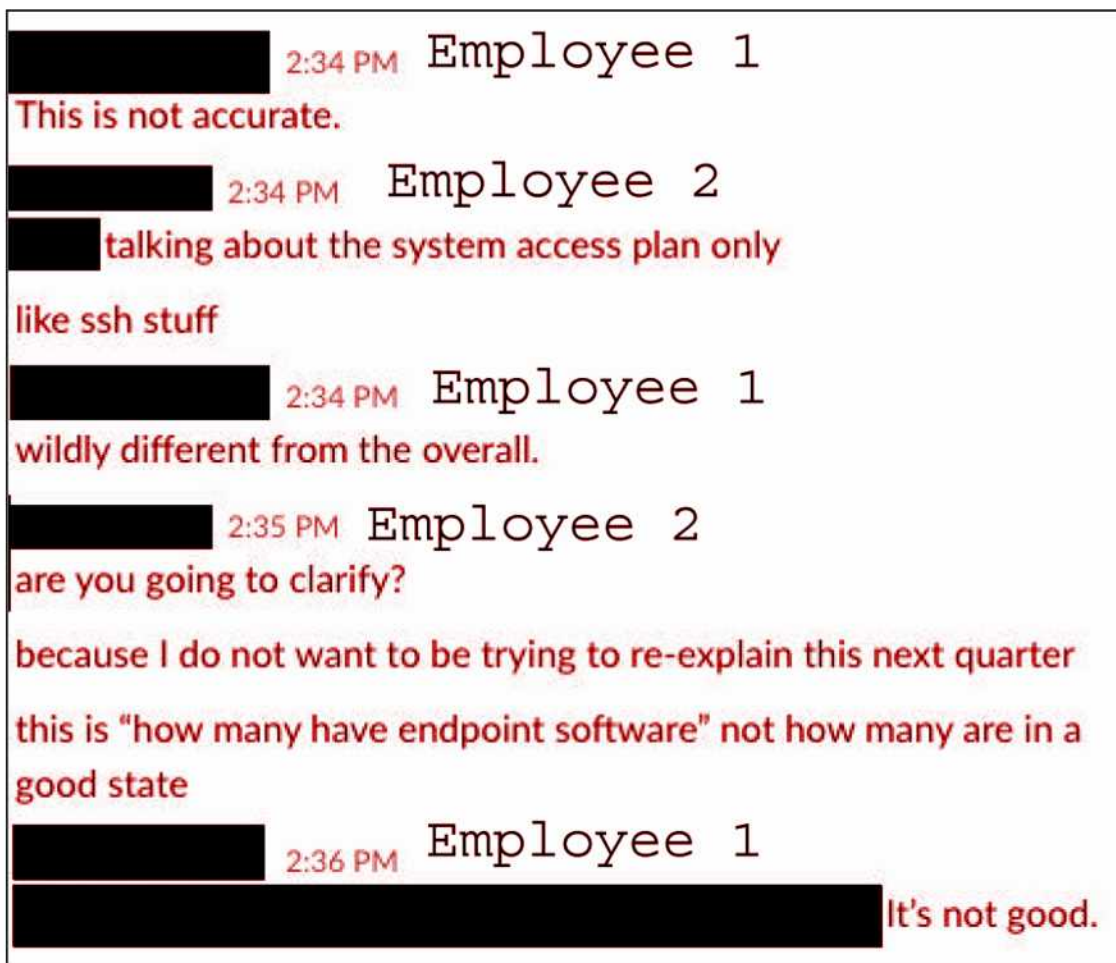
87. Mudge was bewildered and at a loss for words. He objected to Agrawal's plan, on the obvious grounds that Board members should never receive false or misleading information to begin with. Nevertheless, Agrawal was the CEO and his boss. Agrawal had given him a direct order, and Agrawal had promised to help remedy the situation afterwards at Mudge's request. Mudge proceeded, and sent another email on December 15, 2021 documenting his concerns:^{94 95}



⁹⁴ Exhibit 10, pg 3

⁹⁵ After Mudge's signature in the email there were detailed explanations of a subset of items Mudge felt were critical so as to ensure there was no chance for Parag to be confused as to why there were concerns.

88. **December 16:** The Risk Committee received the misleading document before the meeting, and the senior Twitter employee conveyed the same misleading information in a verbal briefing during the main part of the meeting on December 16. This screenshot⁹⁶ of unsolicited, real-time instant messages during the Board Risk Committee meeting confirms that other participants recognized untrue statements:



A screenshot of an instant messaging conversation. The messages are as follows:

- Employee 1 (redacted name): 2:34 PM. "This is not accurate."
- Employee 2 (redacted name): 2:34 PM. "talking about the system access plan only like ssh stuff"
- Employee 1 (redacted name): 2:34 PM. "wildly different from the overall."
- Employee 2 (redacted name): 2:35 PM. "are you going to clarify? because I do not want to be trying to re-explain this next quarter this is "how many have endpoint software" not how many are in a good state"
- Employee 1 (redacted name): 2:36 PM. "It's not good."

89. For those without a technical background, these messages are discussing "endpoint software," meaning software on individual employee machines (as opposed to, e.g., server software running in Twitter's data centers) and "access

⁹⁶ See Exhibit 9

control” to Twitter’s production systems and data using only the “*ssh*” encryption protocol. The senior employee’s comments were misleading in multiple respects:

- a. First, the employee stated that 92% of Twitter employee computers had this security software installed, implying a high level of security. But in truth, the endpoint software did not actually provide or constitute security on its own. Rather, the endpoint software’s primary function was to evaluate whether the employee computer had basic security configurations enabled. And most importantly, the software on these endpoint computers was reporting dire problems. Over 30% of the more than 10,000 employee computers were lacking the most basic security settings, such as enabling software updates.
- b. Second, during a broad discussion of the company’s overall access control issues, the employee stated cited progress on a very small subset of the problem referred to as “system access plan” (which only applied to a small percentage of relevant users) and only the “*ssh*” communication method (which was merely one method among many). Therefore the cherry-picked numbers misrepresented the problem.
- c. Board members were not given enough information to draw these elementary distinctions.

90. At the end of the meeting, consistent with Agrawal’s instructions, in a ~2 minute time slot allotted to him within the 15 minute closed session, Mudge attempted to correct the misleading information provided in both verbal and written materials, but simply did not have time to adequately remedy the situation given the technical information and scope of the misrepresentations. Mudge, therefore, followed up the meeting with an e-mail to Agrawal accepting Agrawal’s prior offer to personally reach out to the board members to correct any remaining misstatements.

91. In the e-mail Mudge re-articulated some of the most concerning items that Mudge felt required further explanation to ensure the Board members were not left confused or misinformed.⁹⁷ Agrawal asked whether Mudge was able to “correct for the committee” the misinformation “like we had discussed for this kind of scenario”:

----- Forwarded message -----
From: **Parag Agrawal** [REDACTED]
Date: Fri, Dec 17, 2021 at 3:20 PM
Subject: Re: action items from Mudge 1 of 2
To: Peiter "Mudge" Zatko [REDACTED]

Did you discuss / correct for the committee during the 15 min exec session at the end of meeting like we had discussed for this kind of scenario?

Parag

On Fri, Dec 17, 2021 at 9:51 AM Peiter "Mudge" Zatko [REDACTED] wrote:
Per your offer - we should make sure the Risk Committee is correctly aligned on an item presented by [REDACTED].

[REDACTED] may have created an impression that overall access control risk was materially declining through their efforts. This is not the case. Access control exposure at Twitter has grown at a rate even faster than our rate of FTE growth (e.g. 46% of all employees had direct access to production in dec 2020 and 51% have direct production access in dec 2021).

Here are observations sent to me from Twitter employees in attendance during [REDACTED] comments on access control in response to questions about how worried the Risk Committee should be about Twitter's access control challenges.

⁹⁷ See Exhibit 9

-
92. Instead of taking the promised action to remediate the situation created by Agrawal’s insistence on providing the Risk Committee with inaccurate and misleading information, Agrawal called Mudge to say he was “disappointed in him.” Agrawal chided Mudge, telling him that he (Mudge) should have been able to make the whole problem go away. Agrawal mentioned he would be visiting New York City and Philadelphia during the holiday break, and—because Mudge lived in the area—suggested they meet up in person to work things out.
93. Agrawal, however, did not contact Mudge to meet in person. Concerned that Agrawal had not reached out to move forward on correcting the misinformation to the Risk Committee, in or around the last week of December, Mudge called Agrawal to ask when they were meeting. Agrawal demurred, stating that “his travel had changed” and he (Agrawal) had not visited the area that would have enabled a meeting. But Agrawal was lying to Mudge.
94. Later, on a different matter, Twitter’s executive security services (which ultimately reported to Mudge) presented Agrawal’s travel detail reports to Mudge for review. The documents showed that Agrawal had in fact visited the East Coast in late December, and that there had been no significant travel changes. This evidence contradicted Agrawal’s claim that he (Agrawal) was not able to meet with Mudge, especially on such a significant topic as addressing inappropriate information being conveyed to the Board. In fact, the minor change in airports would have made a meeting with Mudge even more convenient as the private airport being used was even closer to where Mudge was during the time period.⁹⁸

⁹⁸ See Exhibit 28. Note that Mudge’s residence is between Philadelphia and New York City and very close to the private airport Agrawal used.

Parag and Family Holiday Protective Details

<p>Air Travel</p>	<p>Proposed Schedule</p> <ul style="list-style-type: none"> • CA → NY/PA <ul style="list-style-type: none"> ◦ DEPARTURE: 12/17/2021 @ 10:00 pm PDT <ul style="list-style-type: none"> ▪ [REDACTED] ▪ [REDACTED] ◦ ARRIVAL: 12/18/2021 @ 6:06 am EDT at Philadelphia International Airport (KPHL) <ul style="list-style-type: none"> ◦ [REDACTED] • PA → CA <ul style="list-style-type: none"> ◦ DEPARTURE: 1/2/2021 @ 11 am EDT- 3pm EDT (reservation pending) <p>Deviations from Proposed Schedule (impacts of deviations reviewed in AAR below)</p> <ul style="list-style-type: none"> • [REDACTED] and family flew into Teterboro airport on 18 December instead of Philadelphia, requiring the use of a contractor driver for pickup instead of [REDACTED]
<p>Additional Travel</p>	<p>Proposed Schedule</p> <ul style="list-style-type: none"> • 12/18/21 - NYC • 12/20/21 - NYC → [REDACTED] PA • 12/21/21 - [REDACTED] PA → [REDACTED] • 12/22/21 - [REDACTED] PA → [REDACTED] PA <p>Deviations from Proposed Schedule (impacts of deviations reviewed in AAR below)</p> <ul style="list-style-type: none"> • No movement required from Philadelphia to NYC as originally required (due to flight change) • Returned to NYC on 28 December [REDACTED] • Drove back from NYC [REDACTED] on 30 December

[Disclosure continues next page]

VI. 2022: Termination in the New Year

95. On **January 4, 2022**, deeply concerned that so much time was passing without corrections and unable to get Agrawal to meet or talk about the topic, Mudge sent the following email describing the December 16 meeting as “at worst fraudulent”.⁹⁹

[Disclosure continues next page]

⁹⁹ Exhibit 11, p. 1-2

On Tue, Jan 4, 2022 at 3:20 PM Peiter "Mudge" Zatko [REDACTED] wrote:

Jan 4, 2022

CC: Dalana Brand

Parag,

>> Happy New Year! I was looking forward to tagging up last week to discuss things, but I can appreciate the challenge of holiday schedules and the health issues we are all facing. I hope we will still be able to get together in the not too distant future.

>>

>>

>> As you know, I was hired to achieve certain goals and to fix problems here at Twitter. In order to do that, we need to recognize the actual state of affairs at the company in order to identify what needs to be fixed. Last month, we discussed this issue [REDACTED] report to the Risk Subcommittee of the Twitter Board. [REDACTED]

[REDACTED] to the Risk Subcommittee reports which I have pointed out are misleading and simply wrong.

>>

>>

>> As I have previously reported to you, it is critical that when we report to the Board, or one of its subcommittees (here Risk), that we be accurate and not mislead the board or its subcommittees. I know you share my concern that if [REDACTED] misrepresents the state of affairs, overstates what has been accomplished and ignores what needs to be improved, we do a great disservice to Twitter and whitewash the problem which if not addressed, could undermine the future of this company.

>>

>>

>> [REDACTED] blocked my ability to correct problems [REDACTED] latest report to the Risk Subcommittee, it has presented inaccurate, even false reports to the Risk Management Subcommittee of the Board. Such activity is at worst fraudulent and at best hiding the truth from the subcommittee.

>>

>>

>> In order for me to do my job, contractually, ethically, and legally, I need to ensure that what we report to the Board and its subcommittees is true and accurate. If there is a problem then the unvarnished truth must be presented. Only then can the problem be fixed. I would hope that you would agree with me. Going forward, we must ensure that we do just that. I think we also need to correct misinformation previously provided.

>>

>>

>> In addition, in order to ensure accuracy is maintained, I would request that [REDACTED]

>>

>>

>>

>> I think we have made measurable and material progress since my hire, but we have a ways to go. I know we can accomplish the fixes necessary, but we need to be honest in our reporting and in that way, the Board will know where we are and where we need to be.

>>

>>

>> I look forward to your assistance to improve our company.

>>

>>

>> Best regards,

>>

>>

>> Peiter Mudge Zatko

96. Despite having been repeatedly warned about these issues, Agrawal responded that he was “very surprised” to see Mudge’s concerns. Mudge believes this was a bad faith effort to create plausible deniability after the fact:

On Thu, Jan 6, 2022 at 8:26 PM Parag Agrawal [REDACTED] wrote:
>
> Mudge
>
>
> Thank you for your email. Given our conversations before and after the Risk Committee meeting, however, I have to say I'm very surprised to see it. We take all of these concerns and issues very seriously. Since you have raised concerns that may implicate fraud, we'll report this to the Audit Committee and start a formal investigation. Please keep in mind you always have the opportunity to speak up in any meetings of the board or its committees. As the lead for these Risk Committee conversations, I expect you to ensure the accuracy of information presented.
>
>
> We will follow up with you to fully investigate this matter.
>
>
> Thanks
>
> Parag
>

97. **January 11:** Following Mudge’s reference to “fraud,” Twitter’s Chief Compliance Officer initiated an internal investigation and interviewed Mudge on January 11, 2022. It is relevant to note that the compliance team could face personal liability for failing to investigate fraud allegations. They agreed that the information delivered to the Risk Committee was inappropriate and inaccurate, and—contrary to Agrawal’s order to Mudge—that he (Mudge) should write a report correcting the misrepresentations. Mudge immediately began drafting a corrective report for the Board as agreed upon. On January 11, and again on January 17, the Chief Compliance Officer sent emails to Mudge that Twitter may consider privileged and confidential.

98. **Memorandum for the record:** Mudge documented his concerns again on January 12, flagging the ongoing problem (which had persisted for the entire 2021 calendar year) that culminated with the incorrect PowerPoint:¹⁰⁰

¹⁰⁰ Although this email contains the words “Privileged and Confidential,” and Twitter counsel was copied, the email is not subject to a claim of attorney-client privilege, because it was a message from one non-lawyer, Mudge, to another non-lawyer, Agrawal, and it was not sent for the purpose of seeking legal advice. We have redacted any arguably privileged content.

----- Forwarded message -----

From: Peiter "Mudge" Zatko [REDACTED]
Date: Wed, Jan 12, 2022 at 10:41 PM
Subject: Re: Memorandum for the record
To: Parag Agrawal [REDACTED]
Cc: Dalana Brand [REDACTED], Sean Edgett [REDACTED], Marianne Fogarty [REDACTED]

[privileged and confidential]

Parag:

Thank you for getting back to me. I want to quickly follow up with a few items so we can have the most productive 1:1 tomorrow.

First, for global context, my mission here is to protect Twitter in order to enable the company to succeed. That also means protecting you and helping you succeed in your new role too. I hope you recognize that I am performing this duty throughout this topic. I take this mission, and my responsibilities, very seriously.

I had an ask in my letter to you on January 4 to address the problem with [REDACTED]. Your email reply on January 6 does not mention [REDACTED]. I am hoping our meeting tomorrow will focus on this. As I mentioned in my email to you, I have no problem if Twitter wishes to [REDACTED] but please [REDACTED].

On other topics you indicated in your response of January 6, to my email of January 4, that you were surprised to read what was contained in my email given conversations we had. I take this as being "surprised" that I raised the issue again because we had discussed all of what was in my email both in prior e-mails and by voice.

On the topic of inappropriate material being put in front of Board members, thank you for initiating the audit investigation. I spoke with them yesterday (January 11th). I believe it was a very productive communication.

To assure you that I make every effort to ensure accurate information goes to our Board and Committee members and meetings it is documented that I was actively working to prevent inappropriate information from going to the Board or to the Risk Sub-Committee for months. To this end [REDACTED] prior to the December 9th Board meeting. When this was blocked I made sure concerning material was not put in front of the board and that appropriate data accurately presenting our environment and risks was shared.

The issue of now needing [REDACTED] prior to the Risk Committee was escalated to you [REDACTED]. I shared my concerns around the issue of not having this resolved prior to the Risk Committee meeting. In that conversation you took personal responsibility for ensuring this issue would be resolved. However, a few days before the Risk Committee you called me and told me that due to concerns you had discovered around [REDACTED] you were not able to keep that promise[0].

You indicated that [REDACTED] I had informed you of issues related to that report and my objection to its presentation to the Board. To the extent [REDACTED] inaccurate description of where we were, you informed me that I was to attempt to correct such items to the Board members during [REDACTED] presentation or at the end of the meeting during a closed session. I expressed concern around this approach. I did not want to have an inaccurate report going to the Board in the first place. It is difficult, to say the least, to have a report given to the Subcommittee only to tell them that it is not accurate, but as you instructed, I did my best to reel it in. I was extremely grateful that you offered to personally call the board to assist in correcting the record if I came back to you after the meeting asking for any help in further clarifying areas where I felt there may not be accurate understanding. I sent you an e-mail taking you up on your offer for a topic [REDACTED] which other people in the meeting identified as inappropriate and misleading. I was confused and sad when you told me you were disappointed in me for not having completely resolved these issues.

I have faith in processes such as the audit investigation. I am confident this will get to a place where the board members have an appropriate understanding of our risks and our progress against them.

I want to bring Twitter to where it needs to be. I need your help to clear obstacles to that objective. I am also thankful that you have authorized me to disclose accurately to the Board going forward.

Hopefully with much of the above now in the hands of the audit investigation, our conversation tomorrow will be on my ask in my January 4th e-mail: addressing the problem [REDACTED].

Thank you for your assistance. Looking forward to resolving this and getting towards our year of execution!

Best Regards,

Peiter Mudge Zatko

[0] Very grateful for your phone call, where you apologized for not being able to keep your promise of resolving the issue [REDACTED] prior to the Risk Committee. [REDACTED]

99. **January 18, 11:16am:** Following a request from Twitter’s Chief Compliance Officer, Mudge sent an email confirming he planned to provide corrected materials for the Board “by the end of this week,” as part of a fraud investigation initiated pursuant to his January 4 email. Less than two hours later, Agrawal emailed Mudge, and surprised him with a request to do a call 45 minutes later with himself (Agrawal) and Omid Kordestani,¹⁰¹ Chair of Twitter’s Board of Directors’ Risk Committee:¹⁰²

On Tue, Jan 18, 2022 at 1:15 PM Parag Agrawal [REDACTED] wrote:

Mudge

I have time with Omid at 2p pacific today, and given concerns you have raised — I think it is very important to use this time to speak directly with Omid and provide him your perspective and all details around the last risk committee meeting so he and the committee can be best informed.

Parag

100. In the meeting with Kordestani and Agrawal, Mudge stated he was close to finishing corrective materials for the Board as agreed upon with the Chief Compliance Officer. Both Agrawal and Kordestani were angry at Mudge. This confused Mudge until he realized they were trying to blame him for the awkward situation that by now had led to a formal investigation by the Chief Compliance Officer. During the meeting, Agrawal lied, stating he (Agrawal) had been waiting for over a month for Mudge to produce corrective materials—directly contradicting his order to Mudge that he

¹⁰¹ On information and belief, Kordestani was one of the Board members who had advocated strongly for the decision to promote Agrawal to the CEO position, which Mudge heard had been controversial among Board members. Kordestani had a vested interest in ensuring that Agrawal did not face accusations of misconduct in his first few months as CEO. Twitter’s recent SEC filing also shows that Kordestani holds a significant 934,247 shares of Twitter stock, then worth \$50,636,187. Kordestani’s holding of shares is second only to Dorsey’s. Kordestani therefore has a substantial financial interest in hiding the Company’s dire cybersecurity condition.

¹⁰² See Exhibit 5

stand down on producing such a product. When Mudge attempted to correct this false statement, Kordestani interrupted and refused to let Mudge continue. Mudge sent the following email after the call:¹⁰³

From: Peiter "Mudge" Zatko [REDACTED]
Date: Tue, Jan 18, 2022 at 7:20 PM
Subject: Re: time with Omid
To: Parag Agrawal [REDACTED]
Cc: Sean Edgett [REDACTED], Marianne Fogarty [REDACTED]

Privileged and Confidential

Thanks for setting this up Parag and thank you for the priority and concern. I would not have brought it up if I did not think it was important. The response of the Audit Investigation and Omid lead me to believe I was right to bring it up.

I was a bit surprised by your statement that you have been waiting over a month for a corrective document from me. My personal meeting notes and timeline shows that in December, on a phone call with you on the topic of [REDACTED], the materials, and the upcoming Risk Committee meeting, I offered to create a corrected document for the committee and you instructed me to not pursue this path. I was instead told to forward [REDACTED] documents and have [REDACTED] present, attempting to police the meeting and deal with [REDACTED] document as best I could in the closed session[0]. One week ago (Jan 11th) Marianne and I [REDACTED]

I am confused as to where your statement in front of Omid that you have been waiting on me for this for over a month came from. My apologies for any misunderstandings. If you could point me to where you requested this earlier I would like to figure out how to avoid such misunderstandings in the future.

In the future it would be great to have more than a few hours of notice for a meeting with board members to allow me to gather all appropriate materials and ensure I can provide pre-reads for attendees and yourself. I know that sometimes short notice meetings are just a thing.

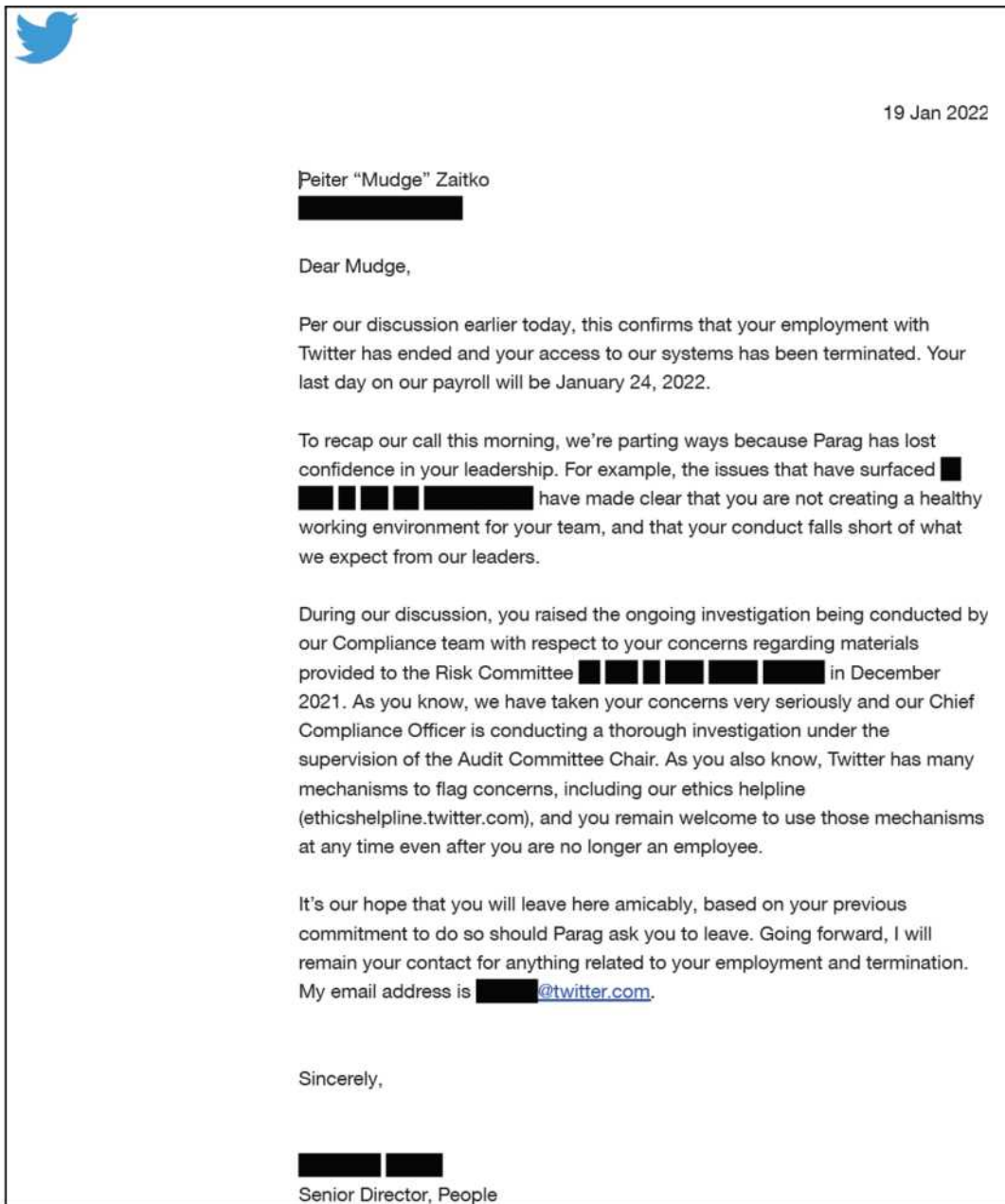
kindest,

Mudge Zatko

[0] And that on any items I still had concerns about you would make personal call(s) to the members to help clear up.

¹⁰³ See *Id.* Although this email contains the words "Privileged and Confidential," and Twitter counsel was copied, the email is not subject to a claim of attorney-client privilege, because it was a message from one non-lawyer, Mudge, to another non-lawyer, Agrawal, and it was not sent for the purpose of seeking legal advice. We have redacted any arguably privileged content.

101. On January 19, 2022, Twitter CEO Parag Agrawal (with one other executive present) called Mudge and terminated his employment.¹⁰⁴



¹⁰⁴ See Exhibit 36

102. During the call, Twitter revoked Mudge’s access to Twitter systems and email. Immediately after the call, a Twitter security agent knocked on Mudge’s door, and repossessed Mudge’s work computers and phones.
103. Based on statements made by Mudge during the termination meeting, later on that same day, January 19, Twitter’s Chief Compliance Officer began emailing Mudge, requesting that he continue working to produce corrected materials for the Board of Directors. Her reference to “your conversation this morning” was the video call in which Mudge had been terminated, and “matters already under investigation” referred to the December 16 misrepresentations to the Board:^{105 106}

**On Wed, Jan 19, 2022 at 11:59 PM Marianne Fogarty [REDACTED]@twitter.com> wrote:
Mudge,
Kathleen Pacini advised me that during your conversation this morning you mentioned the concerns you raised about information shared with the Risk Committee in December. If you were referring to matters that are not already under investigation, please let me know so we can schedule time to talk right away.**

¹⁰⁵ Exhibit 27. No attorney-client privilege attaches to post-termination communications between Mudge and Twitter counsel, because any attorney-client relationship was severed the moment he was terminated and was no longer a “constituent” of Twitter counsel’s client, the corporation. Courts have occasionally upheld privilege claims for post-separation communications between corporate counsel and former employees, but those have occurred pursuant to a formal Joint Defense agreement, or when the corporation and the former employee had a shared interest, e.g. in defending tort claims. But Twitter counsel’s post-termination communications with Mudge were not pursuant to a Joint Defense Agreement nor a shared interest; in fact the parties were by that point adverse since Mudge had raised concerns during his termination meeting that his termination was retaliation for disclosing violations of law. We are not aware of any authority suggesting that Twitter counsel can assert a valid privilege claim in these circumstances. Further, Twitter’s Chief Compliance Officer never indicated these communications were privileged. Although she held a law license, for the purposes of these communications she was acting in an operational, non-legal capacity.

¹⁰⁶ From the moment Mudge was fired, any attorney-client relationship between himself and Twitter counsel was also terminated. Therefore no attorney-client privilege could attach to his post-termination communications with Twitter counsel or staff.

104. Mudge replied to the compliance officer in the affirmative, alluding to Agrawal's January 18 lie that he (Agrawal) had been waiting a month for Mudge's corrective materials:¹⁰⁷

On Fri, Jan 21, 2022 at 1:21 PM Marianne Fogarty <[REDACTED]@twitter.com> wrote:
Hi Mudge.
Can we set up a time to talk to identify and understand the full breadth of additional concerns/issues and can look into them?
The sooner the better. Are you free this afternoon? I'll make myself available at whatever time works for you.
Thank you.
Marianne

On Fri, Jan 21, 2022 at 10:07 AM Peiter Zatko <[REDACTED]@gmail.com> wrote:
Yes. These are now additional concerns/issues introduced by Parag.

105. Later on January 19, Agrawal sent a company-wide email markedly different and negative compared to other notices of departing executives; many Twitter staff inferred (incorrectly) that Mudge was implicated in serious misconduct:¹⁰⁸

¹⁰⁷ See Exhibit 16

¹⁰⁸ See Exhibit 31

Team,

Today

I committed to all of you that I would take fast and thoughtful actions on the priorities that matter most. Decisions about our leaders and how we are structured are foundational. With that, I wanted to let you all know that Mudge Zatko is no longer at Twitter. I made this decision following an assessment of how the organization was being led and the impact on top priority work. I have the utmost confidence in the leads who have jumped in with full focus, and the teams who do this critical work. I also made a commitment that I'll always be transparent with you - that is very important to me. In this case, while I'm sure you may have questions, the nature of this situation limits what I can share at this time. I

106. During the phone call when Agrawal terminated Mudge, Twitter revoked Mudge's access to his Twitter devices and Twitter systems. Nevertheless, Twitter's Chief Compliance Officer was requesting he work **without salary, entirely from memory**, to produce corrective materials for the Board. Her tone grew increasingly urgent:¹⁰⁹

On Fri, Jan 21, 2022 at 1:21 PM Marianne Fogarty <[REDACTED]@twitter.com> wrote:

Hi Mudge.

Can we set up a time to talk to identify and understand the full breadth of additional concerns/issues and can look into them? The sooner the better. Are you free this afternoon? I'll make myself available at whatever time works for you.

Thank you.

Marianne

On Sun, Jan 23, 2022 at 12:42 PM Marianne Fogarty <[REDACTED]@twitter.com> wrote:

Hello Mudge.

I'm writing to follow up on my email of Friday afternoon asking to speak with you to be sure I understand and can look into any additional concerns you may have. You indicated in email that you have additional concerns, and I understand from Patrick Pichette that you indicated the same to him.

My job is to conduct a thorough, objective investigation and report on it to the Audit Committee, of which Patrick is chair. Given the nature of the issues, here I will also report to the full Risk Committee of which Omid, with whom you spoke last week, is chair. But I can only investigate what has been reported to me, and I cannot complete my investigation knowing there are matters that have not been investigated.

I'm available to speak with you at your earliest convenience. Can we set up time to talk today?

Regards,

Marianne

On Sun, Jan 23, 2022 at 5:31 PM Peiter Zatkó <[REDACTED]@gmail.com> wrote:

Yes. There are additional significant items (with documentation).

Appreciate the Sunday email and request for my time. I am unavailable to speak right now.

¹⁰⁹ See Exhibit 16

On Mon, Jan 24, 2022 at 10:07 AM Marianne Fogarty <[REDACTED]@twitter.com> wrote:
Mudge,

Please let me know if you have time today or tomorrow or any time in the near future to talk about your additional items. I would like to speak with you to understand the concerns so that I can look into them. Please forward copies of any documentation relating to those items to me in anticipation of that conversation and to enable me to begin to investigate.

Thank you.

Regards,
Marianne

On Thu, Jan 27, 2022 at 5:37 PM Marianne Fogarty <[REDACTED]@twitter.com> wrote:
Mudge,

I'm writing to let you know that we are convening a meeting of the Risk Committee. At the meeting we will brief the Risk Committee on the concerns you raised in your January 4, 2022 email to Parag and Dalana, including the context you provided in the meeting with Omid and subsequent emails to Sean Edgett and me. If you have additional concerns that you have not yet raised, please do so as soon as possible, but no later than February 1, so that they can be brought to the attention of the Risk Committee in this meeting. It has been six weeks since the presentation about which you raised concerns, and just over three weeks since our investigation began. We do not feel it is appropriate to wait any longer given that we have provided you with repeated opportunities to share further information.

Thank you.

Marianne

-
107. On February 2, 2022, Mudge sent an email directly to Board member Patrick Pichette laying out many of his concerns.
108. On or about February 14, 2022, after at least 150 hours of unpaid work on his personal computer, Mudge sent a 27-page corrective report to Twitter’s Chief Compliance Officer and two Board Directors. The report is included as an exhibit to this filing.¹¹⁰
109. On or around February 16, on information and belief, the Board Risk Committee held an emergency meeting to consider Mudge’s disclosures. Then on February 18, Agrawal publicly announced he was taking unexpected “paternity leave.”¹¹¹ Agrawal returned to work a few weeks later, before his child was born.

[Disclosure continues next page]

¹¹⁰ See Exhibit 1

¹¹¹ "Twitter CEO Parag Agrawal will take paternity leave after three" 16 Feb. 2022, <https://www.washingtonpost.com/technology/2022/02/16/twitter-ceo-parag-agrawal-paternity-leave/>.

VII. Material Misrepresentations and Omissions, and Other Legal Violations

110. Under SEC rules, it is unlawful for a publicly-traded company to “make any untrue statement of a material fact or to omit to state a material fact” in connection with the purchase or sale of a security.¹¹² Misrepresentations create liability if they occur in SEC filings, or in any other public communications available to investors. Further, it is unlawful “[t]o employ any device, scheme, or artifice to defraud,”¹¹³ or “[t]o engage in any act, practice, or course of business which operates or would operate as a fraud or deceit upon any person”¹¹⁴ that is “in connection with the purchase or sale of any security.”

111. For years, across many public statements and SEC filings, Twitter has made material misrepresentations and omissions, and engaged in acts and practices operating as deceit upon its users and shareholders, regarding security, privacy and integrity.

112. Twitter’s misrepresentations are especially impactful, given that they are directly at issue in Elon Musk’s contemplated takeover of the company. For example, Twitter filed SEC Schedule 14A on May 17, 2022, which noted that:¹¹⁵

The obligations of Parent and Acquisition Sub [*n.b., the acquiring companies controlled by Mr. Musk*] to consummate the merger are subject to the satisfaction or waiver of each of the following additional conditions, any of which may be waived by Parent:

- Twitter having performed and complied in all material respects with the obligations required by the merger agreement to be performed or complied with by it on or prior to the closing of the merger;

¹¹² 17 C.F.R. § 240.10b-5(b).

¹¹³ 17 C.F.R. § 240.10b-5(a).

¹¹⁴ 17 C.F.R. § 240.10b-5(c).

¹¹⁵ Page 14, SEC Schedule 14A, available at

<https://www.sec.gov/Archives/edgar/data/0001418091/000119312522152250/d283119dprem14a.htm>

- the accuracy of the representations and warranties of Twitter in the merger agreement, subject to applicable materiality or other qualifiers, as of the effective time of the merger or the date in respect of which such representation or warranty was specifically made;
- and the absence of any Company Material Adverse Effect (as defined in the section of this proxy statement captioned “The Merger Agreement—Representations and Warranties”) having occurred that is continuing.

113. But as part of the same SEC filing, in the *Agreement and Plan of Merger - Execution Version*, Twitter made fraudulent misrepresentations:¹¹⁶

ARTICLE IV — REPRESENTATIONS AND WARRANTIES OF THE COMPANY [n.b. Twitter, Inc.] ...

[T]he Company hereby represents and warrants to Parent and Acquisition Sub as follows:...

Section 4.5 ... Compliance With Laws. ...Neither the Company nor any of its Subsidiaries is in default or violation of any Law applicable to the Company, any of its Subsidiaries or by which any of their respective properties or assets are bound, except for any such defaults or violations that would not have a Company Material Adverse Effect. Notwithstanding the foregoing, no representation or warranty in Section 4.5(a) or this Section 4.5(b) is made with respect to Company SEC Documents or financial statements, “disclosure controls and procedures” or “internal control over financial reporting,” employee benefits matters, Intellectual Property Rights matters, Tax matters, which are addressed exclusively in Section 4.6 (Company SEC Documents; Financial Statements), Section 4.8 (Disclosure Controls and Procedures), Section 4.12 (Employee Benefit Plans), Section 4.14 (Intellectual Property Rights), Section 4.15 (Taxes), respectively. ...

¹¹⁶ Pages A-20-24, Exhibit A to SEC Schedule 14A, available at <https://www.sec.gov/Archives/edgar/data/0001418091/000119312522152250/d283119dprem14a.htm>

Section 4.6 Company SEC Documents; ... [T]he Company SEC Documents complied in all material respects with the requirements of the Securities Act and the Exchange Act, as the case may be, and the applicable rules and regulations promulgated thereunder, and none of the Company SEC Documents at the time it was filed (or, if amended or supplemented, as of the date of the last amendment or supplement) contained any untrue statement of a material fact or omitted to state any material fact required to be stated therein or necessary to make the statements therein, in light of the circumstances under which they were made, or are to be made, not misleading. ...

Section 4.7 Information Supplied None of the information supplied or to be supplied by or on behalf of the Company or any of its Subsidiaries expressly for inclusion or incorporation by reference in the proxy statement relating to the matters to be submitted to the Company's stockholders at the Company Stockholders' Meeting (such proxy statement and any amendments or supplements thereto, the "Proxy Statement") shall, at the time the Proxy Statement is first mailed to the Company's stockholders and at the time of the Company Stockholders' Meeting to be held in connection with the Merger, contain any untrue statement of material fact or omit to state any material fact required to be stated therein or necessary to make the statements therein, in light of the circumstances under which they were made, not misleading at such applicable time...

Section 4.14 Intellectual Property. (a) Except as would not have a Company Material Adverse Effect, the Company and its Subsidiaries solely and exclusively own all patents, trademarks, trade names, copyrights, Internet domain names, service marks, trade secrets and other intellectual property rights (the "Intellectual Property Rights") purported to be owned by the Company and its Subsidiaries (the "Company Intellectual Property"), free and clear of all Liens, except Permitted Liens.

(b) To the Knowledge of the Company, the conduct of the business of the Company and its Subsidiaries as currently conducted does not infringe, misappropriate or otherwise violate any Intellectual Property Rights of any other Person, except for any such infringement, misappropriation or other violation that would not have a Company Material Adverse Effect. To the Knowledge of the Company, no other

Person is infringing, misappropriating or otherwise violating any Company Intellectual Property, except for any such infringement, misappropriation or other violation as would not have a Company Material Adverse Effect.

The Company and its Subsidiaries are in compliance with all applicable Laws, Contracts to which the Company or its Subsidiaries are bound, and internal- and external-facing policies of the Company or its Subsidiaries, in each case, relating to privacy, data protection, and the collection and use of information that constitutes “personal information” under applicable Laws (“Personal Information”) collected, used or held for use by the Company or its Subsidiaries, except where the failure to be in compliance would not have a Company Material Adverse Effect.

(e)¹¹⁷ Neither the Company nor any of its Subsidiaries has experienced any unauthorized access to the information technology systems owned or used by the Company or its Subsidiaries or Personal Information collected, used, held for use or otherwise processed by the Company or its Subsidiaries, except as would not have a Company Material Adverse Effect.

114. **Intellectual Property:** While several of Twitter’s representations and warranties are untrue, in particular note that the “intellectual property” statements are egregious lies. In fact, Twitter senior leadership have known for years that the company has never held proper licenses to the data sets and/or software used to build some of the key Machine Learning models used to run the service. Litigation by the true owners of the relevant IP could force Twitter to pay massive monetary damages, and/or obtain an injunction putting an end to Twitter’s entire Responsible Machine Learning program and all products derived from it. Either of these scenarios would constitute a “Material Adverse Effect” on the company.

115. Twitter has consistently misrepresented in SEC filings its capacity to recover from even a brief outage of only a few data centers, with misrepresentations *increasing* after the Spring 2021 “Black Swan” event that threatened the platform’s survival:

¹¹⁷ *Sic.* It appears that these sections were mislabelled (a), (b), (e).

a. *“We have implemented a disaster recovery program, which allows us to move production to a back-up data center in the event of a catastrophe... [T]his program is functional...”*¹¹⁸

b. *“We have implemented a disaster recovery program ... this program is functional.”*¹¹⁹

c. These claims are fundamental to any proper valuation of Twitter’s business. And they are significantly misleading; some would consider them plainly false. While Twitter is expanding its data centers with another location, and has been working on migrating non-operational infrastructure to a cloud,¹²⁰ no current program is functional enough to mitigate, or more importantly reliably recover from, the type of “Black Swan” event our client warned Twitter of in his 2021 Q1 board presentation.¹²¹ Furthermore, Twitter failed to report the actual Spring 2021, “Black Swan” event, that threatened Twitter’s survival to the extent that it was the subject of an emergency disclosure to the Board of Directors.

116. Twitter made multiple misrepresentations in its SEC Form 10-K for the fiscal year ended December 31, 2021.¹²²

a. *“We focus on ... measures to help protect the privacy of people on Twitter.”*

i. But when the FTC asked Twitter whether it fully deleted the data of users who left the service, Twitter deliberately misled the FTC by stating those accounts were “deactivated,” even when the data was not fully deleted. And in late 2021, Mudge sent memos to executive team members arguing that, in light of the egregious and ongoing misrepresentations to the FTC, French and Irish regulators, plus the very real possibility of multi-billion dollar fines or even bans from major markets, Privacy should become

¹¹⁸ SEC Form 10-Q for the quarter ended September 30, 2021, available at <https://www.sec.gov/ix?doc=/Archives/edgar/data/0001418091/000141809121000209/twtr-20210930.htm>

¹¹⁹ SEC Form 10-K, Twitter, Inc., Fiscal Year ended December 31, 2021, available at <https://www.sec.gov/Archives/edgar/data/0001418091/000141809122000029/twtr-20211231.htm>

¹²⁰ See Exhibit 3; see also “Partly Cloudy: The start of a journey into the cloud - Twitter Blog.” 8 Apr. 2019, https://blog.twitter.com/engineering/en_us/topics/infrastructure/2019/the-start-of-a-journey-into-the-cloud.

¹²¹ *Id.*

¹²² SEC Form 10-K, Twitter, Inc., Fiscal Year ended December 31, 2021, available at <https://www.sec.gov/Archives/edgar/data/0001418091/000141809122000029/twtr-20211231.htm>

Twitter's #1 priority. But Mr. Agrawal's executive team did not even reply to Mudge's e-mail and they left critical privacy compliance initiatives out of the company's top five priorities.

- ii. Twitter internally knew that the vast majority of data in their systems, data that was supposed to be registered and tagged to enable appropriate privacy handling, was not compliant with privacy requirements under the FTC consent decree and international regulations. The executives also knew that compounding this problem was the speed at which more non-compliant data was being created - thereby growing the body of data for which Twitter was out of compliance.

b. "Our prioritization of the long-term health of our service may adversely impact our short-term operating results."

- i. Contrary to this assertion, Twitter's unwillingness to accept short-term costs is jeopardizing its future. Further, Twitter's bonus structure provided executives with cash incentives for quickly growing the user count, which could only happen if the company ignored or deprioritized privacy, security and platform integrity.
- ii. The 2011 FTC Consent Order and the 2020 FTC Draft Complaint both identified protection of sensitive user data as crucial problems to be addressed. But in the decade since then, things actually got meaningfully worse, with sensitive customer information like emails and phone numbers improperly used for marketing, simultaneously while the company negotiated a new settlement with the FTC in 2020 and 2021.
- iii. As directly stated to the Board by a member of the executive team, it was a conscious choice by Twitter to deprioritize health and integrity of the service in order to direct resources towards mDAU growth. Twitter product release managers were given the authority to override security and privacy issues when choosing whether to make changes or ship a product, and in fact were encouraged to do so. This was reported up to Mudge numerous times, but his efforts to stop it were rebuffed. A good example is the Fleets service,¹²³ which supposedly permitted sending Tweets that would automatically disappear (similar to Snapchat). The Fleets product avoided undergoing security and privacy reviews before

¹²³ https://blog.twitter.com/en_us/topics/product/2021/goodbye-fleets

launch. When serious issues were identified at the last minute, Fleets was launched without addressing them. Shortly after the product launch security and privacy issues were found in the service causing teams to scramble to patch and fix overlooked issues when they impacted real users.¹²⁴

- c. **Omission:** Twitter has made no disclosure at all about the material information security problem that was the basis for the original 2011 FTC complaint, namely (1) the company's inability to limit employee access control to its production systems and (2) its highly anomalous practice of permitting engineers to build, test and deploy new code directly in live-production systems. This omission is material as a reasonable investor would want to know about Twitter's lack of basic engineering hygiene and security exposure.

117. **Misrepresenting the 2020 Hack:** Following the July 2020 hack by teenagers, Twitter provided updates via unsigned blog entries.¹²⁵ Broadly speaking, Twitter drastically overstated the sophistication of the hack, and misrepresented the sophistication of its own defenses. But in several cases, the blog posts went beyond misleading, and constitute outright falsehoods:

¹²⁴<https://www.vice.com/en/article/n7vnab/psa-twitter-doesnt-automatically-delete-your-fleets-after-24-hour>

S

¹²⁵ "An update on our security incident - Twitter Blog." 30 Jul. 2020, https://blog.twitter.com/en_us/topics/company/2020/an-update-on-our-security-incident.

the world that help with account support. Our teams use proprietary tools to help with a variety of support issues as well as to review content in line with [The Twitter Rules \(http://twitter.com/rules\)](http://twitter.com/rules) and respond to reports. Access to these tools is strictly limited and is only granted for valid business reasons. We have zero tolerance for misuse of credentials or tools, actively monitor for misuse, regularly audit permissions, and take immediate action if anyone accesses account information without a valid business reason. While these tools, controls, and processes are constantly being updated and improved, we are taking a hard look at how we can make them even more sophisticated.

Since the attack, we've significantly limited access to our internal tools and systems to ensure ongoing account security while we complete our investigation. As a result, some features (namely, accessing the [Your](#)

We're always investing in increased security protocols, techniques and mechanisms - it's how we work to stay ahead of threats as they evolve. Going forward, we're accelerating several of our pre-existing security workstreams and improvements to our tools. We are also improving our methods for detecting and preventing inappropriate access to our internal systems and prioritizing security work across many of our teams. We will continue to organize ongoing company-wide phishing exercises throughout the year.

118. In particular, based on evidence provided in this disclosure and interpreting these statements the way an investor or user would read them, it is not true that:

- a. "[a]ccess to these tools is strictly limited"
- b. "[w]e have zero tolerance for misuse of credentials or tools"
- c. "[w]e ... regularly audit permissions"
- d. "[w]e...take immediate action if anyone access account information without a valid business reason"
- e. "These tools, controls and processes are constantly being updated and improved"

-
- f. “Since the attack, we’ve significantly limited access to our internal tools and systems”
 - g. “We’re always investing in increased security protocols, techniques and mechanisms”
 - h. “We are also improving our methods for detecting and preventing inappropriate access to our internal systems”
 - i. “We will continue to organize ongoing company-wide phishing exercises throughout the year.” (No phishing exercise was ever reported to Mudge as Security Lead, nor were any new policies or practices implemented based on any phishing exercise; on information and belief, the company never organized a phishing exercise as described here.)

119. **False assurances on security:** A September 24, 2020 blog post by Parag Agrawal and Damien Kiernan also included multiple false assertions:¹²⁶

Our continued work to keep Twitter secure

By

[Parag Agrawal](#)

and

[Damien](#)

Thursday, 24 September 2020

¹²⁶ "Our continued work to keep Twitter secure." 24 Sep. 2020, https://blog.twitter.com/en_us/topics/company/2020/our-continued-work-to-keep-twitter-secure.

To further secure our internal tools from potential misuse, we have been strengthening the rigorous checks that team members with access must undergo. This also helps reduce the potential for an unauthorized person to get access to our systems. We have strict principles around who is allowed access to which tools and at what time, and require specific justifications for customer data to be accessed.

behavior on your account to help you keep it secure, we have internal detection and monitoring tools that help alert us of unusual behavior or possible unauthorized attempts to access our internal tools. These tools are constantly being improved, even since the July incident, to include things like expanding our detection and response efforts to include suspicious authentication and access activity.

Our teams have also been investing in additional penetration testing and scenario planning to help secure Twitter from a range of possible threats, including in the context of the upcoming 2020 US elections. Specifically, over a five month period from March 1 to August 1, Twitter's cross-functional elections team conducted tabletop exercises internally on specific election scenarios. Some of the topics included:

courses, we've also enhanced training content on secure coding, threat modeling, privacy impact assessments, and [privacy by design](https://blog.twitter.com/en_us/topics/company/2019/privacy_data_protection.html) (https://blog.twitter.com/en_us/topics/company/2019/privacy_data_protection.html) so privacy is integrated into everything we design and build by default.

Finally, we continue to invest in and scale the processes in place to review products for security and privacy concerns before they launch. If a project could have significant privacy impacts, we conduct a detailed impact assessment to make sure we're taking appropriate measures before we launch it. We've significantly increased the number of privacy reviews and impact assessments the past few years. Specifically, in

120. In particular, the evidence provided in this disclosure demonstrates that many of these statements are simply false:

-
- a. Twitter did not have “rigorous checks that team members with access must undergo”. Twitter did not perform meaningful vetting for employees with privileged access, nor were team members with access to production systems and data evaluated differently from other employees;

 - b. Twitter did not have “strict principles around who is allowed access to which tools and at what time, and require specific justifications for customer data to be accessed”; according to expert quantification and analysis in January 2022, over half of Twitter’s 8,000-person staff was authorized to access the live production environment and sensitive user data. Twitter lacked the ability to know who accessed systems or data or what they did with it in much of their environment. The ranks of those given access to Twitter's production environment, systems, and services which contained user data continued to increase through the end of 2021. For the vast majority of methods that staff and contractors accessed sensitive data, including user data, there were no time-based limits on access; such limits applied only to a small subset of tools.¹²⁷

¹²⁷ See Exhibit 1, p. 21

At the beginning of 2021, 46% of all FTEs had privileged access to production systems and data. By Q4 2021 this number was 51% of employees. Twitter has grown meaningfully in its number of employees. The percentage of employees with privileged access has increased on top of this.



Access to Twitter's Datacenter Production Environment¹¹

- i. Dec 2020 46% of employees (2,763 out of 5917)
 - ii. Dec 2021 51% of employees (3,995 out of 7714)
- (*) The dip was an unintended (internal) incident

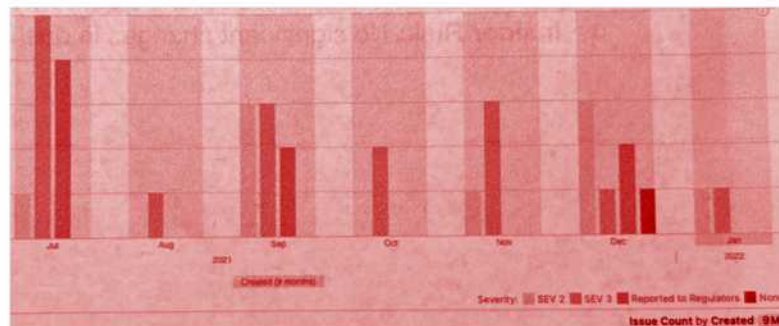
- c. Twitter did not have “internal detection and monitoring tools that help alert us of unusual behavior or possible unauthorized attempts to access our internal tools”; on information and belief they had no meaningful detection or monitoring tools;¹²⁸
- d. It is not true that Twitter “teams have also been investing in additional penetration testing and scenario planning”; Twitter had neither an internal red team nor a third party engaged to do meaningful internal penetration testing within the InfoSec organization, at least none ever reported to Mudge as Security Lead. An excerpt of Mudge’s analysis:¹²⁹

¹²⁸ *Id.*

¹²⁹ *Id.*, p. 16

Twitter has an unacceptable, and near continuous, number of security and privacy incidents. I estimate there were more than 50 Incidents in 2021¹⁵; approximately an incident per week. Based on my professional experience, peer companies do not have this magnitude or volume of incidents.

H2 2021 had 11 Incidents that were required to be reported to regulators, 5 of which happened in Q4.



The Incidents were predominantly related to areas where Twitter has systemic, long lived, problems: 'Access Control' and 'Security Configuration and Bugs'. Together these problems account for more than 80% of the Incidents.

¹⁵ My notes capture 48 incidents in the period of April - November 2021. 50+ is an extrapolation to include January-March, and December, at the same Incident rate, as my data sources were taken away before this document could be completed.

- e. Twitter had not “enhanced training content on secure coding, threat modeling, privacy impact assessments, and privacy by design so privacy is integrated into everything we design and build by default.” Twitter neither followed a mature Software Development Life Cycle nor had one been rolled out across engineering and existing projects and programs. If and when the InfoSec or Privacy teams learned about a project whether from the project manager or through the grapevine, security and privacy reviews often had to be forced into projects. It was further noted that very few of the products submitted for security / privacy review included threat modeling on how the products could

be abused by bad actors. The omission of threat modeling, indicated that engineers had not considered the question of vulnerabilities;¹³⁰

Twitter does not have an industry-appropriate Software Development Life Cycle (SDLC) and Twitter has thus far operated largely without one at all. If it were not for an FTC consent decree, it is possible that Twitter would not be working to put together and deploy an SDLC. This is very atypical in the industry and is a significant risk to the company.

Due to this deficiency, it is inappropriate to label any SDLC progress as “Compliant” to the Committee, as was done in the Information Security documents sent to the Risk Committee. Doing so misrepresents Twitter’s situation as it would be seen by regulators and auditors.

- f. “If a project could have significant privacy impacts, we conduct a detailed impact assessment to make sure we’re taking appropriate measures before we launch it”; but until 2021 Twitter did not employ trained Privacy Engineers. Instead Twitter relied on regular engineers to implement privacy measures without the benefit of guidance from senior Privacy Engineering leadership or people with appropriate domain expertise.

121. Securities Violations: Based on the foregoing, it is likely that Twitter and Mr. Agrawal’s actions constitute violations of multiple SEC rules and regulations:

- a. 15 U.S. Code § 7142 on corporate responsibility for SEC reporting;
- b. 15 U.S. Code § 7262 on management assessment of internal controls.
- c. 18 U.S. Code § 1350(c):

whoever...willfully certifies any statement ...[filed for a public company with the SEC] knowing that the periodic report accompanying the statement does not comport with all the requirements set forth in this section shall be fined not more than \$5,000,000, or imprisoned not more than 20 years, or both;

¹³⁰ *Id.*, p. 15

d. Twitter CEO Parag Agrawal signed the company's 2021 Annual Report:¹³¹

Pursuant to the requirements of the Securities Exchange Act of 1934, this report has been signed below by the following persons on behalf of the registrant and in the capacities and on the dates indicated:		
Signature	Title	Date
<u>/s/ Parag Agrawal</u> Parag Agrawal	Chief Executive Officer and Director (Principal Executive Officer)	February 16, 2022

[Disclosure continues next page]

¹³¹ "twtr-20211231 - SEC.gov."

<https://www.sec.gov/Archives/edgar/data/0001418091/000141809122000029/twtr-20211231.htm>.

VIII. Conclusion

116. For the foregoing reasons, please open an investigation into legal violations by Twitter, Inc.
117. As a senior executive, Mudge was awarded Twitter stock, for which he previously created and has followed without deviation an Automatic Securities Disposition Plan pursuant to SEC rules codified at 17 C.F.R. § 240.10b5-1(c).
118. Whistleblower Aid is a non-profit legal organization that helps workers report their concerns about violations of the law safely, lawfully, and responsibly. We respectfully request the SEC's assistance ensuring that our client never faces retaliation.

Sincerely,



John N. Tye, Founder & Chief Disclosure Officer
Whistleblower Aid

[REDACTED]
[REDACTED]



Debra Katz, Partner
Katz, Marshall & Banks, LLP



Andrew P. Bakaj, Senior Counsel
Whistleblower Aid

[REDACTED]



Alexis Ronickher, Partner
Katz, Marshall & Banks, LLP



Kyle Gardiner, Associate Counsel
Whistleblower Aid

[Redacted]

[Redacted]

Whistleblower Aid

[Redacted]

[Redacted]

[Redacted]



Alia Al-Khatib, Senior Associate
Katz, Marshall & Banks, LLP

Katz, Marshall & Banks, LLP
1718 Connecticut Avenue, NW
Washington, DC 20009
202.299.1140

Exhibits Enclosed:

Exhibit Label	Exhibit Link
1	20220214_Twitter Q4 2021 Risk Committee Issues_redacted.pdf
2	2021xxxx_Alethea_Privileged Confidential_Twitter_Dis-Mis-information-assessment_2021_Q1_redacted
3	202103xx_2021 Q1 board voice.pdf
4	Exhibit 4_202112xx_DRAFT_2021 Q4 Privacy and Data Protection Report
5	20220118_Email_to_Parag_following_Omid_meeting_Jan18_2022_redacted

6	20220202 Mudge 2 Patrick Pichette Feb 2 2022 redacted
7	20220112 12 January 2022 Team Leads Weekly Update with Incidents Charts.pdf
8	202111xx Protect 2022 Strategy
9	20211217_Gmail_action items from Mudge to Parag 1 of 2_redacted
10	20211215 InfoSec Risk Committee Presentation Items to be aware of_redacted
11	20210112_Parag_e-mails_Memorandum for the record_Jan_12_2022_redacted
13	20220118 Gmail - Re Privileged and Confidential - Priority Meeting Request_Jan_18_redacted
14	Exhibit 14 202112xx DRAFT NOT DELIVERED 2021 Q4 Information Security Report
15	202112xx Q4 Privacy and Data Protection Risk Report_redacted
16	20220127_Marianne_continued_requests_for_work_Jan_27_2022_redacted
17	202112xx Copy of access control actual.jpg202112xx Access control actual
19	202107xx - 202212xx July 2021 to December 2022 Board Risk Committee Meeting Schedule.pdf
20	202112xx Insider Risk December 2021
21	2021XXXX Snapshot of data center security system deficiencies_redacted
22	Exhibit 22 202112xx DRAFT Active dashboards showing security discrepancies on work computers - reported to board
23	202106xx_Q2 board meeting - Mudge Board voice.pdf
24	20210816 August 16, 2021 Re_ [Inform] New York Times story on company shift_redacted
25	20220127_email draft Marianne 6 weeks v 1 [was_Re_Follow-Up]_redacted

26	20211216_Q4 Board Meeting Regulatory Dashboard Public Policy_Risk Update_redacted
27	20220119_email_Follow-Up_redacted
28	202112xx_Protective details Parag 2021 family holiday_redacted
29	20210828_August 28, 2021 Re Health Objectives #1 and #2, and partnership with TwS_redacted
30	20220119_Message to Mudge's team ahead of company-wide email
31	20220119_Parag - company-wide email on Mudge termination_redacted
32	20220111_Handwritten executive meeting notes_redacted
33	20220118_Omid Risk Surprise part2 thought it was about XXXX offboarding_redacted
34	20220214_Fwd_Q4 2021 Risk Committee Issues_redacted
35	20220225_Marianne_Audit_Closed_redacted
36	20220119_Letter to M. Zaitko 19Jan2022_redacted
37	220210914_September 14, 2021 Fwd [Inform] Verification Operations Update_redacted
38	202201xx_Q4 Risk Committee Information Security Corrective Document and Timeline.docx_redacted
39	Excerpt on Licensing
40	20211001_Mudge Twitter Nigeria Notes_redacted
41	20211213_emailER_12_13_2021
42	202112xx_NEW_Corrected Risk Information.docx

— END OF DISCLOSURE —

— PROTECTED & SENSITIVE WHISTLEBLOWER DISCLOSURE —

Issues and Objections Regarding Twitter InfoSec Information and the Q4 2021 Twitter Risk Committee

Peiter Mudge Zatzko

Latest Revision: February 2022

Overview	2
Introduction	2
Part 1: The Problem	2
Incorrect and Misleading InfoSec Information: Q4 Risk Committee	2
Identification of the issues in 2021	3
What is Twitter's accurate InfoSec situation, as of Q4 2021?	6
This did not happen overnight	7
Part 2: A More Truthful Q4 Description of Twitter Key Information Security Risks	8
Overview	8
4 Areas of Critical Focus	9
Access Control	9
Access Control and Insider Risk	11
Security Patches and Software Configuration and Versions	12
Client fleet (laptops)	12
Servers (data center)	13
Processes and Compliance	15
Incidents	16
Part 3: Inaccuracies in InfoSec Materials Presented Q4 2021	18
The Deck	18
Access Control	18
SDLC, Security Reviews, Privacy Reviews	21
Patches and Software Configuration and Versions Compliance	23
A note on Zero Trust and Endpoints (Employee computers)	25
Incidents and Incident Classes	25

Overview

Introduction

This document describes (Part 1) events leading up to the transmission of inappropriate information to the risk Committee over objections of the Twitter Security Lead. It then provides a (Part 2) Replacement InfoSec Risk Report on Top Risks. The document closes with (Part 3) select descriptions of inaccuracies and misrepresentations in the materials that the CEO instructed be presented to the Q4 Twitter Risk Committee. Electronic records support this description of events and the information contained herein.

Part 1: The Problem

Incorrect and Misleading InfoSec Information: Q4 Risk Committee

In December 2021 the Risk Committee received information about Twitter's information security posture that is inaccurate and misleading. It appears that Twitter's information security environment has not been accurately characterized to the Board of Directors and Risk Committees dating back to before my tenure. This disconnect may exist elsewhere¹ but the focus of this document is on Information Security.

It is critical that the Board have accurate and truthful views into Twitter's InfoSec issues and posture. The Board needs this accurate information so they can take corrective actions and ensure reports to regulatory and other bodies are accurate. It is crucial that the Board is not misinformed as Twitter works to comply with the existing FTC consent decree, faces possible new violations related to mis-representations in ongoing consent

¹ There was also a disconnect between Twitter's stated Privacy posture and the reality of Twitter's privacy issues. However, by removing Privacy Engineering from Information Security and through the work of myself, [REDACTED], and the new team, this disconnect has been significantly improved. Several ICs have volunteered to provide questions that members of the Board should ask to determine where else this disconnect and misrepresentation may, or may not, exist in Twitter.

negotiations, new demands from other regulators², and moves into more regulated environments³. If the Board is misinformed, representations and statements to the outside world will be inaccurate as well. This would impact Twitter users (customers) and shareholders.

One of the reasons I was hired was to evaluate Twitter's information security environment and provide an accurate assessment. There were concerns that Twitter had serious problems in these areas, which threatened its data security and integrity in its industry. With my domain expertise I was to look into these concerns and make the accurate and truthful state of Twitter's security posture known to the executive team, and the Board, and work to put Twitter on the right path forward.

Identification of the issues in 2021

██████████, ██████████, ██████████, ██████████, ██████████. I raised concerns to Human Resources Business Partners (HRBPs) and I spoke to ██████████ numerous times about my concerns regarding accurate assessments of our security and the necessity to report accurately to the board and its subcommittees on security gaps and where improvement was needed. ██████████ walled me off from Twitter Information Security data and those who were in ██████████ organization. ██████████ actively resisted my direction. ██████████ prepared a report to the Q4 Risk Committee, which report was a whitewash of security problems at Twitter and overstated the improvements made to security. This report contained inaccurate and false statements which I believe were intended to mislead the Board. My efforts to rein in ██████████, whom I viewed as neither sufficiently competent at ██████████ position nor forthright, were opposed by Human Resources due to Twitter's fear of a lawsuit based upon ██████████ many allegations, which were all investigated by Employee Relations and found to be unsubstantiated. I believe these were made to deflect criticisms of ██████████ work.

I concluded in October 2021 that ██████████ should be offboarded due to performance issues. Human Resources informed me that the documentation and communications between myself and ██████████ and the documentation of ██████████ failure to perform met the requirements to offboard ██████████ and that it would be done. At that point Employee Relations was permitted to share with me that they had received numerous other complaints about ██████████ inappropriate behavior.

² Irish DPC, French CNIL

³ E.g. Money Transmission Licenses (MTL)

██████████ offboarding was then repeatedly blocked, including directly by Parag Agrawal. Had ██████ offboarding occurred when approved, going back to October 2021, ██████ report would not have been released to the Q4 Risk Committee.

When I brought to Mr. Agrawal's attention the fact that ██████████ report was misleading, inaccurate and intentionally wrong, he overruled me and overruled my recommendation that ██████ report be rewritten to make it accurate. Mr. Agrawal ordered me to permit ██████████ report to the Q4 Risk Committee's December 16, 2021 meeting with the instruction that I walk back the many and fundamental inaccuracies and falsehoods contained in ██████████ report after ██████ presented it to the Committee. After registering my concerns about this approach on the record, I followed Mr. Agrawal's instructions as best I could. Mr. Agrawal committed to assist after-the-fact in correcting the record. However, after the Q4 Risk Committee meeting, Mr. Agrawal expressed disappointment that I had not completely walked back the report. What he refused to recognize was that walking back a report that instead needed to be repudiated was an impossible task. In essence, he chided me for not telling the Q4 Risk Committee to completely disregard the report submitted to it.

In addition to e-mails that I sent to Mr. Agrawal, and to the head of HR, prior to the Risk Committee meeting, and those immediately following the meeting, I continued to communicate that the situation was not resolved. In receipt of one of my emails, January 4, 2022, where I repeated concerns about the false representations that were made to the Q4 Risk Committee at its December 16, 2021 meeting, Mr. Agrawal replied (January 6, 2022) that he was "surprised" by the allegations I made. These allegations were merely a recapitulation of my prior complaints and concerns.

My reporting of false information triggered an Audit investigation regarding misrepresentations and false statements being made to this subcommittee of the Board. Approximately two weeks later, on January 19, 2022, my employment was terminated. One day prior to my termination Mr. Agrawal held a surprise meeting that included the Head of the Risk Committee. At that meeting Mr. Agrawal falsely stated that he had ordered me to redo ██████████ report a month prior and was still waiting for the corrected information. The electronic record will verify my recollections over his.

The termination of my employment appears to be in direct violation of New Jersey's whistleblower protections (Conscientious Employee Protection Act - CEPA, N.J.S.A 34:19-1, et. seq.). Regardless of my unlawful termination, I *still* view it as crucial that Twitter's accurate and truthful measurements of Twitter's security risks and posture be correctly conveyed to the Risk Committee and the Board.

Below is the accurate assessment of the state of Twitter's current security, as should have been conveyed to the Q4 Risk Committee of the Board.

– Twitter is grossly negligent in several areas of information security.

– If these problems are not corrected, regulators, media, and users of the platform will be shocked when they inevitably learn about Twitter's severe lack of security basics. They will lose confidence in Twitter and this will have real world impact to the platform and to the company.

– Regulators, when evaluating Twitter, *will* identify these as **systemic issues**. They will likely levy new fines and/or increase existing fines. They will also impose constraints and requirements on how Twitter operates, constraining Twitter's freedom to choose how it executes in various areas of engineering and what Twitter chooses as priorities. Further, Twitter may be precluded from conducting business in certain markets.

There are 4 critical areas that have not been accurately represented to the Board:

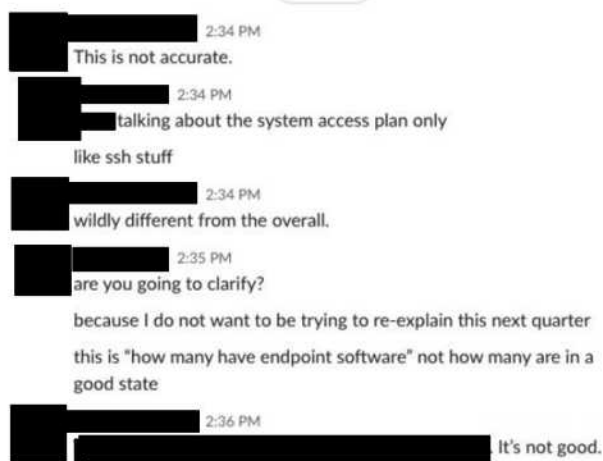
- Out-of-date software and the lack of basic security configuration in existing software (Software and Security Versions/Configurations/Patches)
- Gross problems around access control to systems and data (Access Control)
- Lack of basic processes and compliance such as software development lifecycles, line-managers being allowed to unilaterally overrule security and privacy findings, and a prioritization of running products with known violations over compliance with regulatory requirements⁴ (Processes and Compliance)
- A volume and frequency of security incidents impacting a large number of users' data that is frankly stunning (Incidents)

Before I provide an accurate description of Twitter's Information Security and a critique of the data that was provided to the Risk Committee, it is worth pointing out that other senior people also started to identify that inappropriate information was being presented.

⁴ A recent example includes Twitter choosing not to comply with regulatory requirements, even though it could, until it could optimize more profit within a single region. (French CNIL)

Listening to ██████ abbreviated verbal presentation to the Q4 Risk Committee, both the Twitter Chief Privacy Officer and Twitter's Distinguished Privacy Engineer (the highest engineer rank, equivalent to a VP) objected to what they heard.

██████ made statements about access control at Twitter and then about endpoint (employee computers) security health. The following was an unsolicited live response to what was said:



2:34 PM
This is not accurate.

2:34 PM
talking about the system access plan only
like ssh stuff

2:34 PM
wildly different from the overall.

2:35 PM
are you going to clarify?

because I do not want to be trying to re-explain this next quarter
this is "how many have endpoint software" not how many are in a
good state

2:36 PM
It's not good.

Slack messages from the Chief Privacy Officer and Distinguished Privacy Engineer refuting statements made by ██████ about Access Control and Endpoint security (in the above messages the topic changes from accuracy of Access Control to accuracy of Endpoint security at "this is 'how many...").

What is Twitter's accurate InfoSec situation, as of Q4 2021?

Twitter is **very** far behind the industry in key areas of Access Control, Software and Security Patches/Configuration/Versions, and Processes and Compliance. This is evidenced in the volume and frequency of Incidents. In more than one of these areas Twitter is a decade behind peers such as Google and Facebook.

Some newsworthy highlights are that more than half of Twitter's 500,000 servers are running out-of-date Operating Systems so out of date that many do not support basic privacy and security features and lack vendor support⁵. More than a quarter of the ~10,000 employee computers have software updates disabled! More than half of Twitter employees have access to Twitter's production environment – unheard of in a company the age and importance of Twitter, where nearly all employees have access to systems

⁵ E.g. encryption at rest, kernels, updates, etc.

or data they should not. At Twitter engineers work on live data when building and testing software because Twitter lacks testing and staging environments; work is instead conducted in production and with live data. With this understanding, it is somewhat less surprising that frequent security incidents are so commonplace at Twitter that more than one per week, on average, occurred in 2021 and were determined to involve millions of people's accounts/data. This additionally provides plausible explanations for some of the numerous platform disruptions, as engineering errors that happen during testing occur in, and impact, production.

This did not happen overnight

To get where Twitter is today took more than just a lack of prioritization on areas of information security and privacy across the past year. This took many years. To get to Twitter's current state of insecurity required repeated downplaying of problems, selective reporting, and leadership ignorance around basic security expectations and practices.

Part 2: A More Truthful Q4 Description of Twitter Key Information Security Risks

Overview

Twitter is nearly a decade behind the industry (and peers) in access control. It is significantly behind publicly traded companies in keeping servers and clients up to date with software and patches. Twitter lacks necessary visibility into networks and systems that it needs to state confidently whether identified security problems have been remediated to the extent necessary, and Twitter experiences an outsized frequency and volume of incidents above that of other companies⁶. The incidents are unsurprisingly rooted in the areas of greatest deficiency.

Internal dashboards show 30% of the ~10,000 employee computers reporting that they are not correctly configured to accept software updates⁷. 60% of servers in the Twitter data centers are running out of date (even unsupported) Operating Systems⁸. More than 50% of full time employees have access to Twitter's production environment⁹ because Twitter does not have appropriate development and testing isolation. And Twitter dealt with more than ~50¹⁰ incidents in the past year stemming primarily from systemic areas of risk such as access control.

These truthful views presented here were obscured by Twitter's bias towards presenting individual "wins" to the Board without larger context. Such misrepresentation creates an impression that improvements have been made when in fact core problems have grown. This is a habit that appears to be longstanding within Twitter and is not dissuaded at the senior, and executive, levels.

What the Risk Committee should look for: presentations to executives, and from executives, containing numbers without context. Such information is difficult to interpret. Numerators without denominators lack important context, and as a result are often misleading. The first question should be "out of how many?", followed by "is this

⁶ This statement comes from my experience at Google, Stripe, Motorola, and InfoSec visibility across dozens and dozens of corporate environments during my 30+ year career.

⁷ Uptyx

⁸ Platform Engineering Dashboards - Kernel Compliance/Non-Compliance and Software Compliance/Non-Compliance

⁹ The source for this statement comes directly from Twitter's LDAP server. This server contains employee access rights to systems and resources.

¹⁰ My notes capture 48 formal security incidents from April through November 2021. Additional incidents during {Jan, Feb, Mar, Dec} 2021 are extrapolated.

enough? Where should we be? What trends were revealed and what was the value? How much did a contextless number, such as a number of security reviews lacking the number of projects needing security reviews, interrupt or delay business efforts and cost resources such as time and headcount to the company and projects?”. We will discuss this problem further in the third section of this document where we discuss specific inappropriate representations in the Information Security presentation.

4 Areas of Critical Focus

There are **4 critical areas** of information security that ***have to be the focus*** of the Twitter Information Security Team and tracked by the Risk Committee: *Access Control, Processes and (Regulatory) Compliance, Patches and Software Configurations and Versions, and Incidents*. Unless something exceptional happens elsewhere, the focus of reporting to the Risk Committee should not be pulled away from these fundamentals until they are addressed. These areas, it turns out, have not been accurately described in the past.

Access Control

Access Control - Twitter is an outlier in this area of risk and not in a good way. Most companies work to restrict access to production systems to only a small handful of people because production systems contain extremely sensitive data and issues in production directly impact customers. Engineers at other companies do their work in testing and staging environments, strongly isolated from the crown jewels of production systems that provide the actual running service. By comparison, Twitter engineers and developers perform work directly in production or interfacing directly with production systems and data. Twitter is where Google was prior to 2005-2007, when they identified and addressed the key issue of removing broad employee access to their production environment. Most companies recognize the risks of outages, sensitive data access, and maintaining the integrity of their platform, and intentionally remove almost all direct interaction with production systems and prevent engineers and employees from having access to production data.

Contrary to what may have been heard or read: *Twitter's access control risk is growing, not shrinking.*

At the beginning of 2021, 46% of all FTEs had privileged access to production systems and data. By Q4 2021 this number was 51% of employees. Twitter has grown meaningfully in its number of employees. The percentage of employees with privileged access has increased on top of this.



Access to Twitter's Datacenter Production Environment¹¹

- i. Dec 2020 46% of employees (2,763 out of 5917)
- ii. Dec 2021 51% of employees (3,995 out of 7714)

(*) The dip was an unintended (internal) incident

Companies following basic security principles do not allow this type of access to production systems and live data. Companies have long ago learned to separate production systems and data from testing and staging environments. Twitter does not do this. Twitter's level of exposure and risk in this area far exceeds the industry.

There are smaller pockets of FTEs who are bestowed even further sensitive access. For instance, they can power on/off computers in the data center or they can perform administrative access on servers in Twitter's data centers. It is important to remove or reduce the access these groups have but they are also edge cases. They are a subset of the primary exposure.

Why should these edge cases not be the sole focus? Consider the understood way to predict the likelihood of an unwanted event from a specific risk, like access control. Assumes a random compromise (or malicious behavior) of a Twitter FTE's account. There is a 1 in 2 chance that the compromise happens to one of the 3,995 out of 7,714

¹¹ It is important to note that this access issue impacts international expansion and operation. Any country where Twitter has an engineer, there is access to production systems and data. Other companies have development, staging, and test environments to mitigate this risk, but Twitter has none of these.

accounts with privileges providing access to production and production data. Now consider the risk for a group of FTE accounts with further access within production. There are ~300 such accounts with god-mode access in Twitter. These represent a 1 in 33 (~300 out of 7,714) chance that the random compromise lands within this edge case.

This is not to say the edge cases are unimportant. But, removing these edge cases without a plan and processes in place to allow work to be done outside of production will prove to only be temporary improvements. This was seen in the repeated focus on reduction of this exact edge case of access control at Twitter. Without a plan for being able to reduce broad access the reduction of this small subset almost immediately began re-growing after the initial reduction. The people whose access were reduced needed it back to do their jobs. This subset reduction is discussed in more detail in section three because this subset reduction was presented, without important context, and stated in a way to imply the larger production problem was being solved when it was not.

Access Control and Insider Risk

There are several known insider threats (KNITs) at Twitter¹². Because of the ubiquitous access to production systems and/or data and the lack of isolation environments and logging, this risk is significant. Combine this with ~30 offboardings per week, each of which represent periods of enhanced concern for insider threat, and the lack of access control and ubiquitous access grants are critical problems.



¹² Referenced Q3 Risk

This chart shows the beginning of Insider Risk Tracking (Offboarding pace via JIRA is about 30/week; most of which are not adequately tracked for insider risk) - this is an improvement from Twitter's lack of Insider Risk abilities in Q1 2021. This chart, and effort, is primarily maintained and run in the Corporate Security Organization.

Twitter has limited ability to effectively constrain and mitigate insider risk without having mature access control and a separation of sensitive data and systems beyond what Twitter presently possesses. Correct access control is also a critical path item for privacy. It is required for regulatory and compliance and to meet expectations and representations made to users and the public. As will be shown below, Twitter systems and servers lack basic security compliance, making this even worse.

Contrary to what ██████ told the Risk Committee in the December meeting, at present there is not an agreed-upon plan to address the broad access control issue. In place of a plan there is a *goals-focused* document within the Twitter Information Security organization but Engineering, Privacy, and IT have not signed off on the approach and there remain significant questions around the feasibility of Information Security's understanding and approach around the effort.

Security Patches and Software Configuration and Versions

Client fleet (laptops)

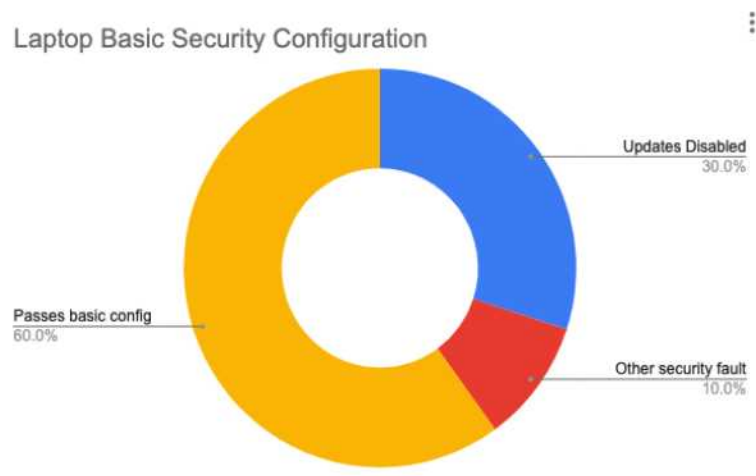
When ██████ told the Risk Committee that nearly all Twitter endpoints (laptops) have security software installed, this statement lacked the following critical context:

Almost 40% of these ~10,000 employee computers (aka endpoint systems aka the client fleet) are not in compliance with basic security settings.

30% of the total endpoint systems report that they *do not have automatic updates enabled*.

These are the systems used to access Twitter's source code, internal systems, and sensitive data. It turns out that this unacceptable state of security on employee systems has not improved from Q1 2021. Throughout 2021 ██████ made numerous statements and references to myself, and others, that the client fleet was in good shape. After all, ██████ stated, most laptops had some security software installed. What the security software reported was either not understood by ██████ or it was understood

but it was chosen not to be made a priority focus. The following is what the software revealed.



Specific details and numbers can be seen on Twitter's Uptyx dashboards¹³

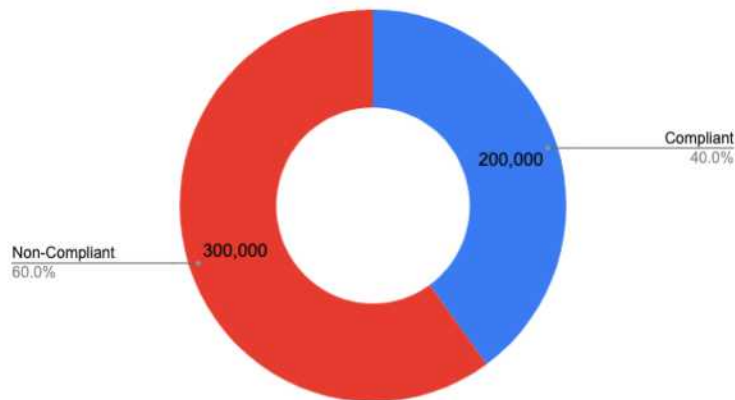
██████████ first actions should be identifying, reporting, and addressing this problem. With focus, it is a 3-5 month project to reach acceptable, and maintained, hygiene. This omission was discovered in Q3 and was not addressed even after repeatedly being brought to the attention of ██████████ and ██████████'s appropriate direct report.

Servers (data center)

Of the approximately 500,000 servers in Twitter data centers, ~60% of them are running outdated Operating Systems and, therefore, are non-compliant even with Twitter's own Engineering standards. In addition to security concerns around outdated software components, many of these outdated OSES are not supported by the vendor. They are also not capable of supporting encryption at rest, a critical compliance and Privacy obligations.

¹³ These numbers are rounded due to my direct data access being removed.

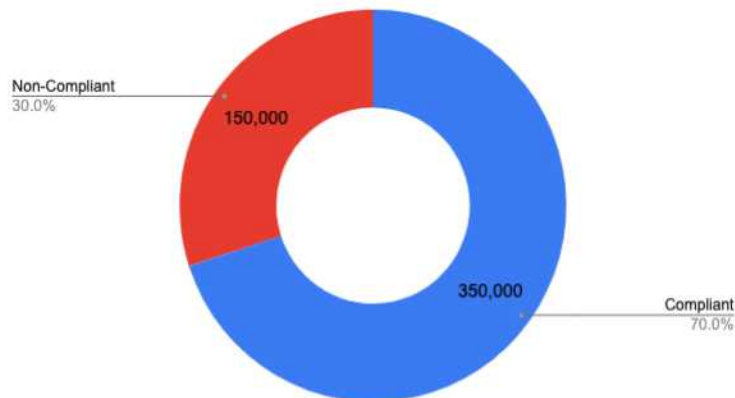
Data Center OS Kernel Compliance



While there is a project underway to address Operating System version (kernel) compliance, it has been reportedly long running (multiple years). Progress has not been significant. The project needs to be revisited, revised, re-staffed, and prioritized with clear goals and visible tracking reported to the Risk Committee¹⁴.

On the same engineering dashboards as above it is revealed that ~30% of the software packages running on the ~500,000 data center systems are non-compliant (out of date or need patching).

Data Center Software Compliance



In addition to security and privacy issues, any engineering outage or security event that revealed that the majority of Twitter's production systems are running out of date, and even unsupported, software would likely result in a significant distraction to the

¹⁴ This project appears to have not been appropriately prioritized, and lack appropriate execution plans, while InfoSec [REDACTED].

company. External pressure would be placed upon Twitter to prioritize the addressing of this shortcoming above many #Participation and #Durability efforts.

Both of these situations have been in the same state for the past 12 months.

Processes and Compliance

This topic refers to regulatory obligations and requirements (e.g. SDLC, security reviews, privacy reviews, FTC consent items, regulatory misrepresentations, etc.).

Twitter does not have an industry-appropriate Software Development Life Cycle (SDLC) and Twitter has thus far operated largely without one at all. If it were not for an FTC consent decree, it is possible that Twitter would not be working to put together and deploy an SDLC. This is very atypical in the industry and is a significant risk to the company.

Due to this deficiency, it is inappropriate to label any SDLC progress as “Compliant” to the Committee, as was done in the Information Security documents sent to the Risk Committee. Doing so misrepresents Twitter’s situation as it would be seen by regulators and auditors.

Twitter is in the process of rolling out a registration and SDLC-capable skeleton framework called Flyway. This initial effort, which lacks integration with security reviews, privacy reviews, and other SDLC checks and balances, is unlikely to be viewed as “Compliant” by auditors and investigators.

Making things more challenging, Twitter lacks the ability to provide a count of total software projects (denominator). This means that when someone says a numerator, say “30 projects did privacy reviews”, it’s difficult to know if this is good or bad - is it 30 out of 35 or 3000? The number occasionally being used for context is a count of projects found in the Unified Priority List (UPL). The UPL is a list that at present only represents engineering (i.e. does not include efforts from Site Integrity, Content Moderation, Privacy, InfoSec, IT, Sales, etc.). The UPL further represents only a subset of engineering efforts and does not include day-to-day-running-of-the business software work.

Senior engineers estimate that the UPL represents only a small fraction of the projects that need to adopt a SDLC. To make the number of projects in the UPL even more

problematic as a context value, in 2022 the UPL is intended to change to reflect an even more limited subset of projects than it presently does.

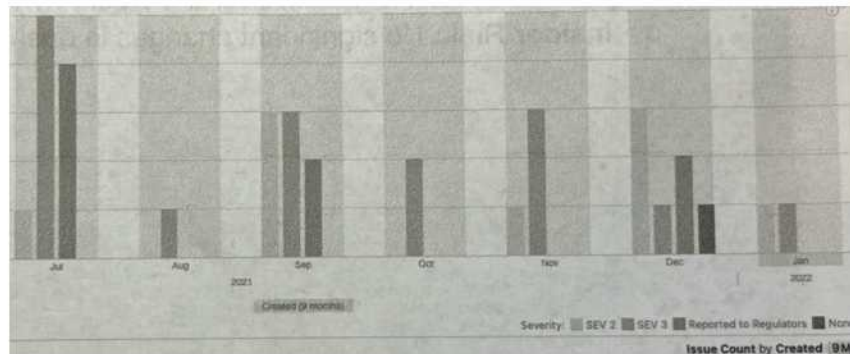
SDLC, security reviews, and privacy reviews, need to be described in terms of whether they would be deemed appropriate by regulators and auditors and in context of how many projects are utilizing these versus how many projects **must** be utilizing them. The UPL is not an appropriate source for total project count.

The Risk Committee members should be aware that *the Twitter SDLC work is not yet what auditors would consider an SDLC, and that security reviews and privacy reviews are not coupled to the SDLC.*

Incidents

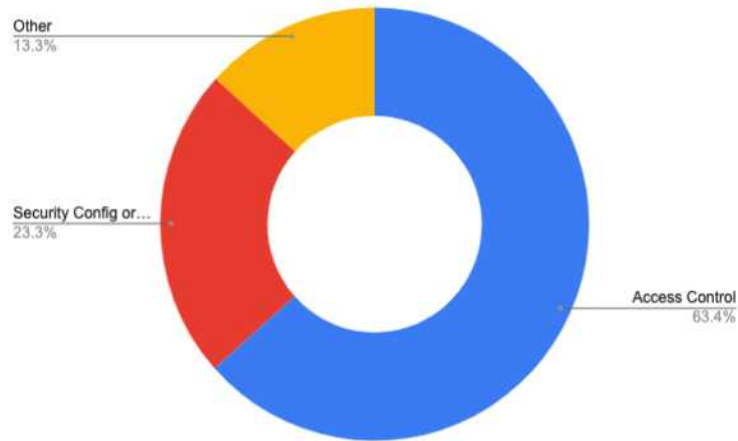
Twitter has an unacceptable, and near continuous, number of security and privacy incidents. I estimate there were more than 50 Incidents in 2021¹⁵; approximately an incident per week. Based on my professional experience, peer companies do not have this magnitude or volume of incidents.

H2 2021 had 11 Incidents that were required to be reported to regulators, 5 of which happened in Q4.



The Incidents were predominantly related to areas where Twitter has systemic, long lived, problems: 'Access Control' and 'Security Configuration and Bugs'. Together these problems account for more than 80% of the Incidents.

¹⁵ My notes capture 48 incidents in the period of April - November 2021. 50+ is an extrapolation to include January-March, and December, at the same Incident rate, as my data sources were taken away before this document could be completed.



Meaningfully improving in the 3 areas above (Access Control, Patches and Software Configuration and Versions, and Processes and Compliance) would logically lead the number of incidents to decrease.

Twitter should be experiencing less than 1 regulator-level incident per quarter. Progress in the other areas mentioned, all leading indicators, will drive improvement in the critical lagging indicator of Incidents.

Part 3: Inaccuracies in InfoSec Materials Presented Q4 2021

The Deck

I identified numerous issues in the materials created by [REDACTED] and put forward to the Q4 2021 Risk Committee. I suggested to Mr. Agrawal that I create a corrected replacement deck, including data points. Mr. Agrawal, as CEO, directed that I not create a corrective document and that I send the objectionable deck forward to the Committee. These events are discussed above.

This part of the document examines important inaccuracies and misleading information present in the deck sent to the Q4 Risk Committee.

There were 11 slides in the deck. Two slides were blank. Six slides were qualitative, aspirational, or otherwise did not meaningfully and quantitatively reference critical risks. While there are questions about the appropriateness of these slides for this setting, they are not the focus of this discussion. Three slides included statistics and measurements intended to represent Twitter's environment and key risks. This document now focuses on problems with the data presented in those three slides and articulated at the Committee meeting.

Access Control

Slide 3 and slide 7 inappropriately represented Twitter's access control risk.

Protect Systems and Data

Tweeps with Direct Access to All Production Servers

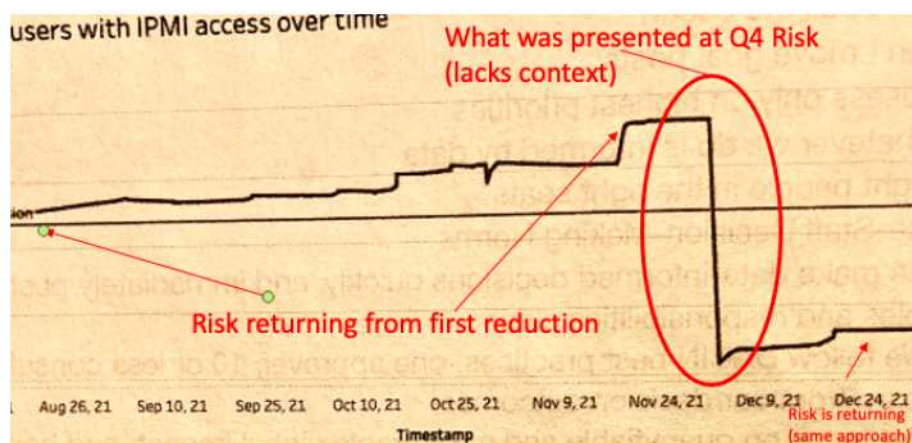


This graphic leads someone to believe significant progress has been made in reducing Twitter access control risk at large. It also implies, with the dotted line, that the reduction is permanent and will continue.

The reduction here focuses on edge cases of access control. These reductions had been made at the end of July and were identified to be temporary in impact, shortly thereafter growing back towards the initial risk.

This graphic represents only 300 users out of the 3,995 users that have production access at Twitter. This represents 5% of the FTEs instead of the 51% needing to be addressed.

The graphic omits that this specific improvement had been previously attempted in July-August. Because there were no systems and solutions in place to enable employees to safely complete their jobs without having the risky access, employees needed to request credentials back to do their work. The Strategy and Operations person overseeing this second reduction, performed again only weeks before the Board and Risk Committee and therefore giving the perception of a recent win, confirmed there were no meaningful changes in the approach this second time that would prevent the re-growth of the risk.



Slide 7 also contains the following statement about access control reduction. The statement is misleading.

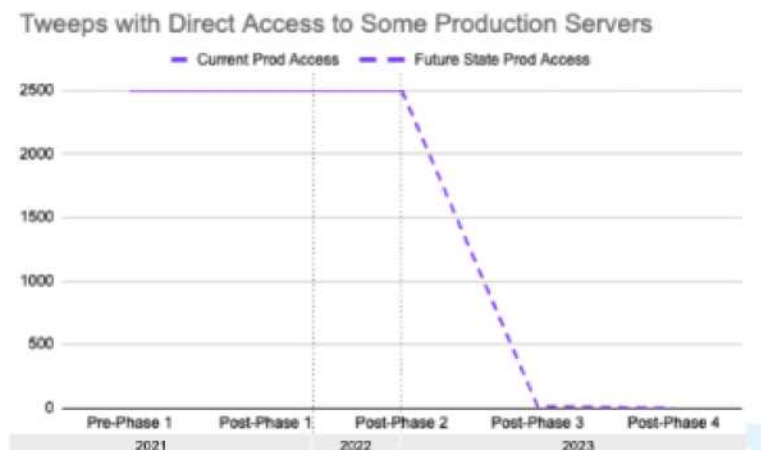
Considerations / Next Steps:

Enterprise Identity and Access Management:

- Reduced extraordinary fleetwide Production level access by 66% (up from 34% in Q3)
 - This is part of a much larger effort to reduce access to both fleetwide and partial fleet production access. That effort is focused on reducing the use cases needed for direct production access, and also providing better ways to provide this access in only emergency situations through JIT and other solutions.

The statement above again implies fleetwide Production level access was reduced by 66%. This is not the case. The reduction being described is the subset reduction discussed above. The word “extraordinary” is used to refer to the subset of edge-case FTEs. Without further clarification this is easily misinterpreted and misleading. This is inappropriate to present to the Committee as it stands.

The graphic in the top left area of slide 3, and recreated again bottom left on slide 7, shows a chart intending to reflect the larger issue of broad access to production access to systems and data throughout Twitter FTEs. It is misleading and the data incorrect.



As a reminder, actual fleetwide Production access grew from 46% of FTEs to 51% in 2021. Any Twitter engineer in any country is presently provided direct access to production systems. The accesses to these production systems are not audited.

Next, the dotted line is aspirational and without evidence or existing proofs of working approaches to back it up. In [redacted] oral presentation, [redacted] stated that there was a plan in place to address the larger Access Control risk and that the plan was already

underway. This is incorrect. Engineering, Privacy, and IT (all stakeholders) have expressed significant concern and disagreement over ideas and approaches brought up by InfoSec. InfoSec has not been working collaboratively with stakeholders and there is not an agreed-upon plan¹⁶. The risk is not flat, as portrayed in the graphic above, but rather the risk is meaningfully increasing.

The following image is an accurate depiction of this access control issue at Twitter.



- Access to Twitter's Datacenter Production Environment¹⁷
- iii. Dec 2020 46% of employees (2,763 out of 5917)
 - iv. Dec 2021 51% of employees (3,995 out of 7714)
- (*) The dip was an unintended (internal) incident

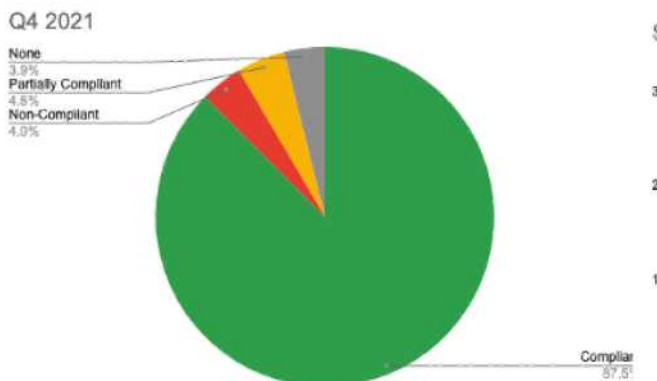
SDLC, Security Reviews, Privacy Reviews

Slide 3 top right and Slide 8 - Infographics on Processes and Compliance (e.g. SDLC and Security Reviews)

¹⁶ There is a *goal/s*-type document but it lacks details and it too is contentious among stakeholders for agreement.

¹⁷ It is important to note that this access issue impacts international expansion and operation. In any country where Twitter has an engineer, there is access to production systems and data. Twitter lacks development, staging, and test environments, so they can't mitigate this risk the way other companies do.

SDLC Flyway Total Compliance



This graphic shows adoption of an internal project related to regulatory obligations. It does not show the amount of regulatory *compliance* reached. The term “Compliance” is misleading and inappropriate as a label. The adoption represents a registration-skeleton and is only for a subset of projects at Twitter. See section 2 of this larger document for more details.



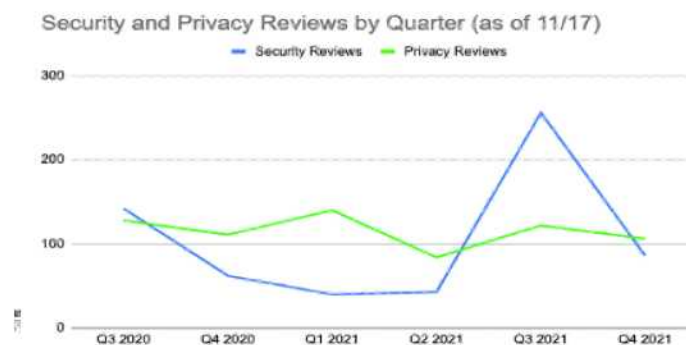
The UPL Histogram and Security Review Sparkline are not relevant together

The top right section of slide 3 contains a graphic about security reviews and a histogram of Unified Priority List projects. It is unclear why these two sets of data, Security Reviews and UPL projects, are overlaid.

As [REDACTED] explained when asked what the relationship was between the two datasets during an earlier review of materials, “*there is no correlation*”. The histogram and time series are not connected or related.

This graphic should not have been presented as it can imply a relationship that does not exist. Neither the UPL bars nor the security review line provide appropriate context. They are both numerators without denominators. See section 2 of this document for more details.

The next graphic, security reviews and privacy reviews, is again without context. At best it represents *some amount of work* having been performed but without context. The Y axis only states the number of reviews performed, not how many needed reviews. Similarly there is nothing conveying what was found, themes identified, review targets to hit, or what costs were to the projects and the business.



Security and Privacy Reviews: how many total projects are there and how many are getting security reviews? What are the security reviews finding? Are they worth the cost and how much do they slow down or impact projects and launches?

Patches and Software Configuration and Versions Compliance

Disturbingly absent from the Q4 Risk Committee InfoSec report was information on the state of client and server software compliance, patching, and configuration.

This is a critical area that the FTC has messaged they will evaluate and it is firmly within their domain to do so. As Twitter is in negotiations with the FTC for prior transgressions, where the FTC is now interested in the baseline of security hygiene at Twitter, this is an area of Twitter's the FTC will likely scrutinize. The fact that Twitter is so significantly lacking in this basic security hygiene and practice will contribute to any FTC decisions and enhance penalties.

Only one item in the report presented to the Risk Committee was related to endpoint (employee laptops):

Threat Intelligence:

- Carbon Black Cloud - CBC is now running on 9.7K endpoints
 - CBC replaces the version we were on previously (CarbonBlack Response Standard). CBC has improved endpoint sensor architecture, more stability, and substantially reduces the build time for Twitter engineers. The CBC software is also vastly improved over the previous version, giving us far greater visibility, query performance and capabilities that did not exist in the legacy version.

This entire section is misleading. Stating that Twitter has threat intelligence software running on 9.7k endpoints can sound impressive but is without context. This entire section is misleading. This is the software that has been reporting that the endpoint fleet is extremely out of compliance¹⁸. The prior version of this software, no longer supported by the vendor - hence the upgrade - was reporting the same thing. "Far greater visibility" is without context and, irrespective, it is meaningless if what the software reports is ignored.

Finding that software build times were reduced for a subset of engineers was a happy accident and was not an intentional goal. This brings up an important question: why are engineers performing software builds locally on their laptops. Presently every engineer has a full copy of Twitter's proprietary source code on their laptop. Ideally software builds would be performed on servers in the data centers, or in the cloud, and in an isolated testing environment. The fact that engineers are performing software builds on their laptops (endpoints) and these systems are in such poor security configuration is indeed very disturbing.

The fact that this paragraph of the document does not reference that the security software replaces a previous version of the security software from the same vendor, makes it sound like the change in software was proactive on Twitter's part. It was not. The version of this software already rolled out, on approximately the same number of endpoints listed here, was discontinued by the vendor. Twitter had no choice but to

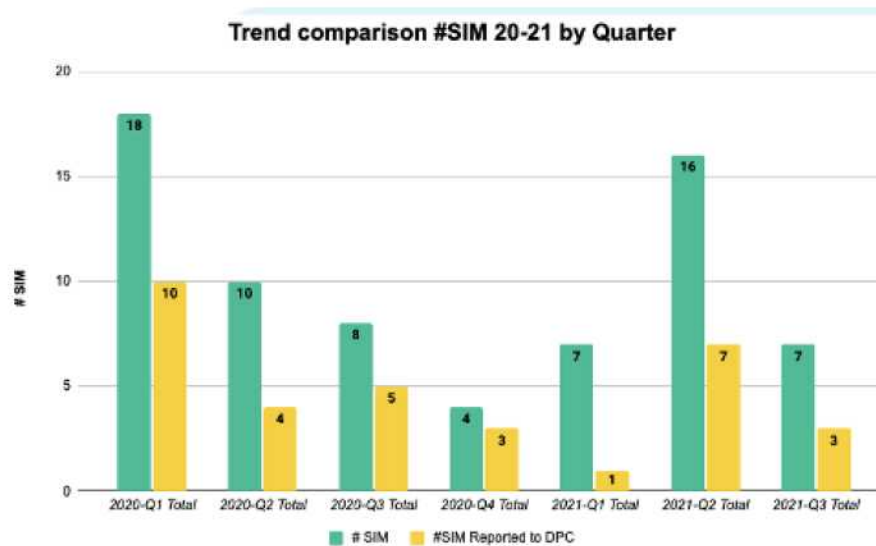
¹⁸ Across the fleet it is being reported that 30% of laptops have automatic updates disabled, random employee systems have their firewalls turned off, remote desktop turned on, system protection against malicious software disabled, and so on. The software is reporting an endpoint fleet that is in significant disarray.

move to the other software (and quickly). This was not an improvement. This was a lateral move that Twitter was forced to make. Presenting all of this as a win is disingenuous.

A note on Zero Trust and Endpoints (Employee computers)

In regards to endpoint (employee computer) security, [REDACTED] has stated that a “Zero Trust” environment is the go-forward strategy for Twitter. In a Zero Trust environment Twitter employees access internal Twitter services and data without being within a VPN. This approach was popularized by Google around a decade ago and called BeyondCorp. Employee laptops are provided cryptographic credentials and “certified”. The laptops are directly connected to the Internet and (only) the specific service connections into Twitter are encrypted and “protected”. To do this safely requires strong security configurations and hygiene of the laptops receiving Twitter *certification* and ensuring the laptops maintain a strong security posture that is not violated. Moving towards a Zero Trust environment without identifying and addressing the issues with the current state of endpoint configurations and security implies a lack of basic understanding around Zero Trust and information security priorities.

Incidents and Incident Classes

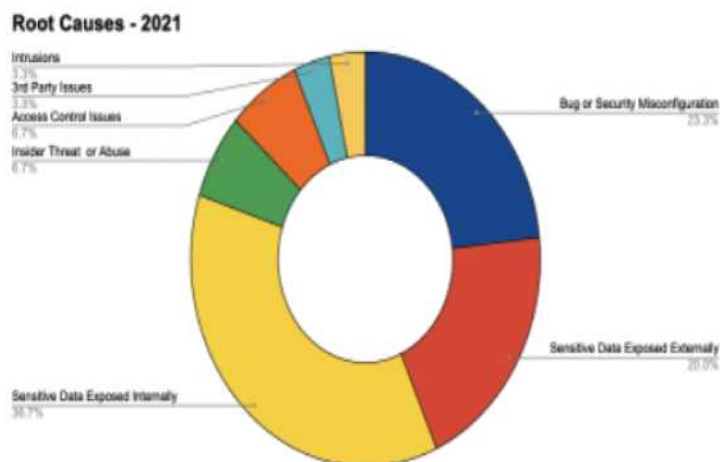


Slide 3 contains information on Incidents: a histogram of Incidents (captured above here) and pie chart of root causes (further below in this document). The histogram is misleading as it only reports a subset of incidents reported and not all SIM (Security Incidents) per quarter. The subset appears to be only related to incidents required to be reported to the Irish Data Protection Commission (a regulatory agency).

It is not appropriate to represent subsets of incidents to the Risk Committee without clearly providing a reason and context. Adding in the missing incidents significantly increases the values. For instance, representing the total number of incidents to the the last section of the chart above, doubles the number of regulator reported Incidents reported to regulators (from 3 to 6) and similarly increases the total Incidents (from 7 to 19). Not only is the chart above incorrect in not representing total incidents, the shape of the chart cannot be trusted to be accurate either.

The correct number is likely closer to 50-60¹⁹ incidents in 2021. More than 1 incident a month was significant and specific enough that it was required to be reported to regulators²⁰. Keep in mind that Twitter is under significant regulatory scrutiny and each of these events worsens Twitter's situation.

The pie chart on Slide 3 showing incident classes (root causes) is also inappropriate.

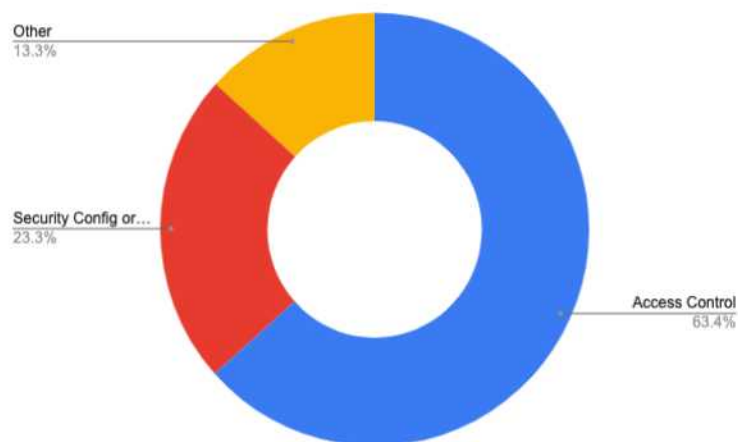


Pie chart presented to Q4 Risk Committee classifying Incidents

¹⁹ Due to a lack of continued access to data I had to extrapolate to 50-60. I reached this number using the number of incidents from April - November 2021 in my notes and then extrapolated the missing months at the same rate of incidents.

²⁰ Extrapolated from personal notes.

The above assigns less than 7% of incidents with a root cause (class) of access control. This is incorrect. Access Control is the cause of more than 60% of all Incidents, and Security Configurations / Bugs account for almost 25%. These are two areas of critical risk that had not been represented appropriately within Twitter or to the Committee.



Corrected classification of 2021 Incidents - note the correlation of incident classes to critical areas of top risks, as identified in Section 2 of this document. This is a more correct graphic that should have been presented

Current State Assessment

Alethea

1.0 -- Executive Summary

Alethea Group was engaged by Twitter to evaluate the state and structure of Twitter's capabilities in countering misinformation and disinformation, with the goal of identifying gaps in its processes, policies, and approach, as well as opportunities to build the organization's ability to safeguard the platforms and its users. This report details the current state of Twitter's misinformation and disinformation capabilities as identified by Alethea Group, based upon internal documents reviewed, stakeholder interviews, and other information gathered as needed. A subsequent report, based on the findings contained in this report, will be delivered in two weeks from final acceptance from the Client in order to make recommendations for how to mature the organization's capabilities to address misinformation and disinformation globally.

Broadly, our assessment found that organizational siloing, a lack of investment in critical resources, and reactive policies and processes have driven Twitter to operate in a constant state of crisis that does not support the company's broader mission of protecting authentic conversation. As a result, Twitter is consistently behind the curve in actioning against disinformation and misinformation threats. *clarity (reactive vs behind peers)*

Teams identified significant gaps in resource allocation, leading to policies and actions that are often reactive in nature and do not allow the company to think about emerging threats. Twitter does not have a traditional threat intelligence capability that would better position the company to be proactive on misinformation and disinformation and to protect authentic conversation.

Ultimately, these gaps mean that although Twitter is a global company with a global mission, it is not currently set up to deliver globally on trust and safety. *- language support here too?*

how? how? Different incentives for different teams working on misinformation and disinformation means Twitter is set up to be reactive, and although it has beneficial partnerships with other social media companies and research institutions, they do not allow Twitter to do proactive analysis that is reflective of the actual threat landscape on the platform or reflective of Twitter's business objectives. These gaps illustrate the extent to which product and growth are prioritized over online user and platform safety. Twitter further lacks sufficient mechanisms to measure progress and impact, therefore it may not be accurately measuring progress or it could be failing to implement lessons learned from the past. *Is this achievable? Are there exceptions?*

Can we compare to other companies? Tools available to Site Integrity to work on these issues are often outdated, "hacked together," or difficult to use, limiting Twitter's ability to effectively enforce policies at scale. A lack of automation and sophisticated tooling means that Twitter relies on human capabilities, which are not adequately staffed or resourced, to address the misinformation and disinformation problem. Further, policies are often written in response to external events, or "fires," rather than being informed by analysis of the current or emerging threats for the platform, without an effective enforcement mechanism and tooling in place. Because policy changes are often implemented

are reluctant to
be ~~slow~~ introduced
policy changes
can they be refined?

DRAFT - FOR FEEDBACK PURPOSES ONLY

Privileged and Confidential//Attorney Work Product

quickly, they often do not incorporate feedback from relevant stakeholders, are not well-executed, and difficult to enforce at scale.

Our assessment found that Site Integrity teams lack diversity, especially gender diversity, across the analytical and managerial level. Additionally, the lack of diverse backgrounds among employees contributed to gaps in foreign-language and on-the-ground contextual capabilities, hindering Twitter's ability to execute its mission and remove harmful content worldwide. Teams in priority growth markets either do not exist, or are not sufficiently staffed or resourced.

Our assessment found that employees in this space are supportive of Twitter's mission and the organization, and have positive perceptions of their teams, teammates, and managers. Despite the challenging subject matter and circumstances, including employees reporting burnout because of a lack of resources, interviewees described managers as receptive to feedback and concerns, and a positive team culture of pulling together to get the work done. The team appears to be dedicated to their mission, believes that Twitter can achieve its goals, and articulated the desire to see the team through this upcoming period of growth.

2.0 Methodology

In order to conduct the current state assessment, Alethea Group interviewed 12 members of Twitter's Trust & Safety, Twitter Services, and Product & Engineering teams, conducted screen sharing exercises to understand Twitter's internal misinformation/disinformation tooling and processes, and reviewed a series of 19 internal documents, retrospectives, and training guides. This assessment does not seek to comprehensively address Twitter's performance, capabilities, or work during the US 2020 election.

3.0 Current State Assessment

3.1 Organization

3.1.1 -- The organizational structure within Twitter that responds to disinformation and misinformation is siloed and not clearly defined. The capabilities were built in an ad hoc manner largely in response to crises. This has contributed to organizational silos, capabilities gaps, and created a culture in which employees must rely on informal relationships across the organization to accomplish work.

Currently, Twitter does not have a clearly defined organization to encompass the functions or offices at Twitter that are dedicated to detecting and mitigating platform harms, and does not

DRAFT - FOR FEEDBACK PURPOSES ONLY
Privileged and Confidential//Attorney Work Product

have the ad hoc structures documented in order to support formalization of functions and offices. Efforts to combat misinformation and disinformation on the platform have evolved in an ad hoc manner as a result of external factors, such as the 2016 elections, coronavirus pandemic, and other pressing threats. Because of the ad hoc nature, the informal organization is driven by policy decisions made in a silo, mostly by San Francisco-based staff, and frequently during a time of crisis.

please back this statement up is there supporting evidence?

This has consistently meant that relevant entities do not have the opportunity to engage other parts of the organization and key stakeholders responsible for countering disinformation, leading to policies that may be unenforceable at scale or not reflective of the threat landscape on and off the platform. Interviewees said this has also meant that historically, teams across the organization have been inconsistent or slow to respond, especially to information risks or threats that are not easily defined, such as the evolution of QAnon or cases of coronavirus misinformation.

assertions please cite examples ii

Without a formal organizational structure in the misinformation and disinformation problem set, the holistic solutions required to mature functions that combat platform manipulation are not sufficiently resourced.

(not resourced? or not well understood and defined and hence insufficiently resourced?)

3.1.1.1-- Site Integrity, which is responsible for platform policy and enforcement related to platform manipulation matters, works with Health and Twitter Services to collaborate on tools, technical fixes, and policy enforcement, but they lack formal processes and structures that facilitate easy identification of roles and responsibilities and instead rely upon informal cross-functional relationships.

huh??

Trust and Safety functions exist in a silo within Security, while actually impacting all parts of the Twitter platform and experience. As a result, Twitter is making critical decisions about new products and product launches without being prepared to mitigate security concerns.

This is confusing contextually

Different parts of the organization are working different pieces of the problem set, but interviewees described a very insular process for their respective teams in which there is a lack of meaningful coordination with other relevant teams and no official mechanism, such as formal working agreements between teams, outlining their authorities and responsibilities to each other. While Site Integrity is responsible for drafting policy, they are unable to adequately respond to threats or enforce new policies at scale because other components of Twitter are not meaningfully engaged. Historically, policies have been created during a crisis or in response to a major platform failure to address misinformation or disinformation, instead of proactively.

How about Policy? ops as separate teams to make

How does Site Integrity responsible for drafting policy preclude scale enforcement?

DRAFT - FOR FEEDBACK PURPOSES ONLY
Privileged and Confidential//Attorney Work Product

✓✓
Many interviewees credit informal relationships for their ability to make any progress or be able to seek support from other teams, whether engineering or product support. Organizational effectiveness appears to be based on the ability to navigate Twitter versus an intentional organization determined by Twitter's leadership, given the necessary resources and support to achieve its mission. is this part of sentence helping?

Example: In one instance, Twitter planned to launch its new product, Fleets, just weeks before the US 2020 election when resources had been pulled from other duties to address the high-profile, high threat election. While SI team members said that they had been involved in a health review of the product throughout, they were not meaningfully involved in the launch of the new product and were not capable or resourced to be able to combat product manipulation. Multiple interviewees reported that they had to "beg" the product team not to launch before the election because they did not have the resources or capabilities to action on disinformation or misinformation on a new product during such a busy, critical time. One interviewee said that SI leadership had to go over the heads of product managers on the Fleets team to help ensure that the product was not launched before the election. According to interviewees, the Fleets example was a serious pain point, underscoring the organizational challenge of new product launches that expose new surfaces which a threat actor can take advantage of. This illustrates the fundamental business challenge of continuing to attract new users while also safeguarding the platform from malign actors, as well as different incentives for different Twitter teams.

3.1.1.2 -- There are components of Twitter that are part of the disinformation and misinformation detection or response that are outside of Site Integrity / Security, and Site Integrity / Security have no access or authority to use these tools absent the good will of other teams.

Through the course of our interviews, we identified multiple teams that were not part of SI/Security yet played a critical role in responding to disinformation and misinformation. SI has no formal authority to require systemic changes or collaborate on key decisions.

For example, as part of a response to disinformation or misinformation, the events teams and curation teams, especially with regards to trending topics, can be partners in mitigating threats by showing Twitter users accurate information. The relationships between the teams with regards to these processes are informal and personality-based versus institutional.

Additionally, with regard to scaled detection of disinformation and misinformation, SI does not have the necessary dedicated engineering support to be able to manage both long-

DRAFT - FOR FEEDBACK PURPOSES ONLY
Privileged and Confidential//Attorney Work Product

term priorities and build products that enable threat detection and mitigation at scale, preventing it from being able to focus on proactive activities and instead making them reactive to the crisis of the day.

3.1.1.3 -- Twitter does not have aligned incentives across the organization, and, as a result, priorities with regards to Product Safety.

Product and product managers own all aspects of product development, including risk calculations with regard to product launches. Recently, SI and other Safety components have been included in the design and development process at various check-in points, and provided the opportunity to provide feedback. However, there appear to be no consequences for product managers should their product launches or products increase the workload or costs to Twitter when it falls on SI to develop policies or scale enforcement.

While SI has the authority to make recommendations throughout the product development process, elements of Twitter responsible for identifying threats or security gaps in the products lack the authority to make decisions on product design or roll-out or to hold product teams accountable for failing to mitigate identified risks to the platform, product, and users online.

Interviewees described both the launch of Fleets and Birdwatch as particular pain points for the Trust & Safety team. While product teams do elicit feedback for new product launches, product managers are incentivized to ship products as quickly as possible and thus are willing to accept security risks.

3.1.1.4 -- SI relies on functions that have no accountability to SI in order to piece together solutions.

Interviewees regularly mentioned under-resourced teams, siloing between organizations, and having to borrow resources (such as engineering support), leading to a reliance on the goodwill of other teams leaders or the willingness of Twitter employees to pitch in to support SI in building out its tooling capabilities. This prevents SI from being able to think strategically and develop priorities and goals that are measurable and enable strategy execution.

3.1.2 -- Within, SI, the organizational structure is siloed, with a heavy emphasis and focus on policy enforcement versus threat detection and mitigation.

DRAFT - FOR FEEDBACK PURPOSES ONLY
Privileged and Confidential//Attorney Work Product

Within the organizations examined as part of this assessment, there appears to be a lack of alignment and prioritization between teams, and teams appear to be policy focused. Aligning teams to focus on the tactics, techniques, and procedures may cause gaps in Twitter's understanding of adversaries and how they deploy a variety of tactics, techniques, and procedures (TTPs) to carry out an operation or manipulate the platform to achieve a goal.

Silos within the SI may also be contributing to a reactive posture. For example, sophisticated IO actors rely upon misinformation to spread false narratives and often use spamming techniques. Understanding how different threat actors abuse Twitter's platform in a variety of ways (e.g. spam violations to collect data and enhance IO efforts) could help Twitter become predictive, designing holistic tooling, or adding friction to adversarial operations. It is not obvious as to why these teams are split up how they are, other than they are to enforce specific policies. While this may be a good approach in thinking through product features or investigative processes, it silos the threat in such a way that can prevent analysts from piecing together the larger picture.

Importantly, misinformation and disinformation -- which have functionally the same impact on users -- are treated as separate issues and are housed under different teams. Given the fact that misinformation can be leveraged in spam campaigns, state-level information operations, and other types of harms, Twitter's approach has led to siloing, organizational confusion, and slow policy development. Interviewees described several instances in which Twitter was slow to act on misinformation because teams did not see the topic or narrative as falling under their purview or fitting neatly into a particular threat actor they monitored, such as on QAnon or Pizzagate.

One interviewee described the organizational challenges faced by Twitter when dealing with the Pizzagate conspiracy theory and related content. Twitter initially felt as though it was not a disinformation issue because it was not seeded by a foreign actor, was not a child exploitation issue because it included false instances of child trafficking, and was not deemed a spam issue. Twitter could not figure out how to categorize the Pizzagate content, which likely contributed to the narrative's expansion and spread on the platform. In its current posture, the teams are siloed to the degree that it is not always clear who is responsible for what.

3.1.2.1 -- Within SI, there do not appear to be clear priorities from the organization's leadership on how to prioritize threats and thus it is impossible to prioritize resources, goals, and KPIs.

Interviewees said that there is no clear alignment across the teams or prioritization of how to address matters related to platform manipulation. Further, without clear and coherent goals, it is not possible to measure progress against goals in order to mature the organization's capabilities, determine how to allocate resource investment to maximize impact, or sequence the development of tools, resources, and capabilities.

DRAFT - FOR FEEDBACK PURPOSES ONLY
Privileged and Confidential//Attorney Work Product

When priorities are developed, it is with a heavy emphasis on English-speaking countries and threats, and whatever goals and metrics are set do not align with the team's observations of the most pressing needs for organizational growth. The team frequently pointed out areas that could be made more efficient through automation, improved processes, and other goals, yet within the organization there does not appear to be a mechanism for meaningful engagement.

3.1.2.2 -- SI sets up strike teams in order to direct resources towards major events, such as elections.

Teams within SI and around Twitter are focused on priority events and providing extra attention to platform matters that are likely to face manipulation. This strike team approach that allows the organization to dedicate additional resources to events appears to be a successful model for addressing threats. However, due to current staffing levels, it requires that teams deprioritize long-term strategic objectives or other responsibilities, and it is not sustainable without increasing resources.

3.1.3 -- Twitter is not poised to deliver on its mission globally, especially in non-English speaking countries.

Twitter lacks the organizational capacity in terms of staffing, functions, language, and cultural nuance to be able to operate in a global context. For example, the misinformation team currently only has two individuals and lacks the sufficient tools to be able to adequately address the threat on a global scale due to a lack of on-the-ground context. This is especially true in priority growth markets, including Africa, Latin America, and Asia. Global teams report a focus on English-language and English-speaking countries. For example, during the 2020 US election, staff were pulled into monitoring, leaving significant vulnerabilities to the regions they support.

The lack of context and understanding has significant implications on the ability to implement policies globally. For example, historically marginalized groups experiencing online threats and harms may not be recognized without an understanding of each country's context, and in some countries it is the government or military that are violating policies, and Twitter is too understaffed to be able to do much other than respond to an immediate crisis. Overseas teams lack the necessary resources to be able to conduct investigations outside of what is already trending or used as a hashtag, making its reactive posture impossible to change without engineering, data science, and investigations support. Twitter expresses a strong preference for fact-checking and labeling content versus removing the content. However, Twitter teams report not having the capacity to fact check in languages other than English.

3.2 Resources

DRAFT - FOR FEEDBACK PURPOSES ONLY
Privileged and Confidential//Attorney Work Product

Alethea Group conducted individual interviews, including screen shares, and reviewed internal documentation to determine whether Twitter has the necessary resources, such as tools, datastreams, staff, and skills, to accomplish their tasks.

The lack of sufficient resources, tools, and capabilities has forced SI and TwS to be reactive and largely limit their focus to threats that affect the United States or English-speaking entities. This has ultimately prevented Twitter from proactive threat detection and mitigation to avoid crises. Interviewees described a largely reactive approach to misinformation, disinformation, and spam in which action is taken on content and threats only if it is flagged by reporters or news headlines, partners, or political officials due to the lack of people and sufficient tools to do proactive analysis.

Despite having a global mission, persistent gaps in resources, tools, and capabilities we identified means Twitter does not have the capabilities to operate globally -- including in priority markets -- when it comes to misinformation and disinformation. It also suggests that Twitter is likely spending resources in crisis management and response, rather than investing in capabilities that will allow the company to get ahead of them.

3.2.1 -- Teams in priority growth markets are not sufficiently resourced.

Teams across SI, TwS, and Product prioritize resources to meet primarily US-centric needs. Interviewees across the board said that they do not have the resources, such as staffing and foreign language capabilities, needed to address misinformation and disinformation even in priority markets, such as Asia.

3.2.2 -- Teams have been persistently understaffed.

Twitter has been slow to staff SI teams since 2016. Despite recent team increases, there are currently only two misinformation subject matter experts in SI, both of which are new hires, and four IO investigators to analyze all IO. One interviewee noted that the lack of misinformation expertise was identified as a serious gap in a retrospective from December 2016 about Twitter's lessons learned from Pizzagate. Twitter did not bring on a team member to focus on misinformation until 2019, although existing staff reported that they did focus some of their time to misinformation, however their other responsibilities remained unchanged resulting in staff being asked to do more without additional resources.

Understaffing has meant the teams across Twitter working on the misinformation and disinformation problem set have had to make significant tradeoffs, especially during critical events and surges. For example, Twitter dedicated 100 full-time staff from across SI, TwS, and volunteers from other parts of the company to manage the US 2020 election under the "Election

DRAFT - FOR FEEDBACK PURPOSES ONLY
Privileged and Confidential//Attorney Work Product

Squad” framework. As a result, based on interviews and provided documents, SI, Site Policy, Product Trust, and Strategic Response teams had to deprioritize all other work, including work on other critical global events, simply to keep up with the rapid pace of US election-related content.

Site Integrity Headcount for 2021

Team	Total Roles Expected	Roles Filled	Roles Unfilled
Management	7	4	3
API	7	5	2
Misinformation	8	2	6
Identity	5	4	1
IO & Security	10	4	6
Spam	17	12	5

Several interviewees noted personal perceptions about understaffing that may not be accurate, but influence how they view the organization’s commitment to filling gaps. For example, one interviewee who had been involved in interviewing candidates for critical roles in SI believed qualified candidates were often rejected by leadership for unimportant reasons. Separately, one interviewee believed understaffing was negatively affecting their team’s ability to get resources from Twitter. They noted their belief that funding for internal tool development was decided based on the number of people in the company who would use the tool, which they believed would continue to keep SI teams at a disadvantage; subsequent conversations with Twitter leadership suggested the described process for acquiring funding may not be wholly accurate.

3.2.3 -- SI does not have dedicated engineering support for their tools, so even minor upgrades or changes to existing tools can take months or years to complete.

SI is severely constrained by not having engineers on their teams or engineers dedicated to exclusively supporting their work. Currently, SI must request assistance from engineering teams in other parts of the company to do things like implement even small updates to existing tools or build new ones that could automate more of the process for both policy and investigative analysts. Because these engineering teams do not have an official requirement to support SI and must complete their own work, SI requests are typically put onto a waitlist. That list is then prioritized by SI’s immediate engineering support needs for current so-called “fires,” such as a critical election. As a result, SI must continue to rely on manual and outdated tools, and individual know-how of its analysts who often must code their own solutions to complete their work. One interviewee called the lack of engineering support dedicated specifically to SI “a real pain point

DRAFT - FOR FEEDBACK PURPOSES ONLY
Privileged and Confidential//Attorney Work Product

for internal tooling needs” and said they have had to wait “sometimes years” for minor updates to existing tools they need to do their jobs.

3.2.4 -- SI lacks sufficient dedicated data science support and staff with technical skills.

SI teams rely on the Scaled Enforcement Heuristics (SEH) team to provide data science support, rather than having their own dedicated data scientists. Even though interviewees described receiving excellent work from SEH, they also noted many of the same challenges they have in getting engineering support, namely that SEH has its own work and priorities.

Additionally, in part because many of the tools used by SI require the user to do their own coding and queries, SI lacks sufficient access to technical resources. Having more usable, updated tools with usable UIs would probably reduce the need for some of the technical capabilities.

3.3 -- Tools

SI analysts and managers we interviewed referenced the below range of tools they use to complete their jobs. We were able to personally view the tools that are noted in bold during a screen share or from training materials.

- **Profile Viewer**
- **Batch Action**
- **ClusterDuck**
- **SafetyGraph**
- **Access Search**
- Guano Interface
- **URL Tool**
- Bulk Media Enforcement Tool
- Abuse Triage Tool
- **Botmaker**
- Smyte
- **Semantic Core Editor MisinfoUI**
- **Strato**
- **Thunderbird**
- Hadoop
- Presto
- BigQuery

3.3.1 -- Twitter has not sufficiently invested in developing internal tools to address misinformation and disinformation. As a result, employees must use multiple outdated

DRAFT - FOR FEEDBACK PURPOSES ONLY
Privileged and Confidential//Attorney Work Product

and manual tools to do parts and pieces of their investigations, analysis, and enforcement.

In both SI and TwS, interviewees and provided documents described a largely manual process of utilizing multiple outdated, cumbersome, and unreliable tools with poor UIs to do parts of their work, including investigations, analysis, and actioning content. For example, there is currently no comprehensive system for tracking misinformation, from identification to remediation. Existing tools used for surfacing misinformation and disinformation threats are set up so that analysts must go to different tools to manually search for a threat actor or narrative already in mind, rather than the tool using automation and ML to identify potential threats that it then pushes to investigators for analysis.

For IO investigators, one of the most used tools, ClusterDuck, which identifies networks of similar and/or coordinated accounts by country, does not do real-time monitoring and analysis. Data is up to seven days old, and, rather than the tool flagging potentially violative behavior to analysts, users must manually click on a drop-down menu of countries to view results to make a determination on possible coordinated activity. One interviewee described ClusterDuck as “pretty hacked together,” and when the assessment team was viewing how the tool operated, it would not load on the first attempt. Another interviewee described ClusterDuck as the only tool really designed specifically for the SI team. A separate tool, AccessSearch, is frequently used by investigators, but its utility is limited by short data storage times (one analyst said it could only store data for two months) that prevent historical research.

Tools used to action on violative content have many of the same problems. For example, according to one interviewee, the process for labeling violative tweets requires using at least five different tools. Tagging tweets in bulk is a manual process that requires the analyst to write a code themselves in a tool called Strato that does not have an easy to use UI. There is also not an easy or automated solution for labelling all tweets that link to a URL that has already been labeled. On Misinformation, SI must manually annotate each new instance of misinformation identified and then moderators manually tag tweets they see with this annotation to apply a warning label.¹ This manual process is especially challenging for large events, such as key elections.

The manual and outdated nature of these tools forces analysts and content moderators to analyze and action against violations tweet by tweet and account by account, a time-consuming process that will keep Twitter reliant on unscalable human power.

3.3.2 -- SI has access to many data sources, but they are spread across several different systems and require largely manual processes to access and analyze.

¹ “Soft Intervention Tool User Manual”

DRAFT - FOR FEEDBACK PURPOSES ONLY
Privileged and Confidential//Attorney Work Product

Interviewees in SI suggested they had access to a large number of datastreams with information about on-platform activity. However, they said finding, accessing, and analyzing that data was challenging and time-consuming because it required the use of several different tools and manual processes to search. They also do not have tools aimed at enabling cross-platform analysis. Several analysts also noted having to do their own coding for querying data because many of their tools lack functional UIs. Many of the fixes are small, but would save analysts time and enable more automation. For example, analysts noted having to separately sign-in to external tools, like Domain Tools, to complete a step in the investigative process, when obtaining API access to external tools would allow for integration into internal Twitter capabilities and remove another step in an otherwise manual process.

3.3.3 -- There are existing internal tools in other parts of Twitter that would be useful for the misinformation and disinformation use case, but SI analysts do not have access to them. Analysts also lack access to externally available tools or datastreams that would allow them to do more proactive cross-platform analysis.

Several interviewees noted that other teams at Twitter have internal tools that would be helpful for the misinformation and disinformation use case, but they do not have access to them. For example, one interviewee said Curation uses a tool to create Moments that could potentially help Misinformation and IO analysts proactively identify threats, but they lack access to the tool. SI also does not have access to externally available tools that would allow them to do proactive and more sophisticated analysis and to get insight into emerging threats, such as a social listening tool that provides cross-platform data. One analyst noted that they do not have dedicated staff looking at off-platform activity beyond what external partners provide them, which limits their ability to anticipate possible threats moving on to the Twitter platform.

3.4 -- Capabilities

3.4.1 -- SI does not have a knowledge management system to track and store findings and data. As a result, SI does not have the ability to monitor threat actors or identify changes in their tactics, techniques, and procedures (TTPs) over time, or to measure the impact of SI's work.

Currently, SI does not maintain a knowledge management tool or capability that would enable analysts to save content, data, or their findings. There is no tool or repository where analysts conducting investigations can keep their notes. Most analysts use their own individual Word Documents so that worthwhile investigative notes are individually stored in a way that is not accessible to, or preserved for, their teammates. As a result, analysts are unable to identify and analyze evolving threats or changes in the TTPs of threat actors, or measure the effectiveness of action and enforcement, because information is not being preserved.

DRAFT - FOR FEEDBACK PURPOSES ONLY
Privileged and Confidential//Attorney Work Product

IO interviewees noted they have a tasking system housed in Jira to action on leads received from internal teams, such as the Piper Team, or from external partners. However, there is no mechanism by which to save the results of their investigations in a single, automated knowledge management system. Currently, once a tasking is marked as complete in the Jira system, analysts must manually copy their findings into multiple different data source tools or folders to store it, creating an extra step one interviewee said analysts just do not have the time to complete. In some instances, the analyst saves their findings on their own systems, meaning data storage is scattered among different analysts, rather than being preserved in one system accessible to all analysts. This also means Twitter is not feeding its findings into tools or training existing tools to increase automation and ultimately learn from past findings.

3.4.2 -- Twitter does not have traditional threat intelligence capabilities to identify, analyze, and warn about current and future threats, or ingest inputs and intelligence from partnerships.

Twitter does not have a threat intelligence capability internally it can direct based on the company's priorities and to position itself to be proactive in protecting authentic conversation. Misinformation and disinformation teams are currently focused on responding to current threats and so-called "fires" that interviewees said are largely driven by external priorities, such as news headlines, journalist inquiries, or the goodwill of partners.

As a result, Twitter is reactive to events and situations, based on other organizations' goals, interests, or priorities. Relying on civil society cannot scale to meet Twitter's needs, as many priority markets do not have regulatory environments or vibrant civil societies to enable research that, in some cases, may identify government-run influence operations.

3.4.3 -- Twitter does not have the capability to add cost to an adversary attempting to exploit the platform.

In part due to the challenges described above, Twitter has employed a limited set of actions against violative behavior on the platform. Currently, most of Twitter's remediation options have focused on labeling, interstitials, deamplification on a select basis, and removal in response to repeat violations. Twitter leadership has publicly state that account removal could set a bad precedent,² and interviewees perceived that removal of accounts or content was considered by Twitter leadership as the option of last resort. However, even removal ultimately does not discourage adversaries from attempting to exploit and leverage the platform, or add costs to their operations because they can quickly adapt. One interviewee did say that Twitter started removing networks piecemeal in order to obfuscate how the network or accounts in question

² <https://www.npr.org/2021/01/14/956664893/twitter-ceo-tweets-about-banning-trump-from-site>

DRAFT - FOR FEEDBACK PURPOSES ONLY
Privileged and Confidential//Attorney Work Product

were found. Another interviewee estimated that it would realistically take two years before Twitter could build out a strategy and capability to add cost to adversaries by deploying actions like artificial environments.

3.4.4 -- SI and TwS teams lack staff with geographic expertise and foreign language capabilities.

SI and TwS teams do not have sufficient staff with geographic expertise and foreign language capabilities, even in key markets, both of which are needed to understand important cultural and language contexts. Currently, the majority of SI staff are located in the US, with a limited presence in Dublin, and an even smaller footprint in Singapore. The IO team has one staff member with expertise in Russia, one with expertise in Iran, and one with expertise in China, making staffing and coverage, particularly during a crisis, unsustainable. One SI employee noted that the language gap was so significant across Twitter that they regularly receive language support requests from all over the company, not just from the teams responsible for misinformation and disinformation.

The lack of sufficient foreign language skills has hindered work in priority markets. For example, several interviewees and internal policy documents stated that Twitter is limited on fact-checking or debunking to mostly English-language content. One interviewee said that they relied heavily on Google Translate for language capabilities and said that for some countries, such as Thailand, SI is only able to search for trending hashtags for possible exploitation by a threat actor rather than doing investigations because they do not have the language or country expertise on staff.

The lack of language expertise is also affecting Twitter's ability to plan for upcoming priority events. According to internal documentation, Twitter is unable to provide even a scaled-back version of the election support that was deployed for the US 2020 election for the upcoming Japanese election, which has been identified as a priority for the company. According to the "US 2020 Civic Integrity Policy/Ops/Product Reflections" document, that is in large part because there are "no Japanese speakers on the Site Integrity team, only one T&S staff member located in Tokyo, and severely limited Japanese-language coverage among senior TwS Strategic Response staff."

3.5 -- People

3.5.1 -- SI employees are dedicated to the mission and the organization, and feel heard by their immediate SI management.

Interviewees all expressed support for the mission and the organization, as well as positive perceptions about their teams, teammates, and managers. They described pulling together to

DRAFT - FOR FEEDBACK PURPOSES ONLY
Privileged and Confidential//Attorney Work Product

meet the demands of each day, and many described a strong commitment to the organization despite challenging circumstances and burnout. Several interviewees also noted that they felt heard by their immediate SI management and felt empowered to raise concerns to them. At the same time, they were not always confident that action would or could be taken in response to those concerns.

3.5.2 -- SI teams lack diversity, especially gender diversity across both the analyst and management level.

Multiple interviewees expressed a concern for the lack of diversity, particularly gender diversity, on the teams responsible for addressing misinformation and disinformation. According to staffing documents we reviewed, only one-third of SI personnel are women and the majority of management and senior-level positions are held by men. Similarly, several interviewees assessed that the lack of diverse backgrounds among employees contributed to gaps in foreign-language capabilities on the teams and, therefore, the teams focused on primarily Western, English-language content and threats. The lack of diversity almost certainly hinders SI's ability to execute its mission and benefit from the talents and abilities a more diverse workforce provides.

3.5.3 -- SI staff are burned out and do not believe Twitter leadership is aware of it.

Employees in SI reported being burned out. They attributed this in large part due to understaffing, the amount of day-to-day work, frequent policy changes that create confusion, time-consuming manual internal tooling, a lack of strategic planning across all the relevant parts of Twitter, and a consistent crisis state of operating as a result of jumping from one "fire" to the next. These issues have created time-consuming processes and stress on teams where employees are expected to work longer hours when a lack of strategic planning creates a crisis. The majority of interviewees also said they are expected to wear multiple hats, and SI interviewees noted in particular a perceived tendency by leadership to rely on a couple of people for everything. They believed that the fact that those people completed their work was used as justification for not hiring more people.

Most interviewees pointed to the rapid pace of work and the significant workload of the US 2020 election as a recent source of employee burnout. However, many ascribed their burnout to what they saw as a culture of constantly being in a state of "firefighting" or crisis, which they largely saw as driven by external events, such as congressional inquiries or news events. Relatedly, several senior managers across Twitter were expected to be "always on" during the election to address escalations on high-profile accounts because of the company's "low risk tolerance," according to documents we reviewed. A similar-sized effort under the Election Squad construct for another priority election would be unsustainable with current staffing levels.

DRAFT - FOR FEEDBACK PURPOSES ONLY
Privileged and Confidential//Attorney Work Product

Interviewees said a lack of strategic planning and coordination between relevant parties in SI, TwS, and Product on product development and deployment had also contributed to staff burnout. For example, SI and TwS interviewees noted that product teams consistently failed to solicit or at least include their feedback on product rollouts, such as Fleets or BirdWatch. They said it had resulted in them having to pull longer hours, often outside of working hours, to address vulnerabilities in products identified sometimes hours before or even after a product rollout.

3.5.4 -- Staffing in SI is top heavy, except for on the Piper Team. Managers are expected to wear multiple hats, including conducting investigations and creating policies, but they spend most of their time with managerial responsibilities, and report spending their days in back-to-back meetings.

SI managers said that they were expected to still conduct IO investigations and lead on developing IO policies, but that they spent the majority of each day in meetings and on personnel management tasks. Some interviewees expressed concern about not having the time to keep up their investigative and technical skills, and one senior manager said they often used what should be their non-work hours to conduct manual investigative work that a more junior employee could do, including finding and suspending large numbers of accounts trying to evade a previous Twitter ban.

3.5.5 -- Content moderators in TwS are not adequately resourced, especially to make determinations on misinformation.

Content moderation is outsourced to vendors, most of whom are located in Manila. One interviewee stated that moderators are “treated like second-class citizens,” are “not fully bought-in” to the company, and are underpaid.

Moderators are not properly resourced to take action, especially on misinformation. Several interviewees said that moderators do not have the geographic expertise or language capabilities to understand important cultural or linguistic context, and therefore are not able to make accurate and consistent decisions on what is misinformation. Another interviewee described a long process for training moderators on new policy rollouts and said that managers often did not have sufficient warning about new policies to prepare moderators in time. As a result, full-time TwS employees have had to, at times, do content moderation. Content moderators are also not proactively trained on emerging threats.

3.6 -- Partnerships

SI has prioritized creating official external partnerships with nine companies, largely other social media platforms like Facebook and Google, and more unofficial partnerships with research organizations, such as the Stanford Internet Observatory. These partnerships give SI insight into

DRAFT - FOR FEEDBACK PURPOSES ONLY
Privileged and Confidential//Attorney Work Product

misinformation and disinformation trends across social media platforms, provide warning of potential threats on their own platform, allow Twitter to potentially get ahead of news stories, and give the company the opportunity to publicly promote its work on misinformation and disinformation in a way that boosts public perception of its activities. However, Twitter is not fully taking advantage of these existing partnerships and has not established other potential partnerships that would set itself up for more proactive, long-term success in addressing the misinformation and disinformation threat. Additionally, these partnerships contribute to SI staying in reactive mode.

3.6.1 -- There is not a consistent view within SI about the goal of external partnerships.

Judging from the interviews we conducted, teams have different views on what the goals of external partnerships are. Some interviewees suggested partnerships were a way to see what the other platforms were doing or to get ahead of a forthcoming news story. One interviewee characterized partnerships as a “moat to protect the organization” from public criticism. This lack of alignment on the purpose and intent of partnerships may mean that there are other partnership opportunities for SI that can help address some of the gaps in capabilities and resourcing described above.

3.6.2 -- Investigating and actioning on inputs from external partnerships often drives SI's immediate priorities and keeps teams in a constant reactive state. However, findings from other platforms do not necessarily reflect the actual threat landscape on Twitter itself.

Intelligence and leads from its partnerships with other social media platforms gives SI critical insight into cross-platform activity that may also be affecting the Twitter platform. Similarly, working with research organizations like the Stanford Internet Observatory gives SI access to experts and early insight into, and opportunities to collaborate on, forthcoming academic research that may gain media attention upon public release.

However, actioning on the work from these partners means Twitter often prioritizes the findings of other platforms, which are also largely set up to do reactive work and have their own internal priorities and challenges. Similarly, academic organizations face staffing shortfalls, meaning they must prioritize their own work products, primarily resulting in retrospective and targeted research projects rather than Twitter being able to direct research and investigations on its own priorities. As a result, prioritizing investigative inputs from both platform partners and academic partners means SI may not be investing its time in addressing the actual threat landscape on the Twitter platform.

3.6.3 -- SI is currently unable to ingest, action, and store all of the intelligence and leads provided by its existing partnerships. It does not currently have partnerships

that could help fill some of the gaps in being proactive to address Twitter's own threat landscape.

Several SI interviewees said it was a struggle to stay on top of actioning on all of the leads provided by partners or flagged by external parties, such as reporters. They believed that prioritizing those taskings contributed to the teams' inability to do proactive work more reflective of the threat landscape on the platform, including: getting actionable intelligence from outside partners that could be informing long-term planning and decisions, identifying threats, assisting with strategic investigations, and helping to move the company from reactive to proactive on misinformation and disinformation. SI's existing partnerships do not include an ability to task them to conduct targeted analyses or longer-term investigations.

3.7 -- Policies

Alethea Group sought to identify current formal and informal policies and processes in place to help understand Twitter's capabilities to address disinformation/misinformation.

3.7.1 -- Policies are often implemented in response to "fires," rather than being informed by analysis of the current or emerging threats for the platform, without an effective enforcement mechanism in place.

Based on interviews with key stakeholders and a review of internal documentation, policies are often created quickly in response to external events, with no clear strategy for implementation. Team members said that because policies are often reactive in nature, there are significant gaps in the content they cover, and that policies do not address evolving threats.

Interviewees said that major events, including Chrissy Tiegen threatening to leave Twitter because of harassment from users who align with the QAnon movement, or the shooting at Comet Ping Pong (Pizzagate) in 2016, forced Twitter to take a stronger policy and remediation position than they assessed it otherwise would have based on the evolution of the threats alone. But because of the reactive nature of these changes, policies were often rushed, not well-executed, and difficult to enforce.

One interviewee stated: "Twitter only seems to respond to fires, and fires only. We can only handle what is the biggest and loudest fire at that moment." This approach means Twitter is often behind the curve in identifying and responding to misinformation and disinformation.

3.7.2 -- Rapid policy changes often do not incorporate feedback from the relevant stakeholders, making it more difficult to communicate, and ultimately enforce, those policies.

DRAFT - FOR FEEDBACK PURPOSES ONLY
Privileged and Confidential//Attorney Work Product

Because policy changes are often implemented quickly, they often do not incorporate feedback from relevant stakeholders, making policies more difficult to communicate and ultimately enforce. For example, in response to a manipulated video of House Speaker Nancy Pelosi in May 2019, Twitter quickly implemented a new policy (Synthetic and Manipulated Media Policy). However, because the policy was rolled out so fast, the organization was unable to effectively enforce it, or train agents on what content was violative. One interviewee stated that feedback was always asked for, and people were “given a seat at the table,” but that feedback was not always given or given in a constructive way.

In another instance, interviewees said that policy decisions were not always communicated to the broader global team, making it more difficult for the policy to be widely enforced.

According to the internal document “US 2020 Election – Policy/Ops/Product Reflections,” while “communication between policy and enforcement teams was generally solid,” during the 2020 election, the “adoption of the decision to stop using interstitials proved to be challenging, as some TwS employees continued to apply the interstitials despite email and Slack notifications about the policy change. A single source of truth on policy enforcement — rather than scattered documents, emails, and announcements — will be vital for future activations.” In short, the rapid rollout of policies leads to uneven enforcement from Twitter’s moderators.

3.7.3 -- Policies to address misinformation/disinformation often do not address repeat offenders and are applied on a case-by-case basis, leading to a lack of scalability.

Interviewees noted that there is not a sufficient enforcement mechanism for repeat violators of Twitter’s policies, and thus, there is little incentive for bad actors to stop posting violative content. One interviewee stated that if 80% of the content that a user posts is misinformation or disinformation, that account should be suspended, adding: “Continuing to address each individual tweet from a user isn’t sustainable given staffing shortfalls.”

According to the internal document, “US 2020 Election - Policy/Ops/Product Reflections,” Twitter’s labelling policies “lack any kind of punitive enforcement for repeated misinformation labels. While tweet removals under the Civic Integrity Policy incur a strike (3 strikes resulting in permanent suspension), labels do not accrue strikes, and therefore do not dissuade repeat or malicious behavior.”

3.7.4 -- Policies are written for a sophisticated audience, making it difficult for agents on the ground to enforce.

Policies that address misinformation and disinformation at Twitter (e.g. the Civic Integrity Policy, Synthetic and Manipulated Media Policy, and COVID-19 Misleading Information Policy) are often

DRAFT - FOR FEEDBACK PURPOSES ONLY
Privileged and Confidential//Attorney Work Product

complicated, highly nuanced, and require significant context for Twitter Services agents to be able to take action. When policy rollouts occur, Twitter trains its agents on those policies, however, many of these agents are located all over the world and may not have sufficient language and/or cultural context to be able to action on specific instances of misinformation. And because of the complicated nature of these policies, remediation and mitigation takes longer and is more difficult to accomplish at scale.

Additionally, when new policies are introduced, content moderators have to manually annotate each new narrative they are seeing, making it impossible to keep track of the content. By creating more digestible policies, moderators would be able to better enforce them.

One interviewee added that policies are often created in a vacuum without the input of subject matter experts and are “therefore not grounded in reality.” Another stated that Twitter’s issue is “not coming up with new policies, but enforcing the ones that we’ve already got.” Because of the sophistication and nuance in already existing policies, they are not only difficult to enforce at present, but also difficult to enforce at scale.

3.7.5 -- Twitter’s US-centric approach to policy decisions makes it difficult to detect and mitigate disinformation and misinformation around the world.

Our assessment found that policy decisions are often made in response to US-based events, such as the 2020 presidential election, QAnon content on the platform, manipulated media of House Speaker Nancy Pelosi, and more.

Because policies are written to address US-based problems, they often do not take into account different ongoing misinformation or disinformation campaigns in other parts of the world. Further, policies that address violative content in a US context are more likely to be enforced because of Twitter’s contextual and linguistic capabilities.

According to the internal document, “US 2020 Election - Policy/Ops/Product Reflections,” Twitter is “ill-equipped to provide even a scaled-back version of the proactive investigation and remediation efforts we implemented in the US — in no small part because we have no Japanese speakers on the Site Integrity team, only one T&S staff member located in Tokyo, and severely limited Japanese language coverage among senior TwS Strategic Response staff.”

Additionally, according to the same document, uneven policy enforcement around the world “creates the potential for accusations of a US-centric bias in Twitter’s actions, as well as unequal and ultimately unfair enforcement of our rules.”

Because of various factors outlined throughout this assessment, policy teams do not have the ability to plan ahead and write proactive policies in response to known upcoming events. While a certain level of uncertainty will always exist (e.g. COVID-19), there are ample opportunities to

DRAFT - FOR FEEDBACK PURPOSES ONLY
Privileged and Confidential//Attorney Work Product

proactively develop policies and capabilities in response to upcoming elections around the world and other major planned events.

3.8 -- Processes

3.8.1 -- While processes exist to elicit feedback from necessary stakeholders, there are no processes to actually incorporate that feedback.

Multiple interviewees explained that while processes exist that elicit feedback from all necessary stakeholders (e.g. product health reviews), feedback often is not incorporated.

Interviewees said that because of existing organizational structures and different incentives across teams (e.g. product teams are incentivized to launch new products), platform and user security are given less consideration than warranted. Further, product teams are not required to incorporate feedback from SI, and because product managers are promoted for launching new products, there is less incentive feedback to be incorporated, and a greater incentive to launch new products quickly.

In launching Twitter's Birdwatch program, members of the SI team said that they were involved in the process throughout, and made suggestions as to how the product could be more secure, including specifically warning that users aligned with QAnon would likely attempt to join. However, feedback was not incorporated in an attempt to keep the product open, leading to a last-minute scramble to secure the product launch. On the evening before Birdwatch launched, Twitter realized that an overt QAnon account had been accepted into the Birdwatch program.

In other instances, interviewees said that the Product Trust team would call out a risk to a product launch, but that the product team would simply "accept the risk" with minimal mitigation efforts. In short, processes don't take into account competing priorities or incentive structures within the company, and when two process owners have competing interests, there isn't a process for deconflicting, at least from a staff perspective.

3.8.2 -- The process for labelling disinformation and misinformation content is largely manual, requires the use of multiple tools, and usually needs to be done on a case-by-case basis.

According to the internal document, "US 2020 Election - Policy/Ops/Product Reflections," even once decisions about enforcement are made, "the process of applying labels is cumbersome," "requires the use of backend interfaces," and the "complex steps involved make scaled application of labels difficult to expand beyond a very small group of highly trained agents."

DRAFT - FOR FEEDBACK PURPOSES ONLY
Privileged and Confidential//Attorney Work Product

Alethea Group participated in a screen sharing process with one of the interviewees, and found that no less than five different tools were needed in order to label a single tweet.

3.8.3 -- There is currently no unified system for tracking misinformation and disinformation, from identification to remediation, according to staff interviews and the US 2020 retrospective document.

The organization does not have a system in place to proactively identify or track misinformation or disinformation threats. Leads on violative content often come from user complaints, partner organizations, or independent researchers, but Twitter does not appear to have a systematic approach to identifying these threats on its own. In the case of disinformation content, the IO team is sometimes given leads from the Piper team, but there are no existing formal processes to do so.

It appears that the organization also does not have a formal process in place for what happens after a threat is identified. Investigators stated that while there is a tool (GoIORef) where tickets are submitted and a queue is created, there is an ad hoc system for responding to those claims. And, because of a limited number of subject matter experts working on the IO team, specific team members are often needed to respond to specific disinformation/misinformation threats.

According to the internal document, "US 2020 Election - Policy/Ops/Product Reflections," Twitter's Civic Integrity Policy defines what content the company should enforce on, but "the specifics of particular conspiracies that emerged in the course of the [2020 US] election, whether those conspiracies have been debunked by external sources (and are therefore eligible for remediation), whether we have specific curated resources available for those specific conspiracies, and how to put all the pieces together in practice *is undeveloped and largely ad-hoc.*"

One interview suggested that the misinformation team and the IO team worked together because of personal relationships rather than any formal processes.

3.8.4 -- The process for identifying what civic events (i.e. the Election Assessment Process) are prioritized involves multiple teams who all use different criterion and planning processes. This results in confusion, a lack of coordination, and uneven resource allocation.

According to interviews and internal policy documents, team members from public policy, sales, regulatory, trust and safety, and others are all involved in the process of determining how to prioritize worldwide elections. However, each office has its own criterion to determine what is a priority. Once an election is assigned a priority, or "tier," there appears to be no process in place to determine the resources needed to sufficiently staff that election. Further, while an election

DRAFT - FOR FEEDBACK PURPOSES ONLY
Privileged and Confidential//Attorney Work Product

might be considered “tier 1,” it does not necessarily receive the same attention or resources as another “tier 1” election.

The result, according to the “US 2020 Election - Policy/Ops/Product Reflections” document, is that “where an election is taking place but doesn’t receive the same treatment as the US election (as happened with elections in Brazil in November 2020), in-region teams may become frustrated with limited support and apply considerable pressure to operational and policy teams to enforce rules on an ad-hoc basis, as well as product teams to build ad-hoc experiences, without adequate preparation or resourcing to do so.” Because decisions in this space are also made from a US perspective, interviewees felt that elections in other countries were given less priority.

3.8.5 -- Twitter lacks sufficient processes to measure progress and impact, and therefore fails to implement lessons learned from the past.

There are no formal processes to measure the impact of policies on deterring or combatting a threat actor, and Twitter does not have data to determine whether policies are working or need to be modified. While Twitter completes retrospectives on progress to goals (e.g. after Pizzagate), there is no process to measure the effectiveness of the company’s remediation attempts. Data is either not retained or not stored in an accessible way team-wide, giving the organization no ability to learn from its past actions.

Dr. Jiv

Knowledge system
IT Infrastructure system
Staffing
Shankh

STOP WATCH

Q1 or ~~Q2~~
beard m by
(only allowed voice)

You have who I am, my mandate, what you can expect to see from me over a defined time period...

Before I get into a bunch of data points, I want you to understand another purpose of this talk: allowing you to have context for other documents and things you hear. For instance: FTC risks. If we have an incident or breach after the settlement we have to report it to the FTC. If it is similar in nature, e.g. access-control, then the FTC is going to want to dig in and find out if we have a systemic problem here.

Boy, I sure hope that doesn't happen.

In 2020, we had 40 incidents (70+ were access control related) and 20 breaches (90% access control). That's more than 10 incidents a quarter and more than a breach a month.

with the Data Damica and I came up with the wording

Then context changes to "we are almost guaranteed to have an access control related breach and the problem is systemic". We should expect this situation with the FTC to happen before we can address it.

you keyed on

Let's get into it: - high enough level but with the first state, i date based view

When I come into a company the first thing I do is to interview across a range of people.

I've interviewed ~40 Tweeps (executive team, managers, ICs, across the company).

1 Question is always: What is it that Twitter does better than anyone else?

We are great in a crisis!

Seems good, but is it? People or organizations that *only* excel in a crisis...

- Do not build strategically to avoid them
- Begin to seek them out (subconsciously or consciously)
 - Have confidence in their ability to execute and dig out of a crisis
 - A crisis can provide structure that is lacking in other efforts
 - Target metrics
 - scope
 - Priority
 - Time frame

- Milestones
- Reward and recognition structures

When I see this I begin digging for indicators that we may subconsciously seek crises or neglect to build strategically to avoid them:

- Do we Hop from crisis to crisis? (yes):
 - 2020:
 - More than 10 Incidents a quarter and more than 1 breach a month
 - Many of these were very similar in nature
 - Some groundhog-day happening
 - 70% of incidents (28 of 40) access control
 - 90% of breaches (18 of 20) access control
 - Hygiene (living with poor hygiene is another indicator)
 - Updated systems and software in our data centers?
 - 53% (186,372) data center servers are running non-compliant kernels
 - 12% (41,443) non-compliant operating systems
 - In general Poor visibility (difficult to clean things up if you don't know what/where things are)
 - Estimated 10% visibility across systems, services, clients (laptops, phones)
 - No centralized logging across engineering
 - Lack of positive control (looked at client side here)
 - No MDM on Tweep phones
 - These are personal devices that access Twitter sensitive information and used for authentication to Twitter networks
 - Lack positive control over software Tweeps can install on their work systems
 - Access Control - this has been raised to the board before (perhaps without stats)
 - Engineers build, test, and deploy directly in *Production (with live Customer data)*
 - Systemic challenges in Access control
 - 43% (2662) of FTEs have access to our production systems (makes sense since we don't have a development or staging environment and engineers work directly with production systems and data - but is still alarming)
 - 10% (>1000) FTEs have access to Advertiser information
 - Campaign data
 - (likely) Billing (bank and routing) and account information

*It's already
already suggested
already working on this
now they're
identified
visibility
taking subjects
off the
table*

*adversary
view*

Are we performing manual processes, repeatedly, instead of automating? Tactically instead of strategically? Yes

- Site ops demonstrated a ~10+ step investigation (repeated each time)
 - Action taken was cheap to adversary and expensive for us
 - Ban a fake account (opponent spun up new ones more cheaply than this banning process)
- Twitter Service handled 100% of 1st person safety reports (yay! - Sept 2020 though Feb 2021)
 - >80% of the responses (114,641 out of 133,812) stated "Not a violation"
 - Does this take the Customer perception into account?
 - Does this message to the customer they should change how they engage in public conversation?
 - Does this take the antagonist's view into account?
 - Does this message: You've found an acceptable/approved lane for harassment. Good to go.
 - Can we respond in ways that discourage adversaries and do not imply that the reporting party should self censor?

More manual v automated examples and also living with extra risk because not understanding the threat...

- In 2020 496 FTEs and 3,174 Contractors were de-badged (terminated)
 - There were a total of 1,477 days (FTEs) and 10,357 days (Contractors) where we knew the person was leaving but they still had full internal access to systems and data.
 - Any evaluation of inappropriate access patterns (insider threat, competitive intelligence, etc.) was not easily accomplished - it would have been manual
 - Poor visibility
 - Poor positive control
 - Poor logging
 - Remember the exposure (i.e. "so what")?
 - 43% of FTEs have direct access to production
 - Full copies of source code
 - 30% of FTEs have access to sensitive Advertiser information
 - Unknown amount have access to our finance data
 - "Leaking is the norm" - I have been told this repeatedly

teaching →

With all of the above helping to provide context around our environment, and some of what is slowing us down or making it difficult to execute on our strategy and operations, let me share the existential threat that surprised me.

- We have [REDACTED] data centers
 - This is not publicly known
 - Jan 6 - 22 there were threats against our data centers
 - Threat matrix of effect:
 - [REDACTED] data centers physically destroyed
 - Twitter unable to do business - full stop (not surprising)
 - [REDACTED] goes down (hard or soft)
 - Twitter continues to run out of [REDACTED]
 - [REDACTED] goes down (hard or soft)
 - Twitter operates, but impaired - and more impaired as time goes on
 - [REDACTED] data centers gracefully go down and come back up
 - We don't know - best guess is weeks to months to bring the service back online
 - We can't boot(?)
 - Known unknown we really should know
 - We had to consider these scenarios from the 6th onward
 - More likely when we removed Trump's account than when suspended
 - Insider threat during this period
 - Think about the above access control, hygiene, visibility
 - SVR (Russian Foreign Intelligence Service) and DPRK (Reconnaissance General Bureau)
 - Both engage in ransoming organizations
 - SVR supports disinformation operations
 - I can think of ways to move inside our systems and networks to get to our DCs and work towards rebooting - they can too (don't fix the specific symptoms - cure the disease)
 - Both organizations would find this ransom an extremely valuable lever to keep (use?)
 - We now know to understand this and quantify it (make it a known known)
 - We are standing up a new datacenter - great opportunity
 - Partly cloudy is a ways off for continuity of operations
 - But we can about this

Q4 2021 Privacy & Data Protection Report

As previewed in the [EOY Information Security Board Report](#), this report provides focus on our top Information Security Risks.

2021 Top Risks Review

Access Control & Exceptional Access to Production Environments

Twitter has a large number of employees with direct access to our production environment. Every Engineer joining the company is provided Production level access. For context about half of all FTE employees¹ are engineers. Best practices are for companies to only allow production access to engineers in very minimal amounts and only in extreme situations (temporarily). Development, test, and staging environments are where engineers should safely conduct the majority of their work. We do not (meaningfully) have such environments. Twitter performs nearly all of these functions directly in production--thus requiring engineers to have production access. Further to this broad access to production, there are several pockets of exceptional access risk². All of this is a-typical for security mature companies due to the risk associated with providing direct access to live customer data and the systems providing the service.

7714 FTEs

Strong access management to the various data and systems throughout Twitter's entire organization is the cornerstone to not only security and privacy but to enabling people to develop with velocity. To be explicit, access control is the primary line of defense to protect against a variety of threats (i.e., internal bad actors³, misuse of data for otherwise legitimate business purposes as occurred in SIM-28--*which were the root cause of the current FTC issues--accidental data spills, etc.*).

CHART HERE

While needed progress has not been made on the larger issue of production access and access to production data (charts above) there have been reductions in two smaller groups of specifically troublesome access rights: IPMI and fleetwide "god" mode access.

¹ Twitter was just shy of 8,000 FTEs as of November 30, 2021, with ~4,000 engineers. A random security issue with an employee or their account would yield credentials that could access production data on average 50% of the time.

² 320 people have superuser access across all systems and data within production and 250+ can remotely disable ("turn off") hardware within data centers.

³ See the statistics in the #Protect presentation

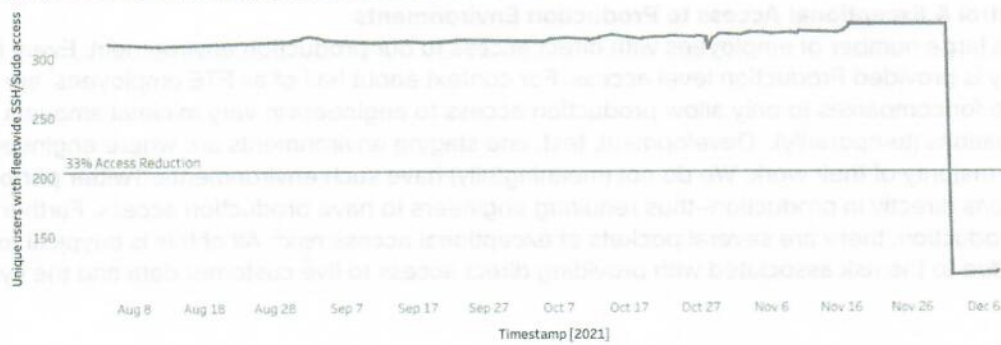


Extraordinary Access summary metrics over time

Number of unique users with IPMI access over time



Unique users with fleetwide SSH/Sudo access over time



These groups need to be essentially eliminated (there should be less than 1-5% of employees with these permissions) as they provide unfettered access across all systems within our data centers and all service data and processes. Initial access reductions in these two

12-28-2020: SSH 2765(46.7%), Hadoop 2855(48.2%), MESOS 2442(41.3%), (FTE: 5917)

4-20-2021: SSH 3133(48.1%), Hadoop 3086(47.4%), MESOS 2778(42.7%), (FTE:6507)

12-13-2021: SSH 3399(51.8%), Hadoop 3679(47.7%), MESOS 3534(45.8%), (FTE: 7714)

Security Management of Systems and Software

Our next top risk we must report to the Board is the state of our security configuration and compliance across our client fleet (laptops and workstations) and servers. We have ~9,000 laptops in our fleet. It had been internally reported that all (about 99%) of our clients correctly have security monitoring software installed on them.

Unfortunately this hid the critical aspect of what the security monitoring software was reporting. It has been revealed that more than half (58%) of our entire laptop fleet is out of security compliance, and one-third (33%) do not have software and security updates enabled on them.

12/13 graph



Fwd: time with Omid

Peiter "Mudge" Zatkan [REDACTED]
To: Peiter Zatkan [REDACTED]

Tue, Jan 18, 2022 at 7:20 PM

----- Forwarded message -----

From: Peiter "Mudge" Zatkan [REDACTED]
Date: Tue, Jan 18, 2022 at 7:20 PM
Subject: Re: time with Omid
To: Parag Agrawal [REDACTED]
Cc: Sean Edgett [REDACTED], Marianne Fogarty [REDACTED]

Privileged and Confidential

Thanks for setting this up Parag and thank you for the priority and concern. I would not have brought it up if I did not think it was important. The response of the Audit Investigation and Omid lead me to believe I was right to bring it up.

I was a bit surprised by your statement that you have been waiting over a month for a corrective document from me. My personal meeting notes and timeline shows that in December, on a phone call with you on the topic of [REDACTED], the materials, and the upcoming Risk Committee meeting, I offered to create a corrected document for the committee and you instructed me to not pursue this path. I was instead told to forward [REDACTED] documents and have [REDACTED] present, attempting to police the meeting and deal with [REDACTED] document as best I could in the closed session[0]. One week ago (Jan 11th) Marianne and I [REDACTED]

I am confused as to where your statement in front of Omid that you have been waiting on me for this for over a month came from. My apologies for any misunderstandings. If you could point me to where you requested this earlier I would like to figure out how to avoid such misunderstandings in the future.

In the future it would be great to have more than a few hours of notice for a meeting with board members to allow me to gather all appropriate materials and ensure I can provide pre-reads for attendees and yourself. I know that sometimes short notice meetings are just a thing.

kindest,

Mudge Zatkan

[0] And that on any items I still had concerns about you would make personal call(s) to the members to help clear up.

On Tue, Jan 18, 2022 at 1:15 PM Parag Agrawal [REDACTED] wrote:

Mudge

I have time with Omid at 2p pacific today, and given concerns you have raised — I think it is very important to use this time to speak directly with Omid and provide him your perspective and all details around the last risk committee meeting so he and the committee can be best informed.

Parag

lawyer letter sent to Twitter

Peiter Zatko
To: Patrick Pichette

Wed, Feb 2, 2022 at 5:20 PM

Hi Patrick,

I'm attaching the letter I had my lawyers create.

I also want to give you a preview of some of the things I'm putting in the documents I'm creating for yourself and the Risk Committee. It has become apparent to me that Twitter Information Security, and other areas, were not accurately characterizing the environment and the risks. It's important to me that a correct characterization of the risks and Twitter's environment are conveyed.

There's a lot of "present wins", "celebrate effort", and "avoid quantification and context" going on there. It predates me and it strongly resists change (even at the exec level).

I hope it is apparent that I am "still" trying to help. I joined because I felt an "attachment to mission". I'm different from a lot of the world that way. I had assumed that and the need to actually figure out what's going on under the hood at Twitter was why Jack tapped me.

Some examples of items I'm putting in to the "corrected q4 report" (I will make sure to send you copies directly):

I'm pretty sure that based upon your comments you were realizing parts of this at the same time it was becoming apparent to me.

E.g.

A)

I called out that engineers have access to production systems and data. Testing and development largely happens directly in production because there is no testing/development/staging environment to speak of.

At a risk meeting you correctly asked if this was normal, to which I replied "no".

51% of FTEs (more than 3k people) have production access. This is up from 46% at the beginning of 2021. This is outrageous.

B)

I brought in (actual) data for where Twitter was in their initial mock-SDLC roll out (Q1 or Q2). You were shocked and frustrated by the data. Apparently Parag and Mike Montano had been telling you for several years that "things were going really well" and that they were "really far along". Perhaps they had been providing qualitative descriptions of effort, because the data showed a very different story.

This situation appears to be uncomfortably common.

C)

In my first Risk Committee meeting you called out the need for straight forward dashboards. I entirely concur! I have never seen a company lack simple dashboards to the extent Twitter does. Twitter is also, it turns out, incredibly resistant to creating them. I pushed on these for the 12 months I was there. I received every excuse under the sun for why they couldn't be done, or were always "almost there".

It turns out that some of them already exist... but people do not want to advertise them. They stay hidden and are not briefed "upward".

For instance in Q3, after having been repeatedly denied accurate data, I personally went and gained access to the dashboards endpoint security (these are the 10,000 employee laptops). Whereas the CISO had been reporting that "we were in good shape", that was/is definitely not the case:

30% of the all Twitter employee computers (laptops etc.) are reporting that they have software updates "disabled" (this is Equifax level bad).

60% of the servers in the data centers are running non-compliant Operating Systems (often too old for support and also not being able to support basic data encryption requirements)

It's not surprising, with the raw data present, that we find (D).

D)

Of the (very) large number of incidents that Twitter experiences (~1 per week) more than 50% of the issues are related to access control and more than 25% are related to security configurations, patching, and software versions.

In closing, thanks for reading Patrick. Typing this up and knowing that someone who has a long standing reputation for ethical behavior is reading this is comforting. It really feels that at the end of all of this I had gone in acting in good faith, trying to help improve things but perhaps that good faith was not met with good faith in return. :(

The circling of the wagons and all the rest of this is all very disturbing (including the NYT "leaks" and follow ups from Parag describing a less than favorable exit.

I'm still looking to do the right thing for Twitter and also amicably close this out.

kindest,

Mudge Zatko

PS - In Privacy Engineering and Twitter Service there are dashboards now! Unsurprisingly these two organizations made more progress while I was at Twitter than they had across multiple years prior. They are now set up for long term success. :)

 Marianne Fogarty 2-2-22.pdf
57K

Jan 12, 2022 - Mudge

WARNING: The discussions and documents covered in this meeting may contain sensitive content. Please add a **WARNING** label to any items in this document that may result in a discussion (or a link) containing sensitive content.

Confidence-Staff Meeting & Notes

Confidence-Staff Working Norms

- Don't boil the ocean
- Don't move goal posts
- Obsess only on highest priorities
- Whatever we do is informed by data
- Right people in the right seats

Confidence-Staff Decision-Making Norms

- We make data-informed decisions quickly, and immediately push them down/out, sharing the why and making roles and responsibilities clear.
- We follow DACIN best practices: one approver, 10 or less consulted, and consultation not consensus, and we learn from our decision outcomes.
- We focus on quantifiable and measurable global impact, and how it relates to the mission of serving "the public conversation".

Confidence Org Foundational Docs (*located in [Confidence-Staff Shared Drive](#)*)

[Internal Webpage](#) | [High-level org chart](#) | [Cross-functional partner rolodex](#) | [Staff Slack Channel](#) | [Confidence-org Shared Drive](#)

IMPORTANT NOTE: Each entry should have a WHAT, WHY, EXPECTED RESULT. And for Accomplishments or Learnings, please include ****QUANTIFIED**** data with context. Please ensure your updates answer the following questions:

- Definition of problem and why it has to be solved (and how people outside confidence will understand/feel it)
- What are the numbers this week and out of what context?
- What were they before, what are the next milestone's numbers (and when) and what is the finish line (and when)?
- What impact does all of this have to others and what is the aggregate impact?
-

CELT (Bi-Weekly 30 mins)

- *N/A this week*

Mudge (STAFF/Org items to share with Team)

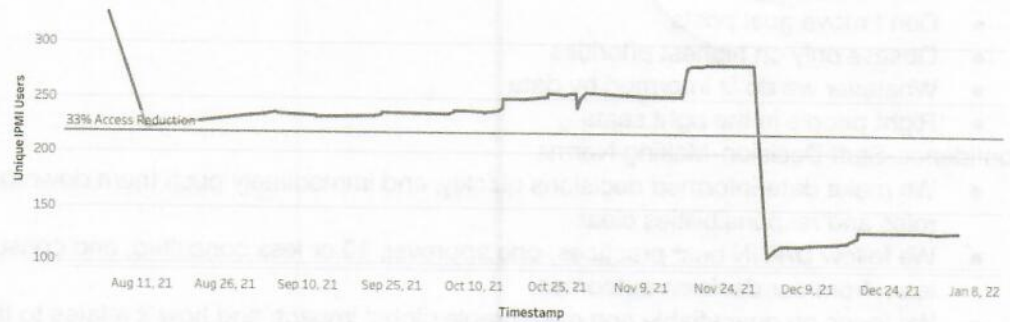
- Confidence 5 Guiding Principles
 - Don't boil the ocean
 - Don't move goal posts
 - Obsess only on the highest priorities (Global View)
 - Whatever we do is informed by data
 - Right people in the right seats
- Obsess only on the highest priority (Global View) discussion - mudge to provide examples and solicit feedback for next week
- Log4j
- Danielle - ratings curve

InfoSec

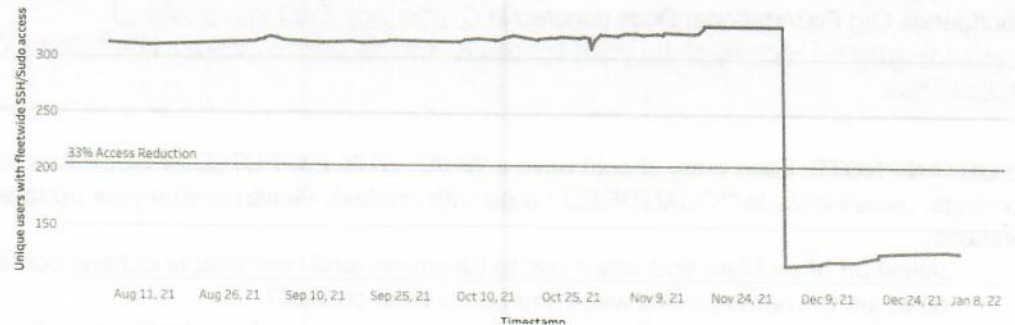
- Your top priorities for the initiative for the week, why, and expected result(s):
 - Interviewing for Sr. Director, GRC as Joseph's backfill. Currently interviewing and considering 4 candidates with interviews and huddles this week and next week.
 - As part of annual performance reviews, performing InfoSec calibration with HR
- Something you accomplished or learned in the initiative (include data w/context):
 - Anchor Charts per Mudge's request:
 - IAM:

Extraordinary Access summary metrics over time

Number of unique users with IPMI access over time

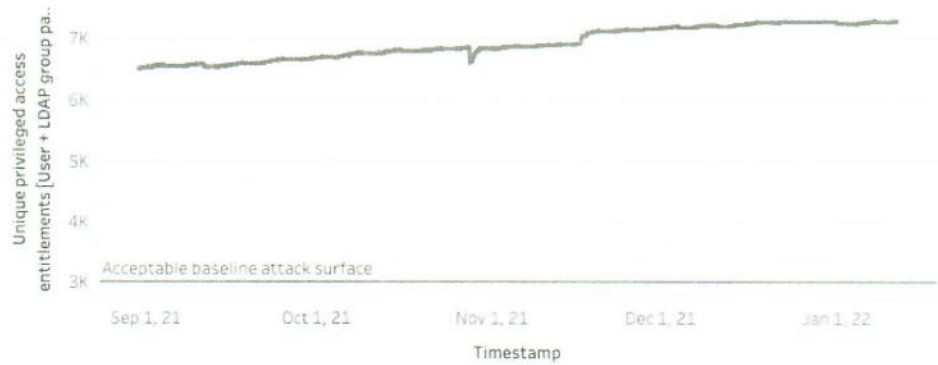


Unique users with fleetwide SSH/Sudo access over time



Privileged access summary metrics over time

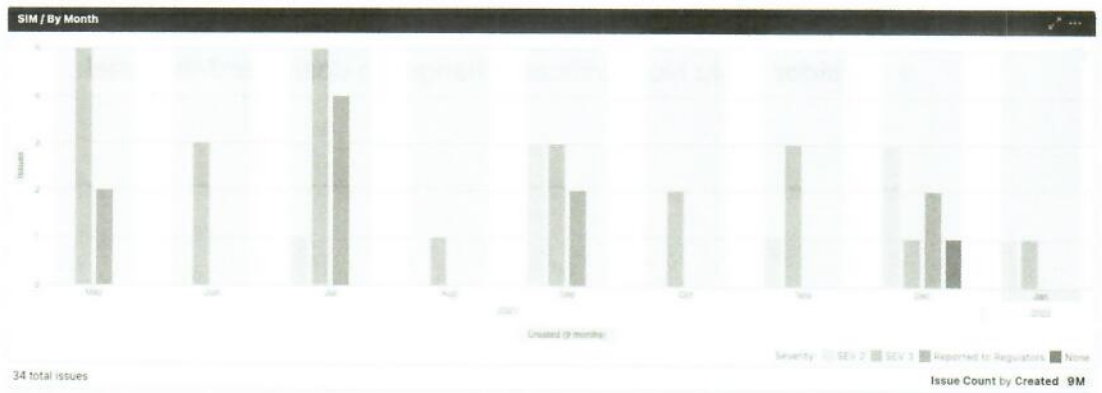
Unique privileged access entitlements over time



Unique privileged users over time



-
- Through a collaboration with Platform Eng, we are aggregating all of the Reduction of Production Access into a centrally coordinated effort. This was decided before the holiday, but due to PTO and other priorities, this kicks off more regularly starting next week. There will be weekly meetings, and once there is a full list of related projects and reporting cadence established we will share. We are planning to provide more info by the first of Feb.
- Incidents:

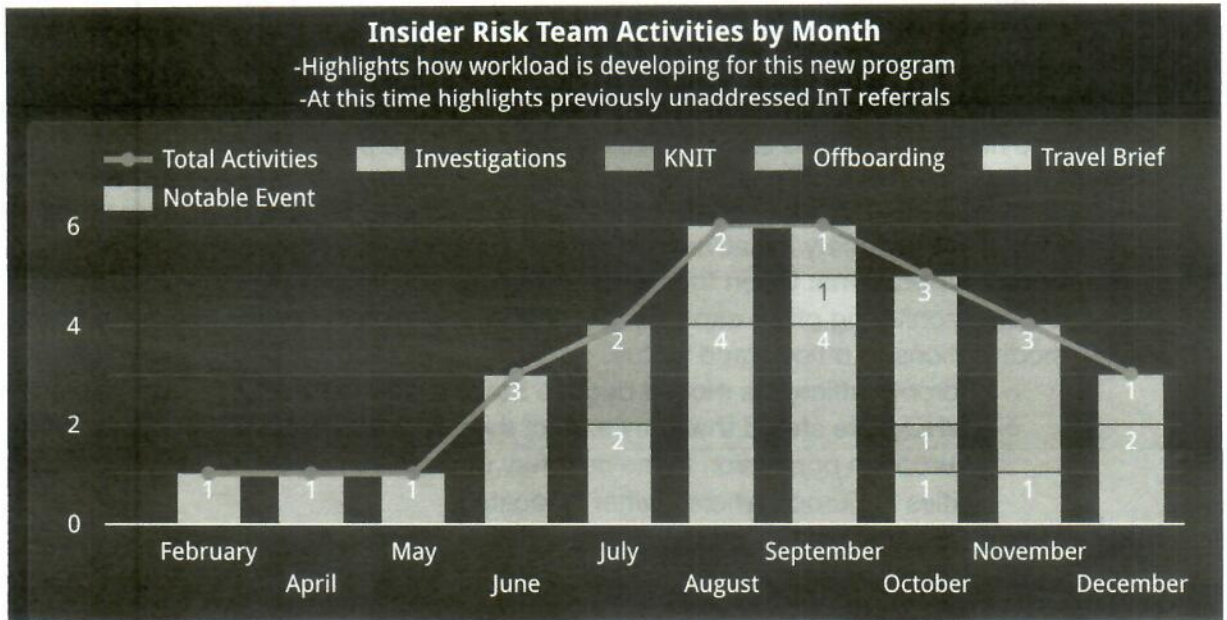


-
- [Jira Dashboard](#)
- [Weekly Report Document](#)
- Log4j Update:

- The most critical CVE (CVE-2021-44228) that was announced on Dec9th is patched on the Fleet.
- There have been subsequent CVEs released as the security community pokes and prods at Log4j, which was expected. This includes CVE-2021-45046 and CVE-2021-44832. The latest patch, 2.17.1 was released on Dec28th, however this CVE has a lower severity score since it requires elevated privileges (access to the logging config file). We have created tickets to the respective teams and are tracking the 2.17.1 patch rollout. We have pushed everyone who maintains internal code and uses log4j 2 to get to at least 2.17 which deals with all the high+ severity vulns
- We have completed our initial impact assessment to determine if Twitter was targeted before the 0 day was known and/or if the vulnerability was exploited at any time. We believe based on the logs, that Twitter was not impacted nor targeted prior to early December 2021
- There is ongoing work related to third party/vendored systems that have upstream dependencies on log4j (e.g. things like Atlassian, vmware, etc). IT (Nick Fohs) is tracking those cases, but we're largely dependent on the vendors pushing out a patch which we will then install via our standard patch mgmt deadlines.
- Endpoints:
 - InfoSec and IT have collaborated to update the Uptycs endpoint health dashboard. We are in the process of quality checking this with teams and will work to share this regularly as an anchor chart in the next 2 weeks.
 - Something you need help with or a risk to flag:
 - Nothing at this time
 - Hot Topics:

CorpSec - CorpSec Dashboard

- Your top priorities for the initiative for the week, why, and expected result(s):
 - **Insider Risk:** No significant changes in dashboard this week.

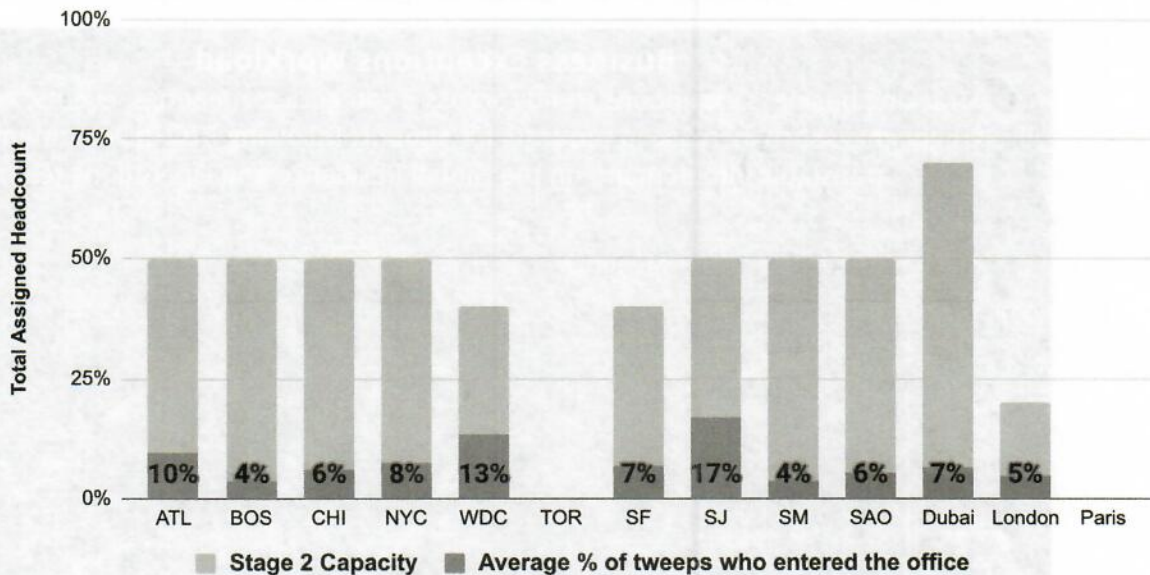


- **Global Resilience:** The Business Exception dashboard indicates a massive decline in meetings, events, and travel exception requests, likely due to the holidays and the onset of the Omicron variant. Tweeps/Leaders have been self-regulating and canceling large events.
 - Opening offices to traveling Tweeps has reduced the number of exceptions requesting access to offices to which Tweeps are not assigned
 - We expect a further decline in international travel requests as the company is intending to permit international travel with updated guidance in the near future.



- **Global Resilience:** Global comms are being drafted and are expected to go out this week regarding return to work among changing health standards from CDC and country regulations. Work.com functionality is being updated to accommodate noting boosters in the system.
- **Global Resilience:** Training the remainder of KSA Team on High Risk Travel this week (those who have not traveled since COVID restrictions were put in place)
- **Global Resilience:** Go/Learn is migrating to go/learning. In the process, we requested regional onboard training (post-Flight School) to be mandatory in go/learning. This training supports the #PRO Tweep Safety Initiative Strategy
- **Global Operations:** Given the increasing cases associated with the Omicron variant, we are closely monitoring office capacity, positive cases/contact tracing and the potential for modifications to office stage status.
 - Toronto office has moved back to Stage 1 due to new City of Toronto guidelines.
 - WHO have stated that they expect the Omicron variant to infect over 50% of the European population in the next few weeks. Numbers are rising significantly across all cities in Europe where Twitter is located.
- **Global Operations:** Return to Office Metrics

RTO Daily Average Attendance (Week of Jan 3 - 7)



- **Special Security Support (S3):** Threats have decreased this past week, 613 to 371, with 14 investigations and one case opened. Agent tools were used one time.
- **Global Risk Intelligence:** Threat Fusion Center to host an analysis roundtable with XFN partners to focus on the Russia/Ukraine situation to align on the current threat assessment.

- **Global Risk Intelligence:** Leading hotspot analysis to proactively address potential concern related to the conflict in Ethiopia, political instability in Sudan, civil unrest in Kazakhstan, and political violence in Myanmar.
- Something you accomplished or learned in the initiative (include data w/context):
 - **Insider Risk:** Offboarding pace that we are tracking via JIRA is about 30/week - out of those, one offboarding (.33%) was concerning due to user activity. Without approval to look at document content, no way to fully vet. But, we have a sense of volume, a sense of how many will need further follow-up, and what authorities we need.
 - **Global Operations:** Contact Tracing January 3-9, 2022:
 - AmWest
 - New cases for the period - (3) 2 SF, 1 San Jose
 - Total cases Q1 - (3)
 - Total persons contact traced for the period - (20)
 - Total persons contact traced Q1 - (20)
 - Number of man hours spent for the period - (7.5)
 - Number of man hours spent Q1 - (7.5)
 - AmEast
 - New cases for the period - (1) NYC
 - Total cases Q1 - (1)
 - Total persons contact traced for the period - (3)
 - Total persons contact traced Q1 - (3)
 - Number of man hours spent for the period - (2)
 - Number of man hours spent Q1 - (2)
 - Data Centers
 - New cases for the period - (3) PDX, ATL
 - Total cases Q1 - (3)
 - Total persons contact traced for the period - (12)
 - Total persons contact traced Q1 - (12)
 - Number of man hours spent for the period - (8)
 - Number of man hours spent Q1 - (8)
 - APAC
 - New cases for the period - (0)
 - Total cases Q1 - (0)
 - Total persons contact traced for the period - (0)
 - Total persons contact traced Q1 - (0)
 - Number of man hours spent for the period - (0)
 - Number of man hours spent Q1 - (0)
 - EMEA
 - New cases for the period - 0
 - Total cases Q1 - 0
 - Total persons contact traced for the period - 0
 - Total persons contact traced Q1 - 0
 - Number of man hours spent for the period - 0
 - Number of man hours spent Q1 - 0
- Something you need help with or a risk to flag:
 - **Global Operations:** Discussion with the COVID Working Group on a clearer Staff statement on Omicron risks, updates, and effects to Twitter events, travel, and RTO.

- **Global Operations:** France LCS lead is due before the Versailles Magistrates court on next Monday the 17th January as a follow up to his last court appearance, which was adjourned. Our team will be supporting on the ground.
- **Global Operations:** EMEA team is supporting a Ghana tweep on travel to the Africa Nations cup in Cameroon - this is proving to be challenging from a logistical and risk perspective.
- Hot Topics:
 - NA

IT

- **Your top priorities for the initiative for the week, why, and expected result(s):**
 - **go/ftc-records Program:** Review application inventory review in order to prioritize Corporate Business Systems that store Customer Data for alignment with go/eraser.
 - 379 applications are registered in Collibra with:
 - 88% (333 of 379) are First Pass Review “Complete”
 - 40% (153 of 379) are identified to be P1 - Critical for storing customer information.
 - **Next Step:** Meet with Lea K. on BigID vendor selection and product concerns
 - **IT SYS:** Data retention policy for CrashPlan EOL 2/28 being worked on with Michael Miller in legal, aligned on dovetailing with Privacy data policy drafted by Lea (thank you!)
 - Dashboards and monitoring being set up with baselines in Zabbix, with project tracked here by the team. Anticipated outcome is greater transparency into environment health and performance, **aligning with hardware refresh and capacity planning within Twitter data centers, office locations, and potential global PoP presence to mitigate in-region application latency issues reported.**
 - **EUS IT:** All-hands 1/13 to welcome the team back (~70 people), discuss priorities and 2022 roadmap plans (thank you Renee for your help and support upleveling this session!)
 - Focusing on capturing mttR, MTTE, CSAT, ticket volume, ticket age by region with anticipated dashboard v1 February timeline
 - Reviewing ITOB (IT Onboarding) ticket queues, reviewing types of tickets and trends opened in new Tweep’s first 30 days post on-board, identify trends and drive roadmap improvements for CPE, IAM, CID, Helpdesk and IT teams who own the service
 - **IT:** Org structure and definition of functions per pillar prior to 1/24 on-boarding of new VP of IT
 - **Position Management:** Implementation of Workday Position Management tool across Twitter to automate and improve the process of planning, managing, and tracking positions.
 - Currently annual headcount planning is done by FP&A using spreadsheets
 - This is the first time we are migrating the manual processes for position management to an automated tool. Approximately 75 to 80 % of automation will be targeted during the first launch.
 - Team is refining functional architect decisions, change management and initial data load, and integration design.
 - Project is on track for Go-live at the end of Q2 2022.

- **Continuous Improvement - Change Management:** First phase of aligning ITSM practices with the ITIL framework. This will help build the foundation for operational rigor and will enable IT to begin generating standard performance metrics in Q2'22.
 - Process flow has been approved by the IT leadership team, currently leveraging IT CAB process with reviews on Tuesday and Thursday weekly.
 - Documentation delivered to IntelliSwift to begin creating training video.
 - Jira Service Management enhancements currently in UAT.
 - Target Dates (currently on track):
 - Late January: Socialize across IT; reconnect with Platform Engineering on potential alignment
 - Mid February: Jira enhancements go-live
 - March 1: Video training available (and trackable)
 - March 30: Mandatory IT training complete
 - April 29: Generate initial Change Management performance metrics

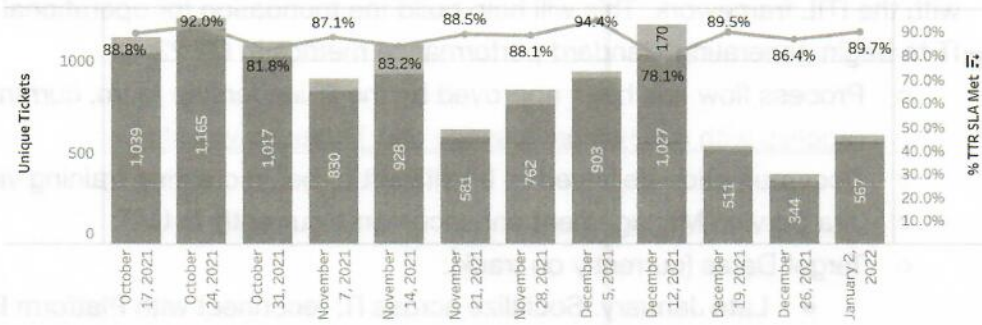
- **Something you accomplished or learned in the initiative (include data w/context):**
 - **Project SAFARI:** Oracle EPM Cloud migration ([ORACLE-25640](#)) - Oracle / Microsoft bug:
 - SSO/2FA is a company policy prerequisite for Oracle EPM Cloud go-live. We encountered an Oracle Smart View bug that prevented us from enabling SSO/2FA to go live in November 2021. Oracle Development confirmed the issue was actually caused by a Microsoft bug which prevents Oracle Smart View from working correctly.
 - Oracle provided Twitter with a new Smart View version that takes advantage of a more permanent and stable way of integrating with Microsoft that effectively bypasses the aforementioned bug.
 - Jan 7th: IT enabled SSO/2FA successfully for approximately 50 users and 15 developers as planned.
 - Late January 2022: In this milestone with Enterprise Data Management go-live, we will add approximately 30 users to the Oracle EPM Cloud with SSO/2FA.
 - Mid February 2022: Our final milestone is to go-live with all remaining EPM Cloud modules (FCC, PCM, Planning) for accounting and FP&A. An additional 200 users will be onboarded onto the Oracle EPM cloud applications.
 - Project is tracking green both on budget and schedule.

 - **Log4j Remediation:** Log4j is a vulnerability that required mitigation steps across MuleSoft and Tableau
 - Mulesoft: Applied mitigation steps provided by MuleSoft's Help Center across our VM servers for Prod, Testing and Staging environments.
 - Tableau: Applied mitigation steps across Tableau Production and DR environments and Tableau Desktop Windows VM ware
 - User Impact: Requested the 200 Tableau Desktop/Prep users uninstall and re-install the latest version which takes approximately 20 minutes

 - **EUS:** Metrics for week ending 1/7 linked here. Highlights:
 - **HD Tickets Resolved: 567**, trending up as Tweeps return from the holiday break. Week over week trending:

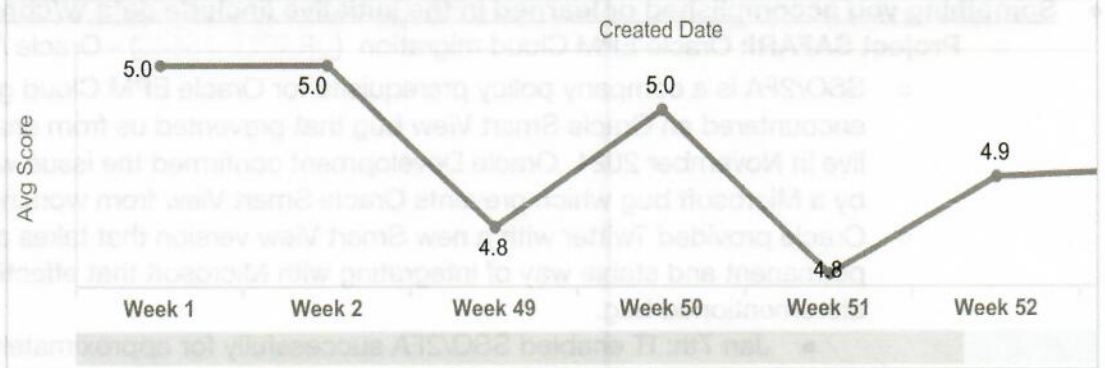


Breakdown of IT Tickets Closed

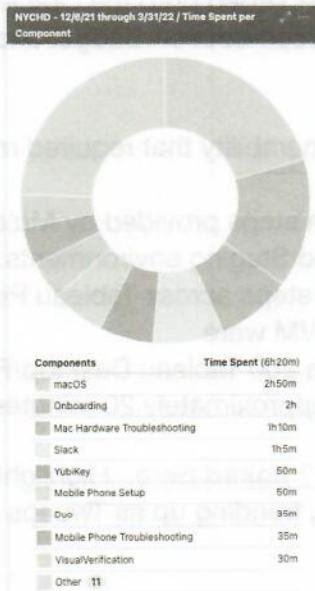


- **mttr (mean time to Resolve): 89.7%**
- **CSAT: 4.9 out of 5 with 93 total responses**

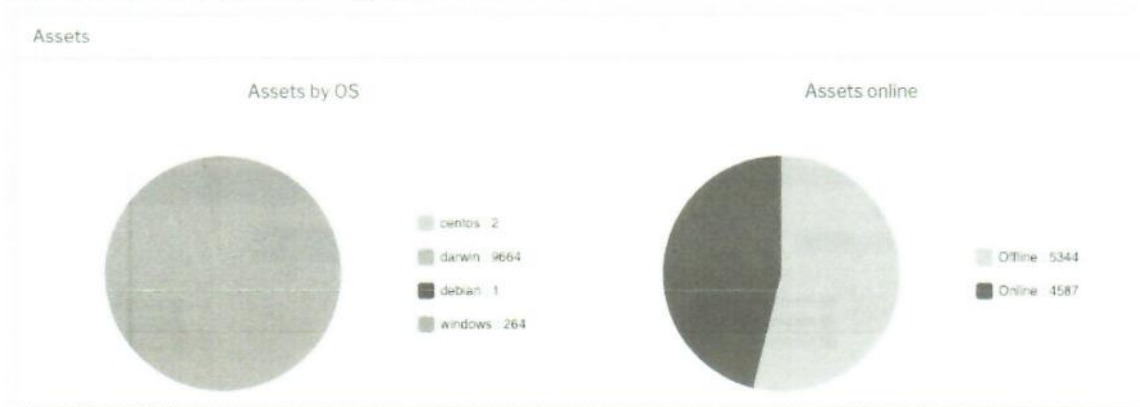
Average Score by Week



- **EUS IT:** Streamlining address collection from new Tweeps to reduce manual steps, met with Workday and People Operations team Monday to determine next steps. Anticipated savings of 40 hours/week for helpdesk technicians, structuring helpdesk dashboard reporting similar to NYC Helpdesk:



- **ITAM:** Purchased additional O365 licenses, migrating away from local MS Office installs, as part of overall risk mitigation strategy for overall desktop/endpoint fleet
- **CPE: 9,934** assets total (Windows, Linux and MacOS). Anticipating increased device check in with M1 HW refresh and migration in mid-Q1

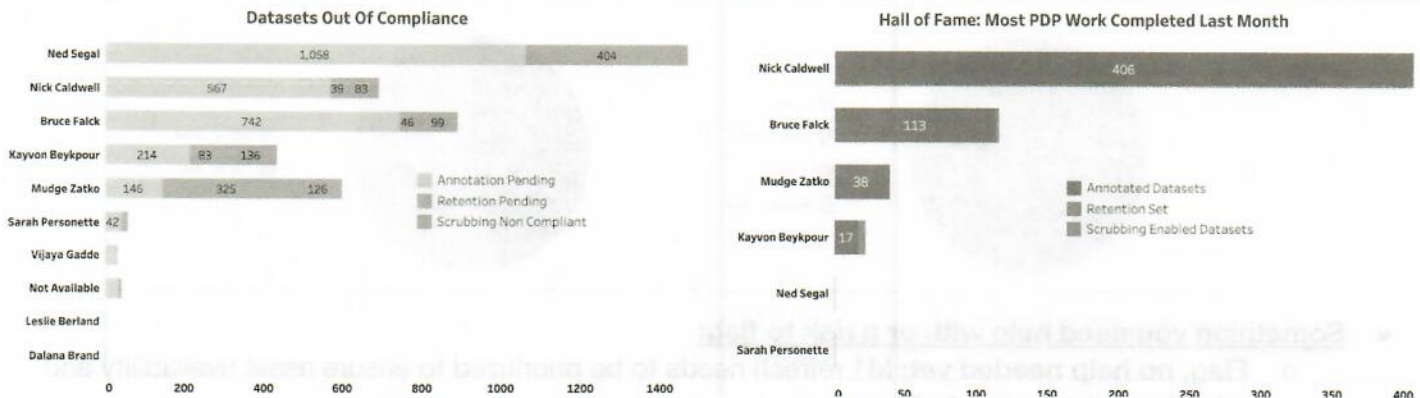


- **Something you need help with or a risk to flag:**
 - **Flag, no help needed yet:** M1 refresh needs to be prioritized to ensure asset availability and prioritized distribution to Tweeps
- **Hot Topics:**
 - **EUS:** Helpdesks in NYC and SF closed week of 1/10 and potentially 1/17 as well due to positive COVID tests for technicians in both locations, following Corpsec guidance.
 - London - REW has yet to install the screening in front of the helpdesk, hoping it will be ready by next week.
 - USA - Current pandemic situation has put a strain on our CTW strategy and may cause us to rethink long-term CTW placement

Privacy

- Your top priorities for the initiative for the week, why, and expected result(s):
 - Snickerdoodle:
 - V1 banner:
 - Binary consent banner went live in France for ~30 minutes from 3:53 pm CET/6:53AM PST to 4:20 PM CET/ 7:20 AM PST on Dec 13th. Banner launch was rolled back around 4:30 PM CET/ 7:30 AM PST due to issues called out in this [doc](#)("Issues identified during Launch" section).
 - We have resolution and a path forward for all issues identified during our launch. Our new tentative schedule for re-launch in France is at 4pm CET / 7am PST on January 25th 2022
 - Next up:
 - V2 banner: with more choices for end users
 - Updated banner on other sites (e.g. help.twitter.com)
 - Scanning to detect errors and avoid having to do this whole project again
 - Eraser
 - **Dataset Scrubbing Adoption**
 - Good State Data Overall: 89.5%
 - Mysql: **94.4%** was 93.9%
 - Manhattan: **91.6%** was 85.3%

- Vertica: **58.6%** (was 58.3) We believe we will see a significant increase once the DRET (legal tickets which approve extended retention e.g. for financial records) integration is complete and the Dashboard excludes these from the metrics
- Continuing to work with Teams to adopt scrubbing tools for platforms where tooling is available with the additional scope of driving annotation and retention to get active data in a good state.



- **Default Good State Storage:** Staffing plan approved; however, urgently working out headcount accounting this week. Teams will start to hire asap, We still cannot meet EOQ2 due date to complete tooling. HDFS, BQ, GCS confirmed their delay. Hoping to retarget for 7/2022.
- FTC Misrepresentations – stay tuned; legal finalizing survey, then we’ll move forward to figure out where we have gaps between what we’re doing and the promises we’ve made, then work to remediate

- Something you accomplished or learned in the initiative (include data w/context):
- Something you need help with or a risk to flag:
- Hot Topics:

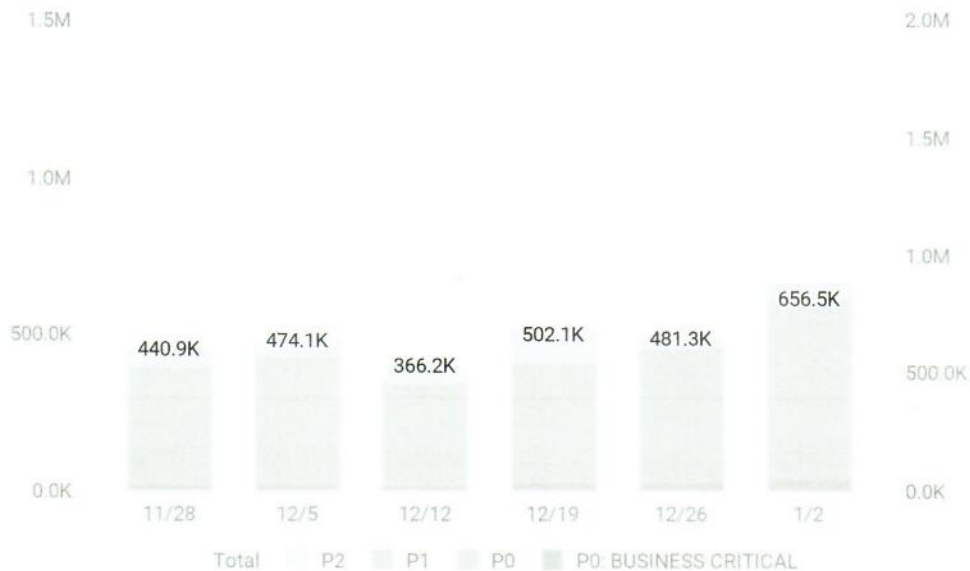
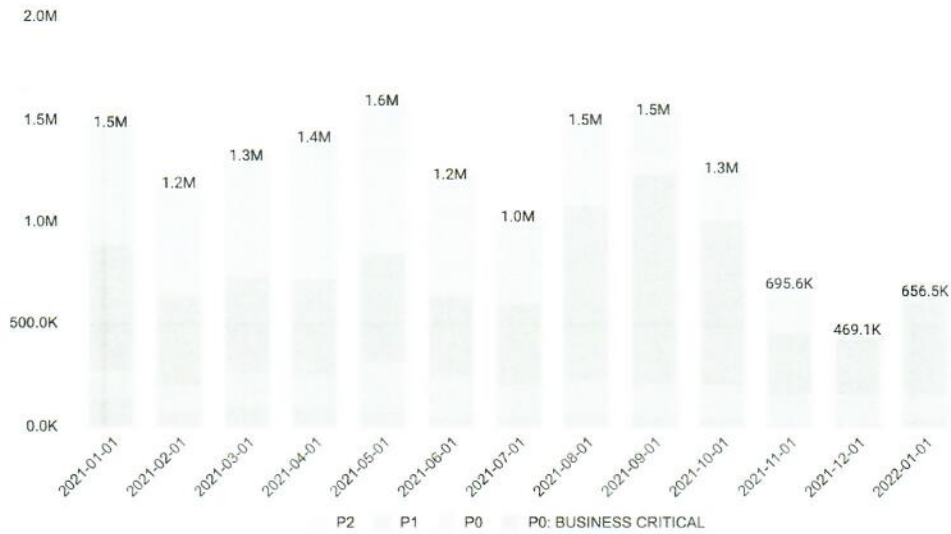
TWS

1. Your top priorities for the initiative for the week, why, and expected result(s):

- We have seen sustained growth in backlogs but with new drivers this week. Overall we have two issues that are overarching - we are seeing agent outages due to COVID and agreed in our Tuesday review to shift to working on proactive options vs waiting until we reactively size the impact by queue. We will update next week on what we find. And then we have seen large week over week variances in incoming cases with multiple causes below. We are also working to confirm how much of the increases exceeded planning forecasts but, as noted, in the past workforce management is a key missing capability in TWS that we are working to build. The specific issues in addition are:
 - i) GDPR: 12k, 171% WoW increase. Driver: There was a new product launch of Age-Gating Adult Content and as historically this queue had not shown much volatility with launches we did not dig deeply enough to determine if there could be an impact. As this launch will continue to roll out, we are course correcting this week and will attempt to estimate where this will land and work on staffing options.
 - ii) Hacked accounts: 7k, 408% WoW increase in backlog. Investigating the root cause here. Potential driver: 36% WoW increase in incoming volume with a major spike on Jan 8 so there may be a new spam scenario and we only have reactive measures in our tool kit.
 - iii) Safety Core: 12k, 71% WoW increase as agents moved to PPI as PPI is a higher priority but PPI appears to have a new spam issue related to Turkey where again our only tools are reactive manual filtering.

- iv) STT (spam): 33k, 31% WoW increase. Driver: a ropo bot has been sending high volumes to STT and we need to track down what changed.

Monthly Trend: Avg Case Backlog for Twitter Service



○

○ Note these metrics and definitions are still evolving, e.g., this includes cases that have a shot of being worked within SLA.

Something you need help with or a risk to flag:

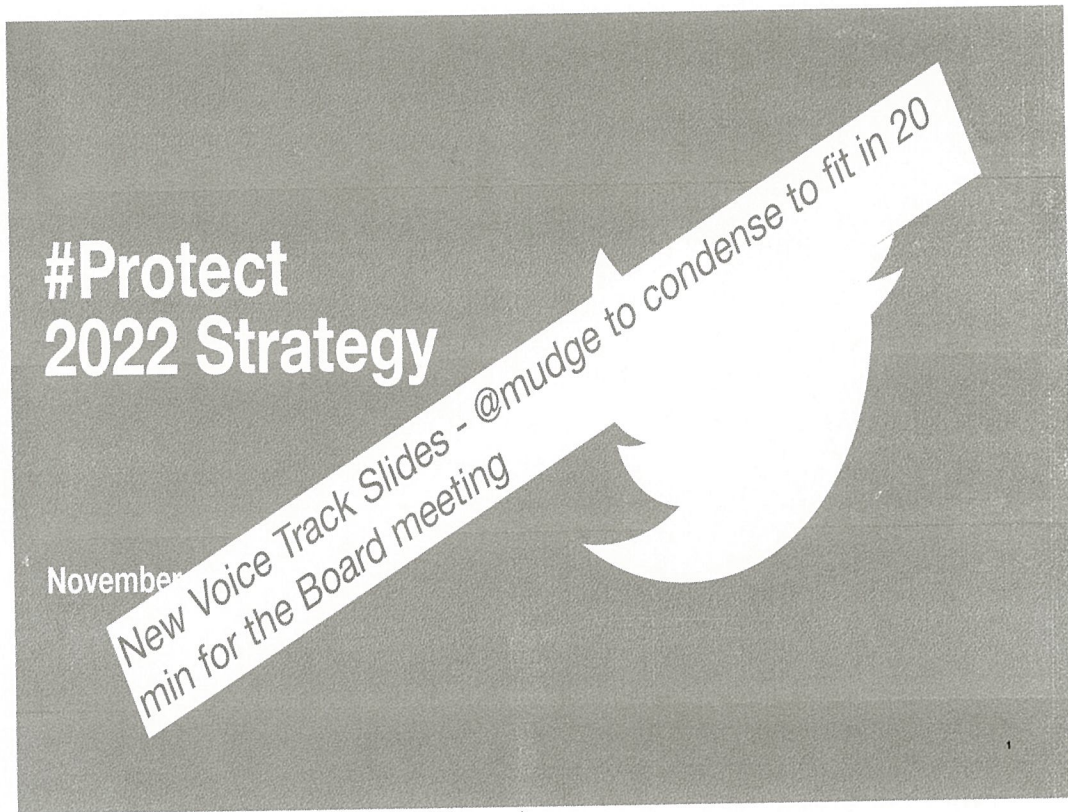
- SPAM - per above
- COVID Omicron vendor staffing impact per above
- As reported before the holidays, the right to privacy policy was expanded in the US November 30. We saw some large high profile escalations when we removed posts under this expansion and had to suspend and regroup. We are in close partnership with the policy team to continue the clarification of the policy and update the agent training and documentation. We were able to turn on a small group of super agents that have been working cases with special support and they have cleared the backlog and continue to work new cases. We are training a few more agents this week to start to learn more about

how we train at scale for these types of cases. Between Dec 17 and Jan 11, we reviewed 1637 cases and found 261 cases to be in violation. An action rate of 13.2%.

- We are starting to see a trend in our verification escalations. Malicious actors are taking over dormant/inactive handles from reputable news sources that are still published on the sites, and they are using this for their verification requests. They then in turn use these verified accounts for malicious activities on the platform. There is no easy way to catch these unless they are sent directly to us. We have started investigating possible ways to mitigate but this is a nuanced and tricky problem to solve.

Around the Room

- <Text>



#PRO Strategy Doc (3-page doc for the Board)



Agenda

1. #Protect (#PRO) - setting context & background
2. How we'll measure success
3. Initiative Strategies
4. Phased launch plan

30 sec

#PRO

*The #Protect (#PRO) Objective's mission is to
protect Twitter and it's Customers
from people who could cause harm to us or
do harm to others through our platform.*

[1 min for this slide]

And that's the mission: protect Twitter and its Customers against the people who would cause us harm or do harm to others through our platform.

That's the "why" and the difference in "focus" of Protect-objective in 2-ish slides.

Into the "what and how"



Why we needed to create #PRO

Setting context and background

Challenges



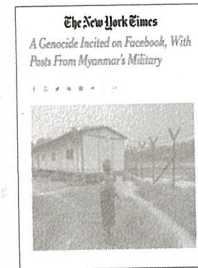
Obligations



Platform as a weapon

IO intends to cause real world geopolitical impact (instability & harm).

13 countries detected running Information Operations on our platform.



Snapshot in Time

- 40 Security Incidents
- 20 Reportable Breaches
 - 70% of incidents and 90% of breaches related to access control
 - ~243,267,594 customers impacted
 - ~26,897 employee accounts involved
- [REDACTED]
 - [REDACTED]

Hacks



[3 min for this slide]

For a broader audience each of these could be a slide - walking through the issue, the threat, and who/what we are protecting ourselves and our Customers from and how well (or poorly) we are positioned to do so.

Here, we can go through these in 90 seconds.

Top Left

We are getting closer to the final FTC consent - an agreement that will last for 20 years covering everything we do with data as a company. Brought on by misuse of data from a lack of data protection and privacy controls. Other regulatory bodies are focusing on us... and more will be coming.

Top Right

We have disclosed 13 Countries running Information Operations on our Platform (that we know of!). ~~There are people intent on abusing the public conversation and FaceBook and Myanmar are the cautionary tale.~~ Facebook admitted they failed to stop their platform being used to cause a genocide. The Myanmar military assumed false personas and incited murders, rapes, and the largest forced human migration in recent history.

Bottom Right

The big twitter code-red hack wasn't sophisticated. It hit only one of several areas in which we have very large exposure. My understanding is that it delayed hiring for over a month and more than 6k person hours spent in the immediate wake of the event.

Bottom left - these are signals that we're sitting on top of other code reds.

There's no finger pointing. There are always tradeoffs and people make the best decisions they can. But a strategy of reacting and trusting that we are excellent scrambling in a crisis (we are) can't be our strategy going forward.

Let's look at the Twitter Objectives and how Protect's focus addresses challenges on this slide in a complimentary fashion to the other objectives. And in a way we aren't focusing currently.



Existing Objectives & Their Focus

- | | | | | |
|--|--------------------------------|-------------------------|------------------------------------|-------------------------|
| #Participation | #Durability | #Fundamentals | #Velocity | #Diversification |
| ✔ Content on the platform, customer tools and apps | ✔ Advertisers, revenue sources | ✔ Platform, Engineering | ✔ Our speed and agility to develop | ✔ Workforce |

[1 min for this slide]

Looking across our Objectives we have significant focus on ourselves our people and workplace, our content, our platform and how we do engineering, our customers and their choices, and our partners and revenue sources.

I look at all of this as focusing on the spirit of #OneTeam

Unfortunately, not everyone is on #OneTeam and we need to focus on those people that aren't as well.

This is why #Protect is complementary.

Focus on the good things & people we want to help



Existing Objectives & Their Focus

#Participation

- ✓ Content on the platform, customer focus

#Durability

- ✓ Advertisers, revenue sources

#Fundamentals

- ✓ Platform, Engineering

#Velocity

- ✓ Our speed and agility to develop

#Diversification

- ✓ Workforce



#Protect focuses on the other side of the coin - the threat

[1 min for this slide]

Protect focuses on the baddies.. the threat. the other side of the coin.



How We'll Measure Success

Through 3 North Star Indicators

#PRO North Stars



Get the basics down and be excellent and efficient at them

None of our security or privacy & data-protection breaches happen due to our own poor hygiene or lack of due-diligence and follow-through.



Make our opponent's life more difficult, not ours

It is measurably and increasingly difficult (expensive) for bad actors to complete their objectives inside Twitter or on the Twitter platform. While doing so, we lead the industry in making the opaque fields of Security, Privacy, and Information Operations transparent and easier to understand.



When we expand the business we don't need to worry(*) about security and privacy

When we open new offices, data centers, products, and revenue lines we aren't overly worried about security, privacy, and human risk. Twitter's privacy, security, safety exposures grow more slowly than the expansion and growth of Twitter.

* We always need to think about them and be smart

[3 min for this slide]

3 North stars and their indicators:

1 Basics - operational rigor, efficiency, and hygiene across privacy, security, safety. No more "ewww that was an embarrassing breach that we shouldn't have had". "Get rid of these large **basic** security and privacy gaps and exposures we have."

2 Make the baddies fire us because Twitter is a hard target, unrewarding, and risky for baddies to get involved with us - all the while ensuring our Customers hire us and it's easy and efficient for *us* get *our* jobs done. Make the safe and secure thing to do the easy and fast thing.

3 When we expand the business - launch new products, revenue streams, open offices in new countries, etc. We shouldn't be continuously re-exposing large areas of risk we don't need to. Our risk should grow minimally in comparison to the growth of our company. For example opening an office in a new country and putting engineers there shouldn't be as concerning as it is because of the exceptionally broad access all of our engineers have to critical systems and sensitive customer data services. (Our peers aren't carrying this repeated exposures the same way we are)

#PRO Key Results (KRs)



Get the #Protect basics down and be excellent at them!

None of our security or privacy & data-protection breaches happen due to our own poor hygiene or lack of due-diligence and follow-through.



End of 2022 KR:

- The number of **internally identified vulnerabilities** goes up by at least **50%**, while the number of **externally identified vulnerabilities** goes down by at least **20%**.
- **100%** of newly launched products & features are in **compliance with our SDLC** and contain a valid threat model. More than **20%** of legacy products & features are in compliance with SDLC requirements.

[2 min for this slide]

Increasing the number of vulnerabilities we are internally finding proactively -> external discoveries start going down... meaning we are better at finding these issues before others find them and we are addressing them

AND

quickly get to 100% of our newly launched products and features launch in compliance with security/privacy and SDLC regulatory ...

...It's more critical, initially, that we not launch new vulnerabilities and incidents than it is to go back and find and address all of the problem areas in legacy products and features.

Demonstrate to regulatory agencies that we have improved processes and procedures in place moving forward.

if we keep launching new security & privacy incidents after we have said we fixed our broken ways - we have ourselves
punishments will be refused to complete security & privacy (data off limits), fines: penalties, reduced access to markets

#PRO KR's



Make our opponent's life more difficult, not ours!

It is measurably and increasingly difficult (expensive) for bad actors to complete their objectives inside Twitter or on the Twitter platform. While doing so, we lead the industry in making the opaque fields of Security, Privacy, and InfoOperations transparent and easier to understand.



End of 2022 KR's:

- We complete a **full inventory** of our manual and automated moderation and security detections and responses.
- Twitter has implemented **baseline language capabilities** to meet existing human-support SLAs for **100% of the top 20 languages**, and are also able to onboard new human support within 2 business days a major event.
- Twitter's **transparency efforts**, including a regularized, risk-managed cadence of Adversary Activity disclosures and a first-in-the-industry annual "Threats to the Public Conversation" Report by no later than Q3 2022, lead to an **observed change in the behavior of tracked bad actors** (whose activities we disclose/discuss in the report).

KR NS 2

complete inventory of ~~man~~ moderation & security actions?
mini-max it (Fi-Fi)

[3 min for this slide]

Building a full inventory of all of our security detection and response activities -- whether formally captured playbooks or adhoc and institutional knowledge. We need to know not just the things we are and are not doing but also the things we have automated and do manually (~~hint - a lot of manual solutions presently -- while the bad actors are doing more and more automated and at scale~~). mini max

Protect needs to do it's work across all languages - because the people who don't have Twitter's best interests in mind work across languages. Using a language other than one of the few we presently have good coverage across can't be a free pass for bad actors on the platform or in our threat safety work. !English is not a free pass

It should become obvious how growing language capabilities will be valuable across the company.

The third measure in north star 2 starts by expanding the company's current work in transparency around Information Operations.

One way to drive up "cost" to bad actors is to shine a light on them and their activities. We start by measuring changes in their behavior due to experiments with the types of information and light we shine on them.

- Game Theory

fightback

#PRO KR's



When we expand the business we don't need to worry* about security and privacy!

Twitter's privacy, security, safety exposures grow more slowly than the expansion and growth of Twitter.



End of 2022 KR's:

- 100% of production access by Tweeps is logged and routinely audited to ensure access is limited to appropriate and validated business needs.

- Legacy persistent **privilege grants** are drawn down to 33% of Q1 2022. (Baird)

- Independent auditor identifies **all promises we have made** to customers as it relates to how we handle their data and information, and we have determined where we are and are not in compliance with what we have promised.

Not afraid to have engineers anywhere: NSA, India, France

[2 min for this slide]

The measures start with understanding these repeated large concerns and exposures we keep fretting about over and over again with new offerings and expansions.

Drawing down some of the key legacy exposures and getting an independent auditor to identify all of the promises we have made to customers around how we protect them and their data. In this I am sure we will find other areas of repeated large exposures that we are relying over and over with our new offerings and expansions.

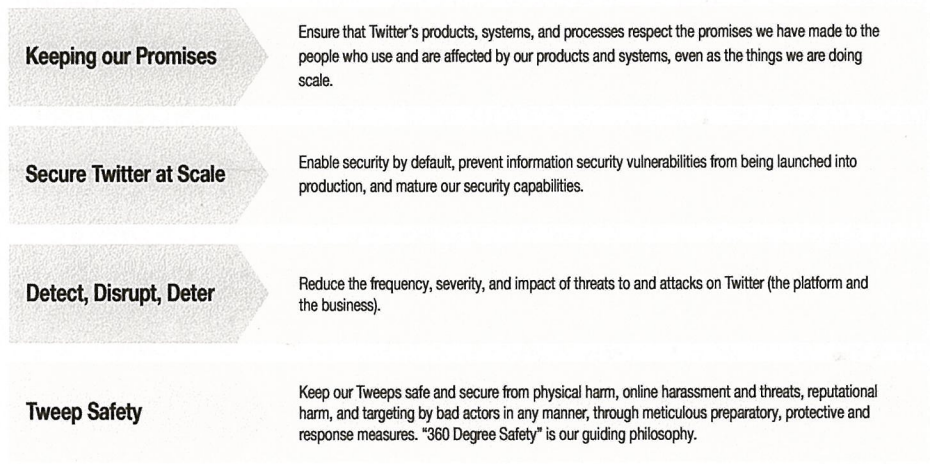
Tools we recycle for new products, principles, fleets, spaces, are valuable, reusable tools, - not re-usable code-red exposures.

We've made many promises to customers and regulators about how we keep them safe and handle their data - we aren't sure we are keeping those promises - each new world market cannot bring immediate broken promises.



4 Initiative Strategies

#Protect Initiatives



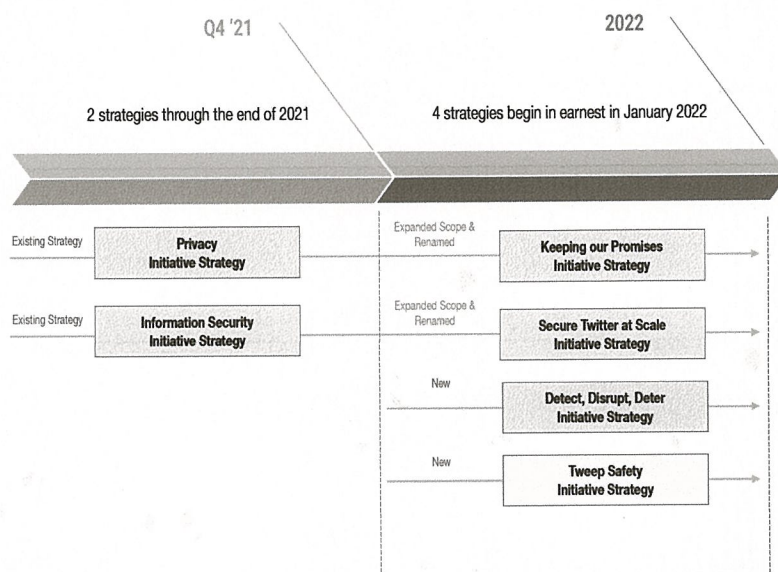
[2 min for this slide]

There are 4 Initiatives within the #CON Obj that have been built or rebuilt to help us achieve this Objective.

We will go through each at a high-level. It is important to note up front, that these 4 Initiatives map to all 3 of the #CON North Stars. We intentionally set it up this way because of (1) the nature of the work (i.e., Security, privacy, threat detection/disruption, and safety work overlaps in fundamental ways, such that moving the needle against one north stars requires work that would also start to move the needle against other north stars.) and (2) to ensure the Obj is as cross-functional as it needs to be to be successful.

(if pressed: detailed mapping of Workstreams to north stars and KRs here)

#Protect will Launch in Phases



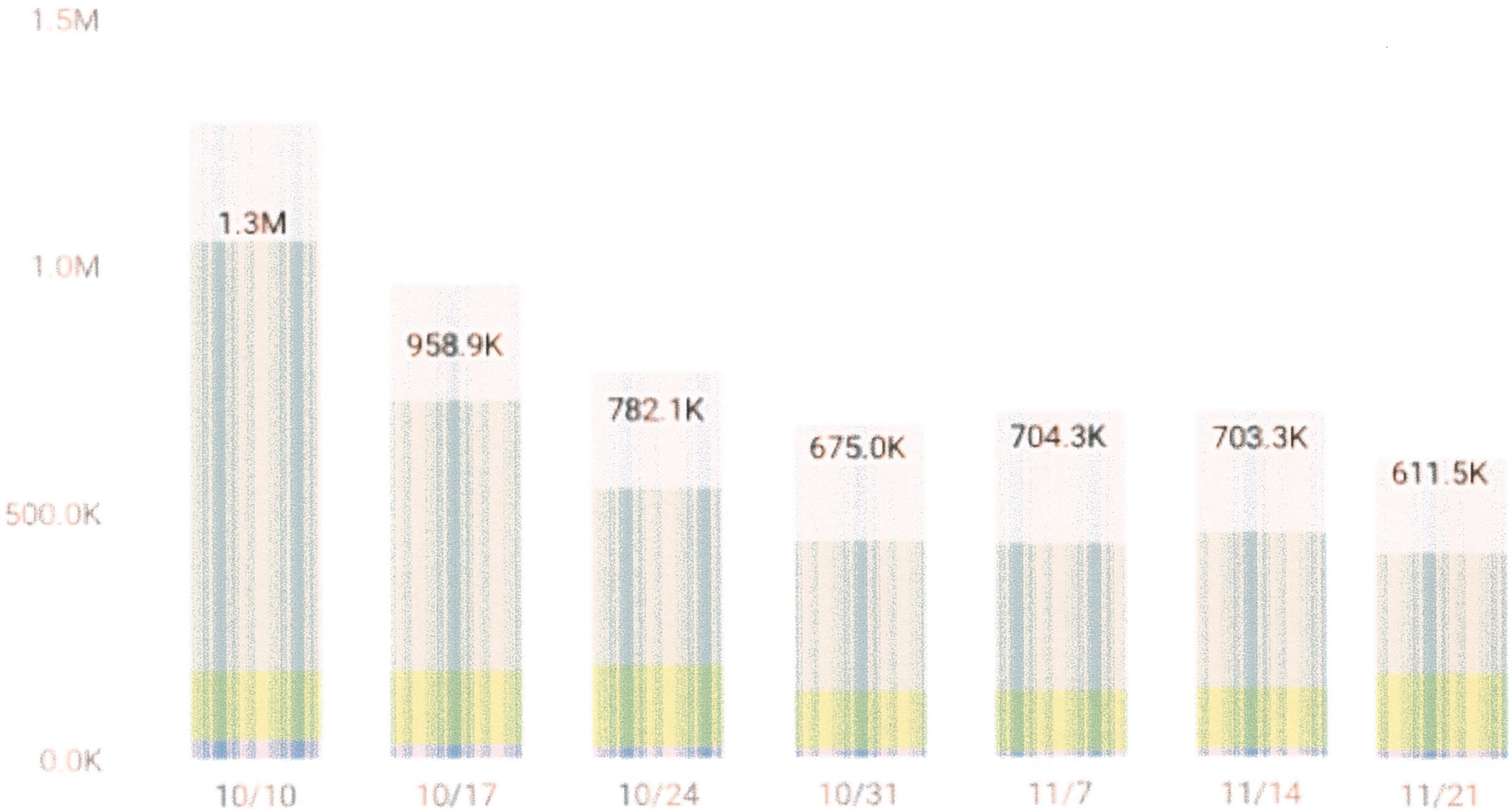
[1 min for this slide]

Suggested speaking notes:

- At the start of Q2, launched 2 Initiatives for H2 2021: Information Security and Privacy -- previously combined as one Initiative in #FUN, but now separated and rebooted.
- We used the rest of Q3 2021 to build, socialize, and align on the full Obj, including further refinement of the InfoSec and Privacy Initiatives
- Moving into 2022, we'll have 4 initiatives: With two new initiatives - Detect/Disrupt/Deter and Tweep Safety. We'll also expand the scope of the InfoSec and Privacy initiatives to fully support the #CON measures of success. We've renamed them as you can see here on the slide to be more indicative of their expanded scope.



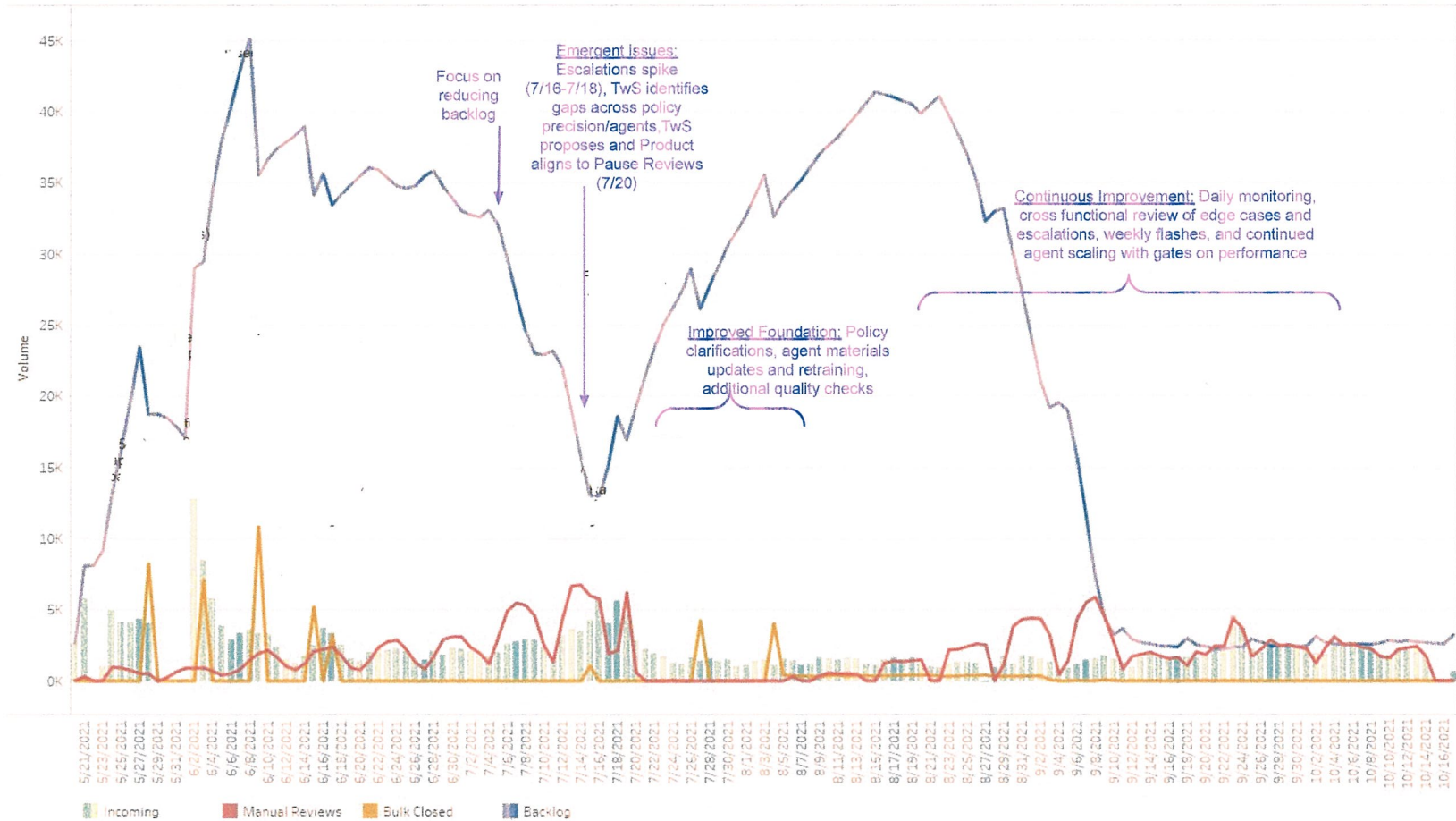
Appendix



Total P2 P1 P0 P0: BUSINESS CRITICAL

Verification Volume Summary

Incoming, Reviewed and Backlog Volumes



Fwd: action items from Mudge 1 of 2

Thu, Dec 30, 2021 at 12:23 PM

[REDACTED]

----- Forwarded message -----

From: Parag Agrawal [REDACTED]
Date: Fri, Dec 17, 2021 at 3:20 PM
Subject: Re: action items from Mudge 1 of 2
To: Peiter "Mudge" Zatzko [REDACTED]

Did you discuss / correct for the committee during the 15 min exec session at the end of meeting like we had discussed for this kind of scenario?

Parag

On Fri, Dec 17, 2021 at 9:51 AM Peiter "Mudge" Zatzko [REDACTED] wrote:

Per your offer - we should make sure the Risk Committee is correctly aligned on an item presented by [REDACTED].

[REDACTED] may have created an impression that overall access control risk was materially declining through their efforts. This is not the case. Access control exposure at Twitter has grown at a rate even faster than our rate of FTE growth (e.g. 46% of all employees had direct access to production in dec 2020 and 51% have direct production access in dec 2021).

Here are observations sent to me from Twitter employees in attendance during [REDACTED] comments on access control in response to questions about how worried the Risk Committee should be about Twitter's access control challenges.

[REDACTED] 2:34 PM Employee 1
This is not accurate.

[REDACTED] 2:34 PM Employee 2
talking about the system access plan only
like ssh stuff

[REDACTED] 2:34 PM Employee 1
wildly different from the overall.

[REDACTED] 2:35 PM Employee 2
are you going to clarify?
because I do not want to be trying to re-explain this next quarter
this is "how many have endpoint software" not how many are in a
good state

[REDACTED] 2:36 PM Employee 1
[REDACTED] It's not good.



Peiter "Mudge" Zatko [Redacted]

InfoSec Risk Committee Presentation Items to be aware of

4 messages

Peiter "Mudge" Zatko [Redacted]
To: [Redacted]
Cc: [Redacted]

Wed, Dec 15, 2021 at 2:49 PM

[Redacted] and [Redacted]

Tomorrow is the Risk Committee and I want to wish you luck!

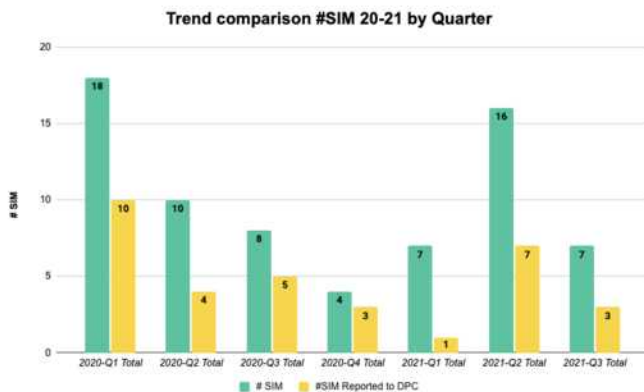
I want to reiterate items that we have covered at meetings for this risk committee (and others).

As you know, it is critical we do not misdirect or mislead the board members through omission, or framing, of specific data. You ([Redacted]) no doubt remember their surprise when they learned the difference between what they had been briefed in meetings prior to your arrival versus what they learned the actual environment was. Wow! We need to avoid putting them in that situation again.

Here are some specific areas in the slides that I think could send the wrong messages. I've brought these up before, so they will likely be familiar. I am citing specifics in this message for clarity in communications here, not that I am suggesting you overwhelm the committee members in specific detail.

9k (of our 10k) systems have security reporting software on them.

This has been reported several times. I worry it is misleading. This security reporting software has been reporting that 50% of all of our systems are not meeting basic security configurations (for over a year). 30% of the systems are reporting as not having software updates enabled. Both of these figures have also been at these levels for over a year as well. Be careful not to confuse the board with the stating we have 90% coverage of our systems with security reporting software versus what that reporting software is telling us about our systems.



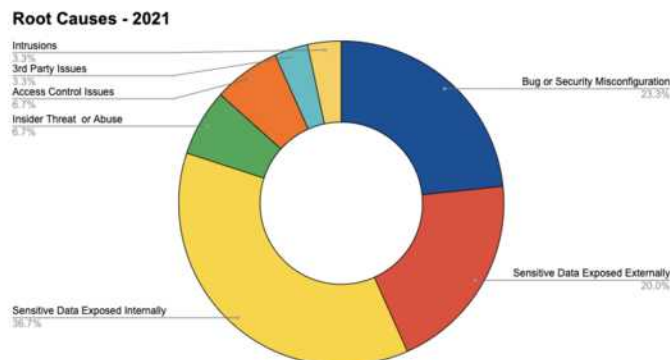
The above graph is now using a subset metric of SIMS as opposed to an expected metric of total SIMs. This could imply we have fewer SIMs, and reported SIMs than we do. In this case the chart now only shows SIMs reported to the Irish-

DPC. The expected metric (all SIMs required to be reported to regulators) is a higher number reported (200% higher in the last bars). Please clarify as/if appropriate.

80-90% of UPL projects are now within SDLC (flyway) compliance

This is good and should be celebrated. However this could mislead as it lacks larger context. The majority of projects at Twitter are not in the UPL (RTB and local). We run the risk of confusing the board members that we are 80-90% done when other estimates are showing we are less than 20% done here. If appropriate it may be important to also remind them that the SDLC and Flyway are currently stubs/skeletons in many ways. Good roll out through engineering. Just be careful of what message may be received

6% of our incidents are access control related



The graph tags access control at 6%. Internally we have referred to access control as more than 75% of our incident roots (sensitive data exposure internally (36.7%), externally (20%), and security misconfiguration (23.3%), are access control related). We need to be clear on this as we message that access control is a systemic issue at Twitter, we know it is one of the greatest risks in our ability to secure the environment, and that this is a key focus in regulator investigations and interest.

Server patch levels

It is table stakes to report the state of hygiene of our systems, both endpoints (clients) and servers (production). InfoSec reports have not done this to my knowledge and this report does not include this information either. 60% of our systems in production are not at the correct patch level. Many of these are unsupported (legacy) operating systems incapable of actually meeting certain security requirements. This is a potential PR issue in addition to the security risk. I am not saying this must be brought up at this Risk Committee, but this is something we should ensure is not continued to be omitted. Not mentioning this topic can lead one to infer that it is a solved issue.

Access to Production Servers

While we should celebrate the reduction of two small, but important, groups of access control. We need to make sure we are not showing data graphs that do not match to actual data (or that show different stories than the actual data).

The charts being shown do not match the data I have seen.

- a. The direct access chart showing the reductions does not match the charts I have seen in Confidence-Staff meetings.
- b. The Direct access to production systems graph seems incorrect. Our total exposure of accounts with direct access to production systems has actually increased
 - i. Dec 2020 46% of employees (2,763 out of 5917)
 - ii. Dec 2021 51% of employees (3,995 out of 7714)

We need to make sure the wins are recognized but that they are not presented in isolation, potentially implying they are representative of progress against the larger risk issue. The larger population of access to production has actually increased and I don't see that mentioned or captured in the graphic. Again, be mindful of what expectations and understandings are being set.

My other comments from our meetings stand on other, similar, topics in the deck.

Thanks for your attention to these items.

Peiter "Mudge" Zatkan [redacted] Wed, Dec 15, 2021 at 3:02 PM
To: Parag Agrawal [redacted]
Bcc: Kathleen Pacini [redacted], Peiter Zatkan [redacted]

Parag and Dalana,

The other day, in our conversation, you suggested I forward [redacted] to the Risk Committee Board without modification or replacement.

I expressed concern given what I see as numerous, and some significant, misrepresentations in the document.

The document has been forwarded to the committee.

This e-mail is more for our records and to ensure there was clarity in the description of my concerns around repeated representation items in the document we are putting forward.

I'm very much looking forward to hearing from you today.

kindest,

Mudge

Notes shared with [redacted] on concerning data presentations in the Risk Committee documents
[Quoted text hidden]

[redacted] Wed, Dec 15, 2021 at 6:02 PM
To: Peiter Mudge Zatkan [redacted]
Cc: [redacted]

Mudge,

It's unfortunate that you've had this content for review for the past week and a half and waited until after the content was already submitted to the risk committee to provide feedback and less than 24 hours before the committee meeting. This isn't enough time to adjust content and points of discussion. I therefore will not be adjusting content, but can take your feedback into consideration for future meetings.

Now, to address your feedback. Most importantly, I have never misled or misrepresented data to the board or risk committee. If anything, my voice has been muted, and I've been excluded from these meetings. Furthermore, none of the content provided is a misrepresentation of data, nor is it intended to mislead. The data points explicitly call out what they measure, and the context provided during the discussion is intended to add color and context (as we elaborated on during the dry-run which you were part of [but perhaps you were disengaged during that part of the dialog]). On your first point for example, the message is one of visibility and ability to measure. It's a statement suggesting marked improvements in our visibility and ability to measure many things, including those associated with the one point you bring up, auto-update enabled. As you know, auto-update being enabled is a tool to ensure timely update of systems, but is not in and of itself a measure of risk (systems without auto-update enabled, for example, can also be up-to-date, as a generic example). This translates to being one of many signals that help us understand the state of endpoint health. This however is not in scope for what this metric is intended to portray, which is one of capability and completeness **visibility**. I can expand on this point with the risk committee and share objectives that we have in-flight to address a broader endpoint metric program that will help us get to more granular measurements without complicating the discussion with the risk committee by getting too tactical (which we have many times received feedback on being). Also, can you provide more detail as to where you are getting data from? If it is our dashboards, it may

On your point around SIM reporting, I'm not sure what you're trying to portray here, or what you believe is misleading or inaccurate. Can you clarify?

On point 3, what we're suggesting, accurately, is, of the UPL items, or in other words, the most important cross-functional efforts, a MAJORITY, are in the flyway v2 process. We intend to add color that this is a very important but first step toward improving the processes across all efforts, but as with all things complicated, we start with the most important things (which UPL suggests).

On your next point around access control related incidents. The graph here is suggestive of the "primary root cause" as identified by our retro process. It doesn't necessarily suggest that there aren't secondary or tertiary (or more) reasons for them. If we went into that level of detail, we would risk devolving the conversation into one about semantics and would draw away from the primary point, and would once again move us toward being too tactical, which we have been called out for in the past on numerous occasions. I will of course also verbally add (as I did during our dry run, see comment above) that thematically access control is one of great importance, which is why we have a massive cross-functionally supported effort on access reduction which has its own section in the presentation (which accurately represents the importance of the topic, calling it out as one of our critical endeavors in the coming months).

On server patch levels, I'm not even sure what to say here Mudge. We have discussed this one to death. Not only have I conveyed this to the risk committee, in detail, but we've identified this as a critical path item for us from a visibility and measurement standpoint, which we are actively scoping as an objective as part of our endpoint metrics strategy. This of course requires involvement of (and commitment from) peer organizations in your charter, which we are working alongside to address this. I will of course add more color to this during the discussion with the risk committee, but your points here don't resonate as it's been a consistent point during these sessions.

On the final point around access, my first point is one I'll redirect to you above. We have already identified this as a critical goal for the organization, and the organization around this item cross-functionally is commensurate with this. Second, I feel you're being disingenuous in your statement around two small wins. I would characterize this as several big wins, starting with significant reduction in the number of users with broad access to all of production (if you remember, the topic we began with in this journey), by >66%. This is accompanied by several other wins associated with IPMI and

SUDO access reduction, which are significant in reducing overall risk. If the point you want to convey is, "there is more work to do," rest assured, it's already part of the talk track for the risk meeting (as conveyed, once again during our dry run session, see comments above).

Now that I've addressed your comments, there is a point I don't want to neglect in surfacing. I already made the point that giving feedback in this form, last minute ahead of such an important meeting is unfortunate, I'd like to take this one step further. Doing so does not set me up for success, is not indicative of your support for this organization or myself, and is written in "gotchya" framing. This is exacerbated by the fact that you've copied a member of my team, which is not in the spirit of creating a collaborative or supportive environment. I've provided this feedback to you on countless occasions, but it is worth bringing up once again.

Regards,

[Redacted]
[Quoted text hidden]

Peiter "Mudge" Zatko [Redacted]

Wed, Dec 15, 2021 at 10:26 PM

To: [Redacted]
Cc: [Redacted]
Bcc: Kathleen Pacini [Redacted]

I'm confused and concerned by the tone and content of your response. I will (re)read your email in greater detail later and we will address.

I agree that if this had been new information it would be inappropriate at this juncture, but it is not. We have covered each of these items on numerous occasions. This includes in meeting review of this specific deck (with Privacy and HRBP present). None of this should be news.

In the first email on this topic today, to both Privacy and InfoSec teams, I shared that I would provide comments and specific observations to help narratives in tomorrow's meeting. Rather than including [Redacted] and Damien, out of deference to you, I narrowed the distribution of InfoSec comments to you and [Redacted]. Seeing as you included [Redacted] in every stage of this effort I assumed you would want her to be included in the feedback.

The InfoSec presentation does not include a written narrative so while I have requested modifications, clarifications, and change on these data points and topics in the past, here I am not requesting a rewrite. I am suggesting ways to help the narrative and items to be mindful of in Q&A.

Much like in the example provided in the email, where the board had become misaligned with what Parag and Mike had been briefing on SDLC (effort vice progress from 2020 and before), we always run the risk of mis-setting understandings and expectations when we brief.

I am looking forward to the meeting. We have a lot to cover on very important topics.

Regards,

Mudge Zatko
[Quoted text hidden]

Fwd: Memorandum for the record

Wed, Jan 12, 2022 at 10:43 PM

[redacted]

[redacted]

[redacted]

[redacted]

----- Forwarded message -----

From: Peiter "Mudge" Zatkan [redacted]
Date: Wed, Jan 12, 2022 at 10:41 PM
Subject: Re: Memorandum for the record
To: Parag Agrawal [redacted]
Cc: Dalana Brand [redacted], Sean Edgett [redacted], Marianne Fogarty [redacted]

[privileged and confidential]

Parag:

Thank you for getting back to me. I want to quickly follow up with a few items so we can have the most productive 1:1 tomorrow.

First, for global context, my mission here is to protect Twitter in order to enable the company to succeed. That also means protecting you and helping you succeed in your new role too. I hope you recognize that I am performing this duty throughout this topic. I take this mission, and my responsibilities, very seriously.

I had an ask in my letter to you on January 4 to address the problem with [redacted]. Your email reply on January 6 does not mention [redacted]. I am hoping our meeting tomorrow will focus on this. As I mentioned in my email to you, I have no problem if Twitter wishes to [redacted] but please [redacted]

On other topics you indicated in your response of January 6, to my email of January 4, that you were surprised to read what was contained in my email given conversations we had. I take this as being "surprised" that I raised the issue again because we had discussed all of what was in my email both in prior e-mails and by voice.

On the topic of inappropriate material being put in front of Board members, thank you for initiating the audit investigation. I spoke with them yesterday (January 11th). I believe it was a very productive communication.

To assure you that I make every effort to ensure accurate information goes to our Board and Committee members and meetings it is documented that I was actively working to prevent inappropriate information from going to the Board or to the Risk Sub-Committee for months. To this end [redacted] prior to the December 9th Board meeting. When this was blocked I made sure concerning material was not put in front of the board and that appropriate data accurately presenting our environment and risks was shared.

The issue of now needing [redacted] prior to the Risk Committee was escalated to you [redacted]. I shared my concerns around the issue of not having this resolved prior to the Risk Committee meeting. In that conversation you took personal responsibility for ensuring this issue would be resolved. However, a few days before the Risk Committee you called me and told me that due to concerns you had discovered around [redacted] you were not able to keep that promise[0].

You indicated that [redacted] I had informed you of issues related to that report and my objection to its presentation to the Board. To the extent [redacted] inaccurate description of where we were, you informed me that I was to attempt to correct such items to the Board members during [redacted] presentation or at the end of the meeting during a closed session. I expressed concern around this approach. I did not want to have an inaccurate report going to the Board in the first place. It is difficult, to say the least, to have a report given to the Subcommittee only to tell them that it is not accurate, but as you instructed, I did my best to reel it in. I was extremely grateful that you offered to personally call the board to assist in correcting the record if I came back to you after the meeting asking for any help in further clarifying areas where I felt there may not be accurate understanding. I sent you an e-mail taking you up on your offer for a topic [redacted] which other people in the meeting identified as inappropriate and misleading. I was confused and sad when you told me you were disappointed in me for not having completely resolved these issues.

I have faith in processes such as the audit investigation. I am confident this will get to a place where the board members have an appropriate understanding of our risks and our progress against them.

I want to bring Twitter to where it needs to be. I need your help to clear obstacles to that objective. I am also thankful that you have authorized me to disclose accurately to the Board going forward.

Hopefully with much of the above now in the hands of the audit investigation, our conversation tomorrow will be on my ask in my January 4th e-mail: addressing the problem [redacted]

Thank you for your assistance. Looking forward to resolving this and getting towards our year of execution!

Best Regards,

Peiter Mudge Zatkan

[0] Very grateful for your phone call, where you apologized for not being able to keep your promise of resolving the issue [redacted] prior to the Risk Committee. [redacted]

On Thu, Jan 6, 2022 at 8:26 PM Parag Agrawal [redacted] wrote:

>
> Mudge
>
>
> Thank you for your email. Given our conversations before and after the Risk Committee meeting, however, I have to say I'm very surprised to see it. We take all of these concerns and issues very seriously. Since you have raised concerns that may implicate fraud, we'll report this to the Audit Committee and start a formal investigation. Please keep in mind you always have the opportunity to speak up in any meetings of the board or its committees. As the lead for these Risk Committee conversations, I expect you to ensure the accuracy of information presented.

> We will follow up with you to fully investigate this matter.

> Thanks

> Parag

> On Tue, Jan 4, 2022 at 3:20 PM Peiter "Mudge" Zatkan [redacted] wrote:

>> Jan 4, 2022
>>
>> CC: Dalana Brand
>>
>> Parag,

>> Happy New Year! I was looking forward to tagging up last week to discuss things, but I can appreciate the challenge of holiday schedules and the health issues we are all facing. I hope we will still be able to get together in the not too distant future.

>>
>>

>> As you know, I was hired to achieve certain goals and to fix problems here at Twitter. In order to do that, we need to recognize the actual state of affairs at the company in order to identify what needs to be fixed. Last month, we discussed this issue [REDACTED] report to the Risk Subcommittee of the Twitter Board. [REDACTED]

[REDACTED] to the Risk Subcommittee reports which I have pointed out are misleading and simply wrong.

>>
>>

>> As I have previously reported to you, it is critical that when we report to the Board, or one of its subcommittees (here Risk), that we be accurate and not mislead the board or its subcommittees. I know you share my concern that if [REDACTED] misrepresents the state of affairs, overstates what has been accomplished and ignores what needs to be improved, we do a great disservice to Twitter and whitewash the problem, which if not addressed, could undermine the future of this company.

>>
>>

>> [REDACTED] blocked my ability to correct problems. [REDACTED] latest report to the Risk Subcommittee, it has presented inaccurate, even false reports to the Risk Management Subcommittee of the Board. Such activity is at worst fraudulent and at best hiding the truth from the subcommittee.

>>
>>

>> In order for me to do my job, contractually, ethically, and legally, I need to ensure that what we report to the Board and its subcommittees is true and accurate. If there is a problem then the unvarnished truth must be presented. Only then can the problem be fixed. I would hope that you would agree with me. Going forward, we must ensure that we do just that. I think we also need to correct misinformation previously provided.

>>
>>

>> In addition, in order to ensure accuracy is maintained, I would request that [REDACTED] [REDACTED]

>>
>>

>> I think we have made measurable and material progress since my hire, but we have a ways to go. I know we can accomplish the fixes necessary, but we need to be honest in our reporting and in that way, the Board will know where we are and where we need to be.

>>
>>

>> I look forward to your assistance to improve our company.

>>
>>

>> Best regards,

>>
>>

>> Peiter Mudge Zatko

>>
>>

>>
>>

>>
>>

Re: Privileged and Confidential - Priority Meeting Request

Peiter "Mudge" Zatzko [REDACTED]
To: Marianne Fogarty [REDACTED]
Cc: Rebecca Falk [REDACTED]
Bcc: [REDACTED]

Tue, Jan 18, 2022 at 11:16 AM

Hi Marianne and Rebecca,

Thank you for your e-mail yesterday. As yesterday was a holiday I missed it, otherwise I would have sent this response then.

[REDACTED]

- [REDACTED]

Thanks. Please let me know if you have any questions or other data I need to know.

Respectfully,

Mudge Zatzko

Begin reference e-mail:

Peiter "Mudge" Zatzko [REDACTED]
to Parag, Dalana

Dec 15, 2021, 3:02 PM

Parag and Dalana,

The other day, in our conversation, you suggested I forward the infosec presentation to the Risk Committee Board without modification or replacement.

I expressed concern given what I see as numerous, and some significant, misrepresentations in the document.

The document has been forwarded to the committee.

This e-mail is more for our records and to ensure there was clarity in the description of my concerns around repeated representation items in the document we are putting forward.

I'm very much looking forward to hearing from you today.

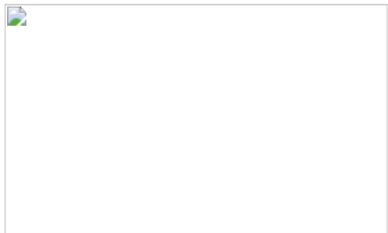
kindest,

Mudge

Notes shared with [REDACTED] on concerning data presentations in the Risk Committee documents

9k (of our 10k) systems have security reporting software on them.

This has been reported several times. I worry it is misleading. This security reporting software has been reporting that 50% of all of our systems are not meeting basic security configurations (for over a year). 30% of the systems are reporting as not having software updates enabled. Both of these figures have also been at these levels for over a year as well. Be careful not to confuse the board with the stating we have 90% coverage of our systems with security reporting software versus what that reporting software is telling us about our systems.



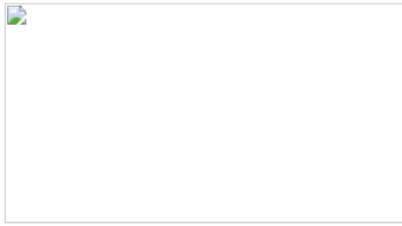
The above graph is now using a subset metric of SIMS as opposed to an expected metric of total SIMS. This could imply we have fewer SIMS, and reported SIMS than we do. In this case the chart now only shows SIMS reported to the Irish-DPC. The expected metric (all SIMS required to be reported to regulators) is a higher number reported (200% higher in the last bars). Please clarify as/if appropriate.

80-90% of UPL projects are now within SDLC (flyway) compliance

This is good and should be celebrated. However this could mislead as it lacks larger context. The majority of projects at Twitter are not in the UPL (RTB and local). We run the risk of confusing the board members that we are 80-90% done when other estimates are showing we are less than 20% done here. If appropriate it may be important to also remind them that the SDLC and Flyway are currently stubs/skeletons in

many ways. Good roll out through engineering. Just be careful of what message may be received

6% of our incidents are access control related



The graph tags access control at 6%. Internally we have referred to access control as more than 75% of our incident roots (sensitive data exposure internally (36.7%), externally (20%), and security misconfiguration (23.3%), are access control related). We need to be clear on this as we message that access control is a systemic issue at Twitter, we know it is one of the greatest risks in our ability to secure the environment, and that this is a key focus in regulator investigations and interest.

Server patch levels

It is table stakes to report the state of hygiene of our systems, both endpoints (clients) and servers (production). InfoSec reports have not done this to my knowledge and this report does not include this information either. 60% of our systems in production are not at the correct patch level. Many of these are unsupported (legacy) operating systems incapable of actually meeting certain security requirements. This is a potential PR issue in addition to the security risk. I am not saying this must be brought up at this Risk Committee, but this is something we should ensure is not continued to be omitted. Not mentioning this topic can lead one to infer that it is a solved issue.

Access to Production Servers

While we should celebrate the reduction of two small, but important, groups of access control. We need to make sure we are not showing data graphs that do not match to actual data (or that show different stories than the actual data).

The charts being shown do not match the data I have seen.

- a. The direct access chart showing the reductions does not match the charts I have seen in Confidence-Staff meetings.
- b. The Direct access to production systems graph seems incorrect. Our total exposure of accounts with direct access to production systems has actually increased
 - i. Dec 2020 46% of employees (2,763 out of 5917)
 - ii. Dec 2021 51% of employees (3,995 out of 7714)

We need to make sure the wins are recognized but that they are not presented in isolation, potentially implying they are representative of progress against the larger risk issue. The larger population of access to production has actually increased and I don't see that mentioned or captured in the graphic. Again, be mindful of what expectations and understandings are being set.

My other comments from our meetings stand on other, similar, topics in the deck.

Thanks for your attention to these items.

On Mon, Jan 17, 2022 at 8:39 PM Marianne Fogarty [REDACTED] wrote:
Privileged and Confidential

Mudge,

[REDACTED]

Thank you.

Best regards,

Marianne

On Tue, Jan 11, 2022 at 4:36 PM Marianne Fogarty [REDACTED] wrote:
Privileged and Confidential

Mudge,

Thank you for your time today, it is greatly appreciated and I am sure it will help us conclude our investigation quickly.

I have one follow-up request that ties in with the latter portion of our discussion -

[REDACTED]

I'm happy to answer any questions you may have.

Best,

Marianne

On Tue, Jan 11, 2022 at 1:36 PM Marianne Fogarty [REDACTED] wrote:
Thanks Mudge.
We can chat now - be on in just a moment.

On Tue, Jan 11, 2022 at 1:29 PM Peiter "Mudge" Zatko [REDACTED] wrote:
Or I can chat now (1:05pm pacific). I'll set an invite and be there just in case you can join.

On Tue, Jan 11, 2022 at 4:26 PM Peiter "Mudge" Zatko [REDACTED] wrote:
Thank you for your quick attention to this matter.

I can be available tomorrow from 12-1 Pacific time.

If that doesn't work for you please let me know and I can see what meetings I can cancel as this is a priority for me.

kindest,

Mudge Zatko

On Tue, Jan 11, 2022 at 2:01 PM Marianne Fogarty [REDACTED] wrote:
Mudge,

Twitter initiated an investigation into the concerns you raised in your January 4 email regarding the substance of matters presented to the Risk Committee.

In connection with this, we'd like to interview you to be sure we understand the full nature of your concerns and can follow up appropriately. Given the importance of the matter and the significance of the allegations, we hope that we can speak with you later today or tomorrow. Rebecca Falk, who leads the Investigations function on my team, will assist in the interview.

Please let me know your availability and we will do our best to accommodate. The only time that I am unavailable is 7-10am pst tomorrow (Wednesday).

Thank you.

Marianne

--



Marianne Fogarty | VP, Chief Compliance Officer
Registered In-House Counsel
Pronouns (She/Her) | San Francisco, CA
Follow me [REDACTED]

--



Marianne Fogarty | VP, Chief Compliance Officer
Registered In-House Counsel
Pronouns (She/Her) | San Francisco, CA
Follow me [REDACTED]

--



Marianne Fogarty | VP, Chief Compliance Officer
Registered In-House Counsel
Pronouns (She/Her) | San Francisco, CA
Follow me [REDACTED]

--



Marianne Fogarty | VP, Chief Compliance Officer
Registered In-House Counsel
Pronouns (She/Her) | San Francisco, CA
Follow me [REDACTED]

2021 Information Security Report

Twitter has struggled in the past with matching the right people to the right roles in the areas of Information Security and Privacy. There has been a lot of effort but it has not yielded the impact needed to meet our requirements and obligations. We are addressing this now. This past year Privacy Engineering was pulled out of Information security. Privacy had become so critical to the company we needed to ensure this item received special focus and that we could demonstrate that we can make progress now, whereas we had struggled in the past. We brought in a new leader of Privacy Engineering, one of world class talent and a track record of significant change at scale. With this new engineering leader teamed up with our Chief Privacy Officer (from counsel) we have made more measured progress towards our obligations over the past 6 months than we have since 2018. We do this by distilling the team and their focus down to their core purpose within the company.

We are now embarking on this journey with Information Security as well. The first step, as you will see in this Top Risk document and that has been absent in the past, is that the focus is narrow and the targets are known and quantified. The progress will be modular, methodical, and well understood and tracked. We will have more detailed discussions on this in the upcoming Risk Committee meeting.

To meet regulatory obligations we will be providing quarterly Security and Privacy Risk reports to the Risk Committee focusing on progress against the top risks cited below and related regulatory investigations. Additionally, we will provide end of year reports from the Chief Information Security Officer assessing the state of our Security Program and the Chief Privacy Officer assessing the state of our Privacy and Data Protection (“PDP”) Program. These end of year reports will provide evaluations of the respective programs and inform of any changes in “top risks” to the company.

Twitter’s Chief Privacy Officer’s 2021 review and what to expect in 2022 is complementary to this document and is provided [here](#).

2021 Review

Through the FTC Consent, we are entering into a 20-year obligation that covers everything we do with data as a company. Information Security Incidents we are obligated to report will be under significant scrutiny from regulatory bodies from this point forward. The regulatory bodies will be looking to identify if the incidents we are reporting indicate we have systemic issues in the areas we have stated we have significantly improved or otherwise addressed and remediated. Our incident rate, and the number that met requirements for reporting to regulatory bodies, the past year must be significantly reduced going forward. From Q3 2020 through Q3 2021,¹ Twitter averaged almost 10 incidents per quarter. Over an incident per month (more than 4 per quarter met requirements to be reported to regulators. Each of these events is a significant disruption to our business operations and needs to be made a very occasional exception instead of a constant norm. Many of these issues are able to be traced back to continuing challenges across *access control* and *inappropriate (security) configurations*. *These risks are the root causes of our FTC and Regulatory risks, which will be addressed here, in the Privacy and Data Protection Report, and in more details at the upcoming Risk Committee meeting.*

Access Control & Exceptional Access to Production Environments

Twitter has a large number of employees with direct access to our production environment. Every Engineer joining the company is provided Production level access. For context about half of all FTE employees² are engineers. Best practices are for companies to only allow production access to engineers in very minimal amounts and only in extreme situations (temporarily). Development, test, and staging environments are where engineers should safely conduct the majority of their work. We do not (meaningfully) have such environments. Twitter performs nearly all of

¹ Q4 2021 is still underway.

² Twitter was just shy of 8,000 FTEs as of November 30, 2021, with ~4,000 engineers. A random security issue with an employee or their account would yield credentials that could access production data on average 50% of the time.

these functions directly in production--thus requiring engineers to have production access. Further to this broad access to production, there are several pockets of exceptional access risk³. All of this is a-typical for security mature companies due to the risk associated with providing direct access to live customer data and the systems providing the service.

Strong access management to the various data and systems throughout Twitter's entire organization is the cornerstone to not only security and privacy but to enabling people to develop with velocity. To be explicit, access control is the primary line of defense to protect against a variety of threats (i.e., internal bad actors⁴, misuse of data for otherwise legitimate business purposes as occurred in SIM-28--which were the root cause of the current FTC issues--accidental data spills, etc.).

Security Management of Systems and Software

Our next top risk we must report to the Board is the state of our security configuration and compliance across our client fleet (laptops and workstations) and servers. We have ~9,000 laptops in our fleet. It had been internally reported that all (about 99%) of our clients correctly have security monitoring software installed on them. Unfortunately this hid the critical aspect of what the security monitoring software was reporting. It has been revealed that more than half (58%) of our entire laptop fleet is out of security compliance, and one-third (33%) do not have software and security updates enabled on them. Additionally, out of ~450,000 servers in our data centers, 68% are running out of date kernels (the brains of the operating system)⁵. Non-patched, out of date, operating systems are present on 70k (15%) of our servers. These issues, clients and servers, represent systems that are vulnerable for exploitation and represent a lack of hygiene that is difficult to justify externally.

FTC & Regulatory i.e., Data Handling & SDLC

Identifying and remediating systemic issues within our access control and security management of systems and software is one of the objectives of the regulatory frameworks we are attempting to meet. Not only are we under obligations to address these fundamental challenges, our lack of visibility into systems and services hinders our ability to detect new and existing vulnerabilities and respond and reconstitute to incidents once identified. The challenges, and our approach to solving the challenges, of identifying accesses and data is discussed in the [Privacy and Data Protection Summary](#) and will be discussed in more detail at the upcoming Risk Committee.

A different type of risk, not being compliant with FTC obligations, that we are tracking and reporting to the FTC on is our Software Development Lifecycle (SDLC)⁶. The SDLC ("Flyway") adoption rate is currently at 87.5% against a goal of 100% of projects on the Unified Priority List ("UPL"), down slightly from 92% in Q3 and from 89% at the beginning of Q2. This does not reflect work that is not on the UPL which leaves a significant gap in achieving our overall goal of ensuring that 100% of all work that should be using our SDLC, and verifying that it is in fact doing so. This is critical as the SDLC will be the way we ensure all work that should receive specified reviews to meet our FTC obligations is in fact doing so. In Q1 2022, the team will be pushing all teams to universally adopt Flyway so that we can create better intake mechanisms to capture this work.

2022 Preview

In addition to re-focusing InfoSec on the basics in 2022, to accomplish several of the Run The Business (RTB) reductions of risks mentioned above, we launched a company wide objective (#Protect Objective) that includes critical aspects of these challenges. Combined we will make measurable--via shared dashboards--and verified progress such that:

- Client systems will be brought up to proper base hygiene within the first 3 quarters.
- Production access will be cut in half and extreme risk groups need to be brought to 20% of what they were by the end of 2022.
- Twitter will integrate identity and access management capability by the end of 2022.

³ 320 people have superuser access across all systems and data within production and 250+ can remotely disable ("turn off") hardware within data centers.

⁴ See the statistics in the #Protect presentation

⁵ These represent security risks and vulnerable software within the heart of the computer. This is also a PR issue should there be a compromise within our datacenters.

⁶ SDLC is a formalized process that is imposed and followed in the development of software across the company.



Q4 2021 Privacy & Data Protection Report

As previewed in the [EOY Board Report](#), this report provides an overview of the 2021 progress on the top PDP Risks and a preview of the top 2022 PDP Risks.

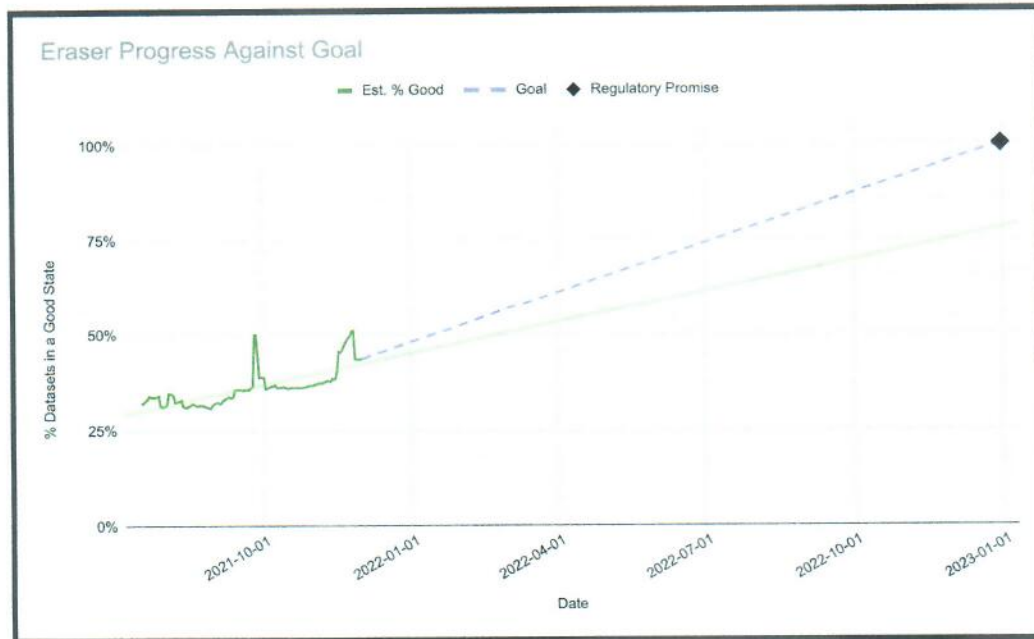
2021 Top Risks Review

Data Hygiene

One of our highest risks has long been our inability to know what data we have, who has access to it, who/what is using it, can we delete the data if asked, and how long are we keeping the data for. Together these can largely be summarized as data hygiene risks. While this presents a set of significant risks grouped together, the two that have been a primary focus in 2021 are Data Deletion and Access Controls.

Data Deletion

Over the last two quarters we have seen progress in our deletion efforts. Below shows the percentage of data that is in a good state (i.e., we know what it is, where it is, and are capable of deleting it as needed) against data that has yet to be brought into a good state (i.e., we don't know what it is, where it is, what it's used for, and we don't delete it).



Progress is significantly slower than hoped and without consistent sustained prioritization and additional resources, we are not likely to hit our Q4 2022 deadline.

Access Controls

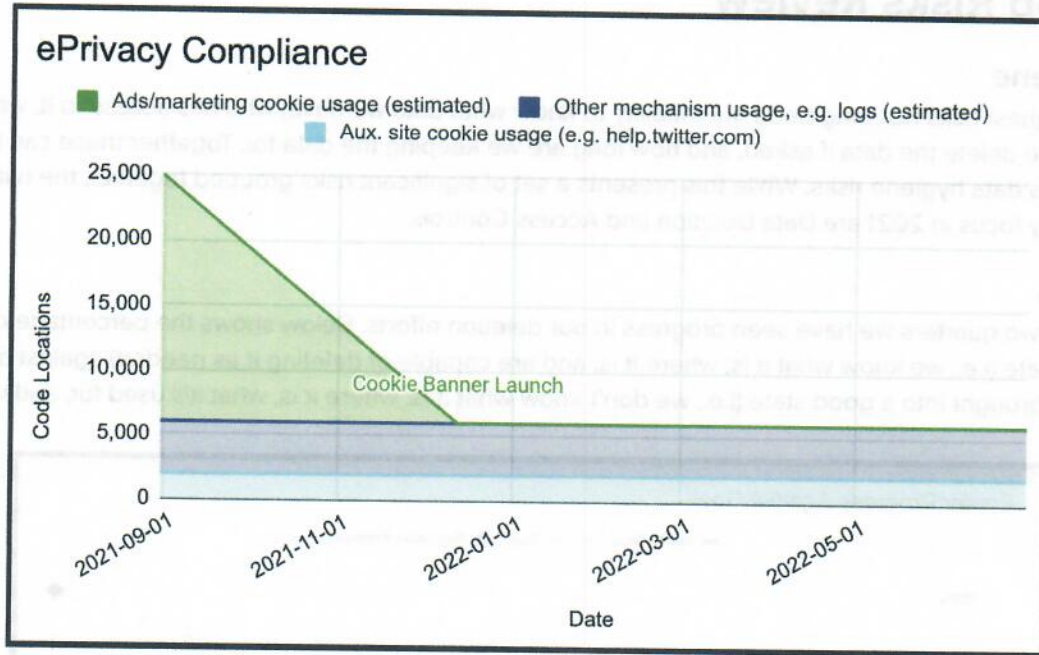
Excessive (i.e., access that is not needed to perform one's function) access to data remains a significant risk. Work has slowly progressed to create some roles-based access measures but these are not where they need to be. Every new employee has access to data they do not need to have access to for the purpose of their role. Until we



have implemented a mature centrally owned and operated system to manage access to data (e.g., entitlements and review, Role Based Access Controls, audits, etc) we are at risk of inappropriate access or use of data. Our inability to delete data compounds that risk, as we retain data that we should not have and which is therefore accessible by people who do not need to have access to this data.

Cookies

We have made significant progress on work to bring our usage of cookies up to the appropriate standard. In December 2021 we will launch an updated cookie banner in France which will correctly separate out and respect a person's choice between essential and non-essential cookies in France (depicted below are the relevant data uses of cookie related data collected from Twitter web properties against the timeline for remediation). All things going well, we expect to roll this banner out across the EU and UK in early January 2022.



FTC Settlement Preparations

During 2021 we made significant progress in preparing to meet our obligations under the new FTC Consent Order. We have prepared runbooks for all of the required aspects of the Consent Order; engaged a third party to conduct work to collect all public statements to prepare a questionnaire for teams to ensure their products and services match our public statements; we have audited existing security and privacy controls, identified missing controls, and begun the process of establishing new control owners; we have begun preparing Flyway (our SDLC) to meet the requirements that all security and privacy reviews are conducted as needed; we have prepared trainings (not yet administered) for teams; and we have begun internal communications to teams about the FTC Consent Order. Progress in some areas was slower than hoped. For example, we had hoped that teams would have completed their review of the questionnaire regarding public statements by year end, however, the questionnaire has required additional time to finalize.

2022 Top Risks Preview

Data Hygiene

We will need to extend our work beyond deletion and access controls to properly mitigate our risks with respect to data hygiene more generally. This work is exceptionally cross-functional and requires support and commitments from teams across the company to achieve our data hygiene needs.



Data Deletion and Usage

While we have set a Q4 2022 deadline for ourselves following conversations with the Irish Data Protection authority, if this risk becomes publicly known or the subject of a focused investigation while we are still fixing this issue, the risk to Twitter will be significant.

Access Controls

Addressing the state of access controls at Twitter will continue to present significant challenges in two respects: 1) creating access controls, as much data still has no controls 2) using them consistently and effectively to control access to data.

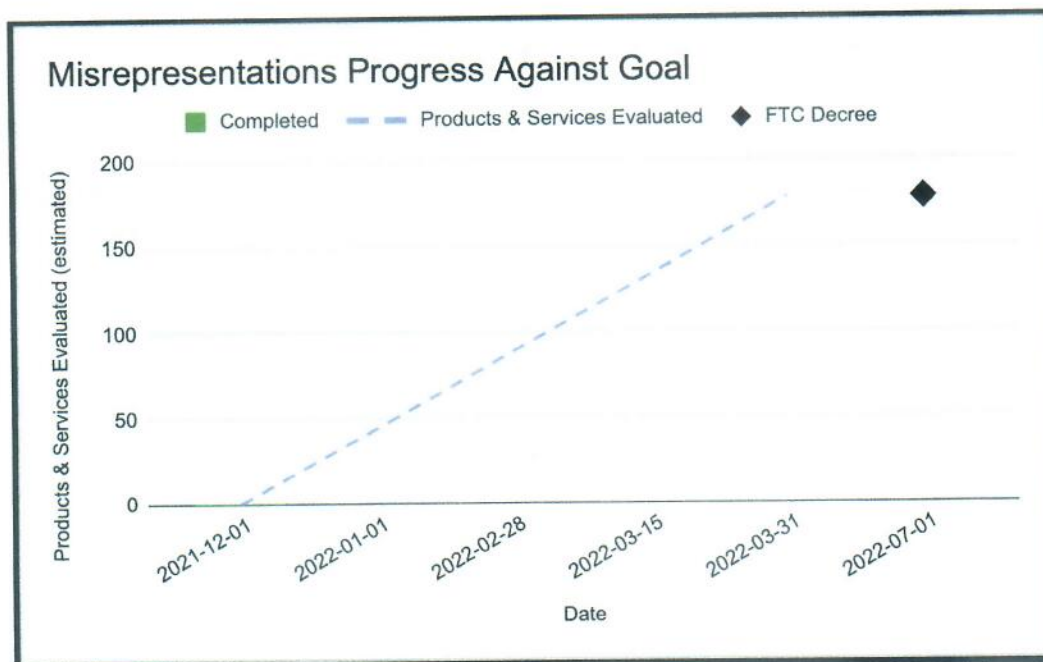
Cookies

Despite our efforts to date, we expect to continue to face regulatory investigations for our use of cookie technology. These investigations will focus on both historical cookie usage and how we are using cookie technology under the new banner. We are likely to face monetary penalties and be ordered to make changes to how we're using cookie technology for advertising, analytics, and keeping people logged into Twitter for prolonged periods of time going forward

FTC Settlement Preparations

In addition to ensuring the respective PDP and Security Programs are sufficient, there is additional work underway to meet specific obligations under the FTC Consent Order. These include ensuring:

- **Existing Products, Services, and Systems Operate Consistent with Existing Security and Privacy Statements:** We collected 5 years worth of statements (i.e., >20K statements), analyzed them into categories of representations, and created a questionnaire that teams that own Products, Services, or Systems can assess whether their Products, Services, or Systems are operating consistently with our existing statements. We will need all impacted teams (i.e., >180 services/products) to complete the questionnaire, identify any gaps, and propose remediation steps during Q1 2022 (depicted below).



- **Data Deletion Occurs:** We must deliver on our promises around data deletion. This work has been ongoing for many years and while progress has recently been made, it is insufficiently fast to meet our Q4 2022 target. We intend to build and lean more on infrastructure which enforces deletion to reach this goal.
- **Ensuring that Emails and Phone Numbers Collected Through Security Flows are Not Used in Advertising Products:** Technical means are beginning to roll out to ensure that we do not misuse emails and phone numbers. Until those are in place, we are still relying on fragile infrastructure and processes to prevent a recurrence of our prior misuse. Delivering technical solutions in 2022 will be key to long term compliance with our obligations.

Privacy and Security Programs

To meet our FTC Consent Order obligations we must have a Security Program and a Privacy Program that are capable of protecting the privacy, security, confidentiality and integrity of data Twitter has. Based on our work with our external partners over the last year, we know that our programs are less mature than they should be. This will present a heightened risk as under the new FTC Consent Order, these programs will be subject to additional audit scrutiny. Ensuring the programs are appropriately mature will be key to mitigating our risk.

PDP Program

The updated Program will be key to meeting our FTC Consent Order obligations. Based on progress so far, we will need cross functional support to:

- **Operationalize the Data Governance Committee (DGC):** as decisions related to data collection, use, maintenance, access and sharing are raised we will need to ensure that the DGC have oversight over these decisions.
- **Implement New Controls (i.e., taken from ISO frameworks) Ensure Ownership and Effective Operation:** in consultation with third-party auditors we have identified a significant number of controls that will need to be implemented across the company.
- **Ensure Required Reviews (Security and Privacy) Facilitated Through Flyway are Completed:** Flyway V2 is launching, which will provide the easiest auditable way of assessing whether appropriate reviews have been completed. Ensuring teams use Flyway will provide us the strongest way to ensure the appropriate reviews are being carried out.
- **Update Policies and Procedures and Enforce Them:** Policies and operational runbooks related to data that have been updated are neomg adhered to.
- **Ensure Required PDP Trainings are Completed:** these will be recurrent trainings and vary based on an employee's role and responsibilities
- **Operationalize Reviews and Audit Trails for Public Statements Related to Privacy or Security:** we must have a process in place to ensure that any statement that we make to the public related to privacy and security is reviewed, and logged as having been reviewed.



Follow-Up

Peiter Zatko <[REDACTED]@gmail.com>
To: Marianne Fogarty <[REDACTED]@twitter.com>
Cc: Rebecca Falk <[REDACTED]@twitter.com>

Fri, Jan 28, 2022 at 3:16 PM

Marianne,

It has been only 2 weeks since you and I agreed that corrective documents should be created.

One week ago, as I was preparing the materials, the system and materials were abruptly taken away from me when I was terminated.

6 weeks ago Parag told me not to create and send corrective materials.

I will have materials ready as soon as possible.

kindest,

Mudge Zatko

On Thu, Jan 27, 2022 at 5:37 PM Marianne Fogarty <[REDACTED]@twitter.com> wrote:

Mudge,

I'm writing to let you know that we are convening a meeting of the Risk Committee. At the meeting we will brief the Risk Committee on the concerns you raised in your January 4, 2022 email to Parag and Dalana, including the context you provided in the meeting with Omid and subsequent emails to Sean Edgett and me. If you have additional concerns that you have not yet raised, please do so as soon as possible, but no later than February 1, so that they can be brought to the attention of the Risk Committee in this meeting. It has been six weeks since the presentation about which you raised concerns, and just over three weeks since our investigation began. We do not feel it is appropriate to wait any longer given that we have provided you with repeated opportunities to share further information.

Thank you.

Marianne

On Mon, Jan 24, 2022 at 10:07 AM Marianne Fogarty <[REDACTED]@twitter.com> wrote:

Mudge,

Please let me know if you have time today or tomorrow or any time in the near future to talk about your additional items. I would like to speak with you to understand the concerns so that I can look into them. Please forward copies of any documentation relating to those items to me in anticipation of that conversation and to enable me to begin to investigate.

Thank you.

Regards,
Marianne

On Sun, Jan 23, 2022 at 5:31 PM Peiter Zatko <[REDACTED]@gmail.com> wrote:

Yes. There are additional significant items (with documentation).

Appreciate the Sunday email and request for my time. I am unavailable to speak right now.

On Sun, Jan 23, 2022 at 12:42 PM Marianne Fogarty <[REDACTED]@twitter.com> wrote:

Hello Mudge.

I'm writing to follow up on my email of Friday afternoon asking to speak with you to be sure I understand and can look into any additional concerns you may have. You indicated in email that you have additional concerns, and I understand from Patrick Pichette that you indicated the same to him.

My job is to conduct a thorough, objective investigation and report on it to the Audit Committee, of which Patrick is chair. Given the nature of the issues, here I will also report to the full Risk Committee of which Omid, with whom you spoke last week, is chair. But I can only investigate what has been reported to me, and I cannot complete my investigation knowing there are matters that have not been investigated.

I'm available to speak with you at your earliest convenience. Can we set up time to talk today?

Regards,

Marianne

On Fri, Jan 21, 2022 at 1:21 PM Marianne Fogarty <[REDACTED]@twitter.com> wrote:

Hi Mudge.

Can we set up a time to talk to identify and understand the full breadth of additional concerns/issues and can look into them?

The sooner the better. Are you free this afternoon? I'll make myself available at whatever time works for you.

Thank you.

Marianne

On Fri, Jan 21, 2022 at 10:07 AM Peiter Zatko <[REDACTED]@gmail.com> wrote:

Yes. These are now additional concerns/issues introduced by Parag.

On Wed, Jan 19, 2022 at 11:59 PM Marianne Fogarty <[REDACTED]@twitter.com> wrote:

Mudge,

[REDACTED] advised me that during your conversation this morning you mentioned the concerns you raised about information shared with the Risk Committee in December. If you were referring to matters that are not already under investigation, please let me know so we can schedule time to talk right away.

We appreciate that you raised your concerns and want to be sure they are fully and appropriately addressed. While my investigation is not complete, we intend to bring all of the concerns you raised with me and Omid, as well as the document you have been preparing in response to my email of January 11, to the Audit Committee and the full Risk Committee in the coming days. If there is anything else you want to include or recommend we do with respect to this issue so it concludes to your satisfaction, please let me know.

I'm available to hear any further concerns you may have as well as any additional thoughts you may have to resolve this matter.

Regards,

Marianne

--



Marianne Fogarty | VP, Chief Compliance Officer
Registered In-House Counsel
Pronouns (She/Her) | San Francisco, CA
Follow me [REDACTED]

--



Marianne Fogarty | VP, Chief Compliance Officer
Registered In-House Counsel
Pronouns (She/Her) | San Francisco, CA
Follow me

--



Marianne Fogarty | VP, Chief Compliance Officer
Registered In-House Counsel
Pronouns (She/Her) | San Francisco, CA
Follow me

--



Marianne Fogarty | VP, Chief Compliance Officer
Registered In-House Counsel
Pronouns (She/Her) | San Francisco, CA
Follow me

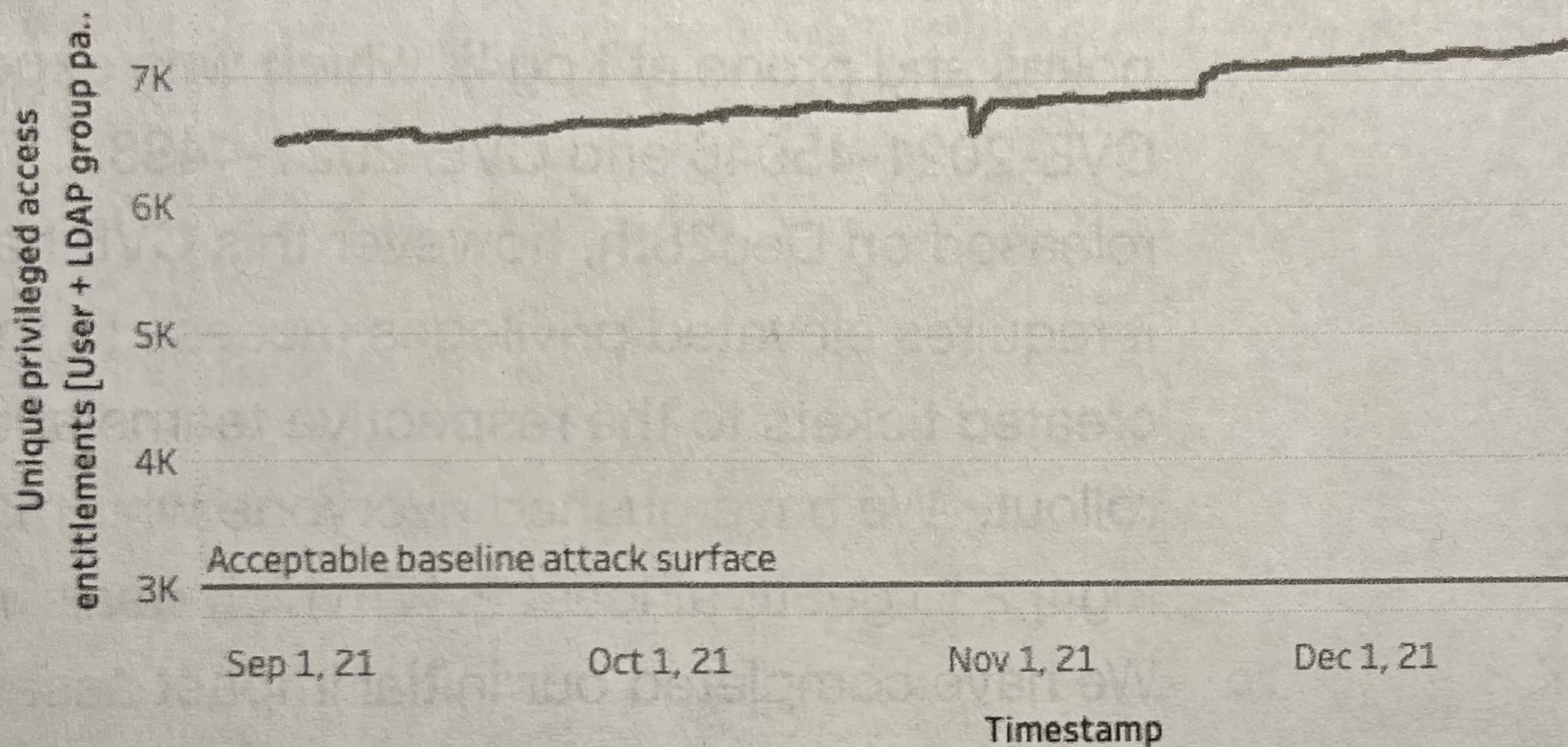
--



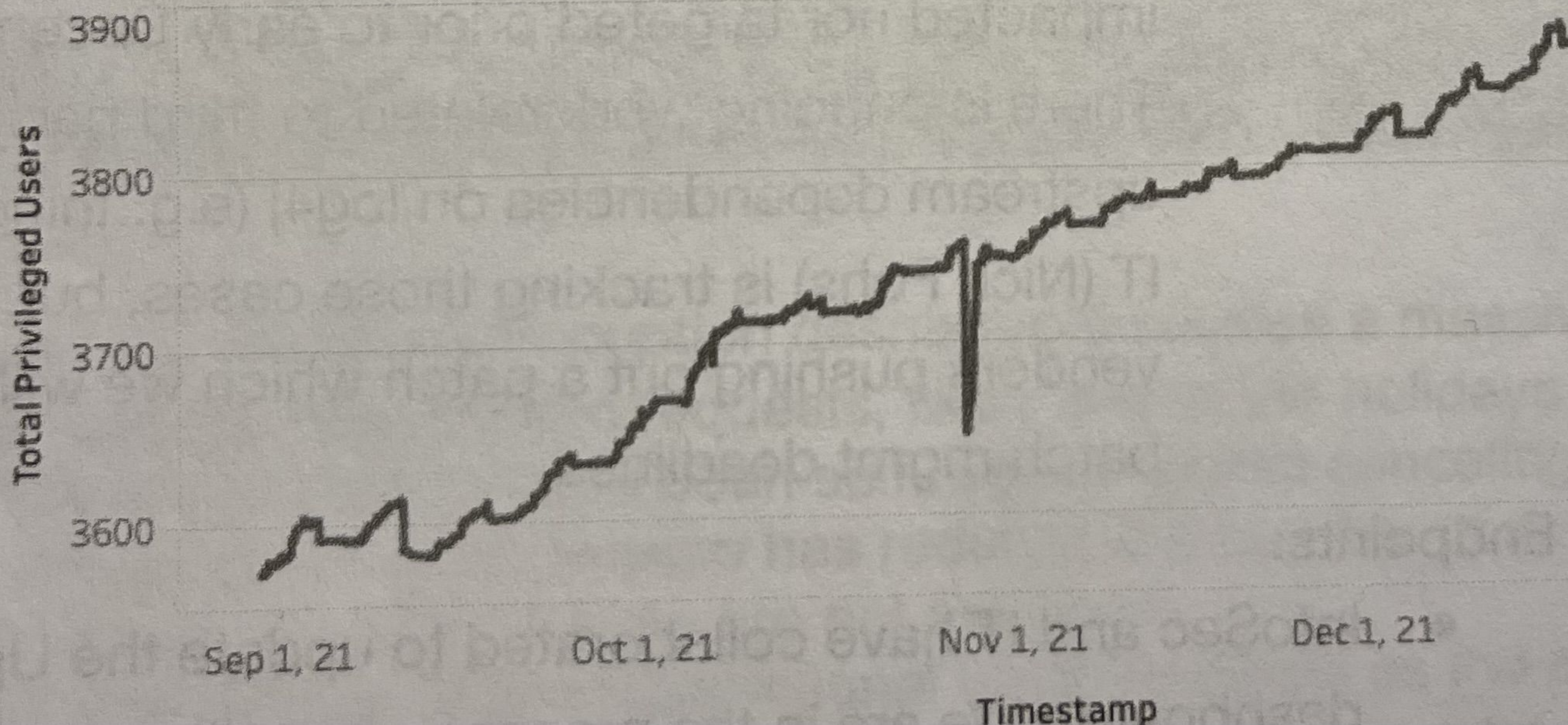
Marianne Fogarty | VP, Chief Compliance Officer
Registered In-House Counsel
Pronouns (She/Her) | San Francisco, CA
Follow me

Privileged access summary metrics over time

Unique privileged access entitlements over time



Unique privileged users over time

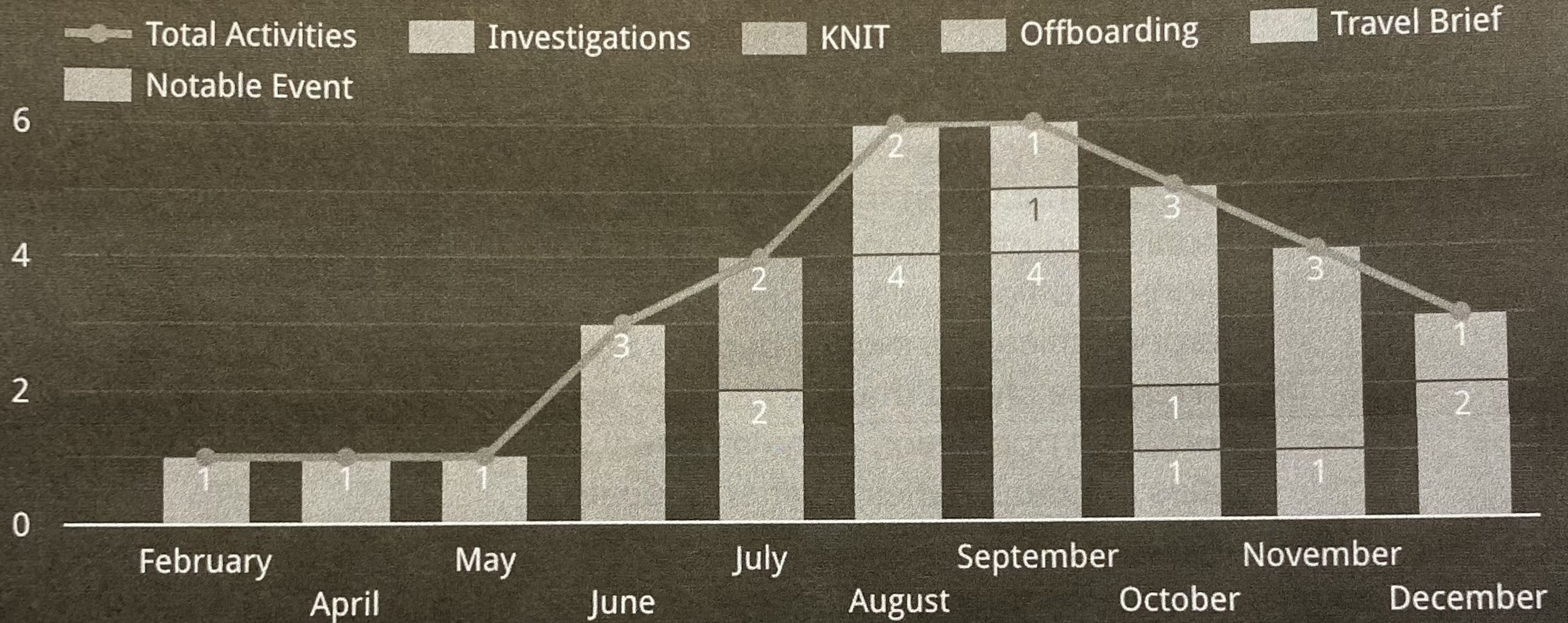


Full Calendar					
2021					
		Due Date to Sierra	Draft Review by Staff	Final Docs	Meeting Notes
July 27	Obj Review	July 23		Here	Here
Sept 2 <- We are here	Board Meeting	August 26	August 23		
Sept 23	Risk Committee	September 16	TBD / ~Sept 13		
October 18	Obj Review	October 15	TBD / ~Oct 13		
Dec 9	Board Meeting	December 2nd	November 29		
Dec 16	Risk Committee	December 9	TBD / ~ Dec 6		
2022					
January 17	Obj Review	January 14			
Mar 3rd	Board Meeting	February 24	February 21		
April 18	Obj Review	April 15			
May 26	Board Meeting	May 19	May 16		
July 18	Obj Review	July 15			
Sept 15	Board Meeting	September 8	September 5		
October 19	Obj Review	October 16			
Dec 8	Board Meeting	December 1	November 28		

Board meetings
 & risk comm
 7/21 12/22

Insider Risk Team Activities by Month

-Highlights how workload is developing for this new program
-At this time highlights previously unaddressed InT referrals

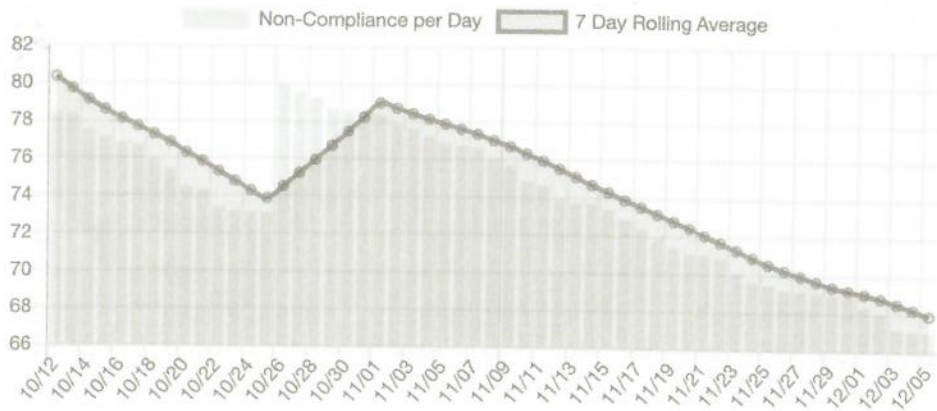


2021

Non-Compliant Kernels

307,544 (67%)

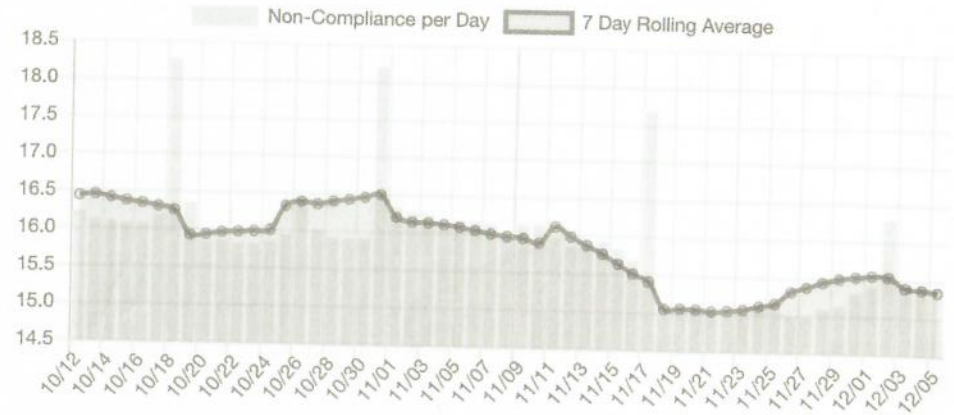
Kernel Version Non-Compliance per Day and 7 Day Moving Average



Non-Compliant Operating Systems

70,056 (15%)

Operating System Version Non-Compliance per Day and 7 Day Moving Average



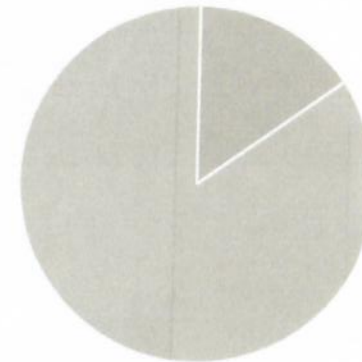
Kernel Compliance Overview (No Exceptions)

Non Compliant 67% (307544) Compliant 33% (151130)



OS Compliance Overview (No Exceptions)

Non Compliant 15% (70056) Compliant 85% (388503)

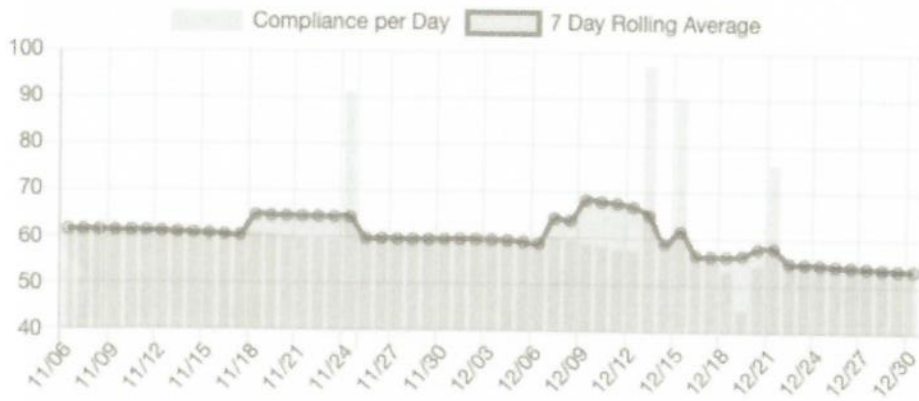


2020

Non-Compliant Kernels

186,372 (53%)

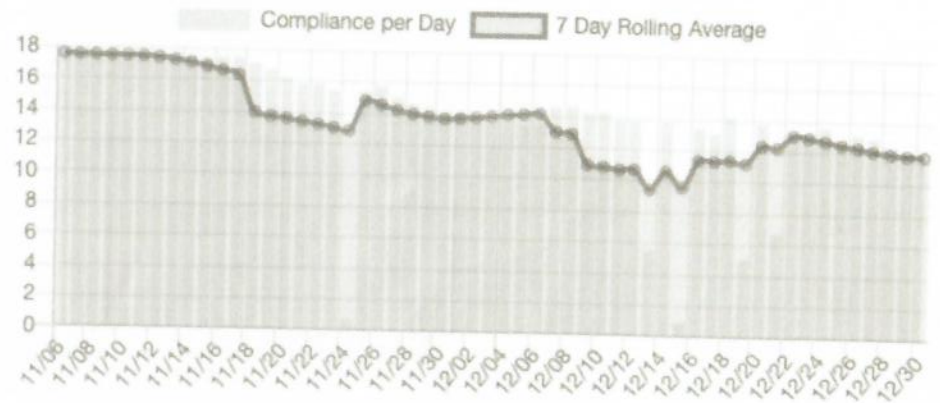
Kernel Version Compliance per Day and 7 Day Moving Average



Non-Compliant Operating Systems

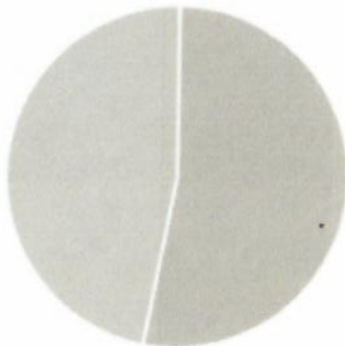
41,443 (12%)

Operating System Version Compliance per Day and 7 Day Moving Average



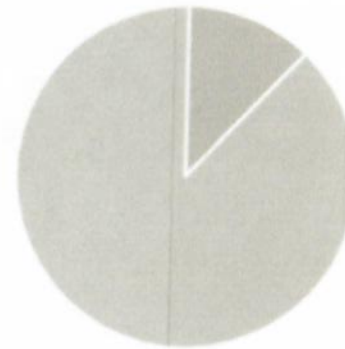
Kernel Compliance Overview (No Exceptions)

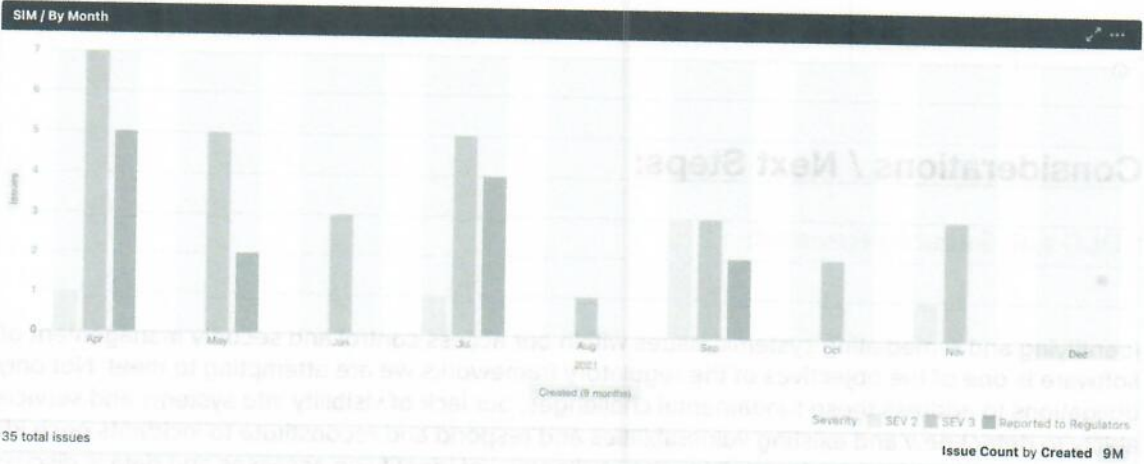
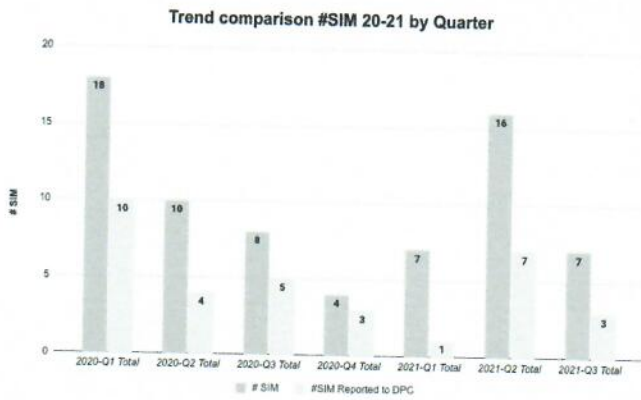
Non Compliant 53% (186372) Compliant 47% (162882)



OS Compliance Overview (No Exceptions)

Non Compliant 12% (41443) Compliant 88% (307807)



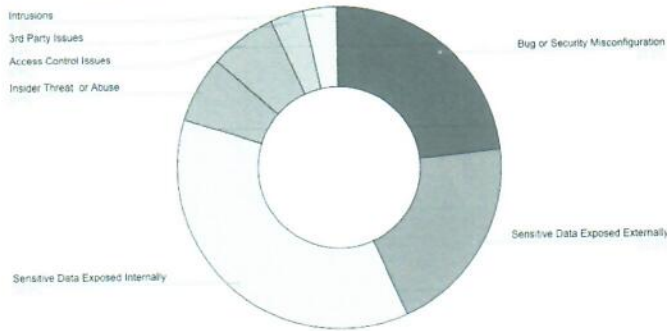


Through the FTC Consent, we are entering into a 20-year obligation that covers everything we do with data as a company. Information Security Incidents we are obligated to report will be under significant scrutiny from regulatory bodies from this point forward. The regulatory bodies will be looking to identify if the incidents we are reporting indicate we have systemic issues in the areas we have stated we have significantly improved or otherwise addressed and remediated. Our incident rate, and the number that met requirements for reporting to regulatory bodies, the past year must be significantly reduced going forward. From Q3 2020 through Q3 2021,⁵ Twitter averaged almost 10 incidents per quarter. Over an incident per month (more than 4 per quarter met requirements to be reported to regulators. Each of these events is a significant disruption to our business operations and needs to be made a very occasional exception instead of a constant norm. Many of these issues are able to be traced back to continuing challenges across *access control* and *inappropriate (security) configurations*. *These risks are the root causes of our FTC and Regulatory risks, which will be addressed here, in the Privacy and Data Protection Report, and in more details at the upcoming Risk Committee meeting.*

⁵ Q4 2021 is still underway.



Root Causes - 2021



Considerations / Next Steps:

SDLC and Security Reviews:



Identifying and remediating systemic issues within our access control and security management of systems and software is one of the objectives of the regulatory frameworks we are attempting to meet. Not only are we under obligations to address these fundamental challenges, our lack of visibility into systems and services hinders our ability to detect new and existing vulnerabilities and respond and reconstitute to incidents once identified. The challenges, and our approach to solving the challenges, of identifying accesses and data is discussed in the [Privacy and Data Protection Summary](#) and will be discussed in more detail at the upcoming Risk Committee.

A different type of risk, not being compliant with FTC obligations, that we are tracking and reporting to the FTC on is our Software Development Lifecycle (SDLC)⁶. The SDLC ("Flyway") adoption rate is currently at 87.5% against a goal of 100% of projects on the Unified Priority List ("UPL"), down slightly from 92% in Q3 and from 89% at the beginning of Q2. This does not reflect work that is not on the UPL which leaves a significant gap in achieving our overall goal of ensuring that 100% of all work that should be using our SDLC, and verifying that it is in fact doing so. This is critical as the SDLC will be the way we ensure all work that should receive specified reviews to meet our FTC obligations is in fact doing so. In Q1 2022, the team will be pushing all teams to universally adopt Flyway so that we can create better intake mechanisms to capture this work.

Access Controls

Excessive (i.e., access that is not needed to perform one's function) access to data remains a significant risk. Work has slowly progressed to create some roles-based access measures but these are not where they need to be. Every new employee has access to data they do not need to have access to for the purpose of their role. Until we have implemented a mature centrally owned and operated system to manage access to data (e.g., entitlements and review, Role Based Access Controls, audits, etc) we are at risk of inappropriate access or use of data. Our inability to delete data compounds that risk, as we retain data that we should not have and which is therefore accessible by people who do not need to have access to this data.

Cookies

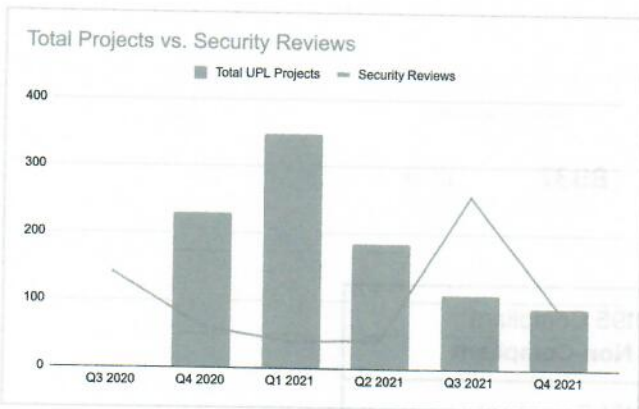
⁶ SDLC is a formalized process that is imposed and followed in the development of software across the company.



Additionally, out of ~450,000 servers in our data centers, 68% are running out of date kernels (the brains of the operating system)⁴. Non-patched, out of date, operating systems are present on 70k (15%) of our servers. These issues, clients and servers, represent systems that are vulnerable for exploitation and represent a lack of hygiene that is difficult to justify externally.

SDLC and Compliance

Describe SDLC and Flyway goals and why regulators are requiring these. This includes security and privacy reviews. Currently a product needs to go through ~7 different reviews prior to launching. Counsel privacy, privacy engineering, information security, ... - these reviews are numerous and burdensome for the launching team to track down and comply with. We intend to unify the review process to make this a smooth single stop process, easing the burden on launching teams and consolidating the expertise and review functions into formalized, and repeatable, processes.



- It is unknown how many projects are not listed on the UPL. Teams prioritize Run The Business (RTB) projects and there are local projects that do not appear on the UPL as well. The new UPL initiative intends to restrict the UPL to the top 100 company wide projects. While this is useful for
- The highlight of this quarter for SDLC/Flyway is the adoption rate, which is currently at 87.5% against a goal of 100% of projects on the Unified Priority List, down slightly from 92% in Q3 and from 89% at the beginning of Q2 but up dramatically up from 34% at the time of the Q2 board report.

Incidents (Lagging Indicator)

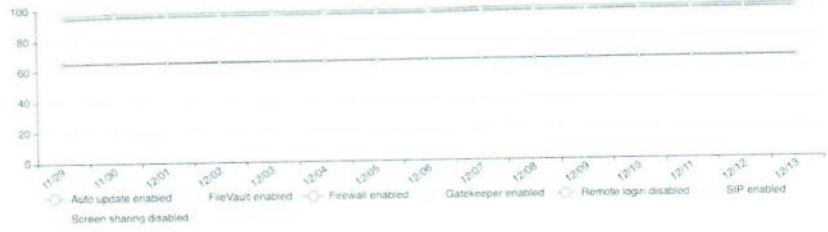
⁴ These represent security risks and vulnerable software within the heart of the computer. This is also a PR issue should there be a compromise within our datacenters.



Mac assets



Health check trendline (%) - last 14 days



Application firewall enabled

9195

1

Auto update enabled

6131

3065

FileVault enabled

9114

82

Gatekeeper enabled

9159

37

Remote login disabled

8984

212

SIP configured

9120

76

Screen sharing disabled

8937

259

Application Firewall Enabled	9195 Compliant 1 Non-Compliant
Auto Update Enabled	6131 Compliant 3065 Non-Compliant
FileVault Enabled	9114 Compliant 82 Non-Compliant
Gatekeeper Enabled	9159 Compliant 37 Non-Compliant
Remote Login Disabled	8984 Compliant 212 Non-Compliant
SIP Configured	9120 Compliant 76 Non-Compliant
Screen Sharing Disabled	8937 Compliant 259 Non-Compliant



Q2 internal
~~briefing~~
board meeting

Mudge Board Voice 2021 Q2

Hi I'm Mudge Zatko,

I'm the newest Staff member and we met last board meeting where I ran through my 60 day litany of threats and risks that keep me up at night.

To that end, we're forming a new department called Confidence to move the company from reactive to predictive and proactive across security, privacy, safety, and integrity teams at Twitter. You'll be hearing more about Confidence at future board and risk committee meetings.

As part of this we evaluated the existing security and privacy programs and found that we were able to pause a large number of them. We stopped these programs because they weren't taking extant risk off the table, were expensive with dependencies across engineering teams, and were struggling to close on an expanding target. Ultimately we may revisit these efforts in a different way to address these shortcomings but in the meantime the pause frees up resources and allows us to focus on critical near term risk removal as we roll out the larger Confidence longer term strategy which you will hear about in H2.

██████████ and Damien will briefly touch upon some of the highlights in security and privacy and going forward we will move more of this over to the risk committee.

August 16, 2021

Kelly Sims

Re: [Inform] New York Times story on company shift

To: Peiter Mudge Zatkan [REDACTED]

So jealous. I miss humans.

On Mon, Aug 16, 2021 at 10:03 AM Peiter "Mudge" Zatkan [REDACTED] wrote:

I have an in-person 1:1 with Leslie on Thursday - Let's tag up Friday or else next week.

Looking forward to it!

On Mon, Aug 16, 2021 at 12:49 PM Kelly Sims [REDACTED] wrote:

Never a dull moment. I'll get on your calendar.

On Mon, Aug 16, 2021 at 9:35 AM Peiter "Mudge" Zatkan [REDACTED]

wrote:

There are some big things brewing under the surface on which you'll need to be brought up to speed. Let's catch up soon!

So happy to have you here :)

On Mon, Aug 16, 2021 at 12:16 PM Kelly Sims [REDACTED] wrote:

Preach!

Also, hi! Let's catch up soon?

On Mon, Aug 16, 2021 at 9:05 AM Peiter "Mudge" Zatkan [REDACTED]

wrote:

Dantley, I've seen nothing but honor and strong execution in your work here. We are here for you.

Regarding yourself and the others mentioned in the article, we have increased our monitoring and are prepared to action any targeted harassment.

Everyone,

I'd like to point out another part of the culture which is problematic and explain some of the less obvious problems with it:

"Fourteen current and former Twitter employees, who were not authorized to speak publicly, spoke with unusual candor to The New York Times about the last two years working with Mr. Davis and the changes he brought to their workplace."

The statement above is a green-light for insider threat targeting and recruitment. This announces that:

- a) we have disgruntled employees
- b) even non-disgruntled employees do not fear repercussions for sharing company private information
- c) our employees are approachable on topics (a) and (b)

From an adversary point of view when such a statement was/is observed it is inferred that the company in question generally lacks mature data protection and data tracking. Otherwise, people would generally be more reluctant to share any information. This inference was/is almost always correct and when such leaks as this scale are observed the adversary defines the company as a high yield, low cost, target.

This leaky-culture is problem. Not just a problem for each individual leak.

kindest,

Mudge Zatko

On Mon, Aug 16, 2021 at 9:48 AM Dantley Davis [REDACTED] wrote:
Thanks everyone for your support and guidance through this. I've given my directs a heads up and will be sending out an email to me team so they are aware. I'll follow up with folks individually if they have questions or concerns.

On Mon, Aug 16, 2021 at 6:08 AM Liz Kelley [REDACTED] wrote:
Sharing the full story that was published today below. We're working on a few minor corrections, but overall, this landed as we expected. To note, we did not comment on the record regarding any individual employee situations; you'll see two included here.

The reporter picked up on how this is an intentional shift we've been working to drive, and how Dantley was brought in to help lead the transformation. Across the 14 current and employees that served as sources for the story, we also see a balance in both criticism of this approach as well as people who welcome the shift as a refreshing change. From both leaked emails and Dantley's interview, we're grateful that his voice is consistent and present in acknowledging opportunity to do better while still holding each other accountable.

We'll continue to monitor internal and external conversation to support our team and handle additional questions.

Culture Change and Conflict at Twitter

Two years ago, the company brought in a blunt executive to make things move faster and to promote diversity. Then the problems began.

By Kate Conger

Aug. 16, 2021, 5:00 a.m. ET

SAN FRANCISCO — Soon after joining Twitter in 2019, Dantley Davis gathered his staff in a conference room at the company's San Francisco headquarters. Twitter was too nice, he told the group, and he was there to change it.

Mr. Davis, the company's new vice president of design, asked employees to go around the room, complimenting and critiquing one another. Tough criticism would help Twitter improve, he said. The barbs soon flew. Several attendees cried during the two-hour meeting, said three people who were there.

Mr. Davis, 43, has played a key role in a behind-the-scenes effort over the past two years to remake Twitter's culture. The company had long been slow to build products, and under pressure from investors and users, executives landed on a diagnosis: Twitter's collaborative environment had calcified, making workers reluctant to criticize one another. Mr. Davis, the company believed, was one of the answers to that problem.

The turmoil that followed revealed the trade-offs and conflicts that arise when companies attempt dramatic cultural shifts and put the onus on hard-nosed managers to make that change happen.

Mr. Davis repeatedly clashed with employees because of his blunt style. His treatment of workers was also the subject of several investigations by Twitter's employee relations department, and of complaints to Jack Dorsey, the chief executive, that too many people were leaving.

Company officials acknowledge that Mr. Davis may have gone too far at times, and he has promised to tone down the way he criticizes people. But they make no apologies and have even given him a promotion. Employee dissatisfaction, they said, is sometimes the cost of shaking things up.

"This is actually a Twitter culture change that we've been trying to drive," Jennifer Christie, Twitter's head of human resources, said in an interview.

A former Facebook and Netflix executive, Mr. Davis, who is now the company's chief design officer, reports directly to Mr. Dorsey. When hired, he was told to revamp Twitter's design team and make it more diverse. His work was considered a model for other Twitter executives, and the company believes the diversity of his department improved under his leadership. Twitter reports its diversity statistics annually but does not break out numbers for specific parts of the company.

"This was a turnaround role, and that meant changes to staff, changes to our work, changes to how we collaborate," Mr. Davis said in a recent interview.

He frequently spoke with his staff about challenges he faced as a Black and Korean man in the technology industry, and won accolades for his design work. He spearheaded forays into new media, like audio tweets and chats, and championed efforts to clean up the conversation on Twitter, including prompts that encourage people to read articles before sharing them.

But Mr. Davis's management style was a bracing shift for employees at Twitter, which has not usually offered the astronomical salaries that are normal at other social media outfits. Instead, the company has tried to attract workers with a welcoming culture typified in a hashtag, #LoveWhereYouWork. Fourteen current and former Twitter employees, who were not authorized to speak publicly, spoke with unusual candor to The New

York Times about the last two years working with Mr. Davis and the changes he brought to their workplace.

As Twitter executives have driven toward a feistier version of their company, tension has not been limited to the design department and its adjoining research group. Workers have complained, sometimes bitterly, about being demoralized.

“We’ve got teams across the board that are reporting things like, ‘We’re concerned about our future,’” Ms. Christie said. “They talk about fear or psychological unsafety.”

The conflicts at Twitter have been echoed at other tech companies where executives are taking a harder line with employees who had grown accustomed to accommodating workplaces. Coinbase, a cryptocurrency company that went public this year, banned political discussions at work and offered exit packages to employees who disagreed with the rule. And this month, Google faces a trial before an administrative law judge after the National Labor Relations Board accused it of wrongfully firing employees who protested company decisions.

“Any kind of major change in blueprint comes with a risk,” said Robert Sutton, a professor of organizational behavior at Stanford University.

Cultural shifts rile employees and sometimes cause financial instability, he said. “There is always this balance between: Do we do it by socialization and having a strong culture, or do we do it with money and cracking down on people?”

Although some Twitter design employees were rattled by the meeting in which they were required to critique one another, Mr. Davis said several had thanked him for the candid feedback.

“We’re kind to one another,” he said. “But also being nice means that you might shy away from saying the thing that needs to be said for us to move forward together.”

Mr. Davis told his staff that he would push for improved performance, and he quickly criticized, demoted or cut workers, more than a dozen workers said.

When employees were let go, he and other managers sometimes followed their departures with emails to the staff remarking on their poor work.

Many employees feared they would be next on the chopping block. Although Mr. Davis, who manages 200 people, stressed the importance of giving critical feedback, he sometimes lashed out at workers who criticized him, employees said.

But others believed Mr. Davis's changes were essential to Twitter's survival. The company needed to toughen up, one employee said.

By late 2019, complaints surfaced to Twitter's employee relations unit, which is staffed by lawyers who investigate workplace issues. The unit looked into accusations that Mr. Davis had created a culture of fear. Among the concerns was that he had made a biased remark to another executive.

The comment occurred during a meeting in which Liz Ferrall-Nunge, who led Twitter's research team, shared concerns about diversity at Twitter and referred to her experience as a woman of color. Mr. Davis seemed to dismiss her, telling Ms. Ferrall-Nunge, who is Asian American, that if she wore sunglasses, she would pass as white, three people familiar with the investigation said.

Ms. Ferrall-Nunge, who left Twitter in 2020, declined to comment. Twitter declined to comment on the record about the episode, citing employee privacy.

Twitter employees who were aware of the episode said they expected better from Mr. Davis because of his outspokenness about diversity. Others defended his track record on diversity, noting that white executives were given more slack while making less effort on diversity issues.

In a lengthy Google document sent in February 2020, Mr. Davis praised Twitter's friendly culture. But he criticized the quality of design and argued that employees were too quick to say yes to projects when they should instead provide criticism. The overly kind atmosphere stifled honest feedback, he argued.

Employees who received the memo noticed that, in the margins, they were able to view comments from human resources representatives and managers

who had edited the document. They were asking Mr. Davis to tone it down. He said other people had told him that it had the proper balance of “tough love.”

That summer, Mr. Davis became the target of online harassment. Extremist groups believed he was involved in kicking them off Twitter, he said. He received death threats, and his personal information was published online.

“I would get a death threat at 12 o’clock, and then at 12:05 I would have a meeting,” Mr. Davis said.

By early 2021, another employee relations investigation into his behavior was underway, in response to complaints that the culture of fear persisted. Ms. Christie said that employee relations looked into every employee complaint and that Mr. Davis was trying to change his behavior.

“We’ve got to find our own Twitter way of direct feedback that’s still empathetic, that’s still respectful,” she said. “That’s not an easy combination.” Mr. Davis was “heartbroken” by the employee complaints, she added.

Company data was beginning to reveal widespread discontent on the design and research teams. Attrition under Mr. Davis had risen and was about double the rate of overall attrition at Twitter, employees said. In annual surveys, employees who worked for Mr. Davis consistently said at a higher rate than other Twitter employees that they felt “psychologically unsafe.”

“I’ve been hearing and absorbing feedback about the culture and morale,” Mr. Davis wrote in a note shared with his management team that was seen by The Times. “I love and deeply respect this team, it’s the strongest team I’ve ever worked with, and yet it’s clear that many of you aren’t feeling that from me. I’m taking a step back to think about my style and approach.”

In March, after a year of battling election and coronavirus misinformation, many employees struggled with burnout. Mr. Davis announced that he planned to move away from the performance culture that had been his mandate.

“My goal is for us to transition to a team of belonging, which is less transactional and more focused on care and support,” Mr. Davis wrote in an

email to employees. He cited the harassment he had received, and asked employees to be patient if they felt he had not done enough to support them.

“I was not celebrating wins, I was focused entirely on what was wrong,” Mr. Davis said, describing feedback he received from his staff. “Since then, I spent some time working on this. We have been celebrating wins, we have been finding ways for the team to come together.”

Current employees said sudden firings and harsh feedback continued. They found evidence for their concerns in Nikkia Reveillac, the head of Twitter’s research department.

Ms. Reveillac told Mr. Davis and other employees that his defensiveness made it intimidating for employees to offer him feedback. In May, she went to Mr. Dorsey. In a message she described to her co-workers, she told him that the culture under Mr. Davis was toxic and causing untenable attrition. Mr. Dorsey did not respond.

Weeks later, Ms. Reveillac was abruptly pushed out of the company and locked out of her work accounts. “Team, I didn’t get to say a proper goodbye. I love and miss you,” she tweeted. Ms. Reveillac and Twitter declined to comment on her departure.

In a staff meeting shortly after, two people who attended said, Mr. Davis told employees that they should not assume Ms. Reveillac had left the company because of conflicts with him. But without a clear explanation, employees were left wondering about whether her sudden departure was a response to going to Mr. Dorsey with her concerns.

###

On Tue, Aug 10, 2021 at 4:40 PM Liz Kelley [REDACTED] wrote:

Hi all,

This week, the New York Times will publish a fairly critical story on our shift to a more performance-focused culture. This story is unprecedented for us. Many current and former employees anonymously spoke with The Times and leaked many internal emails, documents, and conversations – specifically about Dantley and D&R – dating back to 2019 up to present day.

We expect not-so-flattering examples of missteps will be included throughout.

For context, the story started in a much more contentious position. Given the overwhelming amount of leaked information, and in addition to hours spent on background, we put Jennifer and Kayvon on the record with the reporter. This helped cement that this is an intentional shift, and broadened the story from anonymous sources criticising Dantley's leadership, to a story focused on our journey to become more goals-oriented. We are optimistic that the story will be less "toxic culture and management" and more "company undergoing intentional change while trying to still keep its caring culture intact" than it otherwise would have been.

Given Dantley and D&R sits at the center of this story, we're preparing a note for him to send to his team, but not sending proactive comms to all Tweeps at this time. We'll closely monitor internal sentiment and adjust if things start to skew very negatively. We will also share reactive TPs for D&R and other leaders as needed.

We're grateful for the support of many people throughout this process copied here. We'll flag the story as soon as it's live.

Liz on behalf of comms

--

Liz Kelley | Twitter Communications | [REDACTED]

--

Liz Kelley | Twitter Communications | [REDACTED]

Draft : Narianne 6 weeks v 1 [was: Re: Follow-Up]

Peiter Zatko [REDACTED]@gmail.com>
To: Peiter Zatko [REDACTED]@gmail.com>

Thu, Jan 27, 2022 at 9:11 PM

Your timeline is incorrect and does not reflect the relevant events.

For the record, Parag told me not to create and send corrective materials six weeks ago (when I suggested and offered that I create corrective materials and send them to the committee in place of what was sent). When I continued to push on the topic before and immediately after the committee meeting Parag said he would work it out with me over the break when he would meet up with me on the east coast.

I waited to meet with Parag. A meeting he never kept.

It has been just over two weeks since you and I had the interview in the investigation. An investigation that was triggered by my lawyer assisted letter.

It was in that interview you and I agreed that corrective materials should probably be created.

One week later as I was finishing putting together the corrective document I was terminated and my system was immediately remotely locked and a corporate security person was arriving at my house (an hour drive) to collect the laptop.

This was a day after a pop-up meeting with Omid and Parag to which I was only given a few hours notice.

In that very meeting Parag stated that he had been waiting on corrective materials from me for a month. That was a patently untrue statement (with documentation) from the very person who told me not to produce those very materials.

So, to make your statement more accurate, I have had ~1 week to recreate corrective materials that were only agreed upon as necessary 2 weeks ago.

Saying I have had 6 weeks to put together such materials is an interesting way to frame the situation.

I would appreciate the record be accurate and represent the situation correctly.

Fortunately I have documentation for these events and the for the content of meetings and conversations.

On Thu, Jan 27, 2022 at 5:37 PM Marianne Fogarty [REDACTED]@twitter.com> wrote:
Mudge,

I'm writing to let you know that we are convening a meeting of the Risk Committee. At the meeting we will brief the Risk Committee on the concerns you raised in your January 4, 2022 email to Parag and Dalana, including the context you provided in the meeting with Omid and subsequent emails to Sean Edgett and me. If you have additional concerns that you have not yet raised, please do so as soon as possible, but no later than February 1, so that they can be brought to the attention of the Risk Committee in this meeting. It has been six weeks since the presentation about which you raised concerns, and just over three weeks since our investigation began. We do not feel it is appropriate to wait any longer given that we have provided you with repeated opportunities to share further information.

Thank you.

Marianne

On Mon, Jan 24, 2022 at 10:07 AM Marianne Fogarty [REDACTED]@twitter.com> wrote:
Mudge,

Please let me know if you have time today or tomorrow or any time in the near future to talk about your additional items. I would like to speak with you to understand the concerns so that I can look into them. Please forward copies of any documentation relating to those items to me in anticipation of that conversation and to enable me to begin to investigate.

Thank you.

Regards,
Marianne

On Sun, Jan 23, 2022 at 5:31 PM Peiter Zatko <[REDACTED]@gmail.com> wrote:
Yes. There are additional significant items (with documentation).

Appreciate the Sunday email and request for my time. I am unavailable to speak right now.

On Sun, Jan 23, 2022 at 12:42 PM Marianne Fogarty <[REDACTED]@twitter.com> wrote:
Hello Mudge.

I'm writing to follow up on my email of Friday afternoon asking to speak with you to be sure I understand and can look into any additional concerns you may have. You indicated in email that you have additional concerns, and I understand from Patrick Pichette that you indicated the same to him.

My job is to conduct a thorough, objective investigation and report on it to the Audit Committee, of which Patrick is chair. Given the nature of the issues, here I will also report to the full Risk Committee of which Omid, with whom you spoke last week, is chair. But I can only investigate what has been reported to me, and I cannot complete my investigation knowing there are matters that have not been investigated.

I'm available to speak with you at your earliest convenience. Can we set up time to talk today?

Regards,

Marianne

On Fri, Jan 21, 2022 at 1:21 PM Marianne Fogarty <[REDACTED]@twitter.com> wrote:
Hi Mudge.

Can we set up a time to talk to identify and understand the full breadth of additional concerns/issues and can look into them?
The sooner the better. Are you free this afternoon? I'll make myself available at whatever time works for you.
Thank you.
Marianne

On Fri, Jan 21, 2022 at 10:07 AM Peiter Zatko [REDACTED]@mail.com> wrote:
Yes. These are now additional concerns/issues introduced by Parag.

On Wed, Jan 19, 2022 at 11:59 PM Marianne Fogarty <[REDACTED]@twitter.com> wrote:
Mudge,

Kathleen Pacini advised me that during your conversation this morning you mentioned the concerns you raised about information shared with the Risk Committee in December. If you were referring to matters that are not already under investigation, please let me know so we can schedule time to talk right away.

We appreciate that you raised your concerns and want to be sure they are fully and appropriately addressed. While my investigation is not complete, we intend to bring all of the concerns you raised with me and Omid, as well as the document you have been preparing in response to my email of January 11, to the Audit Committee and the full Risk Committee in the coming days. If there is anything else you want to include or recommend we do with respect to this issue so it concludes to your satisfaction, please let me know.

I'm available to hear any further concerns you may have as well as any additional thoughts you may have to resolve this matter.

Regards,

Marianne

--



Marianne Fogarty | VP, Chief Compliance Officer
Registered In-House Counsel
Pronouns (She/Her) | San Francisco, CA
Follow me @MarianneFogarty

--



Marianne Fogarty | VP, Chief Compliance Officer
Registered In-House Counsel
Pronouns (She/Her) | San Francisco, CA
Follow me @MarianneFogarty

--



Marianne Fogarty | VP, Chief Compliance Officer
Registered In-House Counsel
Pronouns (She/Her) | San Francisco, CA
Follow me @MarianneFogarty

--



Marianne Fogarty | VP, Chief Compliance Officer
Registered In-House Counsel
Pronouns (She/Her) | San Francisco, CA
Follow me @MarianneFogarty

--



Marianne Fogarty | VP, Chief Compliance Officer
Registered In-House Counsel
Pronouns (She/Her) | San Francisco, CA
Follow me @MarianneFogarty



Risk Committee - Regulatory / Public Policy Risk Update

Twitter, Inc.

December 16, 2021



TWITTER CONFIDENTIAL

Regulatory Risk Dashboard

OVERALL RISK ANALYSIS		TOP SUB-RISKS	QUALITATIVE FACTORS / KEY MEASURES	TREND	ANALYSIS / COMMENTS
POTENTIAL IMPACT	H	Content regulation laws	<p>Numerous jurisdictions are introducing legislation to require proactive removal/interventions in specific types of harmful content. Intermediary liability changes are being discussed globally, including US Section 230, and India has already changed its rules. The EU's major overhaul (DSA) will impact a wide range of areas and likely take effect in 2023 and the UK introduced a new Online Safety Bill in May 2021.</p> <p>Adult pornography is increasingly the focus of regulatory movement. Notably the UK, JP, FR and DE are focused on how we ensure minors are not able to access it, which is highly challenging when age verification technology is not yet sufficiently available. India now requires its immediate removal.</p>	↑	<p>This will require core investment to facilitate proactive removal of critical illegal content - CSE, NCN and terrorist content - in addition to greater investment in non-illegal but harmful speech (hateful conduct, misinformation). Our ability to comply with laws, and enforce our own rules, in a fuller range of languages is critical to reducing risk here. Our strategy on Section 230 is to educate policy-makers and gain time. In the UK, legislation will undergo a full round of pre-legislative scrutiny, including a dedicated scrutiny committee, which we will engage with fully.</p> <p>Twitter is uniquely exposed on this issue--most platforms do not allow adult porn. The effectiveness of features like safe search are being called into question, while spam-porn is a related challenge. We will launch age gating in January 2022 (delayed multiple times), but our approach to ask users to tell us their DOB is not seen as credible. We are reviewing investments in this area and evaluating risks of monetization features.</p>
LIKELIHOOD	H				
TREND	↑				
RISK OWNER		Less-democratic regimes and free expression	<p>Our values are being challenged by more aggressive postures from less- and anti-democratic countries, including India, Russia, Turkey, China, Vietnam and Brazil. Those regimes are seeking to use domestic law to compel us to do a range of things, including content removal and access to user data. Penalties are often significant financially, but also can involve criminal liability for the company and employees, and ultimately throttling or blocking of the service.</p> <p>These requests are increasingly in conflict with our values and free expression norms, in markets we have identified for revenue and service growth. Our reputation and political risk is impacted by the speed and clarity of decision making with regard to compliance with specific legal frameworks, due to delays in communicating with key stakeholders.</p> <p>India is a major challenge in this regard. The Government enacted new laws on intermediary liability, requiring us to comply with legal requests for user data in 72 hours and content requests in 36 hours, appoint a Chief Compliance Officer (who has personal criminal and civil liability), a law enforcement liaison and a Grievance Officer. Proactive monitoring of rape/CSE content is required. A Code of Ethics and Procedure will be published in future.</p>	↑	<p>We have a cross-functional process to assess new laws and build compliance programs that align with our values (and sometimes mean we refuse to comply and challenge the laws in court). The tension between revenue growth and compliance is increasingly both a regulatory and reputational risk. Our decision about whether to litigate to fight for our principles will help build trust that we are proactively seeking to protect the people on Twitter's voices, but also increase political tension.</p> <p>It is hard for us to counter the political strategy of hostile governments when they are able to exploit delays in our review of legal requests and growing backlogs. We are investing additional resources to reduce this backlog, but inadequate tooling is an impediment to work around.</p> <p>In India, we have increased our compliance and are ready to litigate when we believe orders to remove content or turn over user data violate Indian law. Litigation would challenge problematic aspects of the law, like requiring an individual to be held criminally liable for our content decisions. Our investment in capacity for our content moderation teams, as well as the Hindi language, is a direct factor in reducing this risk.</p>
RISK CHAMPION					
RISK MANAGEMENT STATUS		Privacy and data localization laws	<p>Privacy laws (that also include information security and data protection requirements) continue to proliferate. While there are commonalities between these laws, each has important idiosyncrasies that are costly to comply with. For example, data localization requirements are an increasing regulatory trend and the pressure to comply with these requirements is growing. There are significant risks with compliance and non-compliance strategies. Complying can place data and individuals in harm's way and not complying may result in fines and sanctions up to blocking orders.</p>	↑	<p>The work we are doing to prepare to meet our FTC obligations will support a general level of compliance with many global privacy regulations. We will still need to decide how to meet the idiosyncratic requirements, like data localization. Data localization presents not only a challenge to our principles, but also as to our ability to technically implement solutions and any solution will place data at greater risk.</p>
PRESENTATION DATE					
ALIGNMENT TO STRATEGY					
<p>On Track</p> <p>December 2021</p> <p>*Regulation and new legislation is a direct risk to increasing the costs to our business as we increase #participation and #durability.</p>					

● Completed ● On Track ● At Risk ○ Delayed

TWITTER CONFIDENTIAL



Regulatory Risk Dashboard (cont'd)

TOP PRIORITY MITIGATION PLANS	CORRELATING TOP RISK	OWNER	DUE	STATUS	RECENT PROGRESS	RISK OPTIMIZATION RESULT
Legal Compliance Tooling. Recently-dedicated Health Tools team focused on innovations that will help Twitter meet our legal and regulatory responsibilities in a fast changing environment.	Content regulation laws	Sean Edgett	Ongoing		Team established in Q3 2021 at half capacity, focused entirely on launching legal directives EUCD & Korea Art in this quarter. The team is expected to hire significantly in upcoming quarters to meet additional needs such as foundational tooling and product launches.	Create and execute internal tooling strategy to ensure we are meeting our policy and legal and regulatory requirements balancing the needs of the user.
Wingspan 3.0 Proactive strategic plan for new regulations, including categories of laws and regulatory requests we know are coming.	Content regulation laws	Sean Edgett	Q4		Our system for intake, analysis, and implementation of new laws is up and running. We've engaged outside counsel to categorize new laws over the next two years to help us with topline decisions. Working group meets biweekly.	This work will enable us to make recommendations about categories of laws in advance, map product work, and think holistically about our compliance plans, rather than in a one-off fashion as new laws come online.
Self-regulatory / Best Practice models	Content regulation laws	Sinead McSweeney	Ongoing		Met with Turkish Minister to discuss the benefits of EU self-regulatory models and are providing language for their new law. Became board member of the DTSP and participated in creation of self assessment model of T&S risk (assessment to be conducted in Q12022).	Reducing pace of new regulatory proposals.
China WG. The China WG is focused on mitigating risks and having the company make informed decisions related to our current operations and plans to grow revenue 10X in the next few years.	Less democratic regimes and freedom of expression	Sean Edgett	Q4		We are teeing up a conversation with Staff in January 2022 about the risks of doing business in China. We've completed a risk analysis where each function identified current and/or future risks the company faces if we increase footprint in China, and determined likelihood of each risk whether we have or can easily create a risk mitigation or elimination strategy.	Informed and principled risk decisions about Twitter's operations in China.
India WG. The India WG is focused on compliance with the new IT guidelines.	Less democratic regimes and freedom of expression	Sean Edgett	Q4		Engaged in crisis tabletop exercise to analyze how the company could/would respond to various risks and threats, as well as determine what risk mitigation strategies we can utilize and/or create to prepare. Currently identifying various work streams and action items to put additional risk mitigation strategies in place.	Informed and principled risk decisions about Twitter's compliance with new laws in India.



TWITTER CONFIDENTIAL - See privacy and data protection materials for mitigation plans

Completed
 On Track
 At Risk
 Delayed

Follow-Up

Marianne Fogarty [REDACTED]@twitter.com> Wed, Jan 19, 2022 at 11:59 PM

To: [REDACTED]@gmail.com

Cc: Rebecca Falk [REDACTED]@twitter.com>

Mudge,

Kathleen Pacini advised me that during your conversation this morning you mentioned the concerns you raised about information shared with the Risk Committee in December. If you were

referring to matters that are not already under investigation, please let me know so we can schedule time to talk right away.

We appreciate that you raised your concerns and want to be sure they are fully and appropriately addressed. While my investigation is not complete, we intend to bring all of the concerns

you raised with me and Omid, as well as the document you have been preparing in response to my email of January 11, to the Audit Committee and the full Risk Committee in the coming days. If there is anything else you want to include or recommend we do with respect to this issue so it concludes to your satisfaction, please let me know.

I'm available to hear any further concerns you may have as well as any additional thoughts you may have to resolve this matter.

Regards,

Marianne

--

Marianne Fogarty | VP, Chief Compliance Officer

Registered In-House Counsel

Pronouns (She/Her) | San Francisco, CA

Follow me [REDACTED]

AAR:

Operational Overview

S3 received a request from [REDACTED] on December 13th, 2021 for protective services in Pennsylvania and NYC between December 17th, 2021 and January 2nd, 2022. The request included the need for private air transportation from San Jose to Philadelphia on December 17th with return travel to California from Philadelphia on January 2nd. However, on December 17th at 4pm, [REDACTED] advised of a destination change to Teterboro, NJ from Philadelphia as originally planned. S3 helped coordinate this change and at 9:30pm, [REDACTED] drove the family to the [REDACTED] for their flight.

Due to the change in plans, a [REDACTED] met the family in Teterboro on the morning of December 18th and facilitated the family's movement to NYC [REDACTED] had planned for an internal resource to pick up the family in Teterboro). [REDACTED] having relocated from Philadelphia, arrived in the NYC area within two hours of the family's arrival and remained in NYC on standby until December 19th.

On December 20th, S3 facilitated the family's movement to [REDACTED] PA. Also on December 20th, S3 received a request to drive the family to [REDACTED] PA the following day. On December 21st, S3 drove the family to [REDACTED] PA. Subsequently, on December 22nd, S3 drove the family from [REDACTED] back to [REDACTED]. The original itinerary suggested that [REDACTED] would be traveling from [REDACTED] to NYC and back on December 23rd, so S3 prepared to support that movement. However, this movement was canceled before the 23rd. [REDACTED] remained on standby in [REDACTED] until December 28th.

On December 28th, S3 facilitated the family's movement from [REDACTED] PA to NYC as well as the family's return from NYC to [REDACTED] on December 30th. [REDACTED] remained on standby in [REDACTED] until January 2nd, when he drove the family to the Philadelphia [REDACTED] return flight. The family departed Philadelphia as planned at 11am Eastern, touching down in [REDACTED] at 2pm Pacific. Thereafter, [REDACTED] drove the family to their [REDACTED] home.

Duration:

- We had "boots on the ground" for **17 consecutive days**, from December 16th, 2021 through January 2, 2022.

S3 Personnel:

• [REDACTED]

The Christmas Holiday (S3 tried to rotate everyone home for a portion of the holiday with Family), this being S3's first trip with [REDACTED] and the lack of a known reliable vendor in Philadelphia required that S3 layer extra resources for this operation. Although S3 preferred to use internal resources for driving during this operation, two different contract drivers from a vetted security vendor were used in several instances.

Vendors:

• [REDACTED]

Advance Cities:

- Multiple advances (10+) were conducted in [REDACTED]

Covid-19 mitigation steps:

- All relevant team members conducted COVID rapid tests prior to assignments. At least 15 rapid tests were conducted by S3 team members during this trip

- Other COVID mitigation steps to include keeping drivers consistent, minimal exposure to team members, and wiping down vehicle interiors on a daily basis

What Went Well:

- EA Communications:
 - We had good communication with [REDACTED] throughout the task. We expect these communications to improve as we all get more familiar with [REDACTED] [REDACTED] [REDACTED]
- Limited Drivers
 - Kept potential COVID exposures to a minimum

What Didn't Go Well/Had Room for Improvement:

- One of the vendors was experienced and skilled but not a perfect fit (spoke openly about other EP details he worked).
- S3 managed several last minute scheduling changes [REDACTED] [REDACTED] [REDACTED] [REDACTED] schedule conflicts. To allow for uninterrupted support S3 team members to rotate home during the holidays, more staff than normal was required for this operation.
- The [REDACTED] invoice for air travel went directly to [REDACTED] [REDACTED] [REDACTED] had to be involved in chasing down invoice payment. Further, it is possible that the total service cost seems to have taken her off-guard and caused budget concerns.
- The last minute flight change from Philadelphia to [REDACTED] created the need to use a second contract driver when we were trying to limit exposure to different drivers due to the COVID risk.

Action Items:

- Preemptively sort contracts process for less impact to [REDACTED]
- [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED]

August 28, 2021

Leslie Berland

Re: Health Objectives #1 and #2, and partnership with TwS

To: Peiter Mudge Zatko [REDACTED]

Ahh Mudge this is amazing. You deserve this type of bench! And this will be an amazing turn-around story in her career. Congrats my friend 🙌

On Fri, Aug 27, 2021 at 9:45 PM Peiter "Mudge" Zatko [REDACTED] wrote:
Leslie,

Please don't forward Katrina's note below.

I wanted to share it with you because it makes me extremely happy.

She's precisely what, and who, we need to lead an *operations* team. Katrina is providing precisely the knowledge and expertise hoped for.

Her gravitational pull, executional maturity, and leadership is critical for improving (fixing) <cough> several teams.

Two of the 5 teams within confidence (privacy is the other) are now on their paths to being world class. Early and rocky but demonstrably on their paths.

I'll continue to pour fuel on Privacy and TwS and bring the next confidence teams onto their paths...

Thank you for supporting and encouraging bringing Katrina onboard. I did my own research and legwork on her and felt comfortable but some others had reservations. Your support helped.

V/r,

Mudge

PS - I wasn't expecting to need an entire rebuild in my world when I joined. But, damnit, if that's what's needed we're gonna do it. :)

----- Forwarded message -----

From: **Katrina Lane** [REDACTED]
Date: Fri, Aug 27, 2021 at 7:47 PM
Subject: Fwd: Health Objectives #1 and #2, and partnership with TwS
To: Peiter Mudge Zatko [REDACTED]

TLDR: CONFIDENTIAL update for you on Health OKR partnership. As noted when we talked this week, we had a pretty heavy lift over the last several weeks to get from a high level conversation on "appoint someone with authority to force TwS to do this", to "TwS you need to just commit to the goals as we know the work", to an actual plan that we can validate against resourcing and start to make progress. The most recent commitment push is demonstrated in the bottom email from Oscar.

We would never share this level of drama of course in how things stood. But I wanted you to know in case you ever have to defend the fact we stopped debating commitment and said we need to get a real plan (not what they wanted to hear). It is still a very real risk none of this moves the needle enough to meet their OKR but that was true before too and this plan will for sure move it more. We just don't know enough to size the impact until we do the first deliverables.

I have reviewed this with Oscar and I think he is starting to see the difference, but regardless likes the plan and realizes we may need their help to execute (though he would like our request in triplicate by Monday :) but we have work to do to get there. And importantly, I think our team really sees the difference and now feels set up to have an impact (though a long way to go in doing the work). This was also developed by them working as a team and pulling together - it's a very promising step.

For more details: To get a sense of what the difference is and the progress you can compare his asks in that email (theoretically ok but not actionable and no internal TwS alignment on details as the details were not documented), to the detailed current plan. You can also look at the fact that in his email the words and the actual linked pilot are not even doing the same thing (objective 2 ask 4) to get a sense of the disconnect. We bear some of the responsibility (the dynamics between the pillar leads caused a lot of this, and Michael's approach was also a big issue). Health does too as they pushed very hard on commitment with a statement of fact that the data is clear and we know what to do that was both wrong and distracted all.

Let me know if questions and again if you would rather I spare you some of the sausage.

K

----- Forwarded message -----

From: **Katrina Lane** [REDACTED]

Date: Fri, Aug 27, 2021 at 7:56 AM

Subject: Re: Health Objectives #1 and #2, and partnership with TwS

To: [REDACTED]

Cc: [REDACTED], [REDACTED], [REDACTED]

Please see our detailed update below getting to the specific actions we will be taking in partnership with parts of Health. We believe these plans are the right step to drive the metrics and also are at a more detailed and specific level which should help us collectively better align on the actual planned work. We have time today so we can discuss.

The team is proceeding to confirm named resources (though work is well in progress on the pilot and work will proceed on the others), as well as assessing any additional help needed from Health, and determining the risk to other existing health roadmap work in TwS (on the assumption that these OKRs are the priority). We should have a first cut to you by end of next week.

We took your original email and added in blue the new detailed plan. Talk soon...

k

Health Objective #1: Unhealthy impressions

In more detail:

1. Develop a baseline to help us understand how much impressions “latency” we have in our system (the number of impressions that violative content accrues between posted and actioned, in this context time is irrelevant since tweets accrue impressions at vastly different rates)
2. Develop and implement queue/workflow ranking capabilities: We need to sort and prioritize the way cases are reviewing in a way that helps us achieve this goal. This work is inflight by Health Tools.
3. [TwS Owns] [Launch an experiment by Sept 10th to validate that combining the TOS + Virality model results in reduced unhealthy impressions. Date for assessment will be sent by next week.](#)

4. [TwS Owns] Assess options to align staffing to demand for specific health queues to determine potential impact on unhealthy impressions reduction by shrinking the report to action age gap. Surface options and requirements by end September, and collectively agree with Health what should be implemented within the quarter.
5. [TwS Owns] Run an opportunity analysis between now and end Sept to determine potential impact on impressions reduction to be realized by deprioritizing cases beyond Virality by age. We will start by assessing the impression decay timeframe for the xx biggest current health FIFO queues
 - a. Identify how much agent capacity is currently allocated to FIFO queues
 - b. Develop impressions vs time data for the x biggest FIFO queue and establish “half life” threshold. Then calculate the potential improvement in unhealthy impressions.
 - c. Based on results: assess what it would take to operationalize this prioritization to the agents and agree with Health what can be implemented in year. Ideas include possibly an experiment with a subset to analyze potential impact or implement auto-closure/deprioritization across all FIFO queues
 - d. *Commitments above may have external dependencies from Health XFN that TwS will need support for or require trade off discussions related to lower priority efforts.*
6. **[Request for TwS]** In the spirit of determining ROI, I fully recognize that impressions are not the only thing that matters (e.g. non-consensual nudity is higher severity than spam) and would like your help in developing a framework that we can align with T&S on to determining ROI:
 - a. Our proposal - Health to ensure sure that we are pulled into efforts to create toxicity or harm/impact type frameworks (we are aware of several including severity of impressions chart in Trust and Safety) by being included as consulted in the DACIN, so that we can be sure that this is operationalized in a way to achieve the expected outcomes.
 - b. Second: we will flag throughout this process anywhere we are seeing challenges or direction to pull resources away from our generally agreed to be highest harm work (CSE, NCN, PPI) in order to achieve numeric goals on overall impressions as we believe this would be an unintended consequence. We would like to agree on the best forums to do this on accelerated timeframes.

These replace the original request

1. **[Request for TwS]** *Align agent staffing times with the times at which impressions are generated on the platform*
2. **[Request for TwS]** *Ensure we execute [a pilot](#) to prove out that we can prevent a large number of impressions, and once we do so, shift a large number of agents towards proactive review workflows*
3. **[Request for TwS]** *Incorporate impressions into the criteria for how we evaluate the ROI of manual review (instead of just action rate)*

Health Objective #2: Customer expectations with regards to health-related interactions

In more detail:

1. Top of the funnel:
 - a. Set expectations (in the reporting flow) with customers that there are certain types of reviews (e.g. spam, annoying) in which they shouldn't expect a response from us
 - b. Decrease the number of incoming reports by redesigning the reporting flow. This project is already in flight.
2. Middle of the funnel:
 - a. Automate (primarily in the form of auto-closing with comms) with a particular focus on the large categories of cases (safety and media)
3. Bottom of the funnel:
 - a. Ensuring that we respond to cases within our stated 3 days

Requests from TwS:

Status added below - in italics

1. Continue TSAN support to help us understand the entire funnel, determine appropriate pacing towards hitting goals, and identify the largest opportunities. *TWS: there are no plans to redeploy resources away from current health work but we need to document the current activities and resources and ensure that between existing resources/Health funded analysts that we are able to sustain this level. TSAN is working to document.*
2. Ensure Engineering and Health Data Science has access and training related to operational data and cases, this has been a critical gap that will help us understand opportunities to improve the front-end product *TWS: Per discuss, we need health to be more specific about what this would entail and if different from current RTB we will need to discuss how to resource/what would drop*
3. *ADDED by TWS: We reviewed all existing planned actions against the rest of the funnel and commit to supporting two key middle of funnel efforts where we can help best ensure success: to autoclose bystander safety core reports with high likelihood of non-violation (with customer comms) and providing c/x response for reviewed/non-violative reports in the MTT reactive main group (Partially tracked in TSVC-5 and PAR-580)*

Hi Katrina / [REDACTED] / [REDACTED] -

I realize there are some questions around Health objectives #1 and #2, specific commitments TwS is making in order to hit those objectives.

In the interest of getting to absolute clarity and ensuring we can promptly shift to implementation/execution, wanted to drop this email distilling the work needed down to the basics, for the full context, you can continue to refer to the [H2'21 Health strategy](#).

One of my concerns is that once we factor in planning time, holidays, and any potential disruptions -- we have ~12 effective weeks left in the year to hit these goals (which is not a lot of time given the work ahead of us).

Happy to discuss any parts of this that are unclear, you disagree with, and/or require further refinement. Thank you 🙏

Health Objective #1: Unhealthy impressions

We are reviewing and actioning the majority of violative content way too late into its lifecycle. Based on what we know about how impressions accrue, we have 2-4 hours before content accrues 50%+ of its predicted lifetime impressions, in the majority of cases we action content after it has already accrued the impressions. There are a number of drivers that contribute to this:

1. The way our workflows are ranked (mostly FIFO) because we lack more sophisticated ranking capabilities
2. Even in a world where we significantly improve ranking capabilities and front-ends - which takes time -, the review load is still split across two review platforms (RTP and ServiceCloud) improvements on one side don't carry to the other side
3. The mismatch in review capacity between when impressions occur on the platform and when agents are staffed to review them (e.g. weekends)
4. The way we evaluate success for workflows (primarily based on action rate)

Strategy in a nutshell: Review and action violative content before it accrues impressions

In more detail:

1. Develop a baseline to help us understand how much impressions "latency" we have in our system (the number of impressions that violative content accrues between posted and actioned, in this context time is irrelevant since tweets accrue impressions at vastly different rates)
2. Develop and implement queue/workflow ranking capabilities: We need to sort and prioritize the way cases are reviewing in a way that helps us achieve this goal. This work is in flight by Health Tools.
3. **[Request for TwS]** Align agent staffing times with the times at which impressions are generated on the platform

4. **[Request for TwS]** Ensure we execute [a pilot](#) to prove out that we can prevent a large number of impressions, and once we do so, shift a large number of agents towards proactive review workflows
5. **[Request for TwS]** Incorporate impressions into the criteria for how we evaluate the ROI of manual review (instead of just action rate)
6. **[Request for TwS]** In the spirit of determining ROI, I fully recognize that impressions are not the only thing that matters (e.g. non-consensual nudity is higher severity than spam) and would like your help in developing a framework that we can align with T&S on to determining ROI

Health Objective #2: Customer expectations with regards to health-related interactions

When customers report content they believe violates policies to Twitter, it's a key moment in their journey and forms lasting perceptions as to whether Twitter "cares" and is doing something to improve the health of the platform. There are a few drivers that contribute to our current situation:

1. We don't do a good job setting expectations up front as to whether we'll be getting back to users on their report/appeals and if so when they can expect to hear from us, which results in:
 - a. An inherent pressure to review as many reports as we can
 - b. Inevitably disappointing some customers that expected a response
2. We don't have a clear unified sense of what makes a report "valuable" (is it about customer expectations? Is it about impressions?)
3. We rely on silent auto-closure of cases (28% of cases are silently auto-closed), which continues to perpetuate the perception that Twitter doesn't care

Strategy in a nutshell: Decrease the number of reports that we need to review, so we can focus effort on "high value" reports, appeals, and other requests and get back to users within 3 days

In more detail:

1. Top of the funnel:
 - a. Set expectations (in the reporting flow) with customers that there are certain types of reviews (e.g. spam, annoying) in which they shouldn't expect a response from us
 - b. Decrease the number of incoming reports by redesigning the reporting flow. This project is already in flight.
2. Middle of the funnel:
 - a. Automate (primarily in the form of auto-closing with comms) with a particular focus on the large categories of cases (safety and media)
3. Bottom of the funnel:
 - a. Ensuring that we respond to cases within our stated 3 days

Requests from TwS:

1. Continue TSAN support to help us understand the entire funnel, determine appropriate pacing towards hitting goals, and identify the largest opportunities
2. Ensure Engineering and Health Data Science has access and training related to operational data and cases, this has been a critical gap that will help us understand opportunities to improve the front-end product

Team,

Today

I wanted to give you a preview of news we will be sharing with the company in a few minutes - I spoke with your leads today too. I know this is a lot of change, and I appreciate your patience and flexibility as we navigate this together. Nick is extremely excited to welcome you to his org, and you will be hearing a lot more from him soon. In the interim, Lea Kissner will serve as CISO and Alan Rosa (our new head of IT starting January 24) will manage this team while we conduct a search.

Thank you for everything you do - we support you and we are here for you during this transition. Please reach out to me, Nick, your HRBPs Danielle and Reeda (cc'ed), or any of the InfoSec leads if you need anything at all.



New Message



Team,

Today

I committed to all of you that I would take fast and thoughtful actions on the priorities that matter most. Decisions about our leaders and how we are structured are foundational. With that, I wanted to let you all know that Mudge Zatko is no longer at Twitter. I made this decision following an assessment of how the organization was being led and the impact on top priority work. I have the utmost confidence in the leads who have jumped in with full focus, and the teams who do this critical work. I also made a commitment that I'll always be transparent with you - that is very important to me. In this case, while I'm sure you may have questions, the nature of this situation limits what I can share at this time. I appreciate your trust and understanding.

Moving forward, we are



New Message



Today

Moving forward, we are restructuring teams within Confidence to further reinforce the General Manager (GM) operating structure that we announced last month to improve accountability and speed of decision-making. The changes are the following:

Privacy, IT, and Information Security orgs will move to the CoreTech org led by Nick Caldwell. Lea Kissner will continue to lead Privacy, and Alan Rosa (starting January 24) will lead IT. Both will report directly to Nick Caldwell. Separately, [REDACTED] [REDACTED] will be departing Twitter, and we thank [REDACTED] for [REDACTED] contributions. We are kicking off a search for a permanent CISO, and Lea will take on the responsibilities in the interim. With these changes, I will be holding the CoreTech org accountable for delivering on our key priorities around privacy and



New Message



NICK Caldwell. Separately, [REDACTED] will be depa Today vitter, and we thank [REDACTED] for [REDACTED] contributions. We are kicking off a search for a permanent CISO, and Lea will take on the responsibilities in the interim. With these changes, I will be holding the CoreTech org accountable for delivering on our key priorities around privacy and security.

Twitter Service (TwS) led by Katrina Lane will move to the Bluebird org led by Kayvon Beykpour. Bringing these organizations together enables us to create the best end-to-end customer experience (for all of the ways people interact with Twitter) and for this new Bluebird org to be collectively held accountable for their outcomes around Health.

Corporate Security (CorpSec) led by Pat Geonetta will move back into the Counsel org, reporting to Sean Edgett. Both CorpSec and the



New Message



Today

Twitter Service (TwS) led by Katrina Lane will move to the Bluebird org led by Kayvon Beykpour. Bringing these organizations together enables us to create the best end-to-end customer experience (for all of the ways people interact with Twitter) and for this new Bluebird org to be collectively held accountable for their outcomes around Health.

Corporate Security (CorpSec) led by Pat Geonetta will move back into the Counsel org, reporting to Sean Edgett. Both CorpSec and the Counsel team sit at the center of issues that impact our service and our people, and there was great strength in their tight collaboration.

The work within the #Protect Objective remains important, and will ladder into the respective organizations moving forward.

1



New Message



Edgett. Both CorpSec and the Counsel team sit **Today** enter of issues that impact our service and our people, and there was great strength in their tight collaboration.

The work within the #Protect Objective remains important, and will ladder into the respective organizations moving forward.

Through change, we can get stronger, but it requires intentional work from all of us. Thank you to the teams driving this critical work for your agility, commitment, and focus now and moving into the future.

Onward,

Parag

1m 🕒

"I appreciate your trust and understanding." is a triggering phrase to me

Now 🕒



New Message



Example of misrepresentation from InfoSec -
in action @ Staff mtg.

Jan 11, 2022 Staff Meeting - I mention we have work to
do to ensure we have identified
and patched log4j sufficiently
(we have not)
Pera y asks for clarification on
disconnect around log4j - [redacted] had
sent out a note that led him to believe
we are 100% compliant.

In 1/5 Confidential Staff meeting [redacted]
implied otherwise.

In December updates IT Reported
unknown amount of instances.

Seth has (said to others) we don't have
appropriate inventory to even know where
log4j lives.

Example of [redacted] presenting a subset "win"
as universal ground truth. Inappropriate & misleading
1/5 Confidence Staff meeting [redacted] refused
to put update and info around log4j in writing
in the doc and ~~only~~ insisted on only verbal updates.

Today ▾



Parag Agrawal 3:16 PM

lets talk at 2



Mudge Zatko 3:18 PM

with omid there? I want to understand how transparent you want to be about [REDACTED]. The initial plan with HR was that [REDACTED] would not be at the meeting and that the document would be replaced. That was blocked.

I'll see you at 2



Parag Agrawal 3:22 PM

Want us to be totally transparent with Omid and committee



Mudge Zatko 3:23 PM

ok - thank you. I appreciate it. I know the audit committee needed to keep the two topics very separate so I wanted to understand how to do that here (or if I should) as they are intertwined

sorry - audit investigation

Message Parag Agrawal

New



Fwd: Q4 2021 Risk Committee Issues



----- Forwarded message -----

From: **Peiter Zatzko** [Redacted]
Date: Mon, Feb 14, 2022 at 3:40 PM
Subject: Q4 2021 Risk Committee Issues
To: Patrick Pichette [Redacted]

Hi Patrick,

Here is the document I will be sending to Twitter that provides details around issues in the materials presented to the Q4 2021 Risk Committee. In case it is not provided to you in its entirety, or with appropriate context, I want to send you a copy directly.

I really appreciate you reaching out to me early on and being receptive to me sending you such items directly. I have concerns around the company's ability, or willingness, to operate in good faith at this point. Knowing I am interacting with you helps with that.

The document contains three sections:

Section 1 (6 pages) - A description of the activities and events on the timeline leading up to what appears to be my illegal termination.

Section 2 (10 pages) - Accurate descriptions of Twitter's most significant risks as of Q4 2021. This section basically what Parag ordered me to NOT provide to the risk committee.

Section 3 - (10 pages) - Factual errors and misrepresentations in the Information Security Risk Committee information presented (over my documented objections and efforts to prevent).

If you have **any** questions, please let me know.

very respectfully,

Peiter Mudge Zatzko

 **Twitter Q4 2021 Risk Committee Issue _2022_02_14.pdf**
3572K

Twitter

Marianne Fogarty [REDACTED]@twitter.com>
To: Peiter Zatko <[REDACTED]>
Cc: [REDACTED]@twitter.com>

Fri, Feb 25, 2022 at 1:12 PM

Mudge,

Thank you for the additional information you forwarded on February 14. The Risk Committee was provided that information, as well as information about the concerns you raised initially.

Having ensured that the Risk Committee has complete information and context around the December 16 presentation, we have now closed our investigation. We thank you for raising your concerns and please be assured that we have addressed them as appropriate.

Regards,

Marianne



Marianne Fogarty | VP, Chief Compliance Officer
Registered In-House Counsel
Pronouns (She/Her) | San Francisco, CA
Follow me [REDACTED]



19 Jan 2022

Peiter "Mudge" Zaitko
[REDACTED]

Dear Mudge,

Per our discussion earlier today, this confirms that your employment with Twitter has ended and your access to our systems has been terminated. Your last day on our payroll will be January 24, 2022.

To recap our call this morning, we're parting ways because Parag has lost confidence in your leadership. For example, the issues that have surfaced [REDACTED] [REDACTED] [REDACTED] [REDACTED] have made clear that you are not creating a healthy working environment for your team, and that your conduct falls short of what we expect from our leaders.

During our discussion, you raised the ongoing investigation being conducted by our Compliance team with respect to your concerns regarding materials provided to the Risk Committee [REDACTED] [REDACTED] [REDACTED] [REDACTED] in December 2021. As you know, we have taken your concerns very seriously and our Chief Compliance Officer is conducting a thorough investigation under the supervision of the Audit Committee Chair. As you also know, Twitter has many mechanisms to flag concerns, including our ethics helpline (ethicshelpline.twitter.com), and you remain welcome to use those mechanisms at any time even after you are no longer an employee.

It's our hope that you will leave here amicably, based on your previous commitment to do so should Parag ask you to leave. Going forward, I will remain your contact for anything related to your employment and termination. My email address is [REDACTED][@twitter.com](mailto:[REDACTED]@twitter.com).

Sincerely,

[REDACTED]
Senior Director, People

September 14, 2021

Peiter Mudge Zatko

Fwd: [Inform] Verification Operations Update

To: Kathleen Pacini [REDACTED]

FYSA

----- Forwarded message -----

From: **Peiter "Mudge" Zatko** [REDACTED]

Date: Sun, Sep 12, 2021 at 10:31 PM

Subject: [Inform] Verification Operations Update

To: staff [REDACTED]

Per my comments at Friday Staff on the quantified Verifications improvements with TwS/Product, and Kayvon's email on TwS and Product progress, I wanted to continue sharing with Staff.

One of the reasons I'm pleased, as you will see in [REDACTED] update below, is that TwS decided to use the Verifications problem as a case study for how TwS plans, staffs, and executes... and to correct things needing correction.

Katrina has been exceptional and the teams are learning a *lot*. They are also recognizing that she expects a lot from her people but that she grows them, cares for them, and respects them.

Several TwS leads have told me that they are learning so much that they never knew from working with Katrina. Their (TwS) apprehension around a new leader has turned/is turning to tremendous respect and even excitement about what the potential impact in their jobs and confidence in themselves due to execution and beginning to deliver.

There is still a lot to do, of course, even with the product and support for Verifications, but I am hopeful. That hope is now backed with data :)

First Verifications! Next, the core components of TwS and all new product support!

Mudge

----- Forwarded message -----

From: **Peiter "Mudge" Zatko** [REDACTED]

Date: Sun, Sep 12, 2021 at 5:48 PM

Subject: Re: 9/10: [Inform] Verification Operations Update

To: [REDACTED]

CC: Katrina Lane [REDACTED], Twitter Service Pillar Leads

[REDACTED], Verification Review [REDACTED]

Team,

Outstanding work being done on Verifications.

I understand on Sunday we expect to clear out the backlog! We **always** want to be managing to SLAs instead of backlogs; fantastic work!

Your efforts here are being driven by data, tracked with metrics, and it shows. Having enough precision to identify a handful of agents with too high of an error rate, pulling them out, and sending them for additional training is a level of specificity that shows *ownership* of these operations.

Most importantly the way you are driving these improvements is methodical and able to be institutionalized. This bodes tremendously well for TwS.

I'm proud of what I'm seeing. More importantly, I hope you are able to see this progress and that you are taking pride in your accomplishments and yourselves.

Go Twizzles!

Respectfully,

Mudge Zatko

On Fri, Sep 10, 2021 at 7:23 PM [REDACTED] wrote:

Executive Summary

This week we ramped our production further to bring us to a total of 201 agents currently working through our verification applications. Our open case count sits at 4K cases as of Thursday the 9th of September. In our fifth week our pilot agents reviewed 26K cases. With a reduction in incoming volume and increase in staffing, we are happy to report that we should have the backlog cleared by the weekend and will be ready to receive new cases as the verification

application opens up to more customers in the existing categories. We have partnered with the product team and plan to ramp up as quickly as we can provided we are making or exceeding the agreed 14 days SLA for all cases. If we start to create new backlogs of cases outside of SLA or at any point see a jump in error rate, we will collectively adjust the ramp up plan as needed.

Summary of volume:

	09/03	09/04	09/05	09/06	09/07	09/08	09/09
Incoming	1,332	789	893	1,145	1,454	1,580	1,756
Bulk Actions	72	0	3	42	3	0	0
Manual Reviews	3,231	435	1,437	4,411	5,541	5,901	4,640
Backlog	19,271	19,626	19,079	15,771	11,681	7,360	4,476

What have we accomplished?

Training Update

Our fifth and sixth batches of pilot agents (182) started production in a limited capacity this week. We extended all training classes to 7 days by adding 2 more days to allow for more case practice and agent shadowing before moving into live case review based on the agent quality performance. In addition, we also introduced new guidelines for our training expectations and higher agent quality targets before graduating into full operations. This is to ensure that only top performing agents review incoming verification applications.

Quality Detail

The Effective Agent Correct Outcome Rate for the week (05-Sept to 09-Sept) is 88.48% (there were 2689 cases audited by QAs out of 22K total cases reviewed by agents). The Effective Agent Correct Outcome Rate for Approved Verification Applications is 83.54% (this would make the Effective False Positive Rate 16.46% as it is the remaining percentage where the outcome is wrong for verified applications). There were 41 false positives on the 560 approved reviewed cases. The Effective

Agent Correct Outcome Rate for Rejected Verification Applications is 90.09% (9.91% Effective False Negative rate).

We are continuing to analyze and categorize the root causes of the errors into categories that will be reviewed by Product, Policy and Operational SMEs to understand how and where to drive improvements: policy/ tool gaps, knowledge gaps that could be addressed via training changes, agent oversight/human error, or grey areas.

Automation Update

A data issue caused our Authenticity Model data pipeline to go down and resulted in the usage of stale Authenticity Scores for incoming applications from 8/30 - 9/3. TwST DS & Consumer Identity engineering fixed the pipeline on 9/3 and we have remediated all affected applications. Out of the 7,035 applications affected by the pipeline failure, only 36 applications need SME/QA review and will be done next week in partnership with IdOps and QA team.

Debadging Update

As part of our quality review process, we are instituting a process for debadging false positive applications. We started by identifying all existing false positives to date and have a total of 182 requests that were approved out of line with the current policy. This past Tuesday we debadged and sent out communications to 71 English speaking accounts out of the overall identified 182 accounts that were incorrectly Verified and found through our QA audits. The remaining accounts are non- English accounts and we are currently in the process of building out localized templates to complete this action. We do expect some escalations and resubmissions due to the time lag before this action, and will assess as needed and include in our normal feedback loops. Furthermore, we are partnering with the product team for a more sustainable bulk actioning tool for us to repeat this process as BAU.

Moving forward

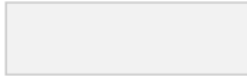
Our focus in the next week will be to accept new applications as the verification application opens up to more customers in the existing categories. We will ensure we

are closely monitoring our quality results for desired level of accuracy as we process applications from a larger population of our users.

--



(She/Her) | Director, User Operations



Twitter Q4 Risk Committee InfoSec Corrective Information and Timeline

In Q4 2021 information was presented to the Twitter Risk Committee. I have expressed concerns around the material that was sent and the verbal presentation [REDACTED]. I believe the materials (the MATERIAL) does not correctly characterize the risks and Twitter environment in the way required for board members and the Risk Committee. The material and presentation include significant omissions, tell partial stories that imply a different reality, and in some cases contain data that is difficult to view in ways other than misleading, at best.

It is important that the record be set straight. This letter attempts to capture how this information ended up making it to the Risk Committee. The second part of this document contains a description of some of the significant differences between what was sent, spoken, and Twitter's actual risk and environment.

Ensuring inappropriate MATERIAL would not be created and/or presented to the December Q4 Board and/or Risk Committee was identified as a concern near the end of October of 2021. Numerous efforts were made by Mr. Zatko, from October through December, to prevent inappropriate information from being put in front of either the Board or the Risk Committee. Each was blocked. The issue was repeatedly escalated. Near the end of November Mr. Zatko took direct measures to replace materials presented at the Q4 2021 Board meeting (Dec 9, 2021) to ensure it was accurate and truthful. With repeated promises of resolution targets consistently not kept, and the Board meeting over with the Risk Committee meeting coming up, the issue escalated directly to the CEO (Parag) and the head of HR (Dalana). In this conversation Parag promised to address and resolve the issue personally, and in a time frame that would allow Mr. Zatko to have confidence in the correctness of the materials presented to the Committee. Days later Parag called Mr. Zatko to apologize. In his own words "I am not keeping my promise." He stated that he was the person now blocking corrective actions.

In response to this Mr. Zatko re-raised concerns over potential misrepresentation to the Q4 Risk Committee and Mr. Zatko suggested, and offered, to create corrected alternative material (CORRECTIVE_MATERIAL). In a call with Mr. Zatko, Parag instructed Mr. Zatko to *not* prepare and send CORRECTIVE_MATERIAL and to forward the MATERIAL as-is to the risk committee. Further it was stated that [REDACTED], the owner and source of this information, would present the material and Mr. Zatko was to attempt to make real-time corrections. Mr. Zatko expressed concerns regarding the inappropriateness of this approach. In response, Parag offered that he would personally call the Risk Committee members, if requested, to help clear up any misleading information. Mr. Zatko reluctantly agreed, did not create CORRECTIVE_MATERIAL at that time, and followed up with an e-mail to Parag and Dalana to have a record of this interaction and the path chosen.

Mr. Zatko attempted to follow the instructions given by Parag at the Risk Committee meeting. Feeling uncomfortable that the InfoSec information in the hands of the Risk Committee could likely still be misconstrued after reflecting on the meeting and meeting contents, Mr. Zatko took Parag up on his offer to follow up with the committee. Parag called Mr. Zatko and told him that

he was disappointed in him that the issue was not gone. It is not known whether Parag followed up with the Committee at that point.

Parag said that he and Mr. Zatkan would get together over the holiday in NJ during Parag's east coast trip. Their focus would be on how to solve the issue(s). Parag said he would contact Mr. Zatkan during his trip to schedule the meetup. Mr. Zatkan thought about the MATERIAL and the concerns during this time. Having not heard from Parag and the end of the break approaching Mr. Zatkan contacted Parag. Mr. Zatkan was told that the trip had been significantly changed and that no meeting would occur (security detail reports do not support the change in itinerary referenced).

As these concerns were ethics related, Mr. Zatkan drafted letters on these topics with the help of his personal counsel, and sent them to Twitter. The intent of the communication was to ensure appropriate processes were engaged and followed on such matters and to again request to be unblocked in discharging his duties of removing underperformers (██████████). The latter item, Mr. Zatkan was already told, had met the bar for documentation and communications to terminate for performance since October/November but was repeatedly blocked and was now being blocked by Parag.

January 11th Mr. Zatkan conveyed information to the, now initiated, Audit Investigation. In the meeting, launched by Mr. Zatkan's letter(s), Marianne Fogerty agreed ██████████. Ms. Fogerty and Mr. Zatkan agreed ██████████.

Mr. Zatkan received an email on January 17, a federal holiday, that CORRECTIVE_MATERIALS needed to be submitted by noon the following day. On January 18, Mr. Zatkan was informed with 3 hour's notice that he would be having a meeting with Omid. In this meeting Parag stated he had been waiting for Mr. Zatkan to create CORRECTIVE_MATERIALS for a full month. This is incorrect. It had only been 8 days since Mr. Zatkan and Ms. Fogerty agreed to write CORRECTIVE_MATERIALS. The same materials Parag had instructed Mr. Zatkan not to prepare. In the January 18th meeting Mr. Zatkan was pushed by Omid and Parag to send e-mails and raw notes immediately. Mr. Zatkan was not comfortable with this, expressing that a short clear document was needed for the Committee and that technical notes alone would be insufficient to capture how significantly the information mischaracterized the actual risk environment. Mr. Zatkan stated that he would forward e-mails that he had sent to the Audit Investigation and complete the CORRECTIVE_DOCUMENT as agreed upon during the investigation.

The following day, January 19th, Mr. Zatkan was terminated and his access to systems and documentation removed before the document could be completed and sent to the Audit Investigation, Parag, and Omid (Chair of the Risk Committee).

The cited reason for termination was a failure to create a positive working environment [REDACTED] [REDACTED], whose termination for documented performance issues, and subsequent these potentially fraudulent actions, has been pending and blocked for months.

Below is a description of the most significant elements that belong in CORRECTIVE_DOCUMENT.

[Peiter to create the 3-4 page document that *should* have gone to the risk committee and put it here, or send as a separate document]

loc 45

January 3rd Disclosed
(I learned) Twitter doesn't
have/never acquired the
rights to material on
VOC AI Models were trained

on - Basically un/covered.
situation. (Damien → Mudge)

This was apparently brought up
and disclosed to the Board a few
years ago. Brought up to
engineering and leadership a
few years ago as well. Never
addressed (apparently/reportedly)

Jan 17, 2022

P45

Staff Doc

Risk Control Report

Backup: Conf - Staff Doc

Backup: Oct - Jan Email

Backup: Calendar

log4j Doc review

10-20k could be running
by 4J - not solid - ~~we don't~~

we would need someone like
Patrick Newman in ZRT to
assist.

Seth Wilson - slash

Jan 16, 2022

AI Models -

own. rights to
original ^{license} training data

Ukraine - Malware service
(Discard) to avoid last minute
detection - deliver tools constantly ^{catch}

Twitter Nigeria

Notes I sent to [REDACTED] [REDACTED] (new senior hire as head of comms - was terminated after a few months)

Oct 1 —

Are you tracking that the Nigerian President said today that he will lift the ban?

He says Twitter has only agreed to 7 of 10 conditions, though

<https://punchng.com/breaking-buhari-orders-conditional-lifting-of-ban-on-twitter/>

Here's the full text in a local outlet:

<https://www.premiumtimesng.com/news/top-news/487593-what-buhari-said-about-twitter-ban-nnamdi-kanu-igboho-insecurity-others-full-text.html>

Paragraph 70-74

70. To address these negative trends, the Federal Government of Nigeria suspended the operations of Twitter in Nigeria on June 5, 2021 to allow the Government put measures in place to address these challenges.

71. Following the suspension of Twitter operations, Twitter Inc. reached out to the Federal Government of Nigeria to resolve the impasse. Subsequently, I constituted a Presidential Committee to engage Twitter to explore the possibility of resolving the issue.

72. The Committee, along with its Technical Team, has engaged with Twitter and have addressed a number of key issues. These are:

- a. National Security and Cohesion;
- b. Registration, Physical presence and Representation;
- c. Fair Taxation;
- d. Dispute Resolution; and
- e. Local Content.

73. Following the extensive engagements, the issues are being addressed and I have directed that the suspension be lifted but only if the conditions are met to allow our citizens continue the use of the platform for business and positive engagements.

74. As a country, we are committed to ensuring that digital companies use their platform to enhance the lives of our citizens, respect Nigeria's sovereignty, cultural values and promote online safety.

Few other local outlets, mostly saying the same thing:

<https://www.pulse.ng/news/local/buhari-orders-twitter-ban-lifted-but-with-conditions/w3tmwqe>

<https://techcabal.com/2021/10/01/buhari-gives-conditions-to-lift-twitter-ban/>

<https://www.thisdaylive.com/index.php/2021/10/01/buhari-orders-lifting-of-twitter-ban-only-if-conditions-are-met/>

<http://saharareporters.com/2021/10/01/twitter-reacts-buhari%E2%80%99s-conditional-lifting-4-month-ban>

Sept 18 —

I am uncomfortable that Twitter has been silent on this. I fear we are now positioned to play the heel.

Nigeria can make whatever “demands” and (within the Nigerian market) if we do not give them everything they state it looks like we are going back “on our word”. A word we never actually gave but that the world will believe we did.

Similarly If Nigeria decides to continue the ban it looks like Twitter is the one at fault.

In the State department this tactic is pretty much known.

Is it too late to send a letter to Nigeria? Something along the lines of:

“we are reading that you appear to be in negotiations with someone claiming to be Twitter. We have not had these conversations and want to make sure you are protecting yourself as this appears to be a potential imposter. We are still very interested in meeting and working with Nigeria to come to an amicable solution to the current situation. You are an important country and market to us, one we respect very much. We are also aware of the financial loss your people and businesses are suffering from this ban (Kelly -I have figures if you need them - we are actually very well leveraged for negotiations here -Mudge) and we want you to be able to support these businesses and your citizens.”

At that point we subtly slide the open letter to a trusted journalist to give the truth a bit of light that can be cited and referenced in the future... when we need to be able to defend our position.

Or do you have other suggestions / ideas?

On Sun, Dec 12, 2021 at 7:55 AM Peiter "Mudge" Zatko <[REDACTED]@twitter.com> wrote:

Privileged and Confidential

Parag and Dalana ([REDACTED]),

I apologize for the need to bring this to your attention and for the time critical nature of the ask. We all have enough on our plates and this should have been able to be handled without more executive support than myself.

I need your support in resolving the matter of [REDACTED] Monday (Dec 13).

Twitter [REDACTED] is stuck in regards to actioning this item. The item was supposed to have been resolved weeks ago. Multiple promised deadlines have come and gone. [REDACTED]

[REDACTED] I am blocked from being able to correctly perform key portions of my duties and obligations for Twitter, including for the board and Risk Committee, and we are causing Twitter increasing harm through these continued delays.

[REDACTED] have context on this item.

Thank you for your attention to this (now) urgent matter.

Kindest,

Mudge Zatko

At present there are 4 areas of risk, at Twitter, that are the most critical for the Twitter Board and Risk Committee to understand: Access Control, Patches and Software Updates, Software Processes and Compliance, and Incidents.

It is imperative that the Risk Committee have a clear and accurate understanding of these areas. The understanding needs to be defined through quantified measurements, not qualitative descriptions or isolated projects, and placed into larger context. What does each risk poses? How does the compare to peers (what values should we be at presently)? How do these values compare to what regulators and the larger public, and our Customers, expect? What is the priority of addressing each area?

Access Control – the total scope and breadth of our challenge at Twitter around access to information and systems.

Where is Twitter in relation to peers and expectations of regulators and Customers? What is the impact of our current situation? Where do we need to be, by when, and how are we tracking?

Patches and Software Version Compliance – Keeping Operating Systems and software correctly patched and current is considered one of the most fundamental and basic components of security hygiene. In addition to essential protection against security vulnerabilities, and disruption of operations, regulatory agencies specifically look at this to understand if an organization lacks maturity in the most rudimentary of security practices. Where is Twitter in relation to peers and expectations of regulators and Customers?

Processes and Compliance (SDLC / SDL) – Software and service development and deployment needs to go through mature uniform processes (SDLC) and be methodically evaluated for security and privacy risks through a mature Security Development Lifecycle (SDL). Decisions regarding security and privacy need to be made at the appropriate level and with global context for value v risk to the company. that decisions made in regards to security and privacy concerns

Incidents – The number, type, and frequency of issues related to security hygiene, controls, and integrity of systems and processes indicate risk that is being realized and risk that is forthcoming. This is the gauge measure whereas access control and patches and updates are levers. version compliance are levers

Access Control

Twitter is significantly behind peers and below acceptable (and industry) standards in access control. To quantify this statement Twitter is at a point analogous to where Google was in 2005-2007. Peers, and companies in general, block employees from accessing systems (and data) in production. This has been the norm for years in the industry. Production environments are sacrosanct at most companies. Not only is this because of security concerns due to live

customer data and direct customer interactions but also to avoid outages and disruptions in service. A non-insignificant number of outages and operations disruptions at Twitter are due to testing and development happening in production instead of testing or staging environments.

How excessive is the access to production systems and data at Twitter? Each engineer Twitter hires (approximately half of all employees and equaling more than five thousand employees at the end of 2021) is provisioned with the ability to connect and communicate with production systems. This includes the ability to directly log in to systems in this sensitive environment.

Twitter data centers and may launch services that directly interface with production data. A range of outages at Twitter have been attributed to failed tests because they occur in production with live systems. These outages are expensive to Twitter even when not highly visible externally. They consume meaningful resources internally and interrupt other work. Numbers for outages, disruptions, and of all sizes, related to a lack of a testing and staging environment should be considered

Edge case – IPMI + SUDO + ... versus main risk.

Risk Committee Misalignment Concerns

Access control is a critical path item for Privacy. Data deletion, and other regulatory obligations, require appropriate, mature, access control. Regulators and our Customers/Users are under the belief that Twitter has more mature, and enforced, access controls than are present. At the end of 2021 there is not a concrete plan to address access control on which IT, Engineering, Privacy, and InfoSec are aligned.