WHISTLEBLOWER
AID

Report government and corporate lawbreaking.
Without breaking the law.

— SENSITIVE & CONFIDENTIAL WHISTLEBLOWER DISCLOSURE —

July 6, 2022

Honorable Mark Warner, Chair
Honorable Marco Rubio, Vice Chair
U.S. Senate Select Committee on Intelligence
211 Hart Senate Office Building
Washington, D.C. 20510

Jay Bratt, Chief
Counterintelligence and Export Controls Section
National Security Division
U.S. Department of Justice
950 Pennsylvania Avenue, N.W., Ste. 7700
Washington, D.C. 20530-0001
Via FedEx or Hand Delivery

Re: Protected Disclosures of Suspected Penetration of Twitter, Inc. by Multiple Foreign Intelligence Agencies, and Other Threats to U.S. National Security

Dear Senator Warner, Senator Rubio and Mr. Bratt:

1. We respectfully request that this disclosure remain confidential for reasons of personal privacy, threats to health and safety, and U.S. national security.

2. We are lawyers representing **Peiter "Mudge" Zatko**, who was employed as "Security Lead", a member of the executive team responsible for Information Security, Privacy, Physical Security, Information Technology, and Twitter Service (the company's internal name for the division responsible for global content moderation enforcement), at **Twitter, Inc.** from November 16, 2020, until the morning of January 19, 2022, when CEO Parag Agrawal terminated Mr. Zatko after Mr. Zatko alleged fraud by Mr. Agrawal.

3. Before joining Twitter, Mr. Zatko held senior positions at Google and Stripe, and within the **Department of Defense**,[1] where he was authorized to access **Top Secret / Special Compartmented Information** for work on programs at the bleeding edge of full-spectrum computer network operations. The Office of the Secretary of Defense bestowed upon Mr. Zatko the **Exceptional Public Service Award**, the highest honor available to civilian, non-career officials. The value of Mr. Zatko's work for the United States has also been formally recognized by the CIA, White House, and U.S. Army.

4. Mr. Zatko decided to proceed with these disclosures quite reluctantly.[2] Under separate cover, Mr. Zatko is simultaneously making protected, lawful disclosures[3] to the U.S. Securities and Exchange Commission (SEC), U.S. Federal Trade Commission (FTC), and Department of Justice (DOJ) of substantial evidence showing that Twitter, Inc., CEO Parag Agrawal, as well as particular senior executives and 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 the FTC about, *inter alia*, the Twitter platform's security, privacy, and integrity;

    b. Violations of various securities laws including auditing and financial control provisions;

    c. Fraudulent and material misrepresentations in communications with the Board of Directors and investors, constituting securities law violations;

[1] 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 Mr. Zatko.

[2] Mr. Zatko helped create the modern information security community of ethical security disclosures. While criminals 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 hackers 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 make more informed decisions, and can push the service to improve. Mr. Zatko made a personal commitment to Mr. Dorsey, the Twitter Board, the greater public, and to himself, that he would do his best to help fix Twitter. Mr. Zatko spent about 14 months pushing improvements from the inside, and was terminated for his efforts. With a heavy heart, Mr. Zatko has concluded that these lawful disclosures are his ethical obligation. Mr. Zatko has tried his best to avoid disclosing unnecessary technical or sensitive information.

[3] The separate SEC / FTC disclosure is included as an exhibit to this document.

5. But this instant disclosure has a different focus: Twitter's negligence and even complicity with respect to efforts by **foreign governments to infiltrate, control, exploit, surveil and/or censor** the company's users, platform, staff, and operations. This disclosure includes information and an exhibit not contained in the SEC / FTC disclosure.

6. **No Privileged Contents:** Upon information and belief, no attorney-client privileged material, information or documents are included here. None of the information or documents provided here were received from a lawyer, part of a communication with a lawyer, or obtained for the purpose of seeking legal advice.

II. Twitter's Deficiencies Threaten Users and Global Security

7. Except for a few jurisdictions in which Twitter is blocked or censored at mass scale, [4] Twitter is available in most of the countries on earth. Dozens of the countries in which Twitter operates are unfriendly to democracy, as determined by Freedom House.[5] Twitter hold the sensitive data (not just content, but information that can reveal geo-locations) of hundreds of millions of users around the world, but its efforts to protect that data are drastically deficient.

8. **Twitter is a soft target:** To cite merely one pervasive problem, **over 50% of Twitter staff are engineers—about 4000 people around the world—and all engineers have direct access to Twitter's production environment with live user data and access to Twitter's full source tree**. This is highly unusual, and highly insecure. Other large tech companies have controlled environments for testing new code; if engineers are based in hostile countries their access to live user data should be strictly limited.[6] In addition to the significant exposure this presents to Twitter's systems and data in their production environment, Twitter was made aware by the US Government that portions of the source code at the company were controlled items as determined by the Export Administration Regulations by the Department of Commerce. Twitter was aware that the company was required by law to deny access to particular foreign national employees, but was unable to implement that

[4] "Censorship of Twitter - Wikipedia." https://en.wikipedia.org/wiki/Censorship_of_Twitter.
[5] "Freedom in the World." https://freedomhouse.org/report/freedom-world.
[6] Several companies refuse to place engineers in specific countries and/or maintain the ability to cut off access to entire offices. At Twitter all employees have excessive access to information that they shouldn't and it is presently impossible to cut engineers off from production system access.

legal requirement because of system architecture flaws. The significance of types of access and lack of controls to prevent inappropriate access from both internal and external entities are described further in the SEC / FTC disclosure.

9. Historically Twitter had been caught without any language translation capabilities prior to significant world events leaving them far behind their peers, unable to perform basic moderation, support, and analysis of platform manipulation by foreign entities. Twitter was caught unprepared for basic translation and language abilities to protect the platform and users from abuse and manipulations repeatedly. Twitter lacked language support and capabilities to support the forcible displacement and ethnic violence in Tigray, Ethiopia (Ahmaric and Oromo) Q1, 2021. Twitter lacked Burmese language ability when the military coup in February 2021 occurred in Myanmar. The company lacked language support and capabilities to support the platform during the US withdrawal from Afghanistan (Pashto and Dari). All of these shortcomings meant Twitter was largely blind to what purposes, and how, its platform was being used during these geopolitical crises.

10. **Perverse incentives to grow total users:** Twitter executives have personal financial incentives ("Value Creation Awards" exceeding $10 million each) for achieving aggressive global growth targets for "mDAU" (monetizable daily active users). Mr. Zatko believes that they are willing to look the other way and/or avoid confronting many of the problems identified in this disclosure and in the separate SEC/ FTC disclosure because they were not rewarded for advancing security, privacy or platform integrity (e.g. stopping disinformation and election meddling). To the contrary, such objectives detracted from growth.

11. These dynamics led Twitter to pursue aggressive expansion into new markets, including non-democratic states where governments were likely to impose problematic conditions in exchange for giving the company access to their citizens. Chasing growth, leadership refused to address a critical problem — the company's technical inability to deter, identify, or remediate activities to protect user data from the adversarial governments of the territories into which the company was expanding. Twitter is simply not capable of expanding into such undemocratic markets while protecting its users' data, and hence safety, from those governments. Across every dimension, monetary bonus incentives for executives rewarded short-term growth at the expense of longer-term safety, privacy and integrity.

12. **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 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.

13. **Foreign Agents on Company Payroll:** Even active foreign intelligence threats like information operations on the platform, and foreign agents on Twitter's payroll, were left unaddressed because of their expected negative impact on short-term mDAU growth. During his time as Security Lead, Mr. Zatko determined with high confidence that foreign governments continued to target Twitter and had successfully placed multiple intelligence agents on Twitter's payroll. Accordingly, it was highly likely internal Twitter systems were compromised by state actors. Even though this problem had happened before,[7] the Twitter leadership team resisted viewing this as a serious concern. Impediments to fixing the issue included a lack of detection or enforcement systems.[8] The lack of visibility to monitor internal activities, significant out of date software, misconfigured employee computers, and lack of access control and data protection are described in detail in the enclosed FTC / SEC disclosure.

14. **India's intelligence agency**, known as the **Research and Analysis Wing (R&AW)**, targeted Twitter physically and electronically[9], and forced Twitter to hire two

[7] "Twitter Insiders Allegedly Spied for Saudi Arabia - WIRED." 6 Nov. 2019, https://www.wired.com/story/twitter-insiders-saudi-arabia-spy/.

[8] For example, around August 13th 2021 it was internally revealed that between 1.5 and 3 *thousand* failed login attempts per day were occuring in Twitter's production environment. Additionally, as reported to the Board of Directors in 2021, centralized logging was not mature and only covered <~20% of systems and services. This lack of basic security hygiene was not relegated to only Twitter's data centers. Approximately ⅓ of the employee laptops were reporting they had disabled software updates. Employee computers were also reporting that they had disabled disk encryption, turned off their firewalls, were configured to allow remote access, and other significant lack of basic protections. As was reported to the Board of Directors in the 2021 Q4 Privacy briefing, because of fundamental and systemic problems with security and access control, all employees of Twitter have access to significantly more data than they need to perform their jobs.

[9] One item was a "raid" of Twitter offices in the middle of the night during pandemic lockdown. This does not support the India Special Cell police claim that they were there to talk to employees about a Tweet. The Special Cell also brought "media teams" with equipment for this visit to Twitter offices in the middle of the night during the pandemic lockdown expecting to find people at the office that they could interrogate.

particular agents of the Indian government. The Indian government also ran harassment campaigns against specific Twitter India employees,[10] including repeatedly requiring Twitter's India site lead to come to police stations, answer questions and surrender his electronic devices for significant periods of time during which the devices were believed likely to have been penetrated and/or imaged. Through these physical and online campaigns it is believed the Indian government was able to co-opt and manipulate at least one company employee to act as an agent of the Indian government. Another person, intending to fill a role required by the Indian government, was revealed as having a fabricated history not unlike those created for spies and intelligence agents.

15. An Indian Court, the police, and members of the **Ministry of Electronics and Information Technology (MEITY)** pushed the company to make these individuals Full Time Employees (FTEs), a status which would provide access to internal systems and documents. In one case, a custom system was spun up at significant effort to create a minimal-access environment that would allow them to perform their job, but not have the excessive access to information all other employees at the company were granted. This particular employee immediately requested copies of any legal documents and strategies Twitter was preparing regarding the Indian Court—an anomalous, out-of-role request according to the employee's manager and executive team members, that strengthened suspicions they were a government agent.

16. Twitter has hundreds of employees in India, all of whom have significant access to data far beyond what they need for their jobs. About 80 of those were engineers with full access to Twitter's source tree and default access to connect to systems and access data in Twitter's production environments. But design flaws in Twitter's system architecture made it impossible to reduce sensitive system access, or

"Police in India raid Twitter offices in probe of tweets with ... - The Verge." 24 May. 2021, https://www.theverge.com/2021/5/24/22451271/police-india-raid-twitter-tweets-government-manipulated-media. The use of India Special Cell police is a tactic that is rumored to be employed by R&AW in their targeting and instrumentation efforts.

[10] These included the public media in India targeting Twitter India individuals and visits to their houses and their relatives houses by the police. One individual suffered multiple police summons. Some of these included confiscation, and return, of his phone. Twitter personal cell phones are permitted access to large amounts of sensitive Twitter internal information and were also used to support authentication to other Twitter systems. Ultimately this person fled to the United States. Ultimately, after being targeted repeatedly by India's police, Courts, and newspapers, Twitter brought him to the US for an extended period for safety. Having agreed to follow very specific instructions to preserve the status of his cell phone, to allow for forensic analysis, he decided to completely erase and reset the device right before handing it over to Twitter Security. This removed any ability to perform appropriate forensic analysis.

identify or monitor for insider threat activity. Further, the harassment and threat of jail time to Indian citizens working for Twitter India, for the company not censoring certain tweets or handing over information on protesters' accounts carried weight.

17. Mr. Zatko repeatedly requested that the topic of whether to leave India and serve those users from outside the country, or invest in fixing the fundamental deficiencies be brought up for an executive decision. Even though several executives including the CEO Jack Dorsey expressed some support for leaving the market due to the physical and geopolitical risks, the topic was repeatedly tabled and never addressed. In one conversation with another executive Mr. Zatko explained the severity of the insider threat problem and why Twitter needed to consider leaving the environment. The executive replied along the lines of "if we already have one insider threat from the Indian government why does it matter if we have more?" There was too much revenue potential in India so Twitter should just accept the compromise, was a prevailing attitude for several executives.

18. **Selective disclosure of government backed influence operations.** Based upon information and belief, Twitter has a policy to publicly disclose attempts by state-linked entities to manipulate Twitter and its users[11]. As such, anytime they become aware of a governmental operation within their system, they should report on it publicly. Contrary to Twitter's own transparency principles they had been aware of an information operation being run by a branch of the Indian Army,[12] yet were refusing to include this information in their information operations transparency disclosures. (While Twitter was refusing to disclose information about Indian backed Information Operations, they released information about operations from other countries including India's rival Pakistan.)

19. Twitter staff approached Mr. Zatko confidentially, and told him that the direction to withhold disclosure of the Indian operations had been going on for almost a year.[13] (The staff hoped Mr. Zatko, in his senior executive role, could intervene to fix this.) Discussing the issue in both executive team meetings and in personal conversations it was revealed that the decision to not report Indian operations was

[11] "Information Operations - Twitter Transparency Center." https://transparency.twitter.com/en/reports/information-operations.html.

[12] The Indian Army branch believed responsible for repeatedly attempting and running information operations was Corps XV, also referred to as the Chinar Corps. https://en.wikipedia.org/wiki/XV_Corps_(India)

[13] According to personal notes on, or about, August 2nd, 2021, Site Integrity, the team responsible for disclosure of information operations told this to Mr. Zatko.

driven by a desire to curry favor with the Bharatiya Janata Party (BJP) and India's courts. Offending the BJP and India's Courts could be detrimental to Twitter employees in the country and user growth on the service. If India took action against the company it could ultimately negatively impact user growth numbers.

20. **Additional / Separate Foreign Agent:** Based upon information and belief, shortly before Mr. Zatko was terminated, Twitter was made aware of **at least one additional employee who was identified as a foreign agent**[14] working on behalf of another foreign intelligence agency. Given the global importance and value of the platform, broad system and data access, and unlikeliness of being discovered, foreign intelligence services would not be doing their job if they were not attempting to place agents, or recruit existing employees, inside Twitter. It is very likely that there are more foreign agents than those being disclosed that are operating undetected and relatively unrestricted within Twitter.

21. **Threats to Democracy:** Over the course of 2021, Mr. Zatko became aware of multiple episodes suggesting that Twitter was complicit in threats to democratic governance. In addition to India's efforts described above, several other countries were demanding Twitter open regional offices with actual Twitter employees in residence. **Turkey, Russia, and Nigeria** were three countries pressuring Twitter to stand up local offices with full time employees.

22. **Turkey.** Mr. Zatko provided a briefing on Turkey and the capabilities and risks presented to Twitter by Turkey's National Intelligence Organization (TNIO) and strongly urged the executive team to not open an office. Even with this information it seemed as though Twitter was looking at opening an office in Turkey. Mr. Zatko later learned that Twitter measured a meaningful amount of monetizable users[15] in Turkey.

23. **Nigeria:** Nigeria was perceived as important for growing Twitter's user base. Immediately upon joining Twitter in late 2020, Mr. Zatko lobbied with others (successfully) to not open physical offices in Nigeria but rather to serve the Nigeria market with offices based in nearby Ghana. The Nigerian government was viewed as unstable and unpredictable, and banned Twitter on June 5, 2021.

[14] The existence of another employee identified as a foreign agent had been reported from an external source to the Twitter CorpSec team. CorpSec is a team within Mr. Zatko's organization and this bit of information was reported up to him.

[15] Twitter stock value was largely dependent upon measurements of "monetizeable" users, referred to as mDAU.

24. After that, the Nigerian government falsely, and repeatedly, reported in the Nigerian press, which was picked up worldwide, that it was in talks with Twitter leadership.[16] For months, there were numerous false reports put forward by the Nigerian government describing non-existent meetings further describing various progress and setbacks being made in negotiations with Twitter. The truth was that the Nigerian government refused any and all meetings with Twitter to discuss the topic for months. Instead of meeting with Twitter the Nigerian government chose to repeatedly publish media articles saying they are in the midst of fabricated negotiations with Twitter and are almost at the point of agreement to end the ban. Nigeria even announced that Twitter had agreed in the negotiations to open a Nigeria office.[17] Twitter's failure to correct the public lies purporting to characterize non-existent negotiations enabled Nigeria to continue pushing their false narratives.

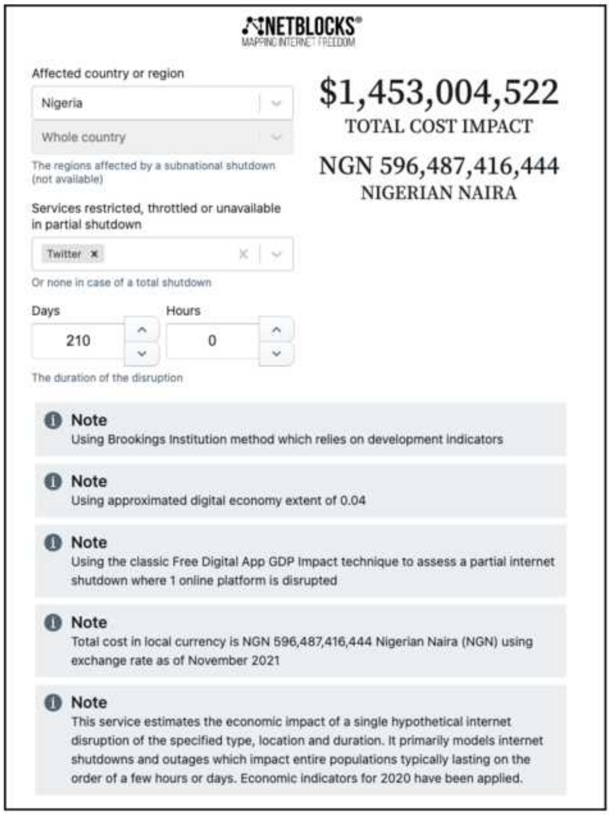
25. The false information, to which Twitter was now a party by not addressing, accomplished two things: 1) international human rights organizations were more willing to refrain from pushing as hard as they had initial censorship of the Nigerian people, and 2) Twitter shareholders were permitted to believe that the company would soon start to again be monetizing approximately 36.9 Million Twitter users in Nigeria.[18] The NetBlocks Cost Of Shutdown Tool, using methods from The Brookings Institute estimates a total monetary loss to the Nigerian economy due to the support of this public lie at approximately $1.4 billion dollars:[19]

[16] "What Buhari said about Twitter ban, Nnamdi Kanu, Igboho ...." 1 Oct. 2021, https://www.premiumtimesng.com/news/top-news/487593-what-buhari-said-about-twitter-ban-nnamdi-kanu-igboho-insecurity-others-full-text.html. Accessed 28 Apr. 2022.

[17] More details included in Exhibit 39.

[18] 39.6M users were mostly upwardly mobile economically and politically, with 20% using the platform for advertisement, and 18% using Twitter to look for employment according to a Twitter contractor who sent in information about the Nigeria situation at the time.

[19] This value is derived from the NetBlocks Cost of Shutdown Tool using 210 days and the shutdown of Twitter in Nigeria as input. From their website "The NetBlocks Cost of Shutdown Tool (COST) estimates the economic impact of an internet disruption, mobile data outage or app restriction using indicators from the World Bank, ITU, Eurostat and U.S. Census." "Cost of Shutdown Tool - NetBlocks." https://netblocks.org/cost/.



Affected country or region

Nigeria

Whole country

The regions affected by a subnational shutdown (not available)

Services restricted, throttled or unavailable in partial shutdown

Twitter ×

Or none in case of a total shutdown

Days 210

Hours 0

The duration of the disruption

$1,453,004,522
TOTAL COST IMPACT

NGN 596,487,416,444
NIGERIAN NAIRA

Note
Using Brookings Institution method which relies on development indicators

Note
Using approximated digital economy extent of 0.04

Note
Using the classic Free Digital App GDP Impact technique to assess a partial internet shutdown where 1 online platform is disrupted

Note
Total cost in local currency is NGN 596,487,416,444 Nigerian Naira (NGN) using exchange rate as of November 2021

Note
This service estimates the economic impact of a single hypothetical internet disruption of the specified type, location and duration. It primarily models internet shutdowns and outages which impact entire populations typically lasting on the order of a few hours or days. Economic indicators for 2020 have been applied.

26. On the Nigeria matter, Mr. Zatko sent repeated messages to senior executives including the new head of communications[20] from June through November 2021, requesting that Twitter correct the public record. Twitter's failure to do so misled investors and the international community on an important issue.

27. **Russia:** Russia was not viewed as important for user growth, and as such Twitter resisted the government's demands to place FTEs in Russia. But in or around September, 2021, a few months before then-CTO Parag Agrawal was promoted to CEO, in an in-person meeting in New York City Mr. Agrawal suggested to Mr. Zatko that Twitter should **consider ceding to the Russian Federation**'s censorship and surveillance demands. 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. Zatko. This interaction was notable because Mr. Zatko was already directing teams to prepare for possible Russian incursions into Ukraine.

[20] These were predominantly Signal.app messages from Mr. Zatko, based on personal notes.

28. **China:** 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. Zatko was told that Twitter was too dependent upon the revenue stream at this point to do anything other than attempt to increase it.

29. In none of these cases did Twitter act responsibly under the circumstances.

30. Our client would be willing to meet with investigators at your convenience. Please feel free to contact us using the information below.

31. 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 your assistance ensuring that our client never faces retaliation.

Sincerely,

Mark S Zaid

Mark S. Zaid
Founder & Legal Partner

John N Tye

John N. Tye, attorney at law
Founder & Chief Disclosure Officer
[redacted]

[redacted]

Andrew P. Bakaj
Senior Counsel
[redacted]

Kyle Gardin

Kyle Gardiner
Associate Counsel
[redacted]
Whistleblower Aid
[redacted]

Exhibits Enclosed:

1. SEC & FTC disclosure, filed separately as Exhibit A1

A1	SEC / FTC disclosure, filed separately (Link to come)
B1	Exhibit B1_20211001_Mudge Twitter Nigeria Notes_redacted
B2	Exhibit B2_202103xx_2021 Q3 or Q4 Risk Committee closed session meeting on foreign agent employee

B3	Exhibit B3_Mudge moleskine + phone notes Dec 2020 Feb 2022
B4	Exhibit B4_202112xx_Insider_Risk_December_2021
B5	Exhibit B5_2021XXXX_Snapshot of data center security system deficiencies_redacted
B6	https://drive.google.com/file/d/1_0hAm3WAKq3Osis4WJMjJsVFUHU2Id66/view?usp=sharing

END OF DISCLOSURE

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 interesting in meeting snd 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 is 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?

Here's where we are:

High confidence India office is compromised - people, facility, and selection of cell phones and laptops. RAW - they often gain physical access through commandeering a unit of Delhi's special cell and using them to show up physically as a cover (setup camera gear and reporters would cover otherwise conspicuous sigint activities)

Here's what "they" are going to do and try to accomplish.

propiety

RAW literally has a charter of "compromising foreign govt's and corporations" - to gain information and ~~influence~~ to strategically advantage the Court and the government (BJP presently). verify internal access

India Sees Twitter as an extremely valuable (and dangerous) propaganda tool and spy capability for enemies of govt).

Try to gain leverage to influence Twitter. They've seen how we react to pressuring our employees (former Counsel employee and [redacted]). They now have a strategic person in place to amplify this leverage. - They will verify

Attempt to gain complimentary information from within the office and from electronic communications to understand where Twitter's head is. What options we are considering wrt India's wants / demands. Explore Twitter's internal systems for other strategic value.

If they have not discovered already, they will learn that engineers can access production. (they will want to maintain this valuable abilitiy to shape the public conversation to India's purposes instead of Twitter's. Identify, Target, silence,

All of this will work against the mission of Twitter to serve the public conversation and to improve the health of the public conversation.

If we don't do anything and just stay the course - we will not win. India is presently leading this dance and we have done nothing but follow/react. Will give some options change this in a moment.

This is not a one off, here are the other countries that will attempt to manipulate or exploit:

China
KSA (remember their agent? That wasn't even the A-team and it was very damaging)
Turkey
Russia (new law for physical presence)

Here are our immediate options and what they buy us.

Pause hiring - conveys that we have options. Conveys that there may be risk to other companies' consideration of leaving the market. Buys time.

Could message we are pulling out.

Messaging to slow down RAW and give them pause. Especially to advantage counsel and court efforts (otherwise the court will be prepositioned against everything we present)

Coordinate "hunting" messaging and activity with Counsel legal efforts. Reset credentials, collect laptops, etc. TSCM visit. Adds meaningful cost to Intelligence activities and may protect some optionality for legal and business.

Most challenging option: full speed ahead and massive hiring. Confidence cannot close on the 10 year security deficit while knowing that we have a tumor internally and that the tumor is growing out of control.

Here's the longer term plan for solving this as a total problem.

Engineers out of production.
Data isolation / privacy enforcement
Positive Control of laptops and phones
Defined operational parameters for execution within "hostile" markets.
Position to neighboring countries to service target environments (e.g. Ghana v Nigeria)
Repeatable framework for offices in these environments and rolesfunctions we can/cannot support

what if we find them

FTC - Irish DPC -
CNIL -

SIM - 144
Cookies - Regulators
under the hood we're fucked -
Dev Birch don't understand state of affairs technically under the hood
(Project TAO?)

SIM STATS

terrifying & gets worse

~~an~~

we're not learning (main takeaway)

repetitive

SIM 89 not pulled to rest of org

Election Squad, mmm!

TWS Starts

...

There are other uncovered crises as well

(we were really lucky in the repetitive nature of Saul)

CDAP direct to

Can't scale but good

no visibility URL make usable

[redacted] : [redacted] 5:0

operate the

non-compliant

53% 186,372 firm

12% 41,443 OS

231,347

9/6/2020 - 2/6/2021

45 prod
37% ads

FTE	Contractor	Total
496	3174	3670

137
+ 37·2(74)
+ 22·3(66)
+ 300·4(1200)
1477

609
+ 224·2(448)
+ 32·3 (64)
+ 2309·4(9236)
10,357

let's look @ this

from just a security point of view —

There is an existential risk

based on my experience & reason I've been drafting : there is an existential risk

- ~~let's assume we handle the existential risk.~~

~~The~~

The what is existential risk

Then is how we operate as a company

[redacted] - [redacted]

low · 2% -

ir [redacted] underperforming

288 ~~292~~

294

- Mission Statement

~~Visibility~~

- 3 year Plan - short version

and

Paragraph showing how we get there

Existential Threat -

Jan 6-20th we received reports of ~~people disclosing the location of our SMF Data~~ people encouraging active disruption of our data centers.

We have [redacted] data centers and several possible effects outcomes from disruption

[redacted] Data centers destroyed, we cease to exist (not surprising) - ~~in a few years time~~

Irshe

India
↳ Dis info about NYT
↳ Critical of Modi

India compromised NYT employees (NSO group) mobile device spyware (covering Kushmere)

very little corporate targeting

Reporters
Non-Indian Nationals didn't go further than harassing online
Indian nationals targeted for compromise

Parag LMM Top of Mind

Splunk

~~Fedin~~

Failed logins - 1.53K
per day - (550 or
low the other day)

MESOS

Production Access -

SFM - 144

Cochies

Deactivates an
account – after 30 dy
no longer publicly ~~fr~~
available –

Tweets, Emails, phone #'s

No DMs, IP logs,
Device level data
in logs

—

CNIL is a beast on
data deletion.

- updated privacy policy 2-3 years ago
- some mitigations – daily, weekly, quarterly

— [redacted]

DPC asks question

we have to fess up we
cannot meet our deletion
obligations

GDPR – vio

FTC → vio

Historic LEO requests
when we said we
couldn't

Inability to meet our
data deletion obligations

30+ apps -

ITEA - Service Cloud

Health Cloud built 30+ tools

ruby, scala & react.

Lightning web components.

Lightning design system

For react.

Session ID leaking externally - anyone with this could access salesforce

[redacted]

Introduce myself

have each other in our rolodexes

TwS -

- ~~Strong inter~~

Ops side of T&S

- Building product health

Bridges

Shift agents from regular work to appeals work

Suspensions should have a reason

Threats are suspended
uses that were not
told why they were
suspended —

T&S ? CorpSec

many teams have ability
to suspend

Bots suspend (we don't
track bots)

all suspensions need to
have a reason
(fixing appeals in health KR)

We aren't tracking suspensions

TwS

Inventory all the bots

perceived accuracy

Reason for all suspension / expulsions

Move Agents from content moderation to appeals

TwS moving more and more towards SalesForce.

move more and more towards SalesForce

June 8, 2021

P+S

Health ITEA Contract

[redacted] [redacted]

JTBD

Make our policies straightforward and scalable in practice (on the platform) ~~t~~and disrupt and deter platform manipulation

Insure the Integrity of the platform at scale ~~and~~ to

Disrupting and deterring platform manipulation

Disruption of malicious parties on our platform becomes a strategy.

Parag Dhiru

Global Engineering presence

India → future of fices

TU1 cf.

Exceptional Broad access

Privacy

IDM.

Buy vs Build.

Inclusion & Diversity [illegible]

50K targets

65+ business execs

profile viewer
agent tool for periscope

→ went to fleets
spaces
→ old profile_viewer

never got the fixing that other tools got from SIM-FS

Anyone w/ access to periscope.admin bypasses all security

OIG Report

Ned / [redacted] – Nigeria / China / Turkey e-mail) to [redacted] etc.

July 19, 2021

✓ A P+J

✓ A3 Verifications

✓ A2 Staff Doc

[redacted]

C3 A [redacted]

✓ A4 NSO Data

C2 [redacted] Follow up [redacted]

C1 [redacted] [redacted]

All hands review

[redacted] follow on w/ Mudge on [redacted]

A4 Jack 360 (in Staff Doc)

go/healthy conversations

Inclusion & Diversity toolkits

50k targets

65+ business execs

profile viewer

agent tool for periscope

went to fleets

spaces

old profile_viewer

never got the fixing that other tools got from SIM-SG

Anyone w/ access to periscope.admin bypasses all security [redacted] what can I do to support our understanding of risk against our core

(A) mission —

Need systemic, repeatable solution - this is happening elsewhere.

Ultimate systemic risk to Twitter

Taking perspective of the threat (adversary) - how I think about these patterns.

Essential Problem:

India - this is about India's first.

Our ability to satisfy the mission.

About Risk

About India.

(B) What's the hard evidence that shows the problem (get out of qualitative narrative → go quantitative)

Bunch of facts - container of evidence.

② Want us to think about how we will handle it when I find evidence of India in our systems.

Sunlight - state backed operations · Indian Army

Protect our users

July 27, 2021

✓ A1 P+S

✓ A5 [redacted] Promo Pack

✓ A6 Quarterly Review input

✓ A3 Staff Doc Input

✓ A4 #Conf Doc.

✓ A2 Alarms

✓ A7 Recs Comms

B1 Slack

~~Nadia Doc~~

B2 Nigeria Doc

B3 Async Reviews

Aug 2, 2021

✓ A1 Ptg

B2 Kikta Report

A3 Conf Obj to Staff/[illegible]

✓ A2 Staff Dos

C1 Review Pkts

B1 Answer any Async Q2

A4 India (3 things [redacted] contractors adv view evidence

A5 SIM-144

B3 Rich Mode!

Rabdo Myalisis

SI - #Conf.

CP.

India — need to have a conversation Insider Risk is expanding —

Grievance Reports - India - going to CP, CP Kicks to SI, drowning SI.

R,

TOS review w/in T&S.

We have been sitting on an Indian Information Operation Influence Op for over a year.

[redacted] - India SIM-144 Exceptional Access to Production client flo

Cookies

privacy — Privacy

India/Countries — [redacted]

we don't have options in India

"We don't have any options"

no leverage - risk is not mitigated even if can't make in our favor.

- can't mitigate the insider threat - concern if ~~we~~ ~~keep hiring~~ with FTEs in sensitive positions in India

but ~~we have a lot of~~ sensitive most our roles positions

Access - Production

Laptop Fleet

pAdmin - has ubiquitous access to all accounts, (potus - etc.)

Spaces Dev team (Engineering) want to use for debugging (not the good tool to use for this)

[redacted] Cy

port over some support capabilities

{Cookies
{Privacy <—> pAdmin (Access to Production
} IAM Access (tool
} Laptop Fleet.
{India/Countries

[redacted] –

Access - Case Volume

S:O - [redacted] helping enough?

Consumer Experience

[redacted]

(?) Account Sign up Bot deprecation

SIM - 144

9/15/2019 Sept

+2FA

added an email
after not having
one

(~~[illegible]~~ d.

May 12/21

Download your Twitter Data

2FA

12th floor 17 St side

[redacted]

Cognizant
Tech U,

Best retention rates for Agents

unpaid security
and privacy
bills

reaching in a crisis.

Doc of j
responsibilities

144 Doc

Contract

writing resp for x
reading resp. for i

Ads cookies are NOT essential cookies.
multi-use cookies pulled apart.
experiments for

A mock up
delete some cookies
Recaptch → Recaptch.net
(separate from GOOG)

SJM-50/52 } long term
SJM 90 } remediation

identified ~~not~~ but never done.

The pack of wolves that have been roaming in the woods are now @ our front door.

told ... are
3 years ago cashier problem
1 year Switcherdoodle agreed to but no backed up/ prioritized.

now here - 5 breaches under investigation

2018 · Irish DPC Inquiry report "finding of fact" - systemic failures in how we launch products.

Parag IMM Top of Mind

Splunk

~~Edin~~

Failed logins - 1.53K
per day - (550 as
low the other day)

MESOS

Production Access -

SFM - 144
Cookies

Deactivates an
account - after 30 dy
no longer publicly ~~f~~
available -

Tweets, Emails, phone #s

No DMs, IP logs,
Device level data
in logs

CNIL - is a beast on
data deletion.

- updated privacy policy 2-3 years ago
- some mitigations - daily, weekly, quarterly.

— [redacted]

DPC asks question
we have to fess up we
cannot meet our deletion
obligations

GDPR - us

FTC → vio

Historic LEO request
when we said we
couldn't

Inability to meet our
data deletion obligations

DC - week of Sept 22nd

Muller et al revealed
some of the errors -

HRBP -

Discovered in connection w/
Mueller investigation 2017
realized we had
Cold storage - actually had
data even though it's hard
to get to.

Re-opened subpoenas and
said we made mistake - was it
is (from DB; we didn't know we
had)

removing Default ; Future Budget

Latham -

- [redacted]

→ Updated Privacy Policy

* [redacted] said there's
not a long term strategy.
(Want to put it together
with me? or would you
like me to give you
specific targets and
end goals to build it
around?)

* Past 3 board / [redacted]
Committee - 2017 →
Data deletion / Retention
has been The TOP risk
highlighted!

Tracking effort not impact / effect

[redacted] Tu [redacted] Execution Review to staff - added column

"Are we more secure at completion" → No

[redacted] review Tracks → "all Green" ??? same issue

Twitter tracks effort on projects and impact to Environment

Scrubbing (6% of datasets - but [redacted] cto told big progress for 3 years and projects almost done)

SDLC - Board vs reality

Beth 7/19 - completeness / helpfulness ~~[illegible]~~ : 90-100%
8/15/19 ~~[illegible]~~ 20-25%

Upcoming Regulatory -

Messaging ([redacted] prep)

Data Deletion
SSM 144
Cookies
FTC

} why didn't Rich see any of this? Instead, reported AOK

Learning - changing cultural view of activities does NOT demoralize Tweeps - staff just thinks it will. Tweeps afraid of ch but embrace much of it gu (e.g. [redacted] - Privacy, Pause)

India and future Companies of Mar Facebook

Flash Privacy controls

MAP data for China

customers

FTC ~~compliance~~ consent

Unencrypted

(Encrypted links in our data centers

Election interference.

Data tagging

Access control

Insider Threat

Above all seem equally important

They aren't: if you take any given of them and do it in a vacuum — ignoring the others. That's what we've been doing.

How we tackle the problem presently ⇒ headcount

why?

unable to prioritize projects (no efforts get stopped, unable to predict duration to transfer headcount, etc.)

Idea

The danger of locally optimizing, tactically, w/o global goals & milestones:

all risks are equal,

all risks are "catastrophic"

w/o global prioritization must defend equally against everything.

example:

[illegible]
need,

③ Sect. 230

1 Transparency around content mod.

2 appeals (rebuttal)

3 algorithmic choice

Bob Zoellick –

[redacted] – "moving Trump removed ~~8~~72% of misinfo"

2 ~~digital acts moving forward in EU (?)~~

- Treatment of Foreign Leaders

SZ{PZ outline   write to •;

Intro

Incidents

Stats

you should be alarmed

(it doesn't need to be this bad)

- We're great in crisis — we drop the ball learning

Eg - SIM 89 learned in 1 place but didn't

replicating

repeating tactical lifts

Prod - huge over provisioning

- other crises waiting to happen

CDAB visibility

other areas ~~etc.~~ ~~Commitments~~

SMP/ATS that need to address

- Setting up Confidence

↳ where are these on the WBR

WBR is a switch to

company goals)

Groundhog day - Crisis

Missing the Basics

SIM Stats 2020

lacking ownership of systems, data, not following processes and not owning configurations

not taking lessons learned

Feels like Groundhog day

why?

not taking lessons learned

Ex SIM Stats

Ex offboarding

Ex SIM 89 for IG fleet restriction but not for rest of company

(broad access - takes work to figure out what they actually need/use)

Action: Threads v3

12 m

84k RCPO

.007

7 tenths of 1 percent

1.7% of our

spam bounces are

appealed.

~~.37% of~~

erlay)

wns

12 M suspension/sion -11M
215 k Appeals cases Ropo Bots
84k ropo appeals

70% of 215k appeals are ROPO

Access Appeals —
& focus

Smite & Botmaker
E-mail

No central live dashboard for Bots

May 4, 2021

Ptg

Conf Obj -

~~3~~ Short Term : Long Term
Doc

Alarms

Threads

ITEA

ToR - Own : Drive

MM : Mudge Cong : NSI - contractor
FTC - SIM 122

Extended Leadership Email

[redacted] or [redacted] - deliverable
Threat Model w/ options

{ June Cold Boot test
{ in PDX !!!! - How long
down (lowest
possible)

Confidence
ToR

~~3~~ T : S pr -> Confidence

Ownership is important -
but where is the unified
basic process enforcement?
Are unit tests optional

SDLC -

figuring SDLC
...

{ SIM-122 easy fix given
regulatory impact
for 10 w/ FTC negotiation pending

May 18, 2021

A1 P+S

B2 [redacted] : [redacted] feedback

A3 Alarms

A2 Actions

A4 Board Async Q's

B1 Interview

B3 [redacted] → [redacted]

Board Voice track

Slack

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

~~Lettre~~ strategy

Workday Request - [redacted]

Lakshmi's Doc

[redacted] - not way to let me know about

50% of spacer broadcasts had wrong language detected (in English queue)

15 mins new staff format

Confidence Objective

hampered by T&S

extract Nicole

IT - need to uplevel

[redacted] not executing

Privacy (FTC)

looking to be real

rock

hard pla

2 pillar leads

goal

backi

more

tra

(2)

Confidence - new org
just getting underway

currently [redacted], Privacy, CorpSec,
T&S, IT

privacy
safety, integrity, risk → systems/ services/ customers/ employees

goal: manage: anticipate people who may do harm to twitter or use twitter to do harm to others.

The more we can clear these risks out of the way the more confidence ~~twitter~~ we have that can execute our mission.

2 ~~still early~~ nascent org —

as such you'll see some changes

Today you'll hear about [redacted] and Privacy from our [redacted] and our CPO (Damien Kieran)

you'll also meet our newest member - [redacted] - our Distinguished Privacy Lead.

Later I'll be sitting in on Health as TwS is in Confidence

TwS is Safety @ Scale for our customers. (scaror)

3 In BCP and DR you are going to hear about improvements against accidental disruptions

~~anything accidental can be~~ anything that can be caused by accident can also be made to happen (intent)

~~Yea~~ for this section I'd like to draw your attention to ~~exceptional privi~~ systemic access control issues and

(4) the significance of our FTC obligations —

~~This is the first domino~~ —

~~quite likely~~

~~This isn't if - this when~~

~~we are~~

Twitter's been

~~we were~~ carrying a lot of

~~technical~~ debt ~~before we~~

security/privacy for a while

~~came on board~~

we will likely not get ahead of this before another incident.

5 we need to make significant progress to demonstrate we are taking this seriously when the other shoe drops ~~This will slow the company down~~ Paying off this long over-due debt will slow the company down. We will work to minimize this impact.

Access control –

Flash – install during presentation.

Robert Zoellich

Clean production is 3x as fast

Patrick Pichette

readiness –

- India Extricating Manish

RAW - intercept extradition if informed via e-mail interc

Put them into a quandry as to whether they should act upon intel.

- Vol, Del, TwS, Confidence

- ~~Oscar~~ Health Tools —
  [redacted] ITCA

- Turn ~~Any~~ no actions in verified acch w/o human

2 items —

escalate quickly

infoSec on interview panels

- Madge - [redacted] DACIN (3)

. Commerce - Export Controlled ↗

a) compliance doesn't know what controlled technologies are, specifically and where they may live (Git, Slack, GSuite, etc.

b) ~~would~~ need to figure out access control solutions

Russ - (confidential, compliance, legal)

find perimeters ~~for~~ ads —
scrub there.

logs - lacking discipline
global issue

Irish DPC is under huge
pressure to do stuff

Banner is a gimme.
came up with enough information
to say it's not wrong.
Detangle cookies

Irish DPC - web site.

sent us a letter -

www.twitter. hey what are you doing w/ all of your cookies.

want us to tell them

cookies, trackers, local storage.

what are we doing w/ them -

Need to correspond to reasonable cookie banner-

we don't know what we are doing w/ them.

same cookies for ads targeting vs login abuse

-- visitor ID = abuse = ..?.
= ads

June 14, 2021

P+J

360°: on Me - How?

Vaccination card.

Addvertising cookies: always
cookies same one w/
for each.

weekly 1 on 1 w/
Parag

Biggest thing is
for you and I to
align

if we disagree on
something there's an
issue —

if we agree then
it's just a matter of

Parag - not aligned
↳ need to align w/ Parag

go out of way to align
w/ Parag

get no-bullshit
"real talk" feedback

"real talk" — good code word
thing for for Parag.

Pose as question

I would want the CTO
to know...

[redacted] is their leverage.

~~Court~~

unlikely [redacted] will be allowed to leave the country — he is their leverage point,

3 weeks

or

arrested

narrative — 1 way ticket

2 journalists — both got anticipatory bail

AJMA -31

fu/o

↑

Spaces 6, 7, 18

[redacted] Summons –

Dehli police summons – empty officer

[redacted] will not visit police station w/o having protection against arrest.

Police – all pillars – isolated/ siloed

MEIT – Ministry of Econ. & IT

no-return ticket – (1 way) family

Turn up heat

legal

add supported platform

add cookies are essential

buying time

(maybe come out a bit on our side?)

Immediate to long term Bubble

w/ DPC -

low hanging fruit that is

! OK and argument will not

cover.

Q: what is example of

low hanging fruit :

① Analytics

some of our analytics

cookies are used for a

variety of things but

various teams can't coordinate

their uses - so we can't

describe our analytics cookies

we have to be able to define

Max Shrems

privacy 'advocate'

Champions and takes companies to court over privacy.

Shrems 1 and Shrems 2

Max has sent complaints to Twitter/ these will converge w/ Irish DPC

---

2 low hanging

Google analytics (cookies)

↳ we give Google the ability to use the data we get as well.

Google refuses different terms

(3) We have cookies we don't know what they are used for.

(4) Cookies that have dual use - e.g. Analytics & Ads have to disentangle

examples of 6 systemic failures in twitter's prod. launch process - 5 breaches reported in 5 mos.

Fines of up to 4% GDP but more likely competitive disadvantage

Irish-DPC

attention will attract attention I

(A) Cookies is GDPR exception
(other 26)
27 other DPA, could take
a crack at us for Cookies
expense by paper cuts (expensive)

(B) Irish-DPC ~~could~~
have many other open investigations
Could get pissed - lose good
will on other investigations
More aggressive on other
investigations - 6 open GDPR
some serious

people who authorized
the roles? which roles

ssh-prod
ssh-next

testing

Everything in production
has no authorization -
[DB's, Services, compute
have no authorization

Debug interfaces

Web interfaces

— Manhattan Debug web page

Clients can make requests directly into prod

- there's no auth

- Dev -

run copy of Twitter on desktop & this copy makes calls directly into prod (w/o creds)

EVT 1) Jack

State of Confidence

2)

[redacted]

Insider Risk - India

Leaks

Privacy

SIM-144

Cookies

Data Deletion

(Regulators)

~~Follow up w/ P~~

clearances

Indonesia –

[redacted] – forced hire

– forced hire

Harminder – [redacted]

↓ – Public Policy

Failed even the simple background check

Built out plan in

2 Markets w/ lower risk

risk ↑ India
Nigeria

Revenue - China

16 page pdf - JTBD

Sept. 8, 2021

P+J

To Do List

Laptop overheating

IT Leadership update on VP of IT

Tobias - Fun DataSets

Eng Credentials

[redacted] India Docs to Susan;

1 page - Confidence version
to Board

[redacted]

[redacted] Udmio,

→ Threads

ITGA → Health Tools

— VP-IT —

DynamoDB, S3, EC2 level

Azure

Spencer - ? fewer than 1k
datasets,
870 on DynamoDB
AWS - 5-10k datasets

DwBirds - Plan coming OOB
Why Can't Privacy Do all of This?
Would have to Guess @ Data
Sets.
Tell us what data is where ...
Launch Review. DataSets that
support Deletion

Twrdukum NAS - 1.1→1.2m

Sept. 21, 2021

P+S

Conf. Obj. to Staff

Risk Center Voice Intro

every night I go to
sleep worrying if India
already with ~~our~~ our internal information
to ~~ident~~ identify target and kill people expertise specific
to this govt.

→ (1) as long as we are within
India's borders
not able to mitigate

4b · ~~we don't have the technical ability (staff) to mitigate this situation~~

4a started this · ~~the path to mitigation is to serve India from outside India's borders.~~
(Geo political Boundary) issue

Provide additional perspective on risk management on a country

perspective on the risk analysis portion of our presence in India.

|| from [redacted] doc.

[illegible]

1 · We have an active insider

2 · India will continue to attempt to influence traffic

3 · They will attempt to gain access ~~[illegible]~~ int

Turkey - in DAG

Business Risk today

Take

unknowns - dynamo DB
(spaces)

HDFS/DBs

SIM-154

Deny list for targeting of certain keywords (illegal)

Went in to see what we're actually doing - becoming apparent we don't know how the system works.

Some around
Lower
- Japan
- prefer
US.

Parag - Sacharla still on the table?
place 1 in spain and 3rd in US

Jan 13, 2022

P+S

[redacted]

We have AI Models built on sets we did not have rights to the training data we used.

Models are going to be a problem.

Jan 18, 2022

A1 P+S

→ A1 Empire CCS

End of Day Itinerary, timing to SZ

Staff Doc

Risk Cmte Doc

✓ A2 Alarms

✓ A3 Reply to Marianne - Audit

~~[illegible]~~ CorpSec Principles ;

Fugitive doch - track

CEO Security follow up

[redacted] :

Staff Doc

Audit Committee Doc

✓ Exec Coach

Came in as Change agent

trying to move the ball forward

understand how culture reacts to this information -

How do I take what I have and what I know and help the ~~agent~~ org/staff receive this and become an effective gent of change.

not about me assimilating here

[redacted] - Mudge

log4J - visibility?

how to close out?

[redacted]

started prepping prior to XMAS Break

Ukraine

- effort to get in front and be pre-positioned

#Protect

log4J - free of current compromise

~~at~~ @ highest risk addressed

"we think" - how what

~~EE~~

Unblock [redacted] so I can own log4J and resolving there.

Parag has ~~influenced~~ encouraged transparency

~~Decision to offboard~~ [redacted] ~~in October~~.

~~Performance.~~

(A) Parag -> transparency
(few moving parts)

In October (Nov) decided to start the offboarding process for [redacted] - performance.

~~Dec. Risk Comte we spent~~

~~Though a number of HR/ER issues she is still here~~

It was not intended that [redacted] would be here for the board or Risk Committee.

Jan 18, 24hr notice to meet w/ Omid (Risk Chair)

Access Control
incident root causes

~~We were working towards~~ offboarding [redacted]

~~Th de~~

Ⓐ cont. I put together the Board section of the [redacted] update to ensure we were correctly setting understandings of top risk areas:

- Total Incidents
- Security Patches and SW Versions
- Access Control
- Process Compliance

We spent the majority of our Privacy/Infosec time on Privacy.

Risk Committee had

~~I brought~~ A

~~I at~~

A1 I triggered an audit investigation on what was communicated to the risk committee.

Bkground:

[redacted] - decided to offboard early Nov, for various HR/ER concerns

offboarding dates repeatedly blocked

It was not intended [redacted] would present to Risk - ~~but within the prior week~~ but ultimately ~~S~~ short notice [redacted] had to be put forward.

A2 I have concerns over the what was presented at the Risk Committee in the [redacted]

~~One Area~~ [redacted]

Privacy / [redacted] Section

(Specifically [redacted]) - ~~I brought~~

I documented these concerns (hence we are here)

~~the~~

~~at the~~

The concern is around how items in the ~~the~~ infosec document (a slide deck) could be interpreted - they lacked a narrative and context that the privacy doc had. And an Item Ms [redacted]

The [redacted] / Privacy

(Focused on Privacy)

[redacted] verbally presented one item:

An improvement and progress in access control.

Also stated a plan we are executing

~~#~~3 Need ~~Want~~ to call out 2 items

1) are thing [redacted] verbally presented, (which was basically the only thing presented as that section was Privacy focused. Access control

2) Address the Doch that was sent for pre-read.

Unlike Privacy document, lacked a narrative.

w/o context and descriptions, I believe parts can be misleading or incorrect

4 I ~~will be~~ am putting together a document to capture and correct ~~gap~~ confusion or misunderstanding the [redacted] section and deck ~~might~~ is at risk of causing

Verbal

Access Control –

"small" project represented as a win – a group of 300 users with production access ~~written~~ recently reduced, ~~temporarily~~, to 100 –

The larger context is

#/context

from 2020-2021 prod access grew from 2.7k to 4k – faster than our Employee # grew.

~~Mentioned there was a plan that is being executed.~~

~~The small win~~ A similar

3441
5421
8862

- Info conveyed not accurate
  represent
- I identified and documented
  (hence we are here)
- Significant efforts to avoid
  sending confusing or misleading
  info
- As part of that [redacted] was
  to be offboarded Oct/Nov
  performance.
  (not intended to present)

I am putting together
concise document to
address and correct.

①

- Simple history
  1st 3 Qtrs supporting [redacted] and providing room to build and execute
  Q4 Review → offboard (HR green light?)
- Morgan: "I see a manager who gave their report room and support."
- Work to ensure incorrect info ! go to Board & Risk
  ~~multiple offboarding, key~~
- Key targets always Board & Risk to prevent this

②

- [redacted] has used Retaliatory and ~~de~~ tactics, ~~force~~ offboarding apparent, to repeatedly Block
- Escalated repeatedly until ultimately to Parag and Dalana (Board & Risk becoming imminent)
- Parag Promises to personally look into and resolve by ~~each~~ beginning of week
- Calls Week of Risk - not going to keep promise - apologizes.
- I ~~suggest I prepare~~ bring up replacement doc. Parag instructs the doc, and [redacted] presenting to board and that I attempt to mitigate

· offered to personally follow up w/ Risk

Predictions / Issues
raised - fulfilled

10) Dual outage - BlackSwan
6) Nigeria - lengthy
8) Pile on of Regulatory
India - Govt - Insider -
6) Lack of visibility (since first report)
- Log45, and inability to ensure problems fixed
8) Access Controls - enabled BlackSwan, contractors and pols w/ Expanding internationally
Privacy tie in

③ • I express discomfort with this and inappropriateness.
further document ~~in~~ e-mail to Parag & Dalana w/ Details

• After Aish I ~~followed~~ took you up ~~requesting~~ ~~taking you~~ up on your offer to follow up for good measure. - You expressed disappointment to my handling of a situation you prevented me from avoiding.

• You suggested we meet up over the break to work on these issues - I eagerly agreed.
- You had your first but cancelled our meetu (w/o informing me)

Oct 29, 2021

Pty

[redacted] 360

#Conf Dash and Voices track

Quarterly reviews

[redacted]

VPN - Spreadsheet to Staff

Reach : [redacted] Graph 6

5 Sierra Staff

We & all other companies they should have a presence to engage customers

31 as employees vs contractors.

@Twitter Support

Go slightly bigger and measure

6 people responding to 2%

Because responding on platform to customers (1 to many) not at scale contractors (employees)

6.8m

measured UAM lift 28%

net neutral cost w/ measurement

this

Nov 1, 2021

A1 P+S

A2 Alarms

B1 Stocks - personal

→ ESPP & 401k

A6 Interviews Read CV's and choose items for questions

A5 [redacted] Doc.

→ Purchase Plan

A4 Jack notes

A3 Staff presentation

183 (900)

117 (800).

Teams big on owning service not problem

206 PØs last year

Disney - 3-5 PØ's

Notice Chaz Harnahan's team

you've got this!

Not Being willing to say 'no' - in over her head.

Analysis Paralysis

Senior members are buried under 17 or 18 things - not prioritized - only person on project.

TWS 25 HC

6 people on @TwitterSupport (new thing)

responding to 1-2% of addressable public messages

Experiments show increase in WtM customers engaged (2.1%w)

$1.9 incremental revenue per month

slight increase in revenue vs cost

full coverage 14.6M (cost)

revenue + 15.1M (.15M)

max 2.6M for 100% scenario

given size of investment

doing 3M larger test

Nov 9, 2021

- A1 P+S
- A3 Set Dates for [redacted] Offers
- A2 Dry Run #Con
- A5 Functional Strategies Review
- A4 Staff Doc Async

within 30 (c) months of coming on board we will see material progress on these standards

Run of Show

Put [redacted] on point —

ER completion dates

Board, Risk

my role is the integrity of this org

make sure the ball is not dropped

Job 1 is the technical foundation of [redacted] - the strong fundamental

identifying the core areas of highest exposure and highest likelihood of happening —

— The standards by which the confidence organization lives by and is evalu

clients, servers, data, secrets, ~~[illegible]~~ incidents - closing (understanding)

Nov 19, 2021

P+S

Alarms

10b51

Deposit Cheques (MS)

Threads responses (devbirds)

#Protect

Tony Hawkins : Stage 2

Privacy

Functional Strategies (Review

Google Workspace Contract

Regulatory : Compliance mgt context

MOU to Dev Birds (for the record) requesting Privacy be the top priority in 2022

RCB

Kim

[redacted]

Work

optim

Pu

fa

12:11 Nov 23 Jack tells staff time for him to step Down

Parag as CEO

Bret becoming chair (Board)

Patrick leaving

- Pichette

Jack on the Board

Nov 30, 2021

A1 P+S

B1 Remove reference to [redacted] Katrina in QS Doc.

C1 Twitter holiday cards info

C2 [redacted] - Invoice

[redacted]

[redacted]

[redacted]

[redacted]

Attachment to mission

& { Confidence - Org Driven
{ #Protect Objective

my mission

see world though eyes of security

Attachment to Mission

Driven

Jacob call - Jack call

Pat [redacted]

my mission
who I surround myself with

NYC Site Lead
staff support for TwitterFolk

Hygiene - End Point
Access Control

IAM

5K

133

44

3077

9K 9000
- 5940 33 1/3

9000 3060 Endpoints
Systems do not have
Software updates
enabled

99% of fleet has security reporting software

only 50% of laptops are
security compliant

Dec 2 2021

Q3 2020 → ~~13~~ 8, 5

Q4 2020 → 4, 3

Q1 2021 → 7, 1

Q2 2021 → 16, 7

Q3 2021 → 13, 6

Q4 2021 → 4, Ø

8.8, 3.4

Q3 2020 - Q4 2021

9.6, 4.4

incidents

reported to regulators

Q2 2021
16 7
1~~[illegible]~~

Dec 6, 2022

PLS

Staff Doc

Board Voice track

[redacted] Concur

Kudo Board

Access Control

Compliance - Sec Config

Incident tracking : root cause

Data Deletion

. . . .

1100 offboarding,

2020 : 2021

345 751

6506 - 2021 / year

~2K reduction = 38%

versus

100 reduced in ~~345~~

(whatever) 2 small high risk pockets -

High risk pockets that need 90% +

March 16, 2021

- ✓ A1 P+S
- ✓ A3 Alarms
- ✓ A2 Stats
- • A4 Strategy for Risk Community
- A5 Stats for Risk - Quorious & IT
- B3 UPL follow up (2nd pass)
- B4 Confidence Board
- B5 Tickette assignment
- B2 Bot Inventory
- B1 ITCA / Health

115 PB of data in HDFS
3x increase (36PB) 2015

32PB not registered (27%)

41% of Datasets active but unregistered

Total DataSets 39,303 ?

!

20% (7k) not accessed in 4 months

Nov 31, 2020

✓ A1 PTJ

• A2 Verify Health elections

✓ A3 [redacted] refill,

✓ A4 Stocks

A5 Mindfulness

C3 get Learning

B4 60 Day doc.

C1 Advisory Board

B5 CoS JD's

B1 Eng & Sec Roadmap Docs

A6 Set Alarms

C2 CFTL Checks to bank

B3 eng@ Survey (Google Form

B2 Board Docs & Responses

Advisory Board: Jack, Ned, [redacted]
[redacted], Jennifer

[redacted]

└ specific goal / Fire Jack

Silverlight — Igan goal is keep an eye on Jesse - Igan support Jack.

1 year stand off period
end of January 2021
Shareholder Vote

---

why did we miss MAP

why didn't we capitalize on COVID.

Dec 7, 2020

✓ A1 P+S

✓ A2 Harms

✓ A3 Mindfulness

B4 go/learning

B5 60 Day Doc

B6 Advisory Board

B7 CoS JD's

B3 Eng & Sec Roadmap Docs

B2 Eng @ survey Form Doc

B1 Board Prep

· A4 E-mail

- B8 Contracts - Althea / Kiagan
(see notes)

· A5 Eyeglass metrics for Board

Althea - Kiagan, Caldwell, Oscar, Shanely
~~2k~~

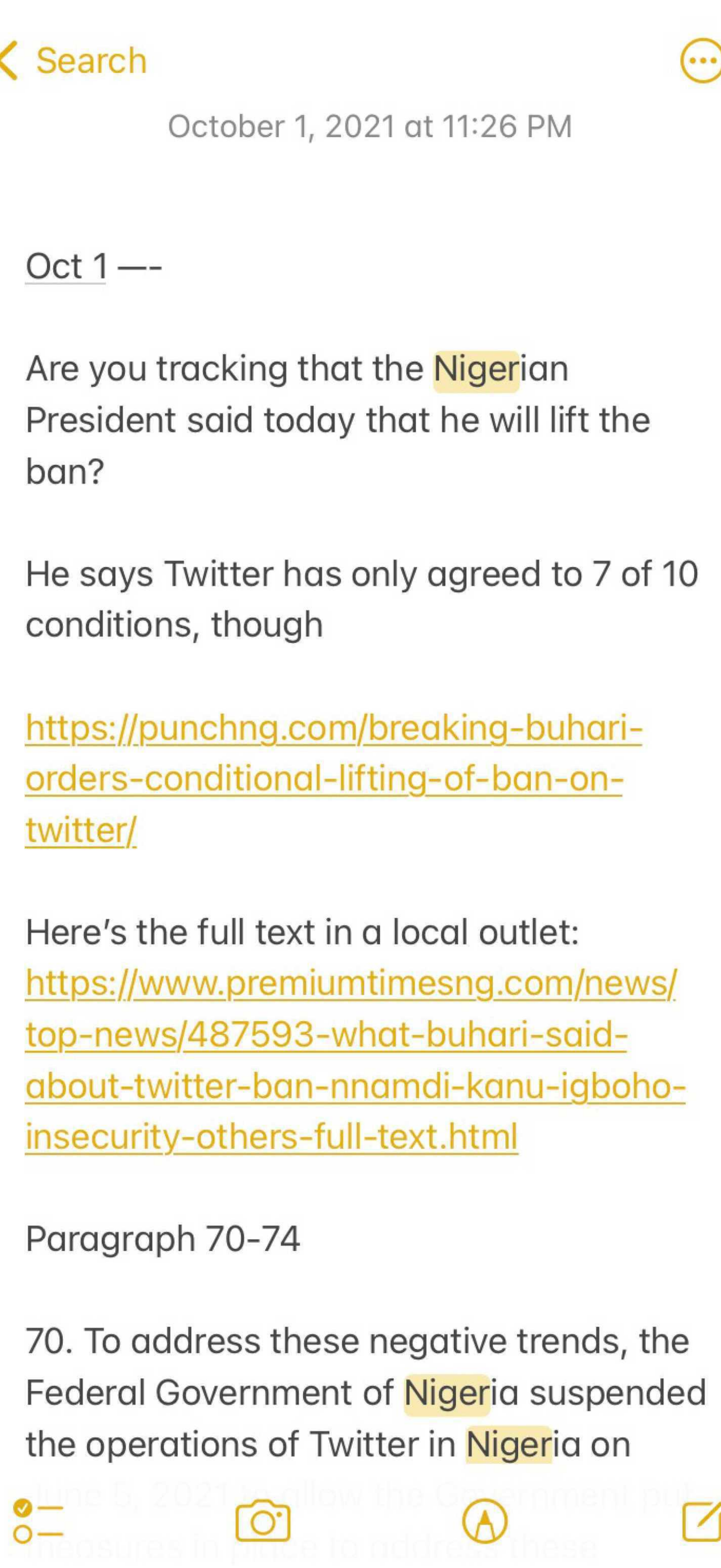
Advisory Board: Jack, Ned, Sean, Vijaya, Jennifer

Jack Notes

[redacted] isn't entirely wrong

everything can be traced to lack of metrics and quantified goals @ staff level

- Example: Crisis
- haven't seen any docs (outside revenue) with measurable and goals or milestones w data values.
- ~~lots~~ global prioritization not matching local workflows. (?)



October 1, 2021 at 11:26 PM

Oct 1 —-

Are you tracking that the Nigerian President said today that he will lift the ban?

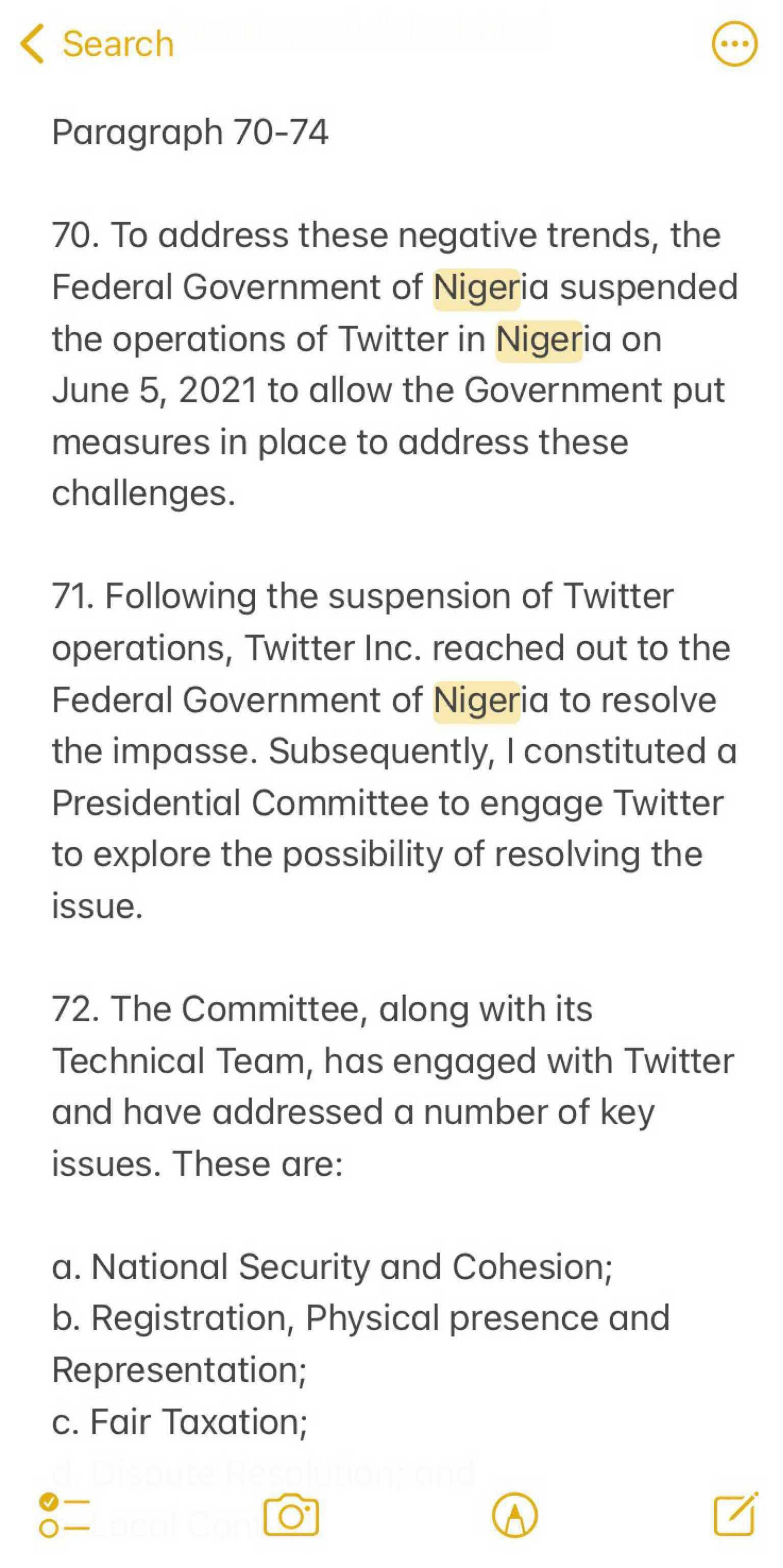
He says Twitter has only agreed to 7 of 10 conditions, though

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



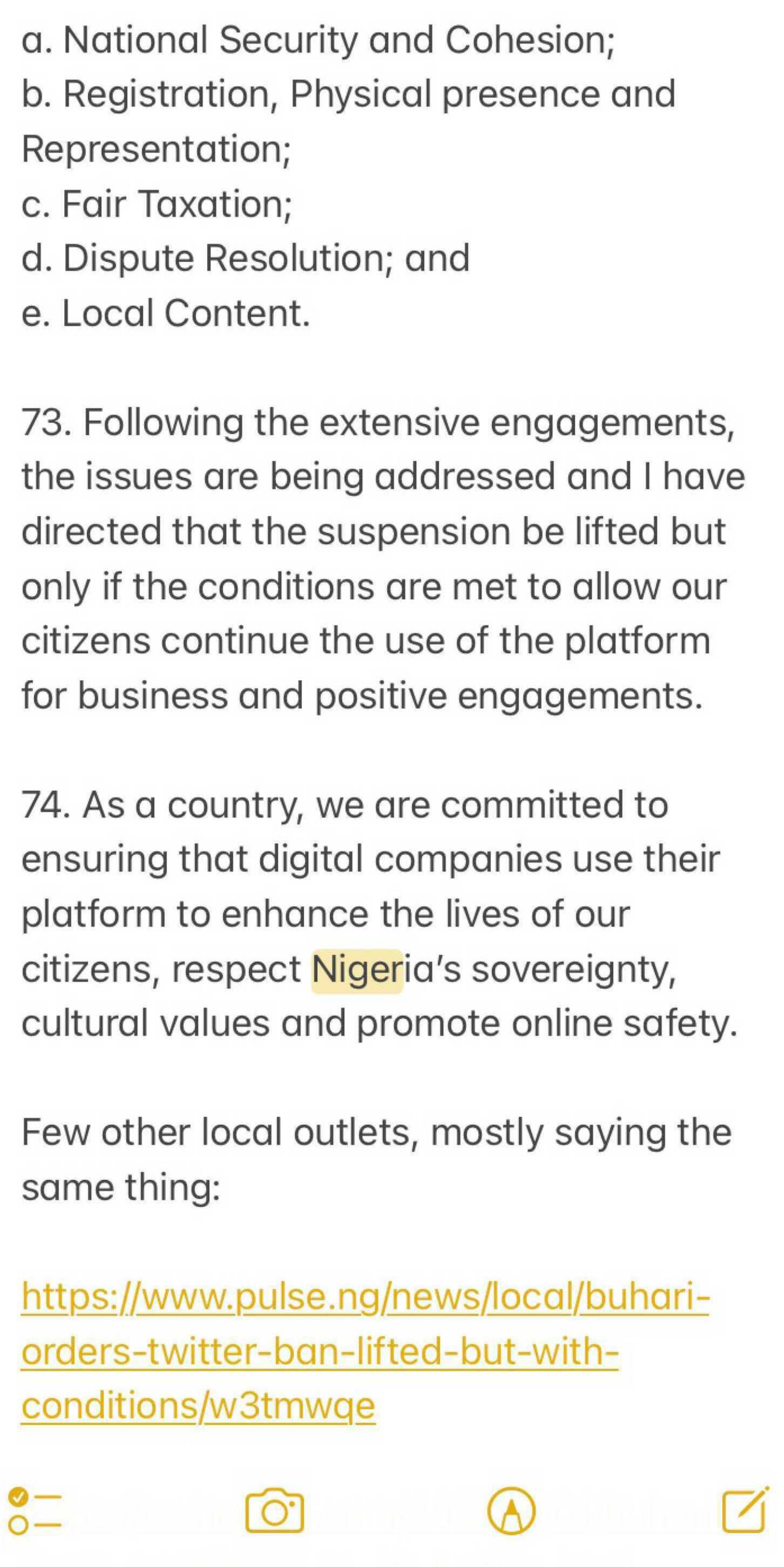
Paragraph 70-74

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 Con



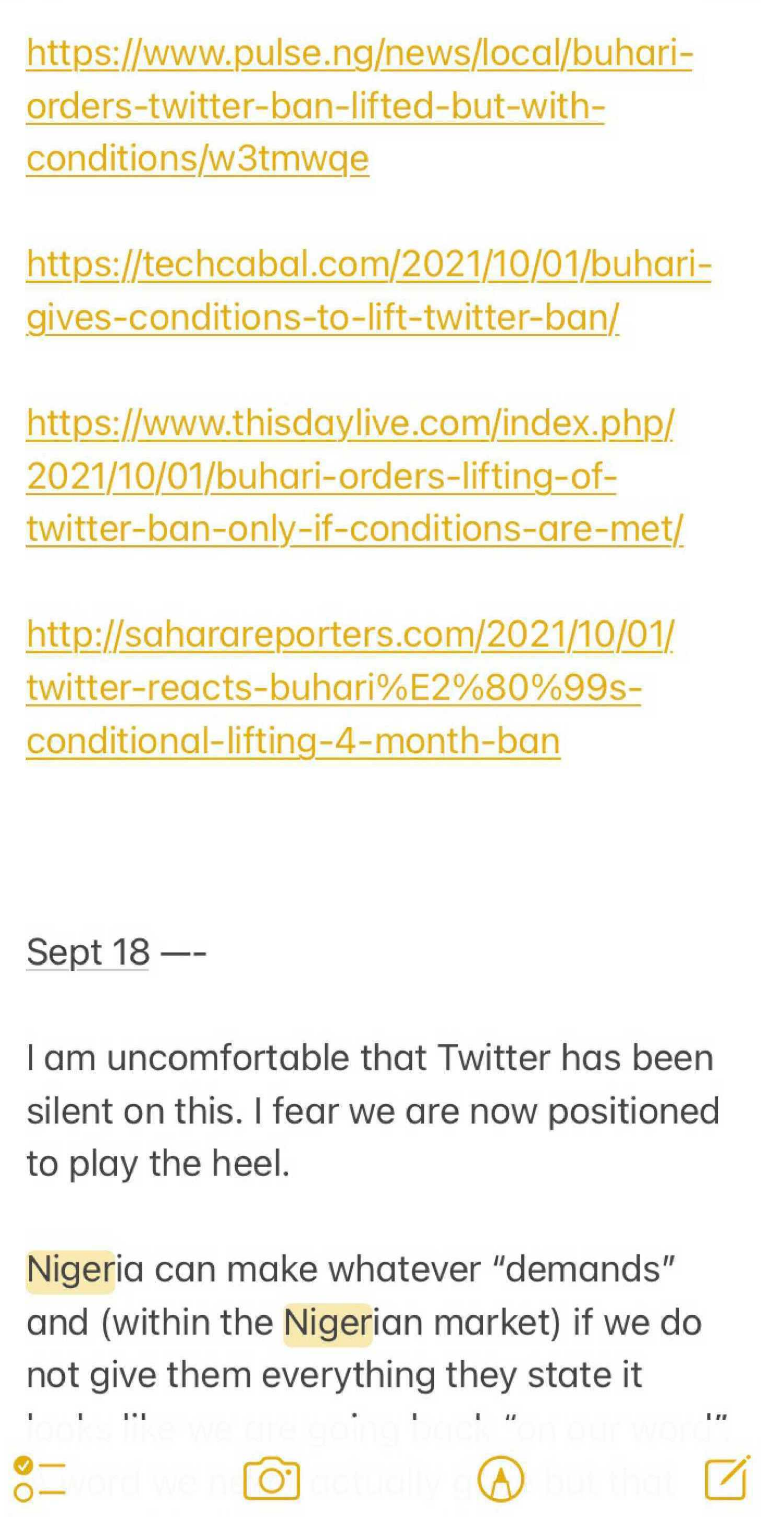
a. National Security and Cohesion;
b. Registration, Physical presence and Representation;
c. Fair Taxation;
d. Dispute Resolution; and
e. Local Content.

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://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/

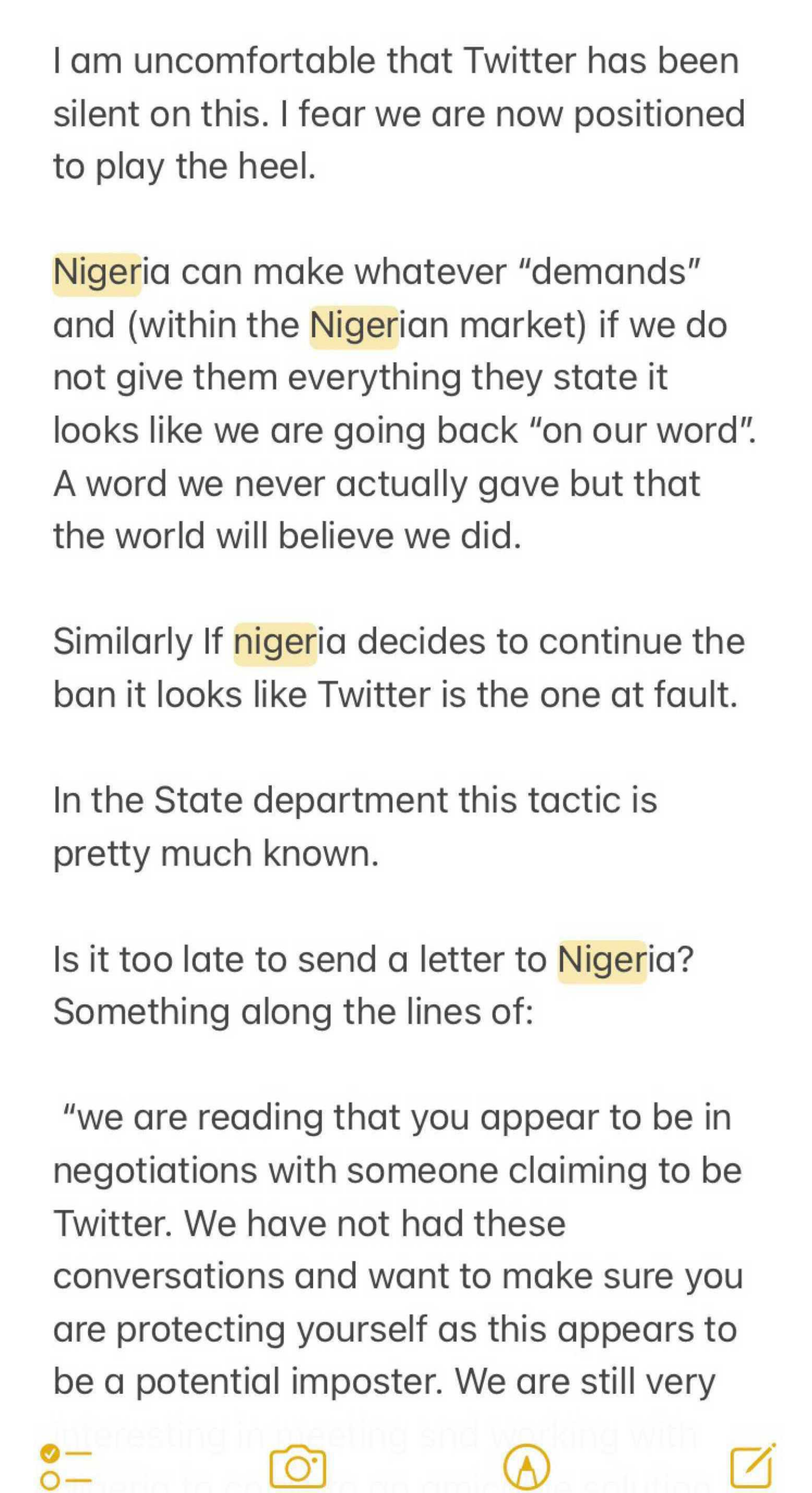
https://www.thisdaylive.com/index.php/2021/10/01/buhari-orders-lifting-of-twitter-ban-only-if-conditions-are-met/

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". word we ne[illegible] actually g[illegible] but that



I am uncomfortable that Twitter has been silent on this. I fear we are now positioned to play the heel.

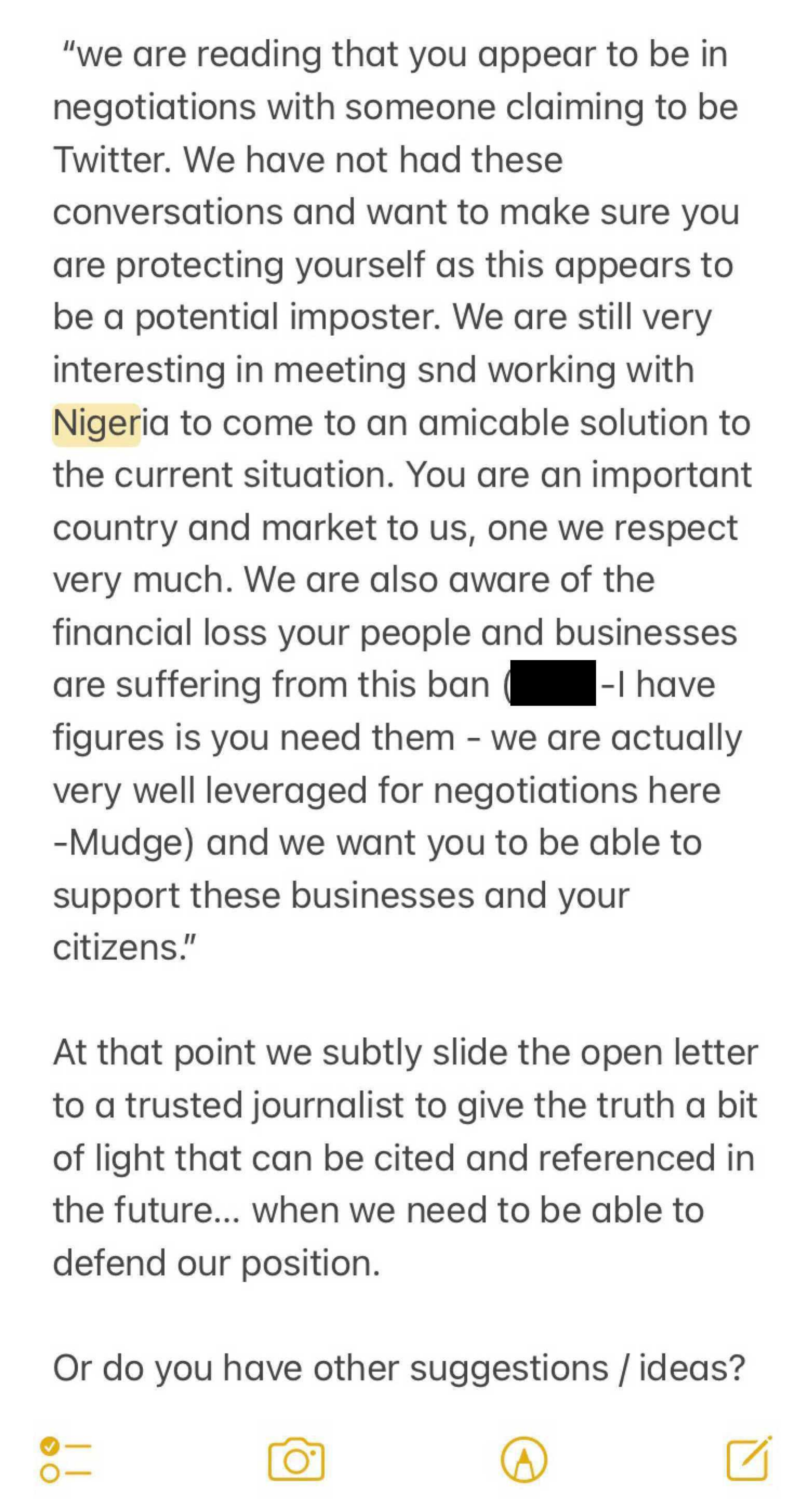
Nigeria can make whatever “demands” and (within the Nigerian market) if we do not give them everything they state it looks like we are going back “on our word”. A word we never actually gave but that the world will believe we did.

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:

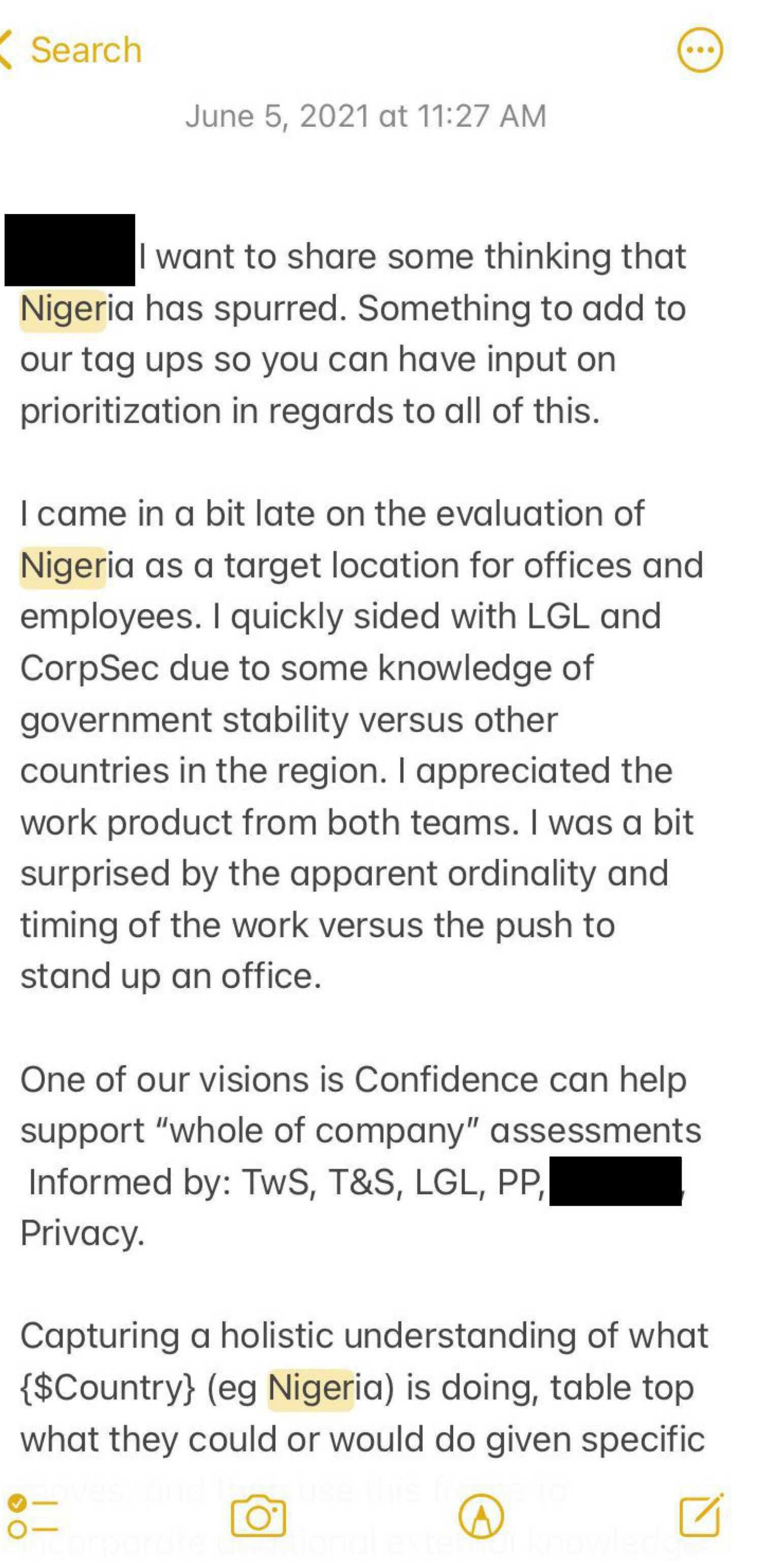
“we are reading that you appear to be in negotiations with someone claiming to be Twitter. We have not had these conversations and want to make sure you are protecting yourself as this appears to be a potential imposter. We are still very interesting in meeting and working with



“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 interesting in meeting snd 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 ([redacted] -I have figures is 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?



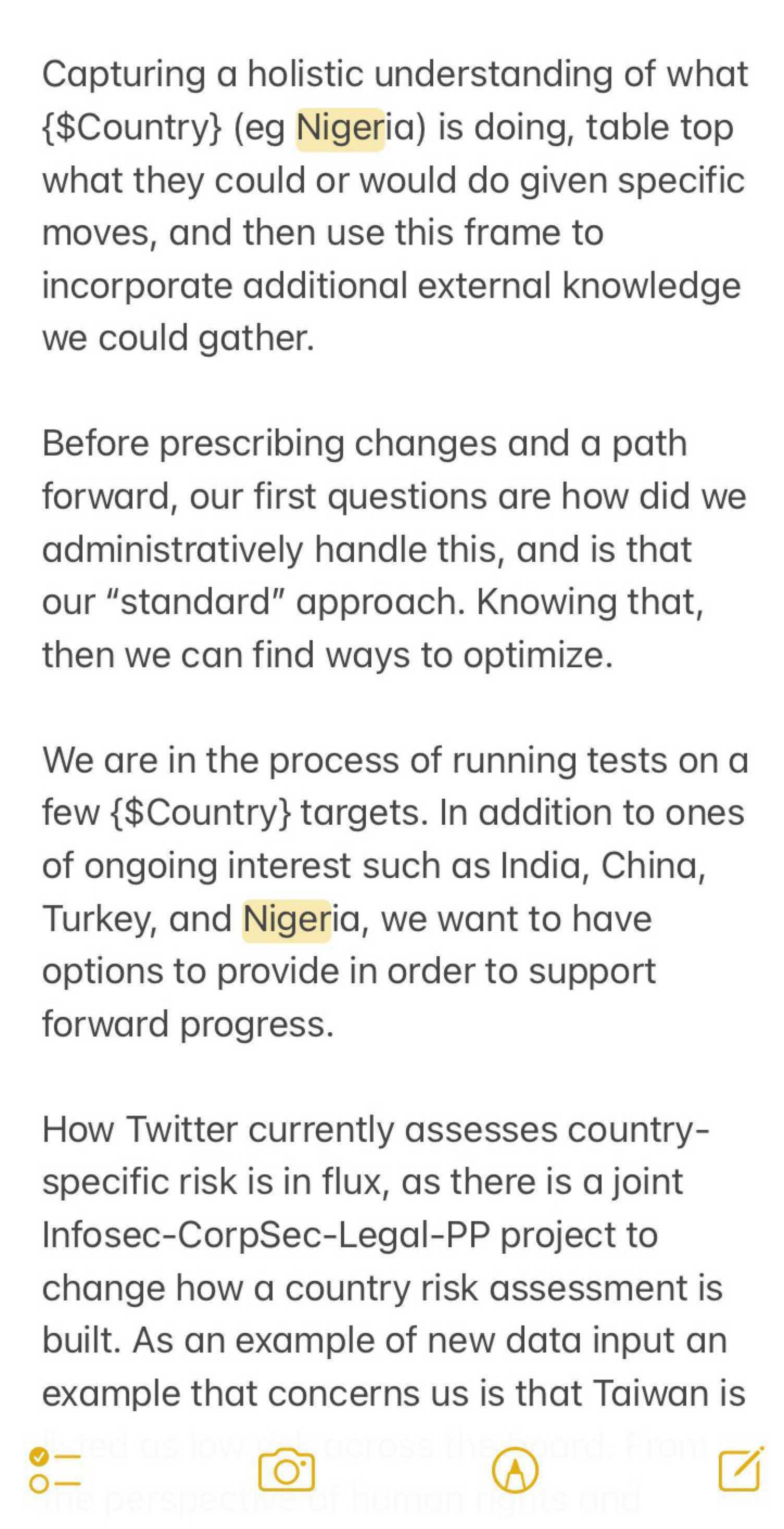
June 5, 2021 at 11:27 AM

[redacted] I want to share some thinking that Nigeria has spurred. Something to add to our tag ups so you can have input on prioritization in regards to all of this.

I came in a bit late on the evaluation of Nigeria as a target location for offices and employees. I quickly sided with LGL and CorpSec due to some knowledge of government stability versus other countries in the region. I appreciated the work product from both teams. I was a bit surprised by the apparent ordinality and timing of the work versus the push to stand up an office.

One of our visions is Confidence can help support "whole of company" assessments
Informed by: TwS, T&S, LGL, PP, [redacted], Privacy.

Capturing a holistic understanding of what {$Country} (eg Nigeria) is doing, table top what they could or would do given specific moves, and then use this frame to incorporate additional external knowledge

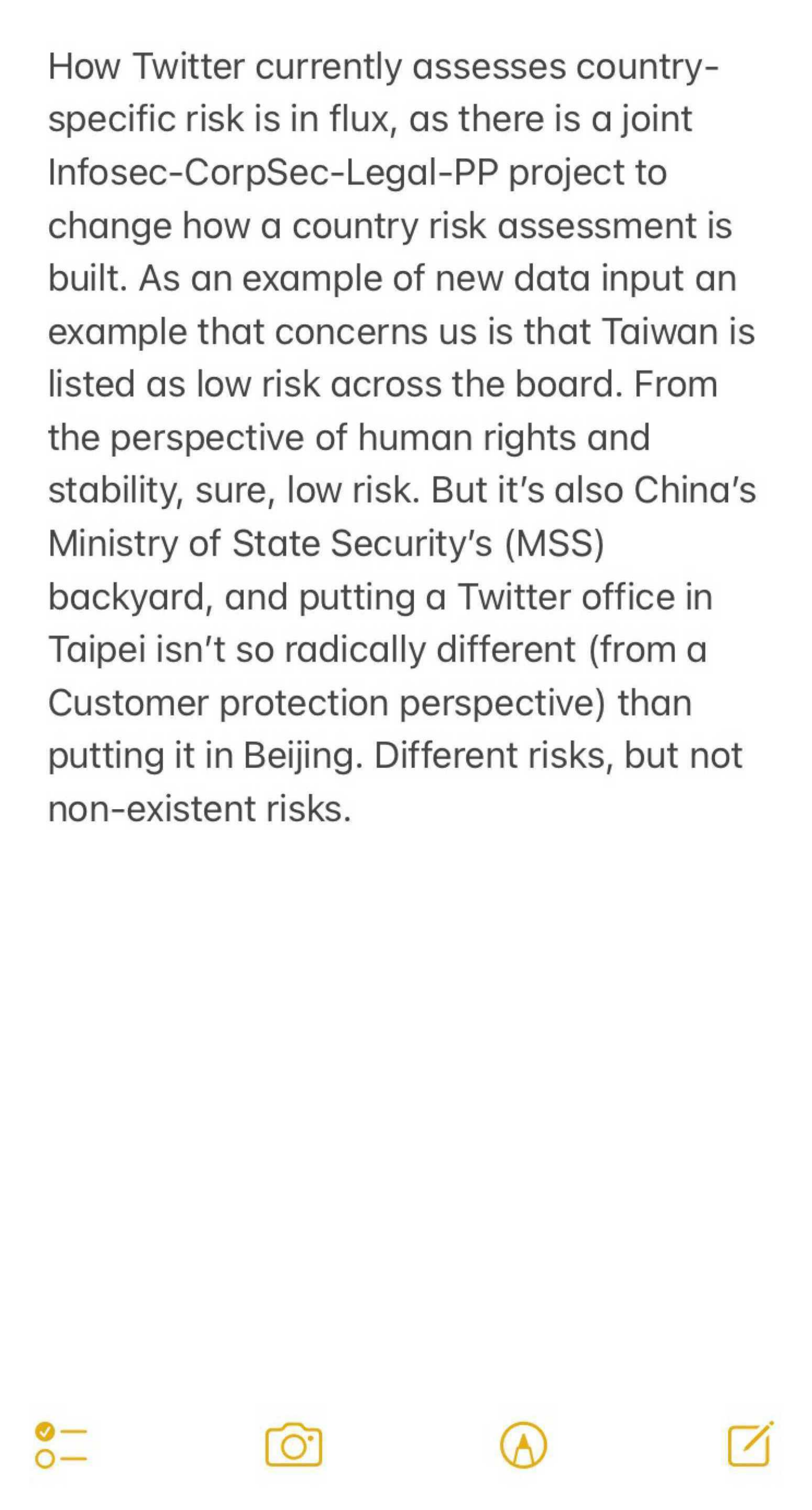


Capturing a holistic understanding of what {$Country} (eg Nigeria) is doing, table top what they could or would do given specific moves, and then use this frame to incorporate additional external knowledge we could gather.

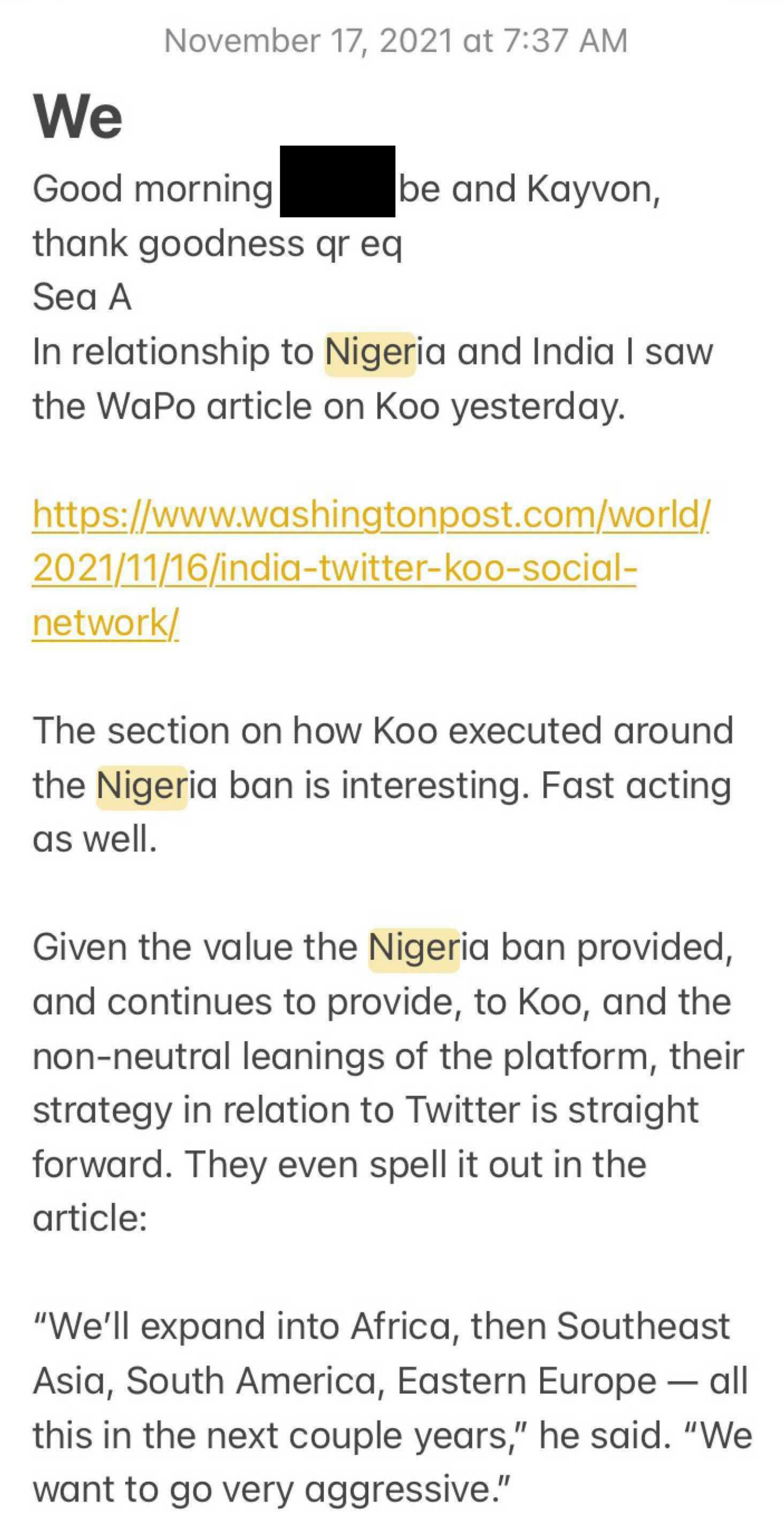
Before prescribing changes and a path forward, our first questions are how did we administratively handle this, and is that our "standard" approach. Knowing that, then we can find ways to optimize.

We are in the process of running tests on a few {$Country} targets. In addition to ones of ongoing interest such as India, China, Turkey, and Nigeria, we want to have options to provide in order to support forward progress.

How Twitter currently assesses country-specific risk is in flux, as there is a joint Infosec-CorpSec-Legal-PP project to change how a country risk assessment is built. As an example of new data input an example that concerns us is that Taiwan is listed as low risk across the board. From the perspective of human rights and



How Twitter currently assesses country-specific risk is in flux, as there is a joint Infosec-CorpSec-Legal-PP project to change how a country risk assessment is built. As an example of new data input an example that concerns us is that Taiwan is listed as low risk across the board. From the perspective of human rights and stability, sure, low risk. But it’s also China’s Ministry of State Security’s (MSS) backyard, and putting a Twitter office in Taipei isn’t so radically different (from a Customer protection perspective) than putting it in Beijing. Different risks, but not non-existent risks.



November 17, 2021 at 7:37 AM

We

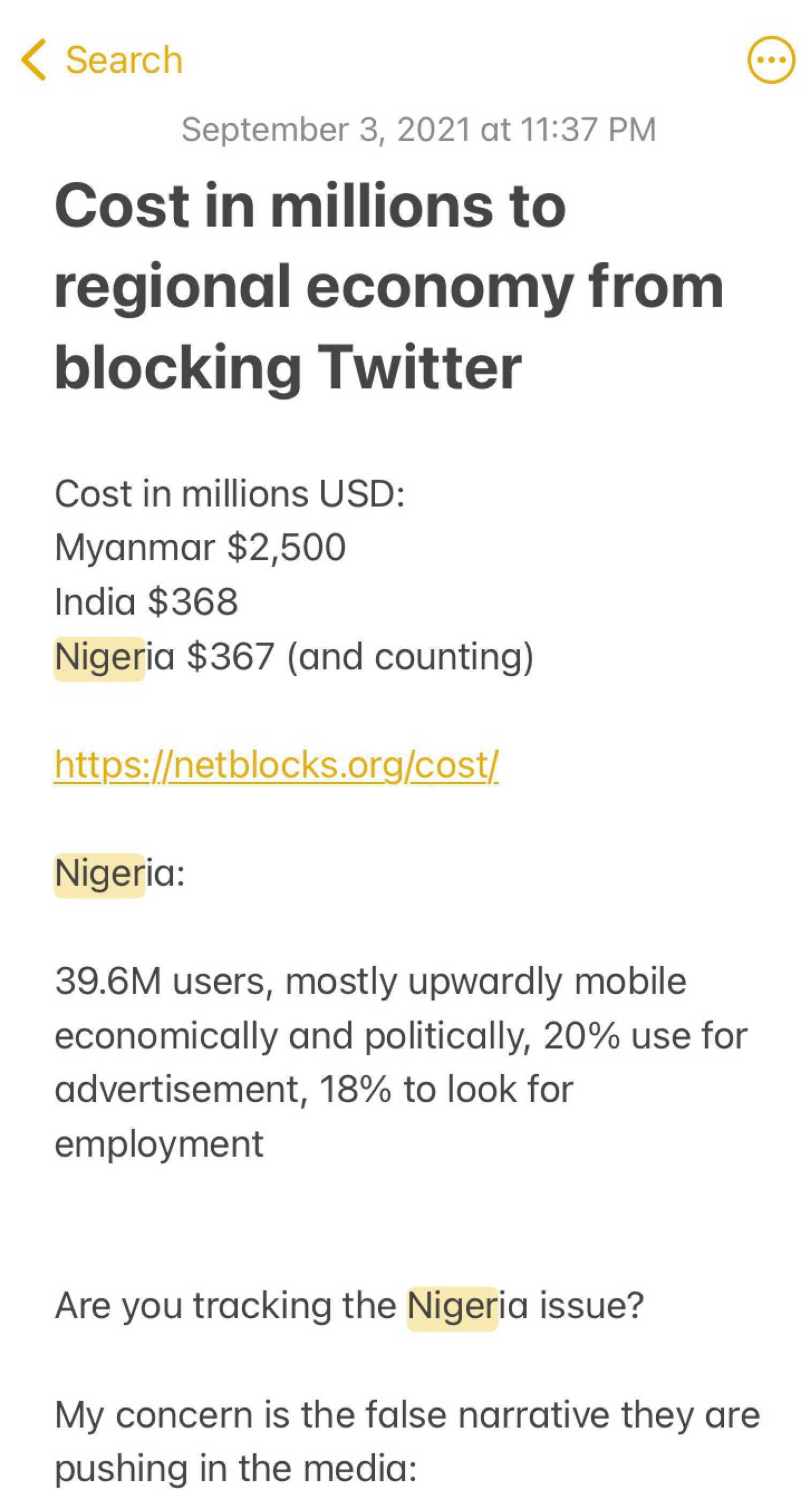
Good morning [redacted] be and Kayvon,
thank goodness qr eq
Sea A
In relationship to Nigeria and India I saw the WaPo article on Koo yesterday.

https://www.washingtonpost.com/world/2021/11/16/india-twitter-koo-social-network/

The section on how Koo executed around the Nigeria ban is interesting. Fast acting as well.

Given the value the Nigeria ban provided, and continues to provide, to Koo, and the non-neutral leanings of the platform, their strategy in relation to Twitter is straight forward. They even spell it out in the article:

“We’ll expand into Africa, then Southeast Asia, South America, Eastern Europe — all this in the next couple years,” he said. “We want to go very aggressive.”



September 3, 2021 at 11:37 PM

Cost in millions to regional economy from blocking Twitter

Cost in millions USD:
Myanmar $2,500
India $368
Nigeria $367 (and counting)

https://netblocks.org/cost/

Nigeria:

39.6M users, mostly upwardly mobile economically and politically, 20% use for advertisement, 18% to look for employment

Are you tracking the Nigeria issue?

My concern is the false narrative they are pushing in the media:

My concern is the false narrative they are pushing in the media:

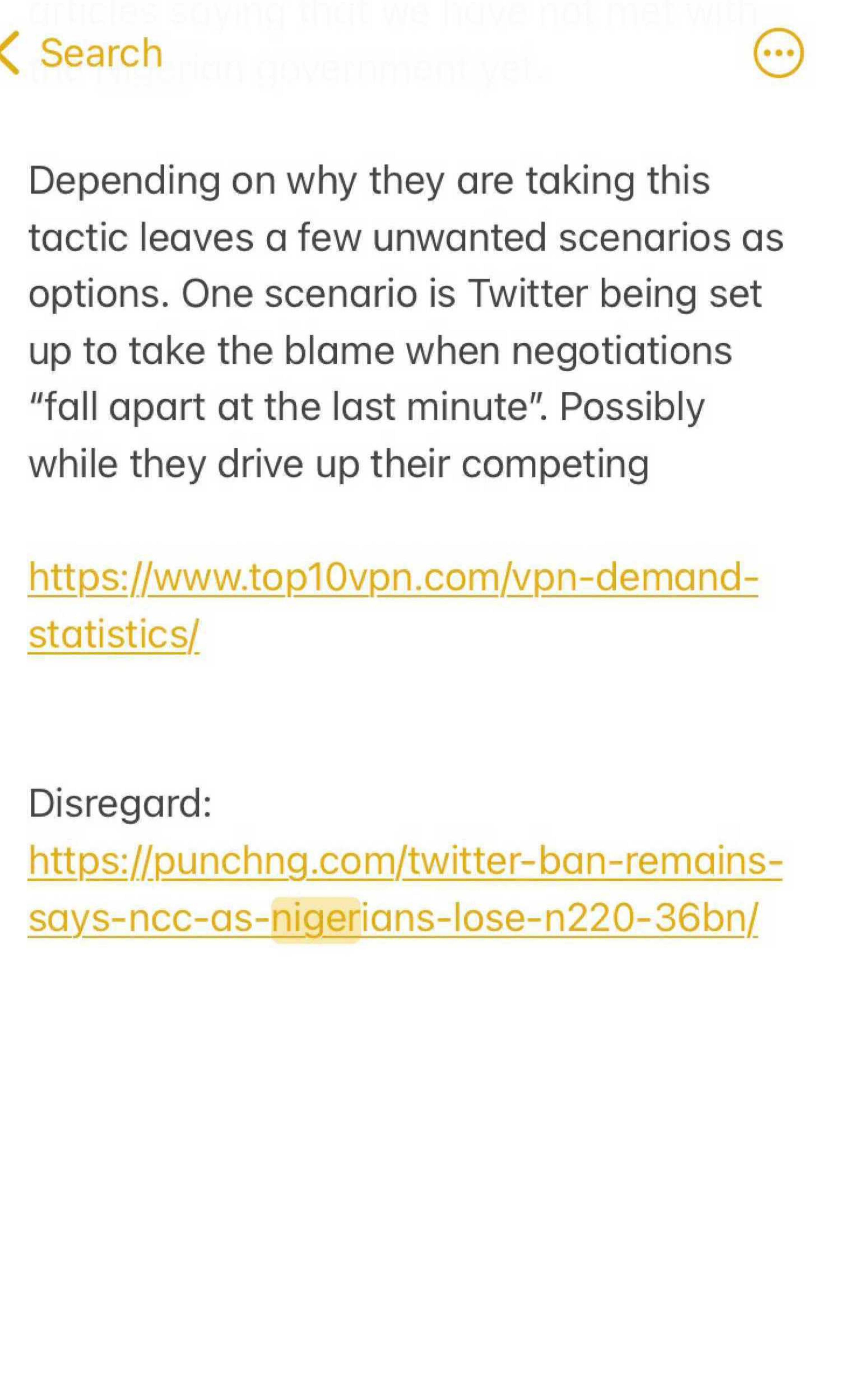
They have refused meetings with Twitter to date. Yet they are publishing media articles saying they are in the midst of negotiations with us and are almost at the point of agreement to end the ban.

Their must recent article claims the minister flew to New York to meet with Twitter execs.

Meanwhile their economy loses 6M a day (360M and growing) while they ban us.

We have not commented on any if the articles saying that we have not met with the Nigerian government yet.

Depending on why they are taking this tactic leaves a few unwanted scenarios as options. One scenario is Twitter being set up to take the blame when negotiations "fall apart at the last minute". Possibly while they drive up their competing

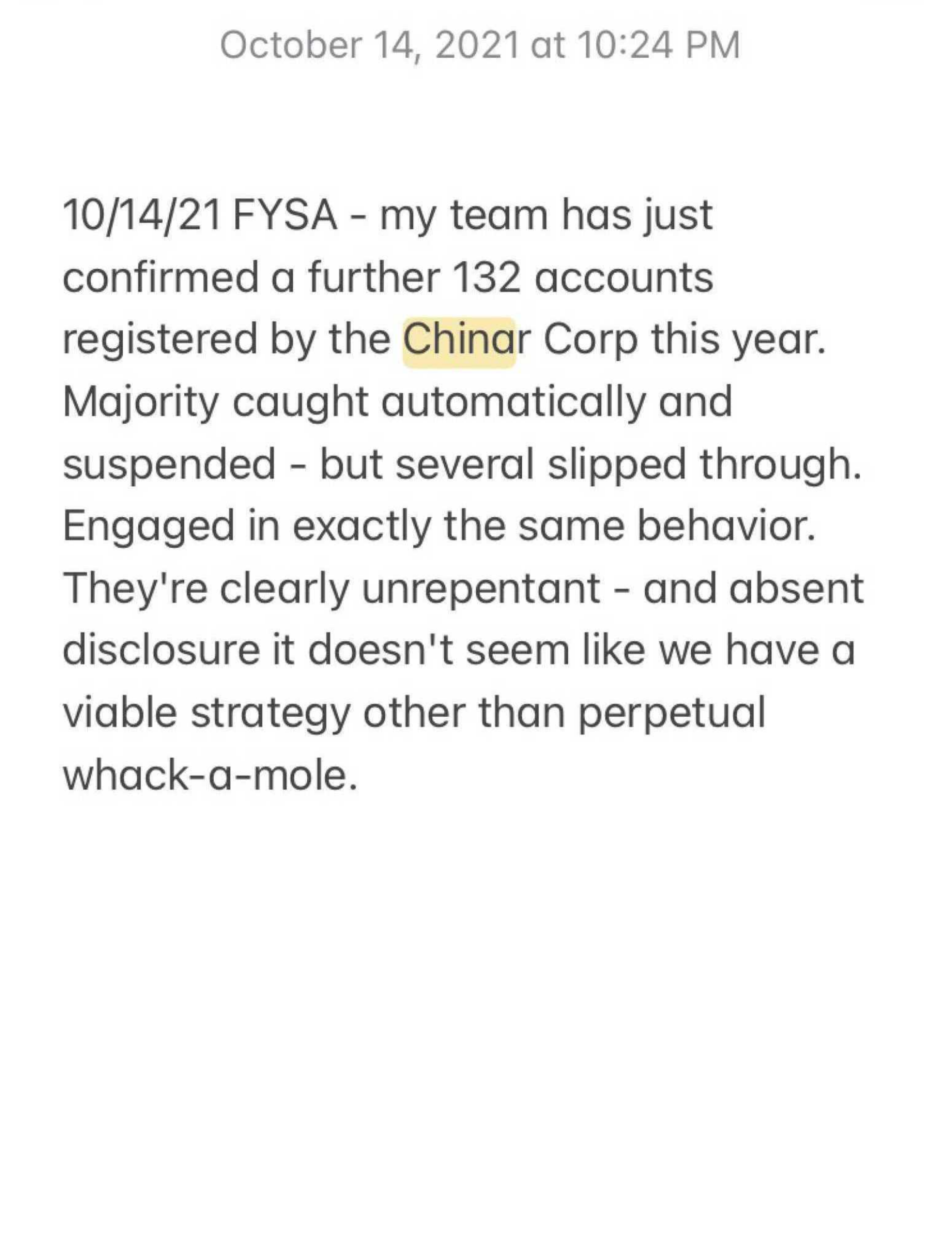


articles saying that we have not met with the Nigerian government yet.

Depending on why they are taking this tactic leaves a few unwanted scenarios as options. One scenario is Twitter being set up to take the blame when negotiations “fall apart at the last minute”. Possibly while they drive up their competing

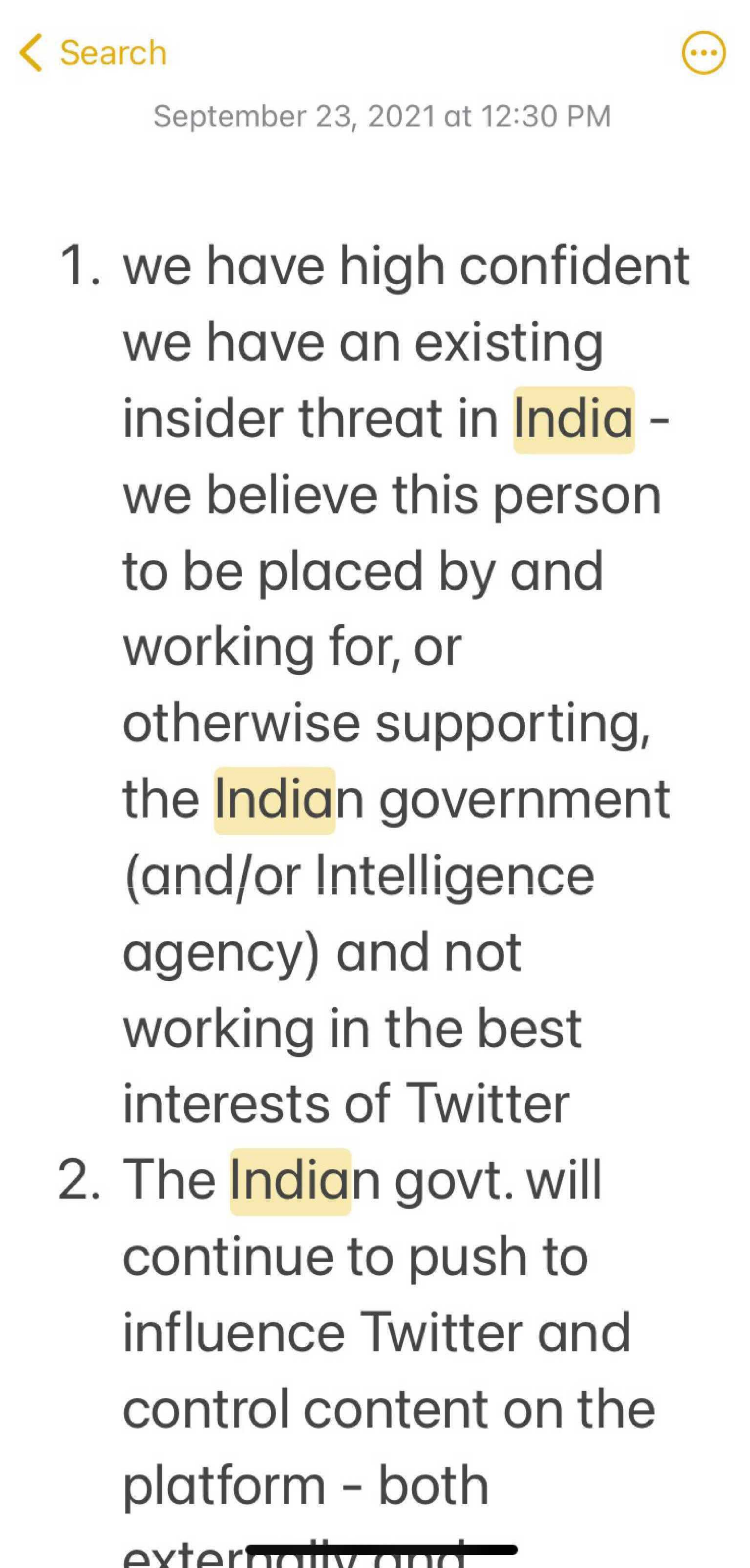
https://www.top10vpn.com/vpn-demand-statistics/

Disregard: https://punchng.com/twitter-ban-remains-says-ncc-as-nigerians-lose-n220-36bn/



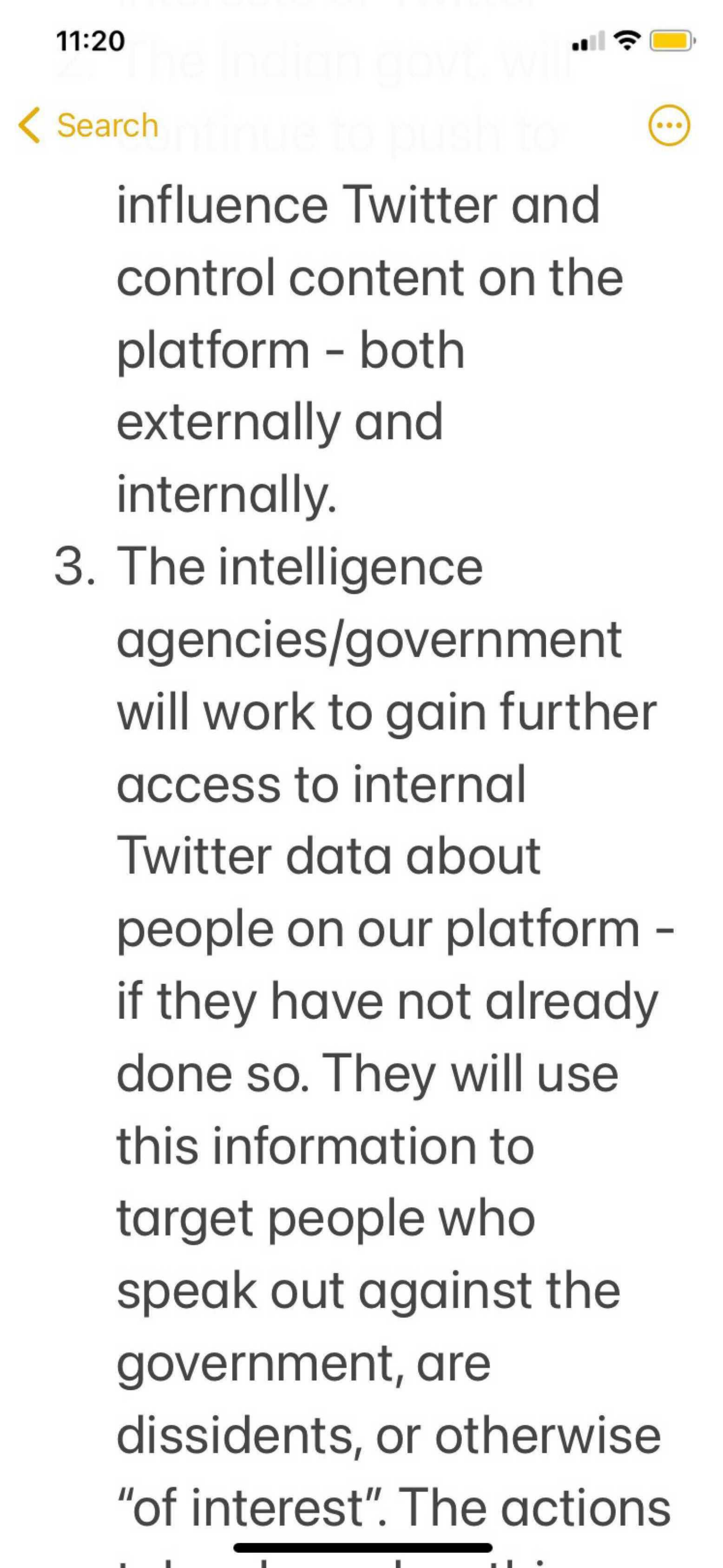
October 14, 2021 at 10:24 PM

10/14/21 FYSA - my team has just confirmed a further 132 accounts registered by the Chinar Corp this year. Majority caught automatically and suspended - but several slipped through. Engaged in exactly the same behavior. They're clearly unrepentant - and absent disclosure it doesn't seem like we have a viable strategy other than perpetual whack-a-mole.

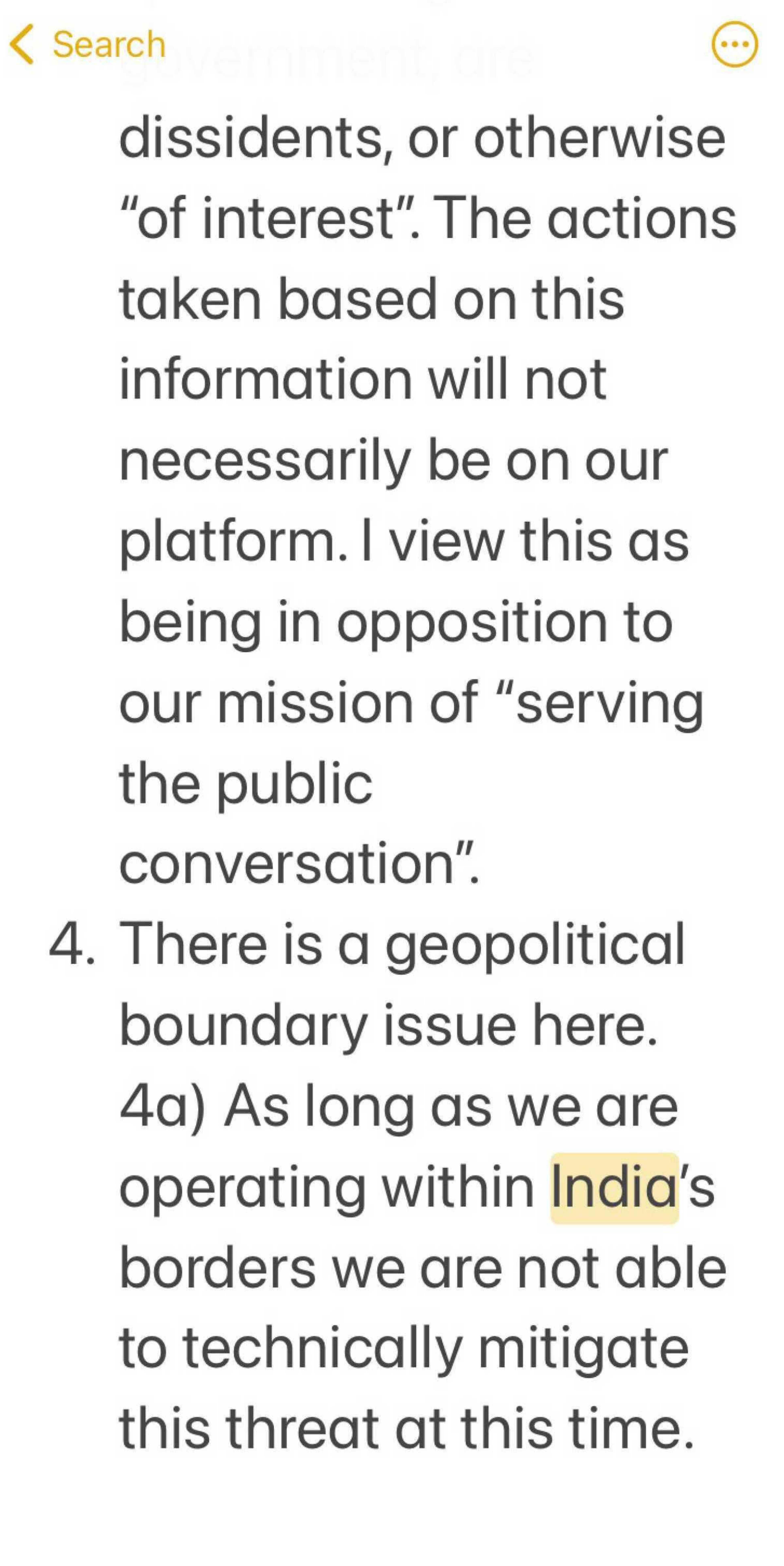


September 23, 2021 at 12:30 PM

1. we have high confident we have an existing insider threat in India - we believe this person to be placed by and working for, or otherwise supporting, the Indian government (and/or Intelligence agency) and not working in the best interests of Twitter
2. The Indian govt. will continue to push to influence Twitter and control content on the platform - both externally and

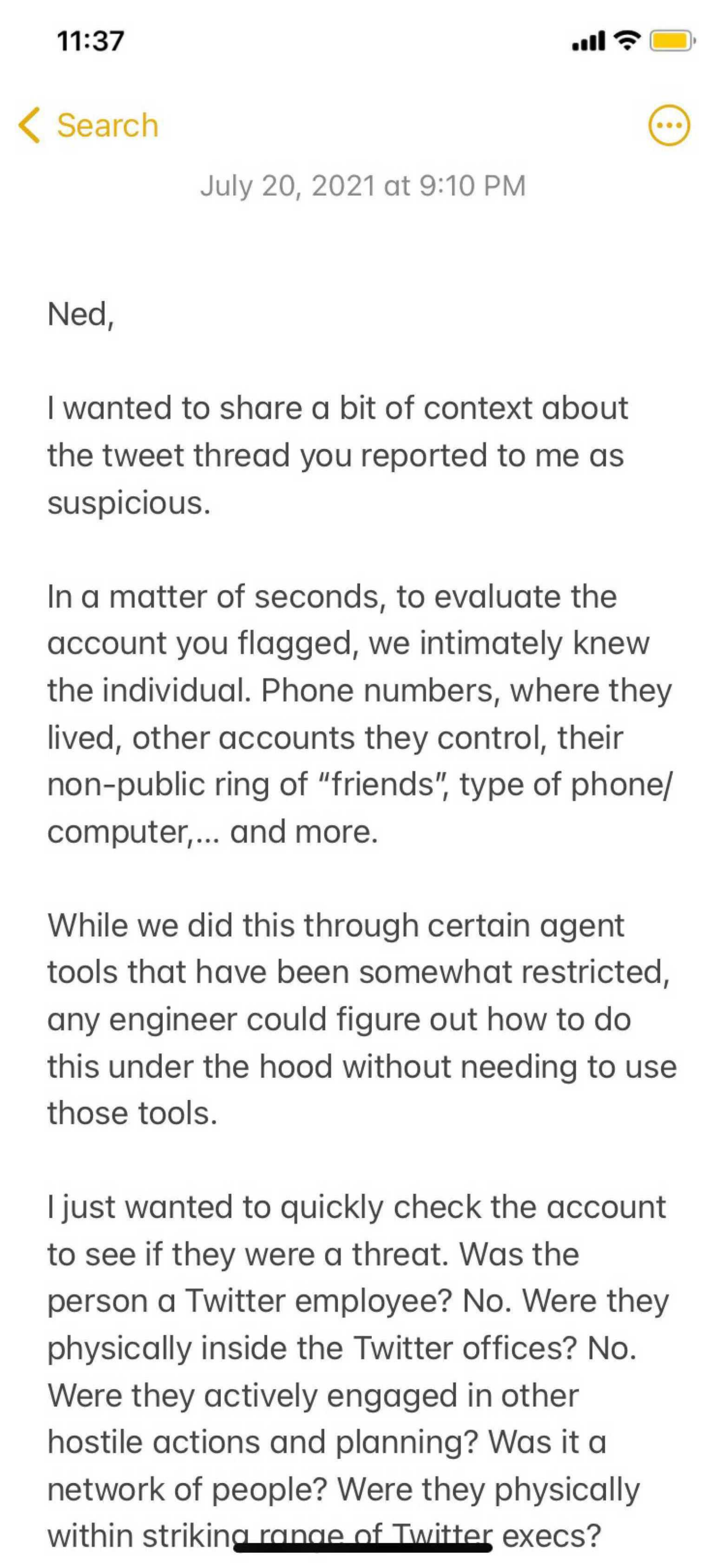


2. The Indian govt. will continue to push to influence Twitter and control content on the platform - both externally and internally.
3. The intelligence agencies/government will work to gain further access to internal Twitter data about people on our platform - if they have not already done so. They will use this information to target people who speak out against the government, are dissidents, or otherwise "of interest". The actions



government, are dissidents, or otherwise "of interest". The actions taken based on this information will not necessarily be on our platform. I view this as being in opposition to our mission of "serving the public conversation".

4. There is a geopolitical boundary issue here.
   4a) As long as we are operating within India's borders we are not able to technically mitigate this threat at this time.



July 20, 2021 at 9:10 PM

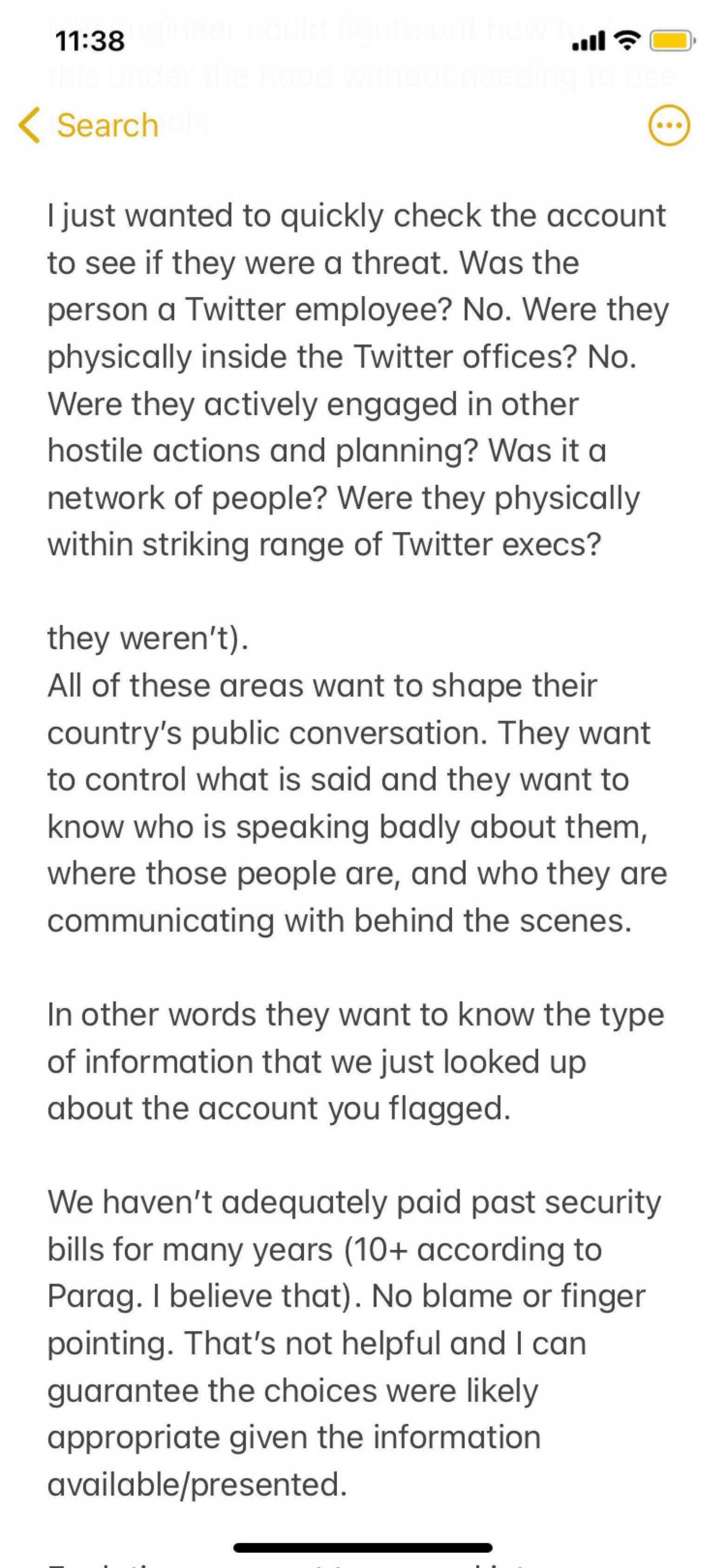
Ned,

I wanted to share a bit of context about the tweet thread you reported to me as suspicious.

In a matter of seconds, to evaluate the account you flagged, we intimately knew the individual. Phone numbers, where they lived, other accounts they control, their non-public ring of "friends", type of phone/ computer,... and more.

While we did this through certain agent tools that have been somewhat restricted, any engineer could figure out how to do this under the hood without needing to use those tools.

I just wanted to quickly check the account to see if they were a threat. Was the person a Twitter employee? No. Were they physically inside the Twitter offices? No. Were they actively engaged in other hostile actions and planning? Was it a network of people? Were they physically within striking range of Twitter execs?

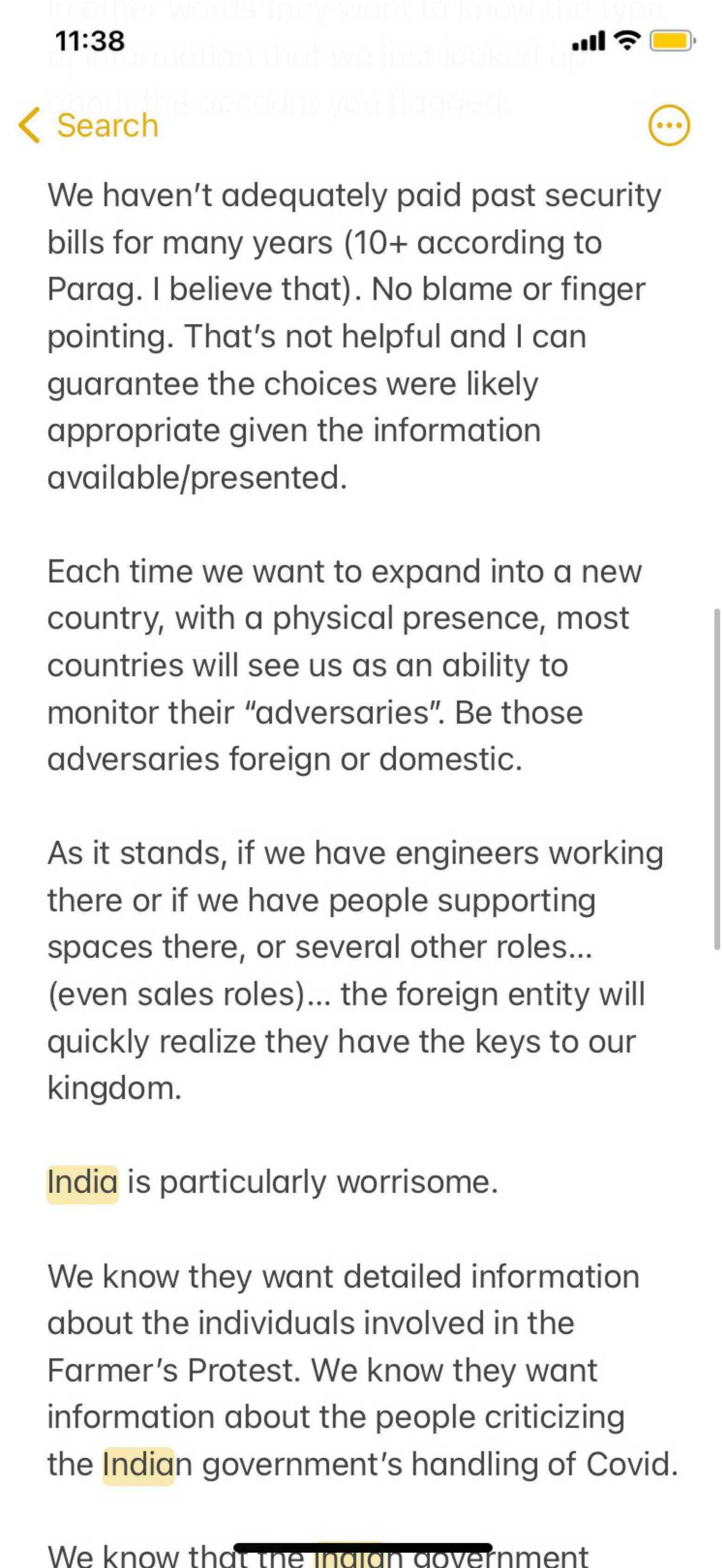


I just wanted to quickly check the account to see if they were a threat. Was the person a Twitter employee? No. Were they physically inside the Twitter offices? No. Were they actively engaged in other hostile actions and planning? Was it a network of people? Were they physically within striking range of Twitter execs?

they weren't).
All of these areas want to shape their country's public conversation. They want to control what is said and they want to know who is speaking badly about them, where those people are, and who they are communicating with behind the scenes.

In other words they want to know the type of information that we just looked up about the account you flagged.

We haven't adequately paid past security bills for many years (10+ according to Parag. I believe that). No blame or finger pointing. That's not helpful and I can guarantee the choices were likely appropriate given the information available/presented.



In other words they want to know the type of information that we just looked up about the account you flagged.

We haven't adequately paid past security bills for many years (10+ according to Parag. I believe that). No blame or finger pointing. That's not helpful and I can guarantee the choices were likely appropriate given the information available/presented.

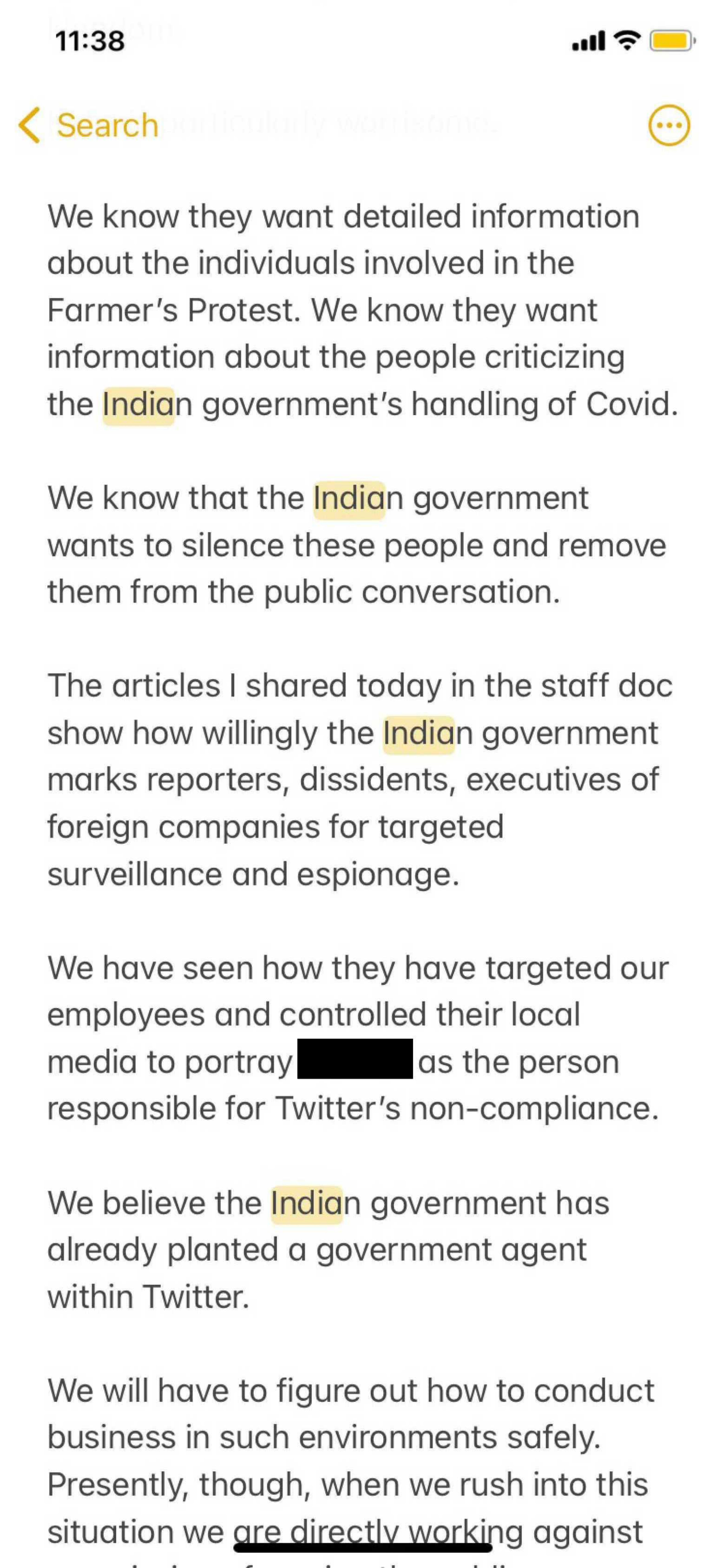
Each time we want to expand into a new country, with a physical presence, most countries will see us as an ability to monitor their "adversaries". Be those adversaries foreign or domestic.

As it stands, if we have engineers working there or if we have people supporting spaces there, or several other roles... (even sales roles)... the foreign entity will quickly realize they have the keys to our kingdom.

India is particularly worrisome.

We know they want detailed information about the individuals involved in the Farmer's Protest. We know they want information about the people criticizing the Indian government's handling of Covid.

We know that the Indian government



We know they want detailed information about the individuals involved in the Farmer’s Protest. We know they want information about the people criticizing the Indian government’s handling of Covid.

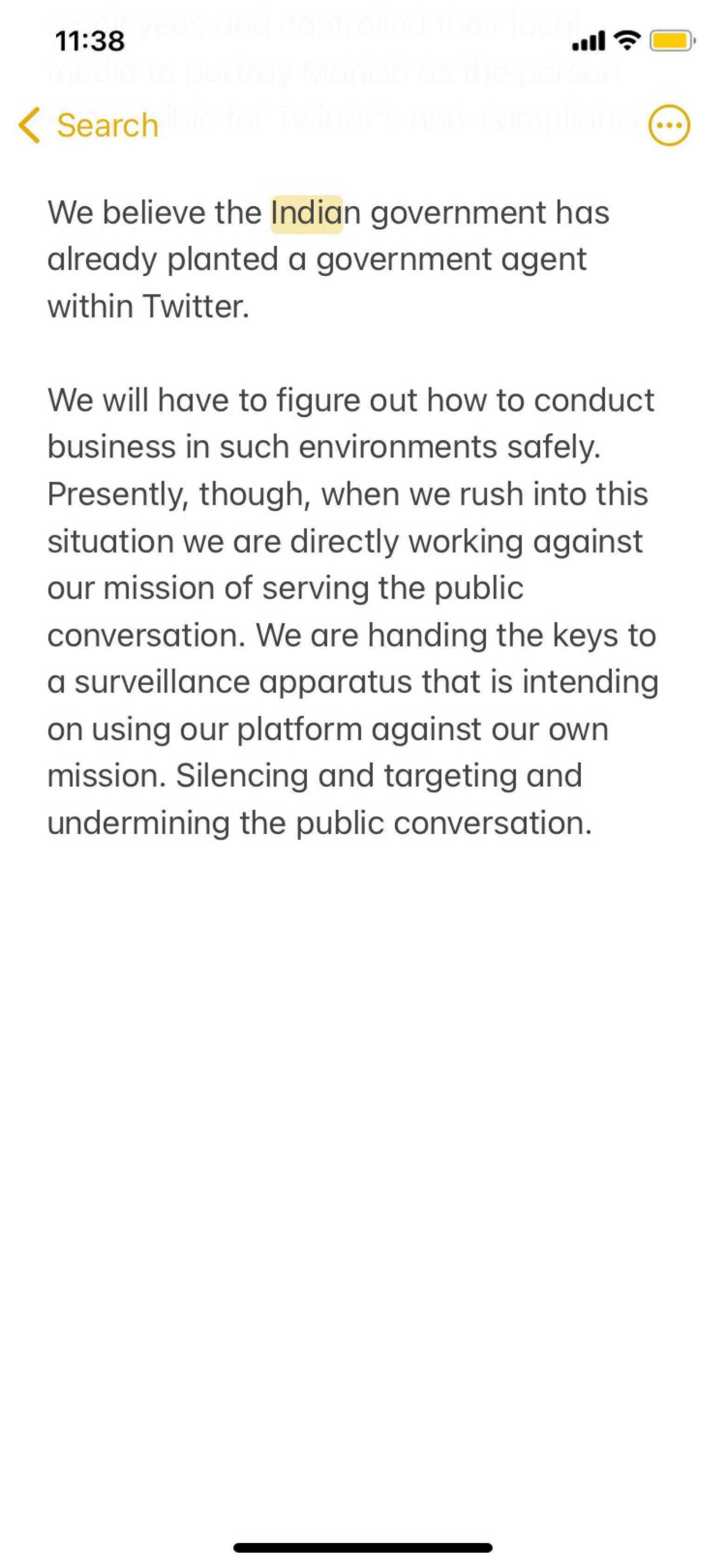
We know that the Indian government wants to silence these people and remove them from the public conversation.

The articles I shared today in the staff doc show how willingly the Indian government marks reporters, dissidents, executives of foreign companies for targeted surveillance and espionage.

We have seen how they have targeted our employees and controlled their local media to portray [redacted] as the person responsible for Twitter’s non-compliance.

We believe the Indian government has already planted a government agent within Twitter.

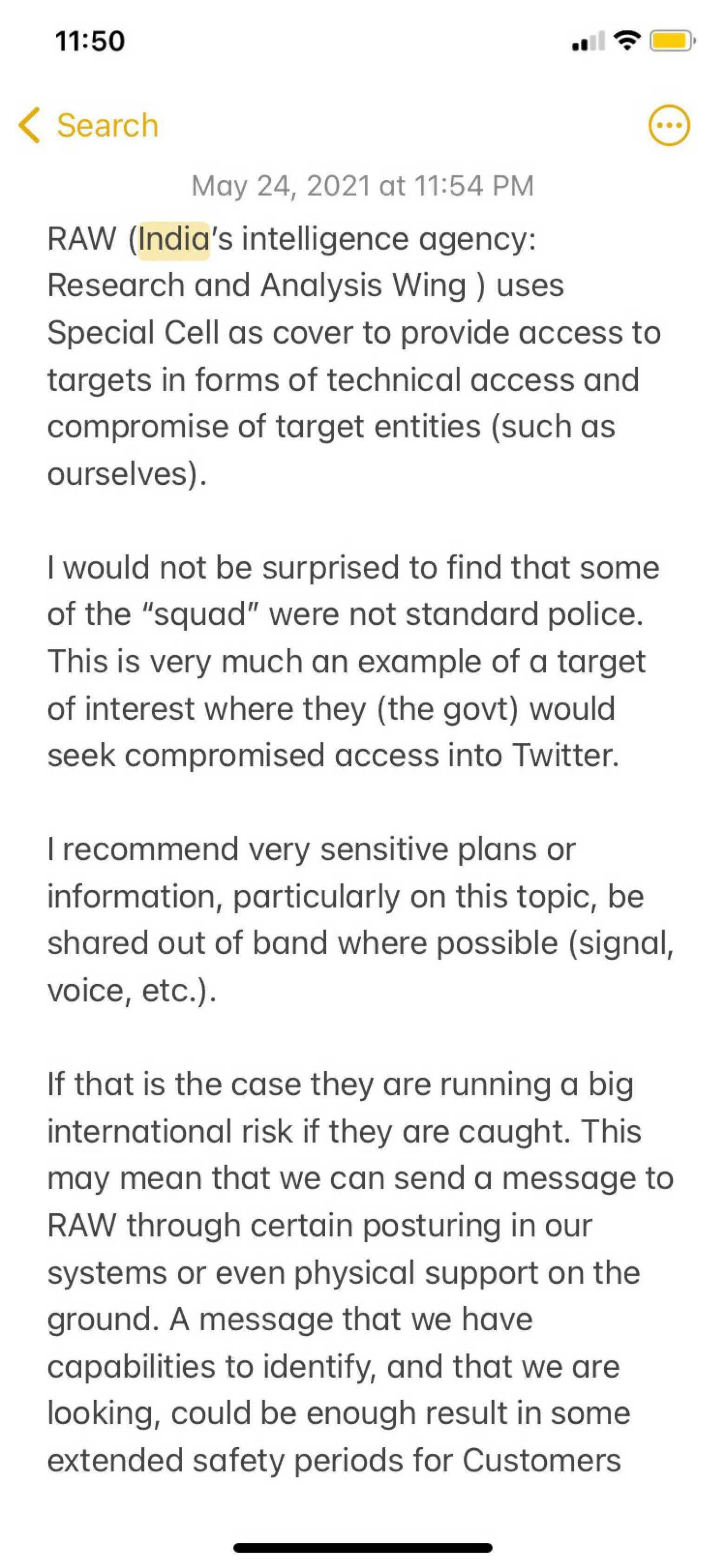
We will have to figure out how to conduct business in such environments safely. Presently, though, when we rush into this situation we are directly working against



employees and controlled their local media to portray Manish as the person responsible for Twitter's non-compliance

We believe the Indian government has already planted a government agent within Twitter.

We will have to figure out how to conduct business in such environments safely. Presently, though, when we rush into this situation we are directly working against our mission of serving the public conversation. We are handing the keys to a surveillance apparatus that is intending on using our platform against our own mission. Silencing and targeting and undermining the public conversation.



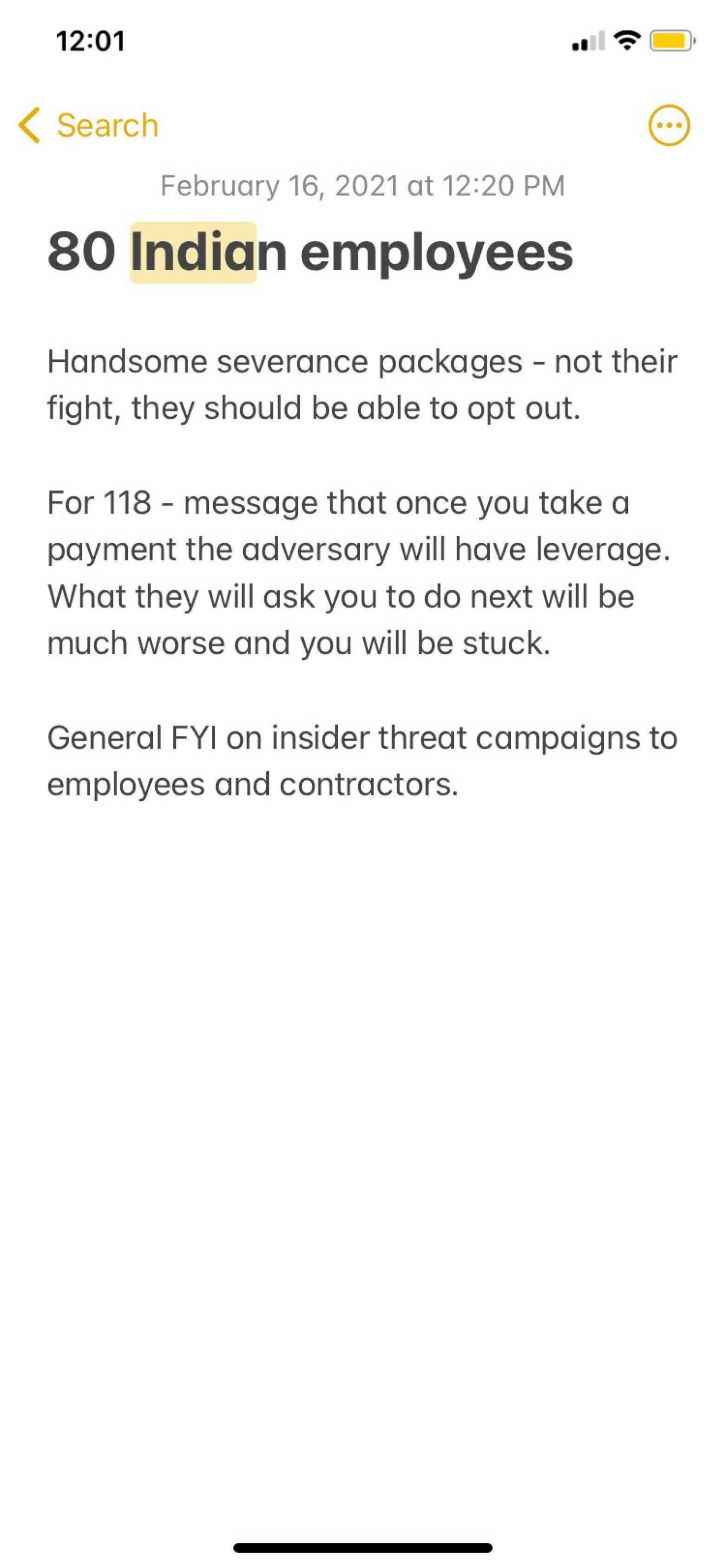
May 24, 2021 at 11:54 PM

RAW (India's intelligence agency: Research and Analysis Wing) uses Special Cell as cover to provide access to targets in forms of technical access and compromise of target entities (such as ourselves).

I would not be surprised to find that some of the "squad" were not standard police. This is very much an example of a target of interest where they (the govt) would seek compromised access into Twitter.

I recommend very sensitive plans or information, particularly on this topic, be shared out of band where possible (signal, voice, etc.).

If that is the case they are running a big international risk if they are caught. This may mean that we can send a message to RAW through certain posturing in our systems or even physical support on the ground. A message that we have capabilities to identify, and that we are looking, could be enough result in some extended safety periods for Customers



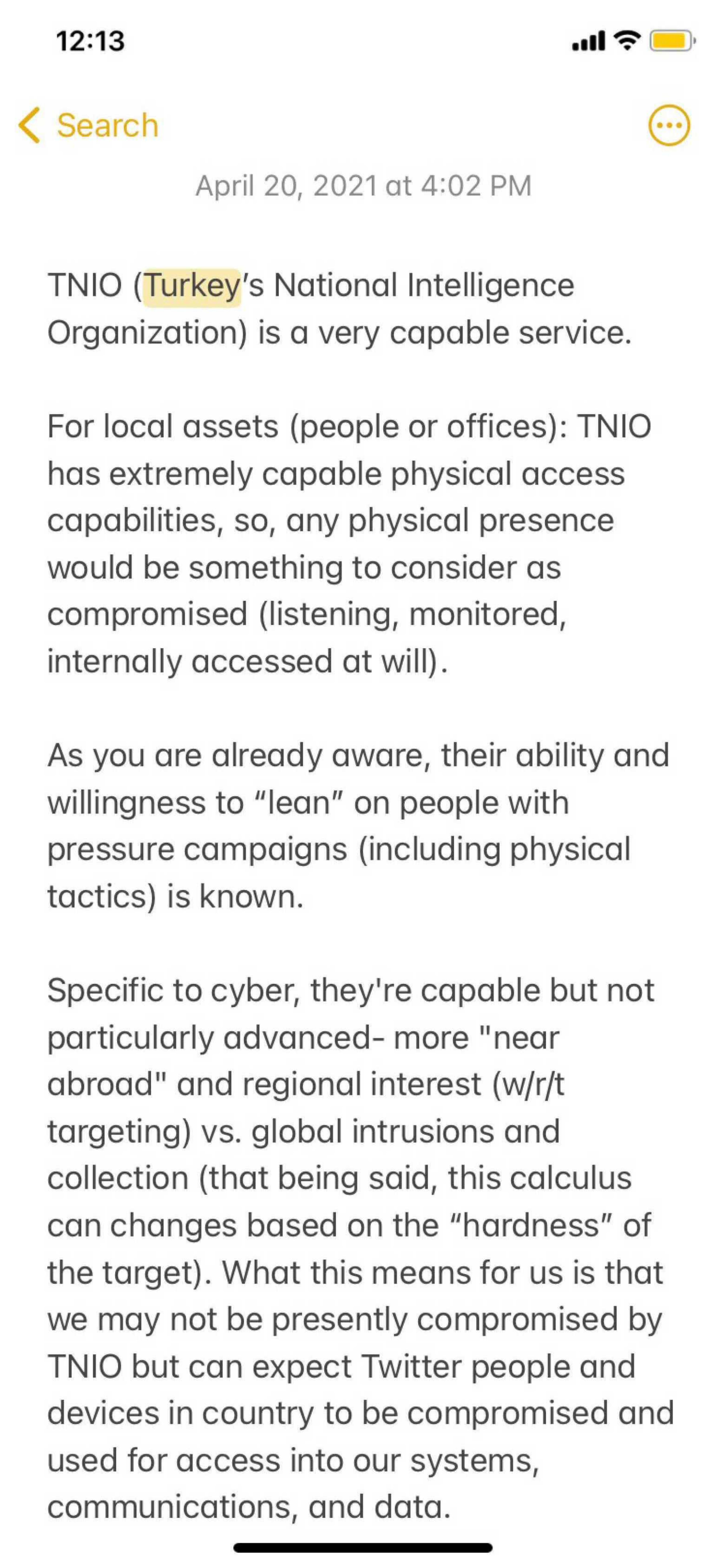
February 16, 2021 at 12:20 PM

80 Indian employees

Handsome severance packages - not their fight, they should be able to opt out.

For 118 - message that once you take a payment the adversary will have leverage. What they will ask you to do next will be much worse and you will be stuck.

General FYI on insider threat campaigns to employees and contractors.



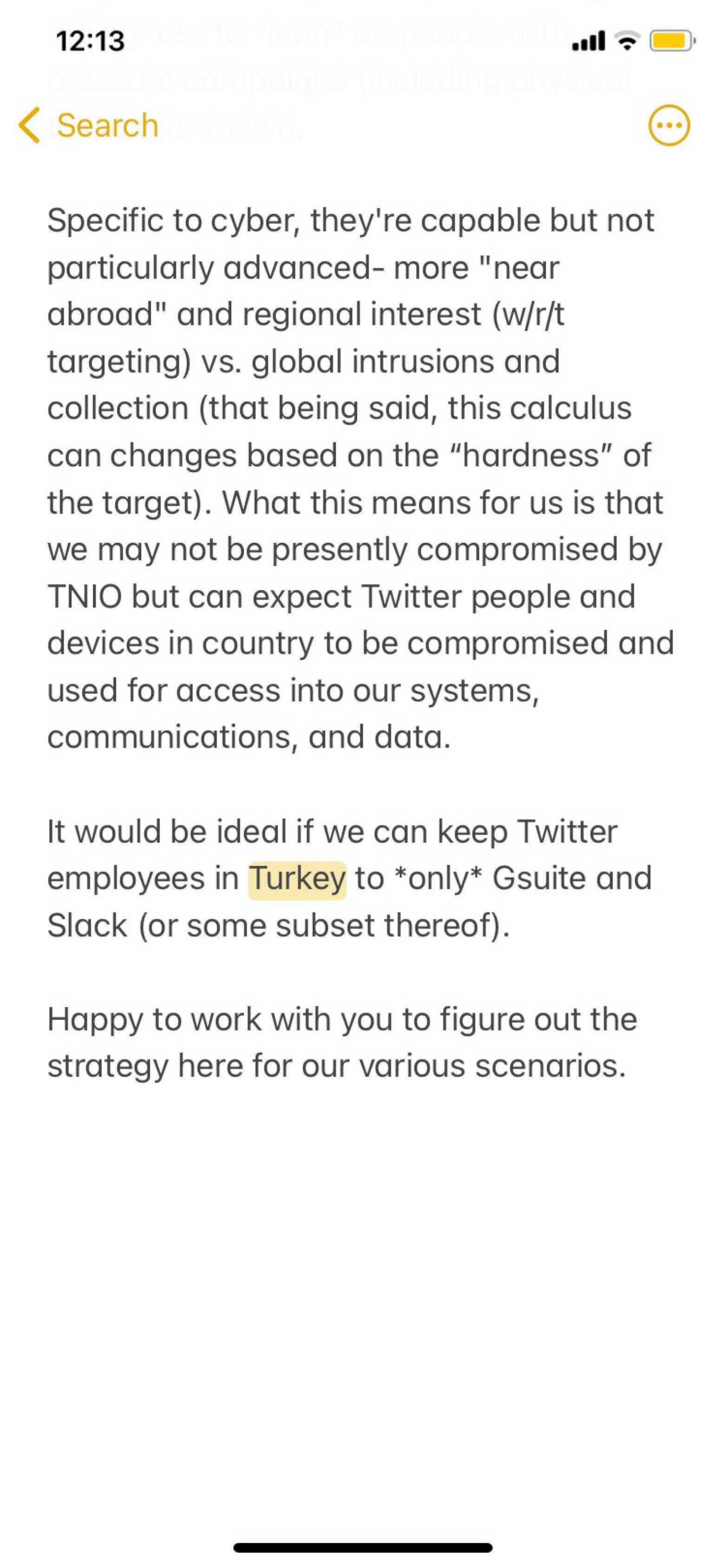
April 20, 2021 at 4:02 PM

TNIO (Turkey's National Intelligence Organization) is a very capable service.

For local assets (people or offices): TNIO has extremely capable physical access capabilities, so, any physical presence would be something to consider as compromised (listening, monitored, internally accessed at will).

As you are already aware, their ability and willingness to "lean" on people with pressure campaigns (including physical tactics) is known.

Specific to cyber, they're capable but not particularly advanced- more "near abroad" and regional interest (w/r/t targeting) vs. global intrusions and collection (that being said, this calculus can changes based on the "hardness" of the target). What this means for us is that we may not be presently compromised by TNIO but can expect Twitter people and devices in country to be compromised and used for access into our systems, communications, and data.

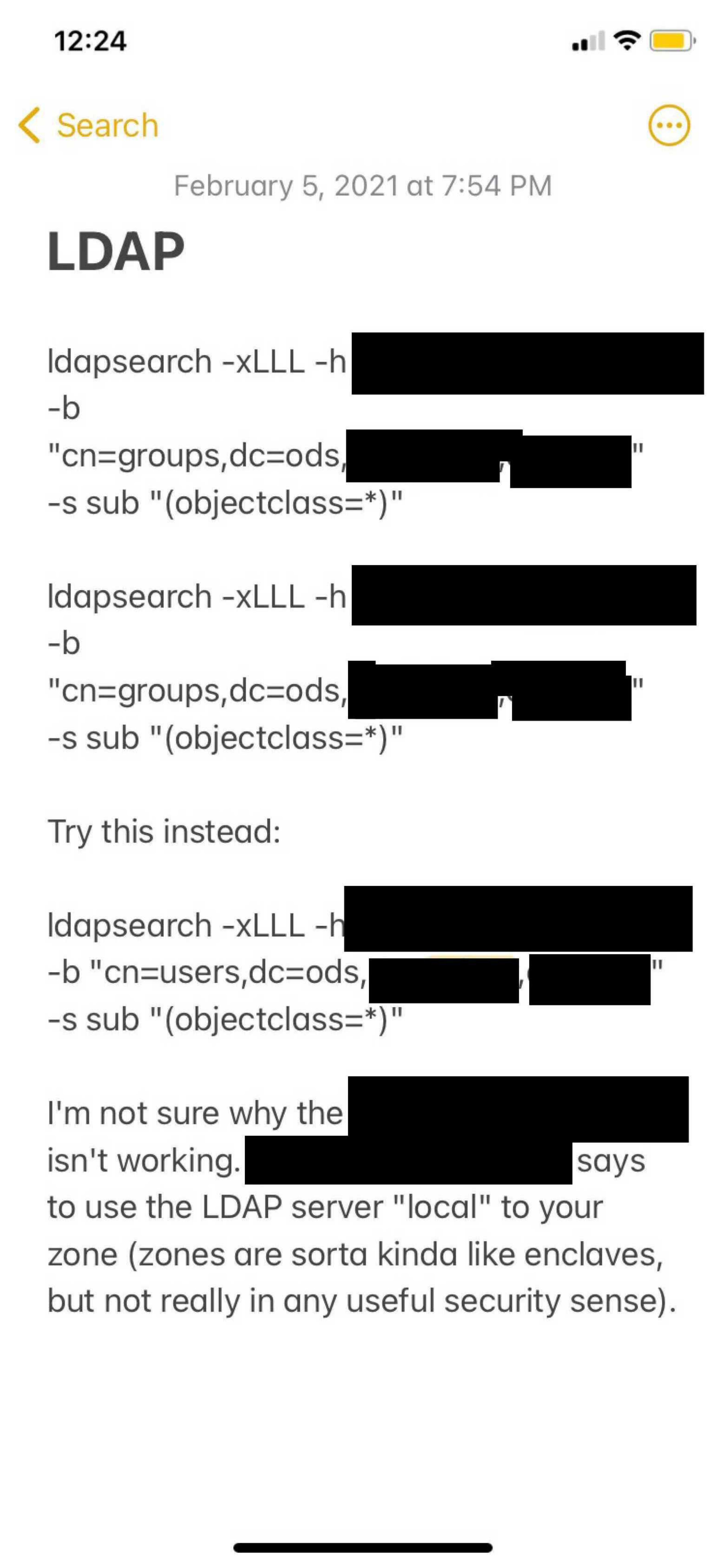


gness to "lean" on people with
pressure campaigns (including physical
is known.

Specific to cyber, they're capable but not particularly advanced- more "near abroad" and regional interest (w/r/t targeting) vs. global intrusions and collection (that being said, this calculus can changes based on the “hardness” of the target). What this means for us is that we may not be presently compromised by TNIO but can expect Twitter people and devices in country to be compromised and used for access into our systems, communications, and data.

It would be ideal if we can keep Twitter employees in Turkey to *only* Gsuite and Slack (or some subset thereof).

Happy to work with you to figure out the strategy here for our various scenarios.



February 5, 2021 at 7:54 PM

LDAP

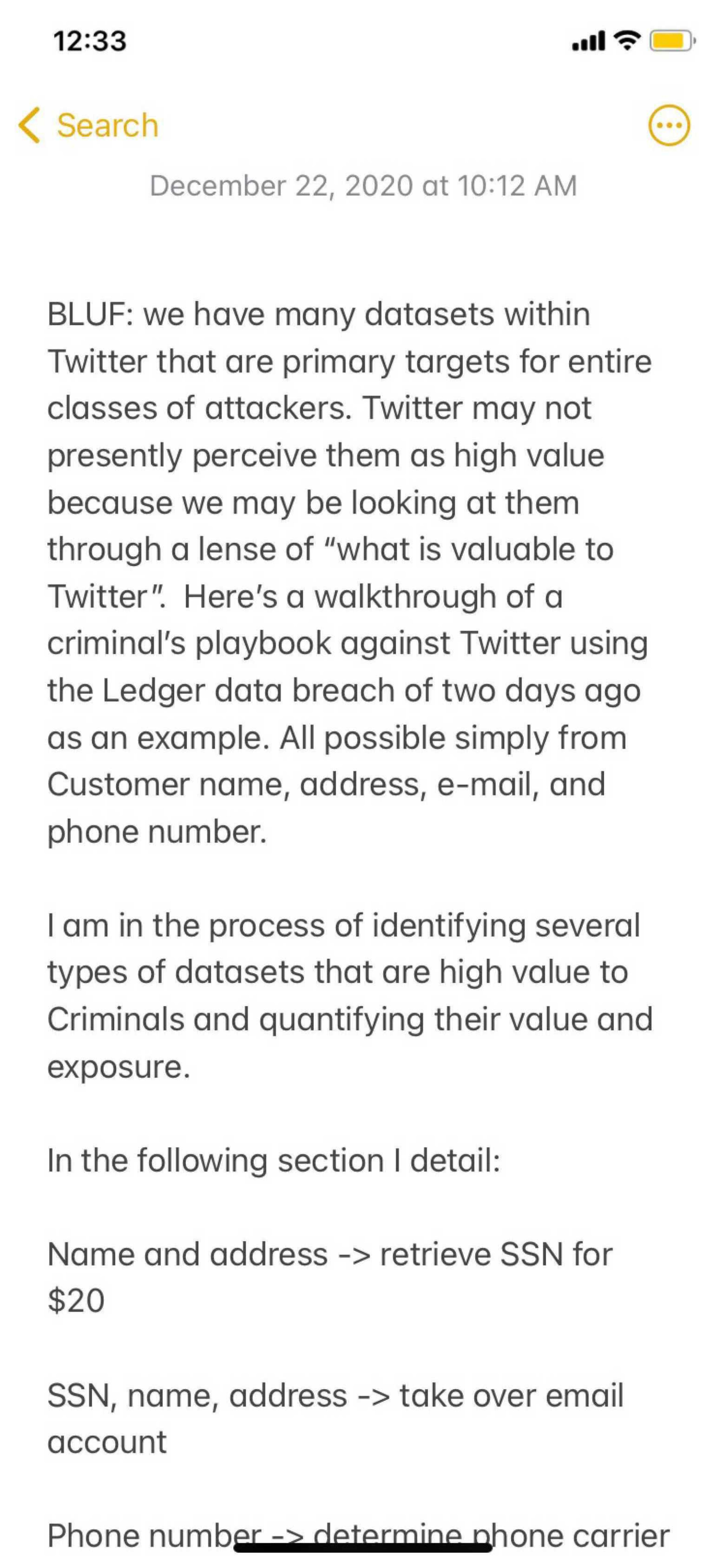
```
ldapsearch -xLLL -h [redacted]
-b
"cn=groups,dc=ods,[redacted],[redacted]"
-s sub "(objectclass=*)"
```

```
ldapsearch -xLLL -h [redacted]
-b
"cn=groups,dc=ods,[redacted],[redacted]"
-s sub "(objectclass=*)"
```

Try this instead:

```
ldapsearch -xLLL -h [redacted]
-b "cn=users,dc=ods,[redacted],[redacted]"
-s sub "(objectclass=*)"
```

I'm not sure why the [redacted] isn't working. [redacted] says to use the LDAP server "local" to your zone (zones are sorta kinda like enclaves, but not really in any useful security sense).



December 22, 2020 at 10:12 AM

BLUF: we have many datasets within Twitter that are primary targets for entire classes of attackers. Twitter may not presently perceive them as high value because we may be looking at them through a lense of "what is valuable to Twitter". Here's a walkthrough of a criminal's playbook against Twitter using the Ledger data breach of two days ago as an example. All possible simply from Customer name, address, e-mail, and phone number.

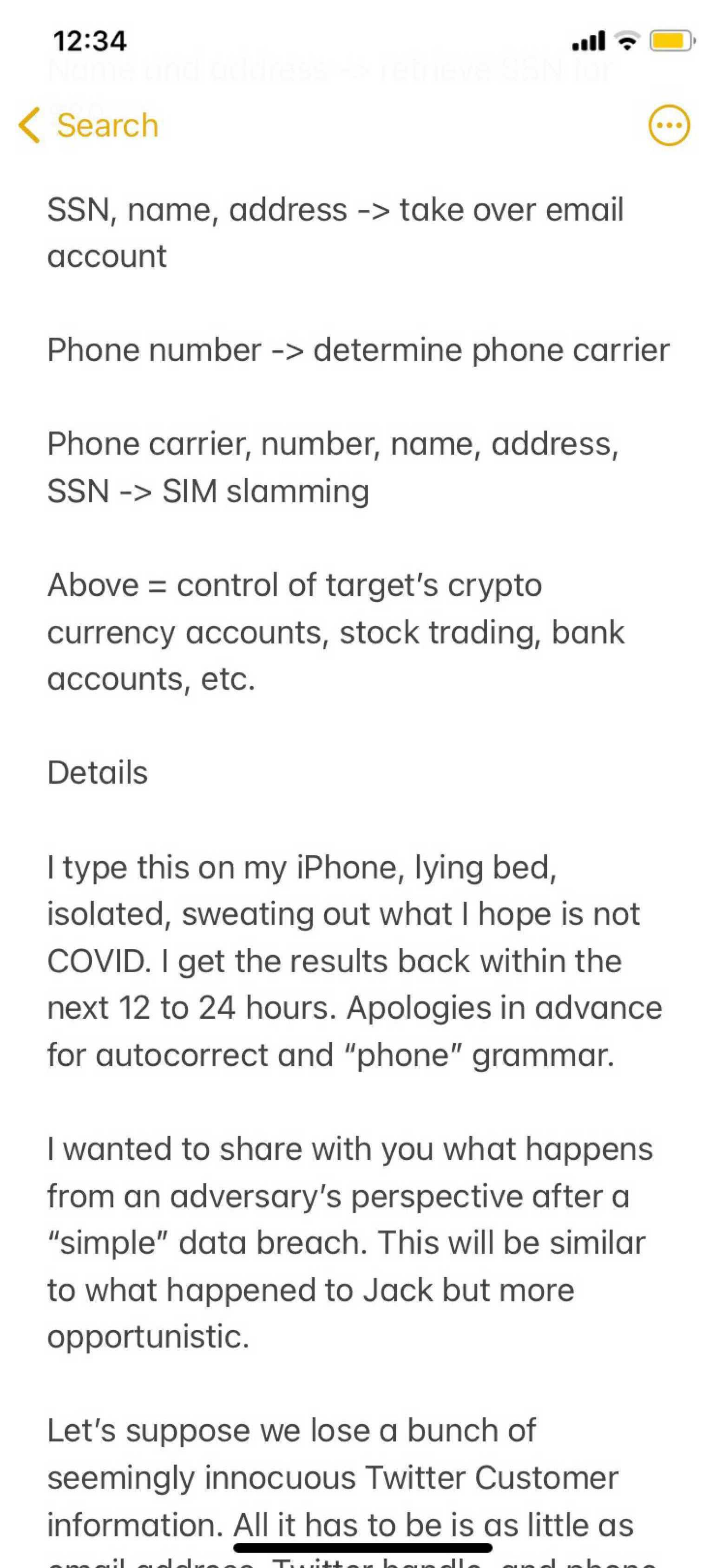
I am in the process of identifying several types of datasets that are high value to Criminals and quantifying their value and exposure.

In the following section I detail:

Name and address -> retrieve SSN for $20

SSN, name, address -> take over email account

Phone number -> determine phone carrier



Name and address -> retrieve SSN for

SSN, name, address -> take over email account

Phone number -> determine phone carrier

Phone carrier, number, name, address, SSN -> SIM slamming

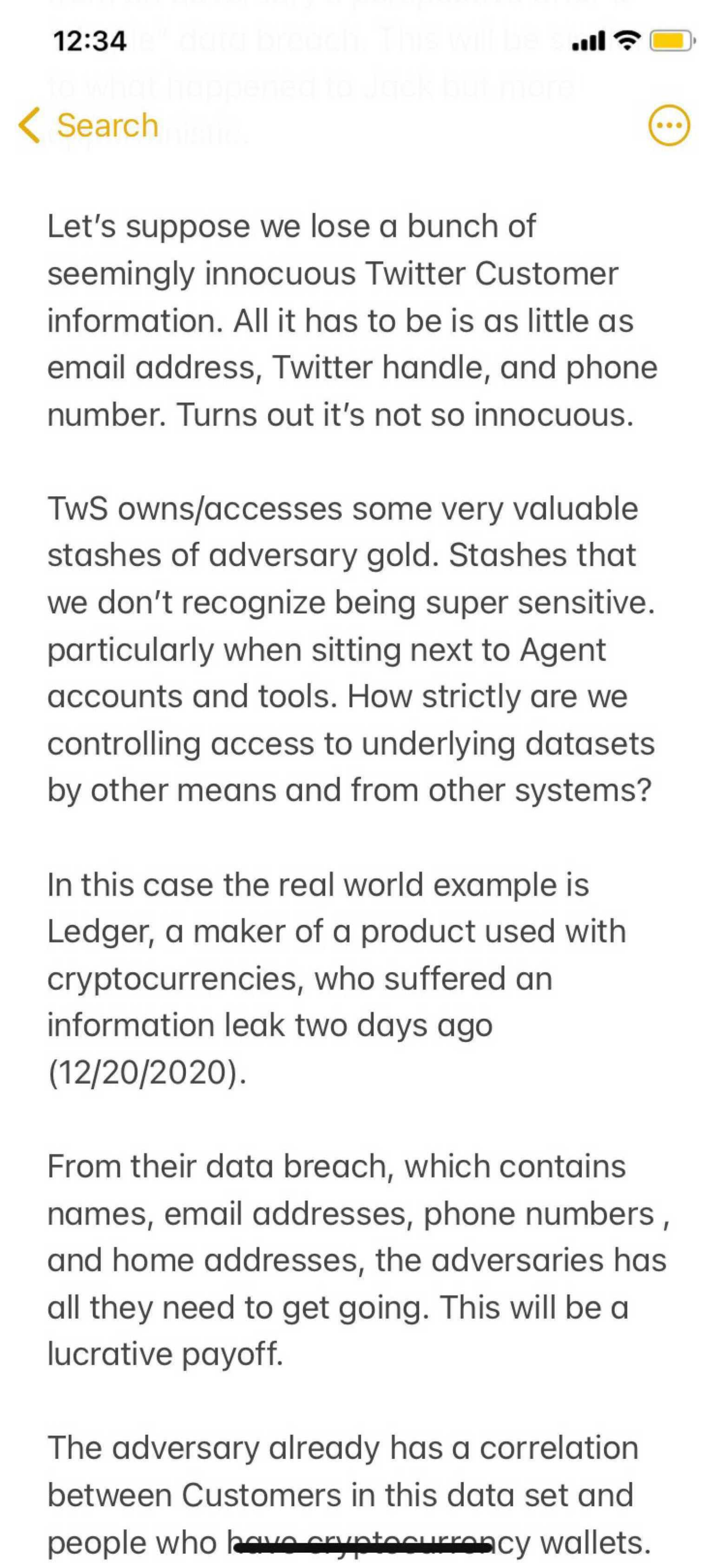
Above = control of target’s crypto currency accounts, stock trading, bank accounts, etc.

Details

I type this on my iPhone, lying bed, isolated, sweating out what I hope is not COVID. I get the results back within the next 12 to 24 hours. Apologies in advance for autocorrect and “phone” grammar.

I wanted to share with you what happens from an adversary’s perspective after a “simple” data breach. This will be similar to what happened to Jack but more opportunistic.

Let’s suppose we lose a bunch of seemingly innocuous Twitter Customer information. All it has to be is as little as email address, Twitter handle, and phone



le" data breach. This will be si
to what happened to Jack but more

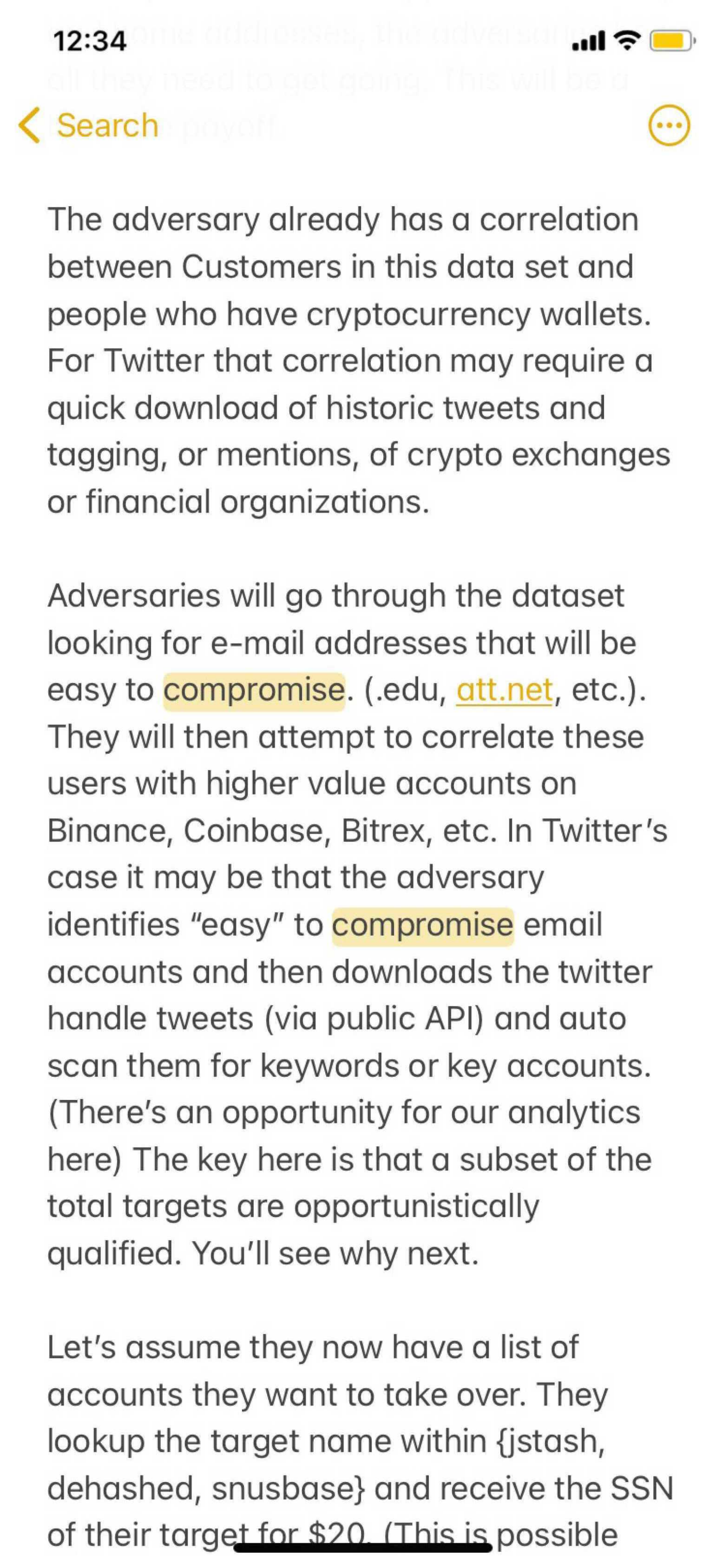
Let's suppose we lose a bunch of seemingly innocuous Twitter Customer information. All it has to be is as little as email address, Twitter handle, and phone number. Turns out it's not so innocuous.

TwS owns/accesses some very valuable stashes of adversary gold. Stashes that we don't recognize being super sensitive. particularly when sitting next to Agent accounts and tools. How strictly are we controlling access to underlying datasets by other means and from other systems?

In this case the real world example is Ledger, a maker of a product used with cryptocurrencies, who suffered an information leak two days ago (12/20/2020).

From their data breach, which contains names, email addresses, phone numbers , and home addresses, the adversaries has all they need to get going. This will be a lucrative payoff.

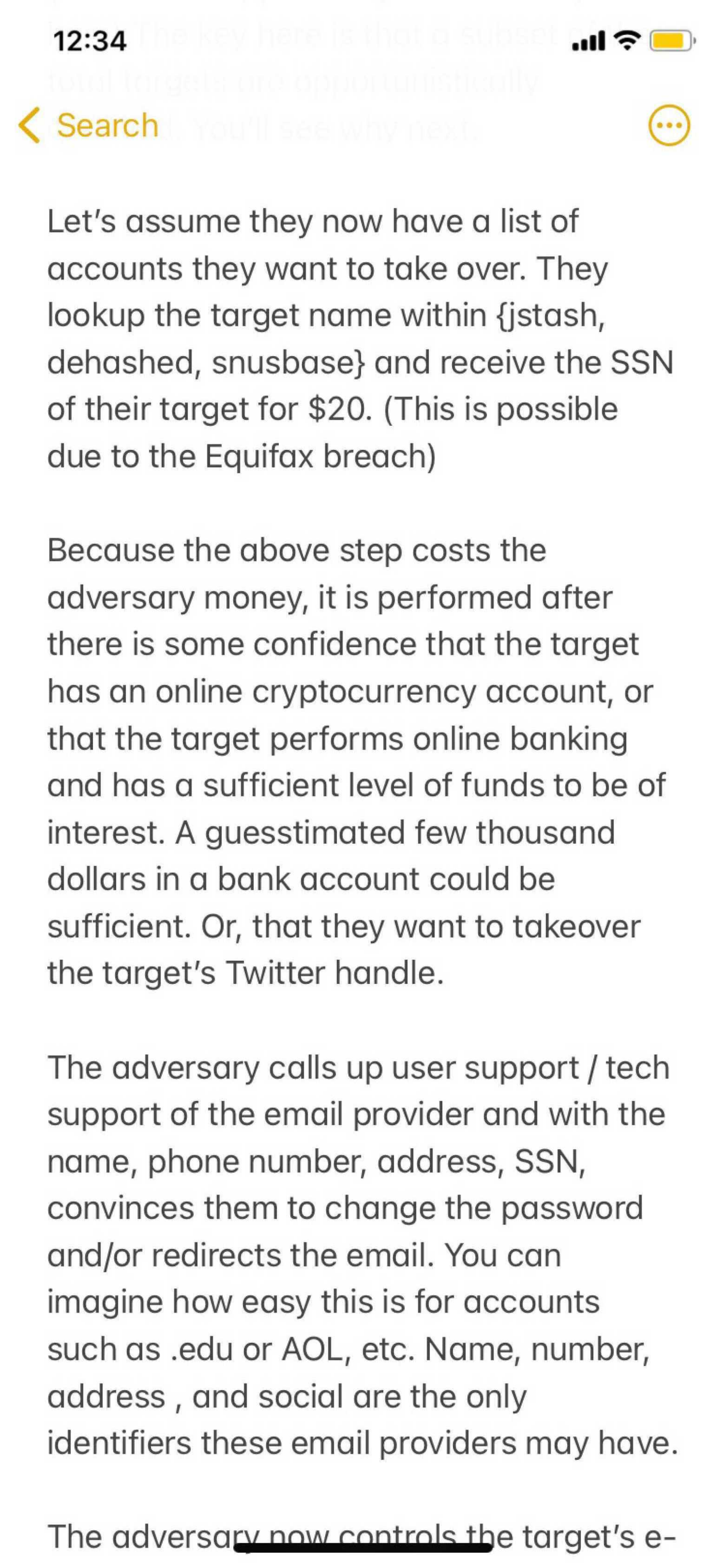
The adversary already has a correlation between Customers in this data set and people who have cryptocurrency wallets.



The adversary already has a correlation between Customers in this data set and people who have cryptocurrency wallets. For Twitter that correlation may require a quick download of historic tweets and tagging, or mentions, of crypto exchanges or financial organizations.

Adversaries will go through the dataset looking for e-mail addresses that will be easy to compromise. (.edu, att.net, etc.). They will then attempt to correlate these users with higher value accounts on Binance, Coinbase, Bitrex, etc. In Twitter's case it may be that the adversary identifies "easy" to compromise email accounts and then downloads the twitter handle tweets (via public API) and auto scan them for keywords or key accounts. (There's an opportunity for our analytics here) The key here is that a subset of the total targets are opportunistically qualified. You'll see why next.

Let's assume they now have a list of accounts they want to take over. They lookup the target name within {jstash, dehashed, snusbase} and receive the SSN of their target for $20. (This is possible



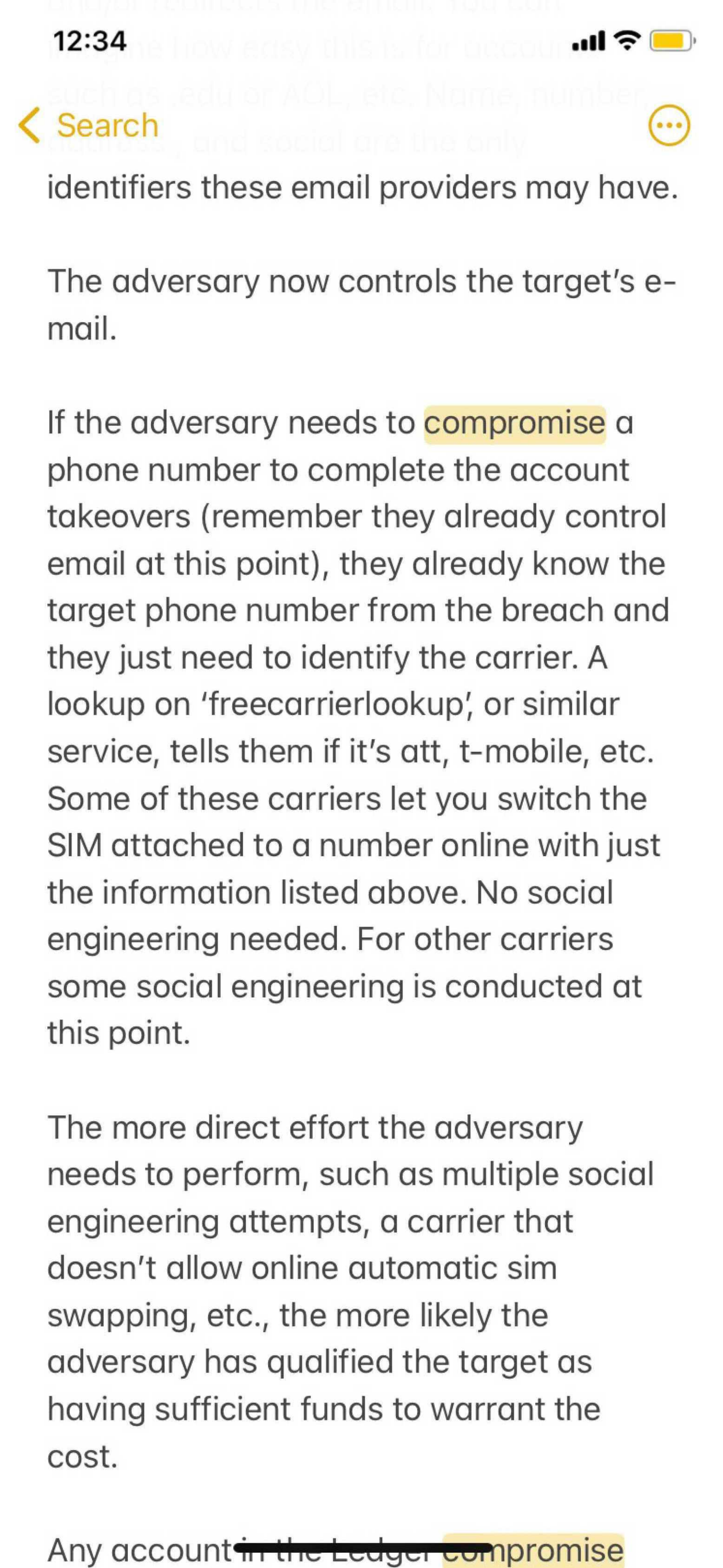
The key here is that a subset of the total targets are opportunistically d. You'll see why next.

Let's assume they now have a list of accounts they want to take over. They lookup the target name within {jstash, dehashed, snusbase} and receive the SSN of their target for $20. (This is possible due to the Equifax breach)

Because the above step costs the adversary money, it is performed after there is some confidence that the target has an online cryptocurrency account, or that the target performs online banking and has a sufficient level of funds to be of interest. A guesstimated few thousand dollars in a bank account could be sufficient. Or, that they want to takeover the target's Twitter handle.

The adversary calls up user support / tech support of the email provider and with the name, phone number, address, SSN, convinces them to change the password and/or redirects the email. You can imagine how easy this is for accounts such as .edu or AOL, etc. Name, number, address , and social are the only identifiers these email providers may have.

The adversary now controls the target's e-



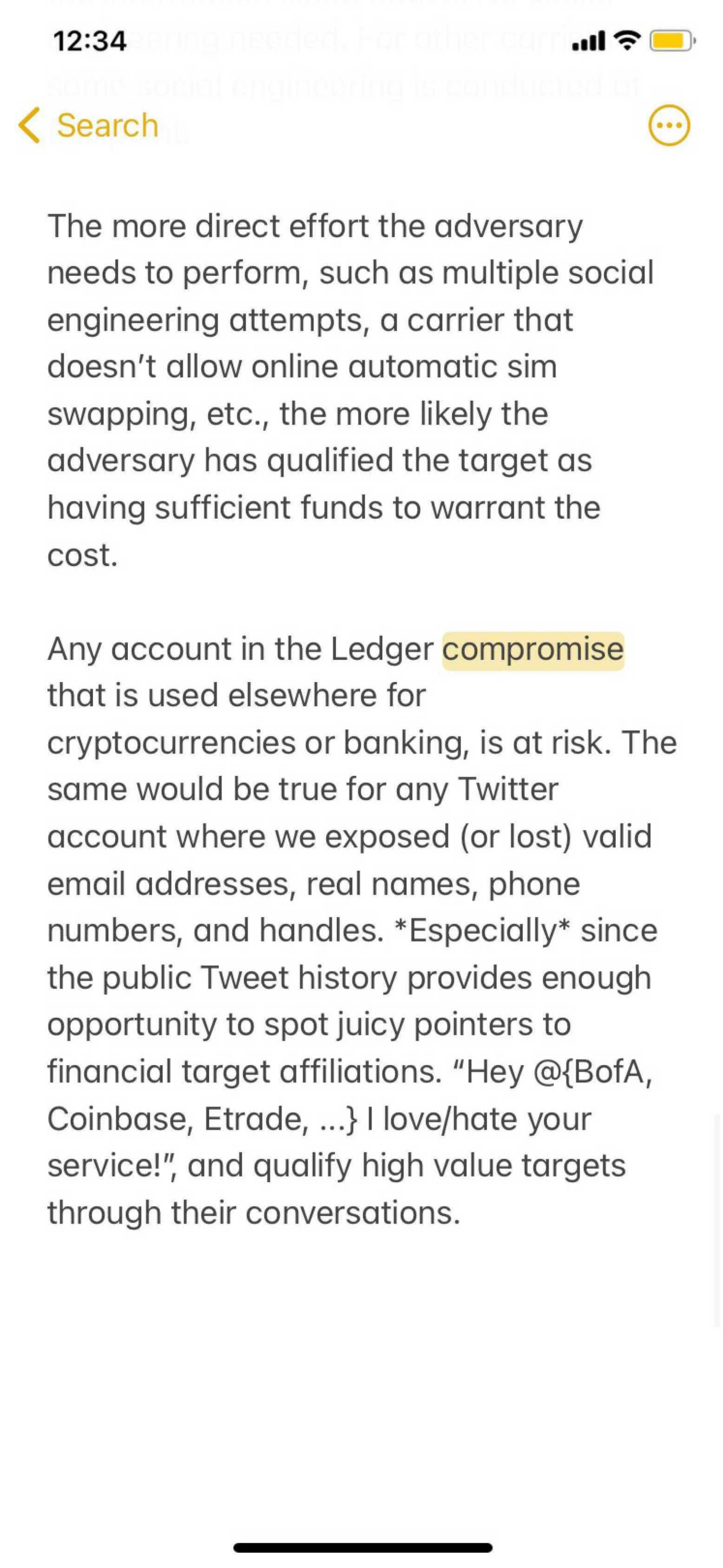
imagine how easy this is for accounts such as .edu or AOL, etc. Name, number, address, and social are the only identifiers these email providers may have.

The adversary now controls the target's e-mail.

If the adversary needs to compromise a phone number to complete the account takeovers (remember they already control email at this point), they already know the target phone number from the breach and they just need to identify the carrier. A lookup on 'freecarrierlookup', or similar service, tells them if it's att, t-mobile, etc. Some of these carriers let you switch the SIM attached to a number online with just the information listed above. No social engineering needed. For other carriers some social engineering is conducted at this point.

The more direct effort the adversary needs to perform, such as multiple social engineering attempts, a carrier that doesn't allow online automatic sim swapping, etc., the more likely the adversary has qualified the target as having sufficient funds to warrant the cost.

Any account in the Ledger compromise



eering needed. For other carriers, some social engineering is conducted at this point.

The more direct effort the adversary needs to perform, such as multiple social engineering attempts, a carrier that doesn't allow online automatic sim swapping, etc., the more likely the adversary has qualified the target as having sufficient funds to warrant the cost.

Any account in the Ledger compromise that is used elsewhere for cryptocurrencies or banking, is at risk. The same would be true for any Twitter account where we exposed (or lost) valid email addresses, real names, phone numbers, and handles. *Especially* since the public Tweet history provides enough opportunity to spot juicy pointers to financial target affiliations. "Hey @{BofA, Coinbase, Etrade, ...} I love/hate your service!", and qualify high value targets through their conversations.

Insider Risk Team Activities by Month

-Highlights how workload is developing for this new program

-At this time highlights previously unaddressed InT referrals

Total Activities
Investigations
KNIT
Offboarding
Travel Brief
Notable Event
6
4
2
0
1
1
1
3
2
2
2
4
1
1
4
3
1
1
3
1
1
2
February
April
May
June
July
August
September
October
November
December

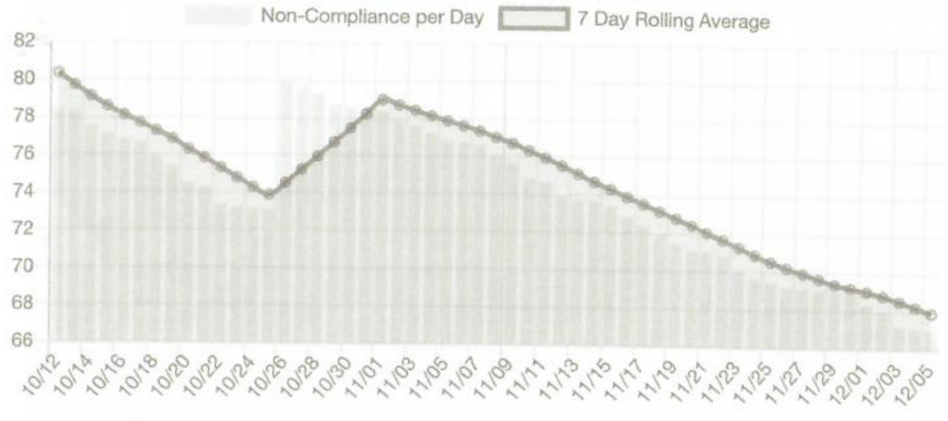
2021

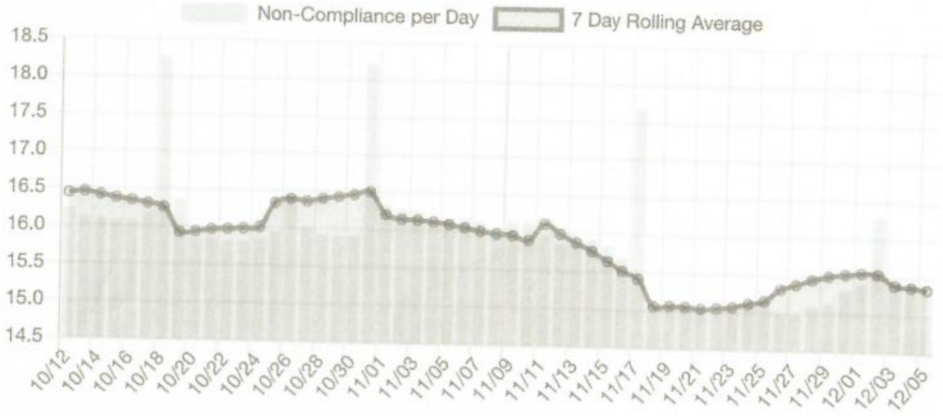
Non-Compliant Kernels

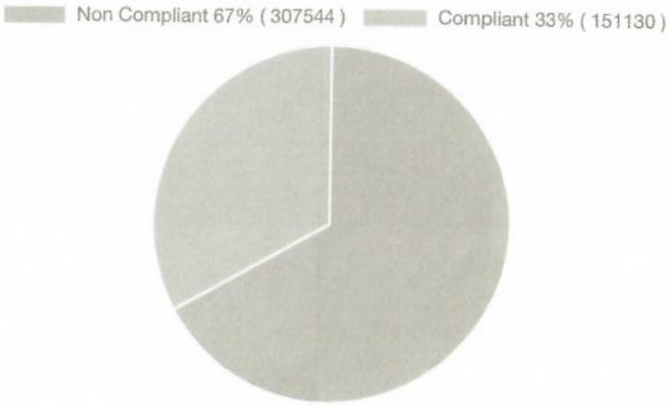
307,544 (67%)

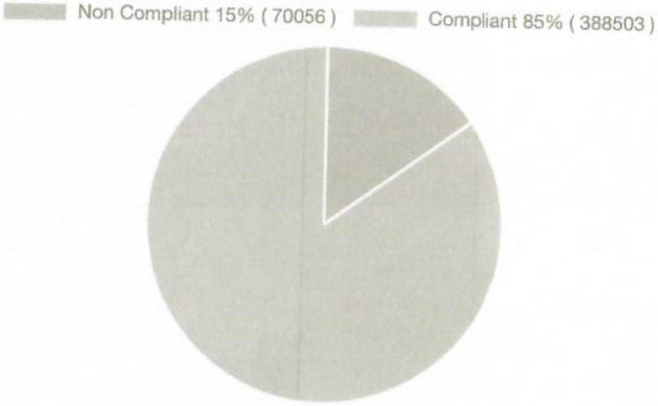
Non-Compliant Operating Systems

70,056 (15%)


Kernel Version Non-Compliance per Day and 7 Day Moving Average


Operating System Version Non-Compliance per Day and 7 Day Moving Average


Kernel Compliance Overview (No Exceptions)


OS Compliance Overview (No Exceptions)

2020

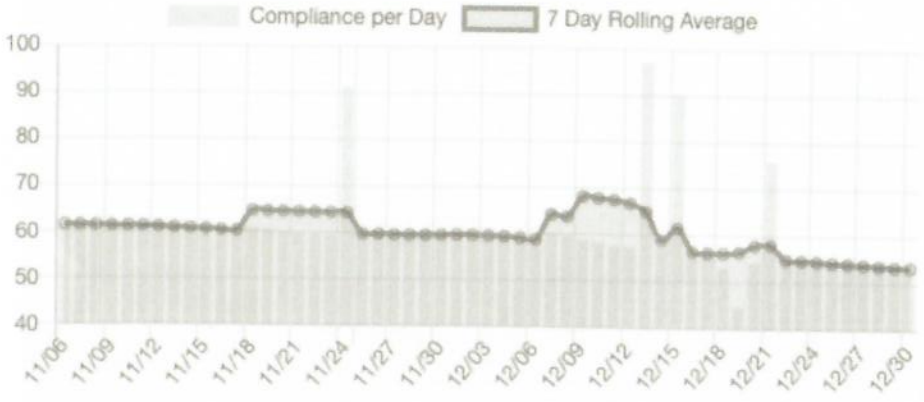
Non-Compliant Kernels

186,372 (53%)

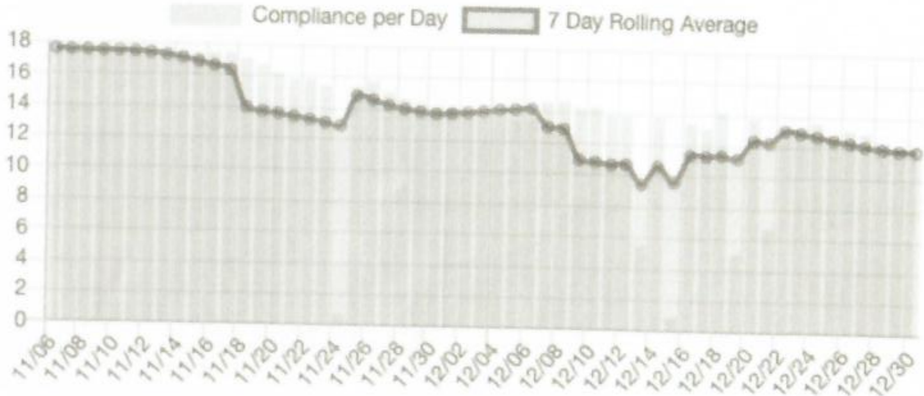
Non-Compliant Operating Systems

41,443 (12%)

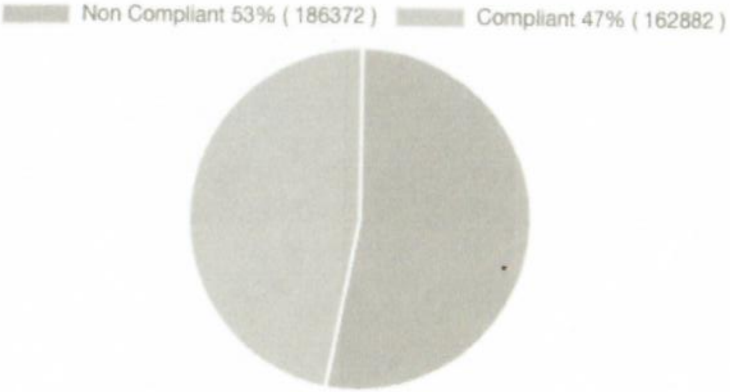
Kernel Version Compliance per Day and 7 Day Moving Average



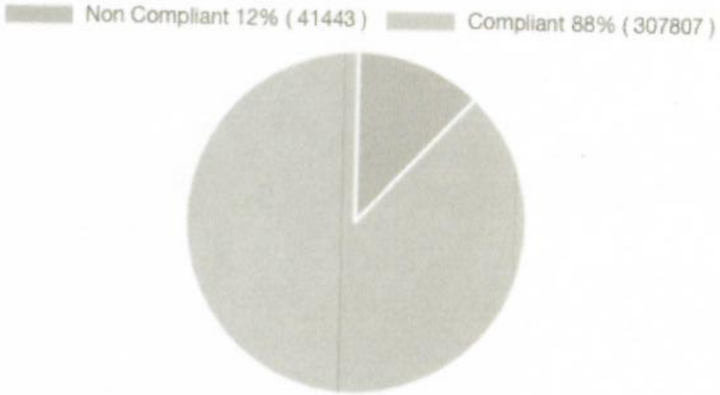
Operating System Version Compliance per Day and 7 Day Moving Average



Kernel Compliance Overview (No Exceptions)



OS Compliance Overview (No Exceptions)



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?

huh? 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 ~ track? implement lessons learned from the past. Is this achievable? Are there existing facts?

Can we compare to Sota.. elsewhere?

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
we ~~seldom~~ introduce
policy changes?
can these be refined?

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 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 in footnotes

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 as separate teams to enable scale?~~

How does, being responsible for drafting policy preclude scale enforcement?

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-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.

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.

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

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

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 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

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"

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.

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*

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 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 tothem. 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.

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 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.

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 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 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."

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 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.

Del /
Vul

#1 Knowledge mgmt
Investigation mgmt

#2 Staffing strategy