

— ANONYMOUS WHISTLEBLOWER DISCLOSURE —

[REDACTED]

SEC Office of the Whistleblower
Via Online Portal & Fax

Re: Supplemental Disclosure of Securities Law Violations by Facebook, Inc. (NASDAQ: FB), SEC TCR # [REDACTED]

Facebook misled investors and the public about bringing “the world closer together” where it relegates international users and promotes global division and ethnic violence.

To the SEC Office of the Whistleblower:

1. The instant letter is one of multiple disclosures related to the above-captioned matter. Our anonymous client is disclosing original evidence showing that **Facebook, Inc. (NASDAQ: FB)** has, for years past and ongoing, violated U.S. securities laws by making **material misrepresentations and omissions in statements to investors and prospective investors**, including, *inter alia*, through filings with the SEC, testimony to Congress, online statements and media stories.
2. **Summary.** Facebook claims it has 2.8 billion users across the globe, nearly two-thirds of which use a language other than English; and Facebook has repeatedly reassured the public that it is committed to international issues; however, Facebook's internal documents show that Facebook lacks adequate systems, and facilitates polarizing misinformation and ethnic violence, across the world.

FACEBOOK'S KNOWLEDGE OF MATERIAL ISSUES

3. **As an Initial Matter, Facebook Targets and is Used by People Around the World.** For example, in its last 10-K,¹ Facebook outlined:

¹<https://sec.report/Document/0001326801-21-000014/>. Emphasis is added throughout this letter by bold/underlined text.

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

<https://WhistleblowerAid.org> — Anonymously via Tor Browser:
<http://p6ufg73qskew53cglxt6hktyt35rbl46yultzyuytq3tvicywa3pclid.onion>
Contact via **SecureDrop** over Tor: <http://whistlebloweraid.securedrop.tor.onion> — via **Signal App**: +1 201-773-1371

*“The aggregate market value of the voting and non-voting stock held by non-affiliates . . . was \$563 billion. . . **Facebook daily active users (DAUs) were 1.84 billion on average for December 2020 . . . Facebook monthly active users (MAUs) were 2.80 billion.**”*

“Daily Active Users” worldwide were estimated as 1,845 million at the end of December 2020, with “Daily Active Users” in the United States and Canada combined as 195 million in the same time period [i.e., 10.57% of users are in those countries]. Europe was listed as 308 million, Asia-Pacific as 744 million, and the “Rest of World” as 598 million during this timeframe.

4. Moreover, as summarized by the Federal Trade Commission’s recent amended complaint against Facebook for antitrust violations:

*“Last year alone, Facebook generated revenues of more than \$85 billion and profits of more than \$29 billion . . . [and] in earnings calls, Facebook COO **Sheryl Sandberg described Facebook . . . as the ‘world’s first global platform** that lets marketers personalize their messages at unprecedented scale,’ and called Facebook Blue and Instagram the ‘two most important mobile advertising platforms’ in the world.”²*

5. Facebook Has Admitted Potential Investor Risks With its International Operations. For instance, Facebook’s recent 10-K³ and 10-Q⁴ acknowledged:

*“We have significant international operations and plan to continue expanding our operations abroad . . . **If we fail to deploy, manage, or oversee our international operations successfully, our business may suffer. In addition, we are subject to a variety of risks inherent in doing business internationally, including . . . enhanced difficulty in reviewing content on our platform and enforcing our community standards across different languages and countries** . . . geopolitical events affecting us, our marketers or our industry, including trade disputes and pandemics.”⁵*

6. In addition, Facebook recognized in internal records earlier this year (2021):

*“We frequently observe **highly-coordinated, intentional activity** on the FOAS [n.b. Family of Apps and Services] by **problematic actors**, including states, foreign actors, and actors with a record of criminal, violent or hateful behaviour, **aimed at promoting social violence, promoting hate, exacerbating ethnic and other societal cleavages**, and/or delegitimizing*

²https://www.ftc.gov/system/files/documents/cases/ecf_75-1_ftc_v_facebook_public_redacted_fac.pdf.

³<https://sec.report/Document/0001326801-21-000014/>.

⁴ <https://sec.report/Document/0001326801-21-000049/>

⁵<https://sec.report/Document/0001326801-21-000014/>.

social institutions through misinformation. This is particularly prevalent--and problematic---in At Risk Countries and Contexts."⁶

7. **Whistleblowers and News Articles Have Highlighted Facebook's Involvement in Violence and Genocide as well as Global Political Misinformation. As just one example, in Myanmar,** the New York Times⁷ and BBC News,⁸ among others, reported that military used the platform to promote fake news and other content facilitating what is now recognized as a **genocide**. For instance, BBC News reported back in 2018 and also earlier this year in 2021:

*"Facebook . . . agrees with a report that found it had failed to prevent its platform being used to 'incite offline violence' in Myanmar. The independent report, commissioned by Facebook, said the platform had created an 'enabling environment' for the proliferation of human rights abuse."*⁹

*"It is widely said that, in Myanmar, Facebook is the internet . . . [and] UN human rights investigators have . . . concluded that hate speech on Facebook played a key role in fomenting violence in Myanmar. . . Facebook was complicit in a genocide. There were already signs and strong calls for Facebook to handle the incitement of violence on the platform but their inaction really contributed to the fanning of violence in Myanmar. . ."*¹⁰

8. The Myanmar events led not only to Congressional questioning,¹¹ such as by Maria Cantwell in October 2020, but to a Facebook report and further promises, such as:

*"Facebook stands against hate and violence, including in Myanmar, and supports justice for international crimes . . . we have invested heavily in people, technology and partnerships to examine and address the abuse of Facebook in Myanmar, and [an independent] report acknowledges that we are now taking the right corrective actions. . . [e.g.,] use of artificial intelligence . . . If. . . content violates our policies, we will remove it."*¹²

9. **Further, in late 2020, a widely publicized whistleblower--a former Facebook data scientist--spoke out against Facebook regarding global political manipulation** on the platform and was subsequently fired. As reported by BuzzFeed and other news sources, as well as first-hand statements in the enclosed documents, stating:

⁶ [REDACTED] *Coordinated Social Harm*, p. 4.

⁷ <https://www.nytimes.com/2018/10/15/technology/myanmar-facebook-genocide.html>.

⁸ <https://www.bbc.com/news/world-asia-46105934>.

⁹ <https://www.bbc.com/news/world-asia-46105934>.

¹⁰ <https://www.bbc.com/news/world-asia-55929654>.

¹¹ <https://www.rev.com/blog/transcripts/tech-ceos-senate-testimony-transcript-october-28>.

¹² See <https://about.fb.com/news/2018/11/myanmar-hria/>.

"she'd identified dozens of countries, including India, Mexico, Afghanistan, and South Korea, where . . . abuse was enabling politicians to mislead the public and gain power. It also revealed how little [Facebook] had done."¹³

"I've found multiple blatant attempts by foreign national governments to abuse our platform on vast scales."¹⁴

"I personally discovered . . . a massive CIB [Coordinated Inauthentic Behavior] operation that used thousands of inauthentic assets to boost President Juan Orlando Hernandez of Honduras on a massive scale to mislead the Honduran people . . . Yet despite the blatantly violating nature of this activity, it took me almost a year to take down his operation . . . similar to Honduras [was Azerbaijan] . . . [But] We focus upon harm and priority regions like the United States."¹⁵

FACEBOOK'S MISSTATEMENTS AND OMISSIONS

10. Facebook Made Misstatements Regarding its Commitment to Global Issues, Including in Testimony Before Congress. In particular, in 2018, before the Energy and Commerce Committee,¹⁶ Mrs. Brooks asked Mark Zuckerberg:

"I want to make sure there is appropriate attention on . . . how you help us in this country keep our country safe from terrorists . . . So I need you to help assure us, as well as the American people, what is Facebook's role, leadership role, in helping us fight terrorism and help us stop the recruitment? Because it is still a grave danger around the world."

11. Mark Zuckerberg responded:

*"Terrorist content and propaganda has no place in our network, and we have developed a number of tools that have now made it so that **99 percent of the ISIS and al-Qaida content that we take down is identified by these systems and taken down before anyone in our system even flags it** for us. So that is an example of removing harmful content . . ."*

12. Mrs. Brooks followed up:

¹³<https://www.technologyreview.com/2021/07/29/1030260/facebook-whistleblower-sophie-zhang-global-political-manipulation/>.

¹⁴<https://www.buzzfeednews.com/article/craigsilverman/facebook-ignore-political-manipulation-whistleblower-memo>.

¹⁵ ██████████ - big post, p.1-2, 4, 8, 19; see also ██████████ - exit post and team debate.

¹⁶<https://energycommerce.house.gov/sites/democrats.energycommerce.house.gov/files/documents/20180411-FC%20Fac ebook%20Transparency%20and%20Use%20of%20Consumer%20Data.pdf>.

"I have heard you say 99 percent . . . but yet, as recently as March 29, ISIS content was discovered on Facebook, which included an execution video . . . on April 9, of Hezbollah content and so forth. And so what is the mechanism that you are using? Is it artificial intelligence? Is it the 20,000 people? . . . this is just within a week."

13. Mark Zuckerberg replied:

***"[I]t is a combination of technology and people.** We have a counterterrorism team at Facebook which is -- . . . Two hundred people -- just focused on counterterrorism. And there are other content reviewers who are reviewing content that gets flagged to them as well. So those are folks who are working specifically on that. **I think we have capacity in 30 languages that we are working on. And, in addition to that, we have a number of AI tools** that we are developing, like the ones that I had mentioned, that can proactively go flag the content."*

14. As another example, in a November 2020 statement¹⁷ before Chairman Graham, Ranking Member Feinstein, and other Members, Mark Zuckerberg testified:

"Our products enable more than 3 billion people around the world to share ideas, offer support, and discuss important issues, including politics, public health, and social issues. We think that community is particularly important now when the COVID19 pandemic has disrupted so many aspects of our daily lives. We believe we have a responsibility to keep people safe on our service and to protect free expression, and we work hard to set and enforce policies that meet this goal."

15. Moreover, in May 2021, Facebook responded to questions following a March 2021 hearing entitled "Disinformation Nation: Social Media's Role in Promoting Extremism and Misinformation."¹⁸ For instance, Honorable Schakowsky asked:

*"Phrases and symbols associated with minority communities are systematically more likely to be flagged as hate speech on social media platforms, regardless of how benign the content is. The Sikh community has noticed the targeted removal of benign posts that reference Sikh religious artifacts, political speech, and community organizing. How will Facebook create channels of recourse that will prevent the structural and systematic suppression of content from minority communities? . . . [and] **What is the role of international policy teams in determining global policies?** How do*

¹⁷<https://www.judiciary.senate.gov/imo/media/doc/Zuckerberg%20Testimony.pdf>.

¹⁸<https://docs.house.gov/meetings/IF/IF16/20210325/111407/HHRG-117-IF16-Wstate-ZuckerbergM-20210325-SD004.pdf>.

*international policies impact the policies of other nations? **What safeguards does Facebook have . . .***

16. Facebook responded:

*“The conversations that happen on Facebook reflect the diversity of a community of **more than two billion people communicating across countries and cultures and in dozens of languages**, posting everything from text to photos and videos. We recognize how important it is for Facebook to be a place where people feel empowered to communicate. While Facebook is unequivocally opposed to hate speech and committed to removing it from our platform, we’re also committed to getting better at addressing content moderation issues, including improving specific policies, our review process, and community reporting. Our policies are enforced both automatically and by means of human review . . .”*

*“Our Community Standards are a living set of guidelines—they **must keep pace with changes happening online and in the world**. The core of our policy development process is a twice-monthly, global meeting where we debate and discuss potential changes to our Community Standards . . . members of our content policy team reach out to internal and external experts, analyze data, conduct research, and study relevant scholarship to inform our policy proposals. **This multi-step effort allows us to account for a range of perspectives and opinions across the globe**, as well as unintended consequences and efforts to thwart our policies . . .”*

17. To emphasize its goals, Facebook represented:¹⁹

*“Our mission is to give people the power to build community and bring the world closer together—a **mission that is inherently global and enhanced by a global scope** . . .”*

18. In the same May 2021 statement of answers for the record,²⁰ Facebook outlined:

*“Moving fast to find and remove dangerous organizations, including terrorist and hate groups, takes significant investment in both people and technology. At Facebook, **we have tripled the size of our teams working in safety and security since 2016 to over 35,000 people**— including teams that review reports of hate speech and content that praises, supports, or represents hate groups. We also have several hundred people who exclusively or primarily focus on countering dangerous organizations as their core responsibility. . .”*

¹⁹<https://docs.house.gov/meetings/IF/IF16/20210325/111407/HHRG-117-IF16-Wstate-ZuckerbergM-20210325-SD004.pdf>.

²⁰<https://docs.house.gov/meetings/IF/IF16/20210325/111407/HHRG-117-IF16-Wstate-ZuckerbergM-20210325-SD004.pdf>.

*“Four years ago, we developed a playbook and a series of automated techniques to detect content related to terrorist organizations such as ISIS, al Qaeda, and their affiliates. **We’ve since expanded these techniques to detect and remove content related to other terrorist and hate groups.**”*

*“**We continue to proactively investigate coordinated inauthentic behavior (“CIB”) campaigns targeting public debate in the US and around the world.** When we find CIB campaigns, we will investigate them, disrupt them, and make our findings and attribution public, as we’ve repeatedly committed to do . . . In August 2019, for example, we removed a network that originated in China and focused on Hong Kong. As another example, in September 2020, we removed a network that originated in China and focused primarily on the Philippines and Southeast Asia more broadly, and also on the United States.”²¹*

19. More recently, in August 2021 in response to the events in Afghanistan, Nathaniel Gleicher, Facebook’s head of security policy, tweeted:²²

“We’re working closely with our counterparts in industry, civil society and government to provide whatever support we can to help protect people . . . Over the past week, our teams have been working around the clock to do everything we can to help keep people safe,”

20. Further, in the recent Notice of Annual Meeting and Proxy Statement,²³ Facebook’s shareholders proposed having a human/civil rights expert on the board, stating:

*“In September 2020, **a Facebook employee reported Facebook ignored global political manipulation from foreign governments** seeking to ‘abuse our platform on vast scales to mislead their own citizenry.’ [see also below]*

*Children’s rights organization Plan International found **online attacks against girls globally** are most prevalent on Facebook.*

*The Christchurch **terrorist attack in New Zealand, livestreamed on Facebook**, led to a global call to limit the spread of extremist content . . .*

In Myanmar, where violence against the Rohingya ‘bears the hallmarks of genocide,’ a Facebook commissioned human rights report showed the company ‘created an enabling environment.’ In Ethiopia, Facebook’s platform amplified ethnic tensions and calls for genocide, inciting

²¹ <https://docs.house.gov/meetings/IF/IF16/20210325/111407/HHRG-117-IF16-Wstate-ZuckerbergM-20210325-SD004.pdf>.

²² <https://twitter.com/ngleicher/status/1428474001630785540>.

²³ <https://www.sec.gov/Archives/edgar/data/1326801/000132680121000022/facebook2021definitiveprox.htm>.

violence. *The United Nations' Human Rights Council called on Facebook to 'conduct periodic reviews of the human rights impact of their activities in Ethiopia' noting 'a clear responsibility to integrate human rights . . .'*"

21. In rejecting that shareholder proposal, Facebook represented:

"We recognize the need to protect and respect both civil and human rights and we have made, and continue to make, significant progress on both of these fronts and fight abuse across our services. We believe that implementing this proposal is unnecessary because of our continued progress in this area and our efforts to fight abuse across our services. . ."

22. Facebook Made Key Misstatements About its Language Capabilities, Including in Securities Filings and Congressional Testimony. For example, in Facebook's recent 10-K²⁴ and 10-Q²⁵ forms, it represented:

*"We have significant international operations and plan to continue expanding our operations abroad where we have more limited operating experience, and this may subject us to increased business, economic, and legal risks that could affect our financial results. We have significant international operations and plan to continue the international expansion of our business operations and the translation of our products. **We currently make Facebook available in more than 100 different languages, and we have offices or data centers in more than 30 different countries.**"*²⁶

23. Likewise, in its May 2021 statement to the Committee on Energy and Commerce,²⁷ Facebook responded to a number of questions regarding concerns about combating information in different languages. For example, Honorable Eshoo asked:

*"Spanish, Cantonese, Mandarin, Tagalog, Vietnamese, French and French Creole, Korean, German, Arabic, and Russian . . . **What are you doing to combat misinformation in these languages?**"*

*"**With regards to Spanish-language misinformation** and Spanish-speaking content moderators, **how does Facebook account for cultural differences of content moderators** that come from diverse cultural experiences (e.g., Spain and South Texas)?"*

²⁴<https://sec.report/Document/0001326801-21-000014/>.

²⁵ <https://sec.report/Document/0001326801-21-000049/>

²⁶<https://sec.report/Document/0001326801-21-000014/>.

²⁷<https://docs.house.gov/meetings/IF/IF16/20210325/111407/HHRG-117-IF16-Wstate-ZuckerbergM-20210325-SD004.pdf>.

“Given the violence in Myanmar, please provide the number of Burmese-speaking content moderators (employees and contractors) for each of Facebook, Instagram, and WhatsApp.”

24. Facebook replied to these questions as follows:

“We work with over 80 independent third-party fact-checkers around the world, covering more than 60 languages, all of whom are certified through the non-partisan International FactChecking Network (IFCN) to help identify and review false news.”²⁸

*“Our machine learning technology is constantly developing as slang, colloquialisms, and other idioms are fed into it to improve accuracy. We conduct Spanish-language content review 24 hours per day at multiple global sites . . . There is quality control as well as management on site . . . We also have leadership, supervisors, trainers, and market specialists on site . . . We have extensive training for our reviewers . . . We have clear rules that are designed to eliminate bias and subjectivity and strict implementation standards and operational guidelines . . .”*²⁹

“We have dedicated significant resources to the safety of our platform in Myanmar. . . This includes people who spend significant time on the ground working with civil society partners who are advocating on a range of human and digital rights issues across Myanmar’s diverse, multi-ethnic society . . . Our goal is always to have the right number of people with the right native language capabilities to ensure incoming reports are reviewed quickly and effectively . . . we are supplementing our hiring with investments in technology and programs. Following the military coup in Myanmar . . . Facebook is adapting to meet these events . . . We are committed to doing our part to keep people safe on our services . . . We will continue to invest extraordinary resources into content moderation, enforcement, and transparency.”³⁰

25. In the same statement, and in response to Honorable Cárdenas, Facebook said:

“Generally, our machine learning is built by gathering what we call training data in local languages. Our classifiers run in a specific language, not on translated content. We also build our classifiers by using local language data . . . our classifiers are constantly being retrained and improved based on real data, which is also helpful as content and language use evolves in the real

²⁸<https://docs.house.gov/meetings/IF/IF16/20210325/111407/HHRG-117-IF16-Wstate-ZuckerbergM-20210325-SD004.pdf>.

²⁹<https://docs.house.gov/meetings/IF/IF16/20210325/111407/HHRG-117-IF16-Wstate-ZuckerbergM-20210325-SD004.pdf>.

³⁰<https://docs.house.gov/meetings/IF/IF16/20210325/111407/HHRG-117-IF16-Wstate-ZuckerbergM-20210325-SD004.pdf>.

*world. We're taking significant steps to fight the spread of misinformation . . . Our machine learning models to find potentially violating COVID-19 and vaccine content are **trained to surface content in 19 languages using a data set of known violating content originally written in each of the 19 languages** . . .*³¹

26. Moreover, Facebook's Public Page on its "Machine Translation Technology" boasts milestones about its language capabilities, stating:

*"Breaking language barriers through machine translation (MT) is one of the most important ways to bring people together, provide authoritative information on COVID-19, and keep them safe from harmful content. Today, we power an average of 20 billion translations every day on Facebook News Feed . . . Typical MT systems . . . [don't] scale effectively on Facebook, where people post content in more than 160 languages across billions of posts. . . We need one multilingual machine translation (MMT) model that can translate any language to better serve **our community, nearly two-thirds of which use a language other than English.***

*"[W]e're excited to announce a major milestone: the first single massively MMT model that **can directly translate 100x100 languages in any direction** without relying on only English-centric data . . . we built the first truly "many-to-many" data set with 7.5 billion sentences for 100 languages."*³²

27. This follows a critique published by Reuters back in 2019 that stated:

*"At least **652 million people worldwide** speak languages supported by Facebook but **where rules are not translated**, according to data from language encyclopedia Ethnologue. **Another 230 million or more speak one of the 31 languages that do not have official support.**"*³³

28. Facebook Made Misstatements Regarding its Efforts to Combat Harmful Content and Misinformation in International Countries. For example, in the recent Notice of Annual Meeting and Proxy Statement,³⁴ shareholders outlined:

"The Facebook brand has been diminished in recent years due to the platform's use as a tool for gross disinformation, hate speech, and to incite racial violence. What was envisioned as a tool to connect people has been

³¹<https://docs.house.gov/meetings/IF/IF16/20210325/111407/HHRG-117-IF16-Wstate-ZuckerbergM-20210325-SD004.pdf>.

³² <https://about.fb.com/news/2020/10/first-multilingual-machine-translation-model/>.

³³<https://www.reuters.com/article/us-facebook-languages-insight/facebooks-flood-of-languages-leaves-it-struggling-to-monitor-content-idUSKCN1RZ0DW>.

³⁴ <https://www.sec.gov/Archives/edgar/data/1326801/000132680121000022/facebook2021definitiveprox.htm>.

co-opted for dissemination of disinformation and violent extremism, which has **led to many instances of human suffering and death**. Management and the board have failed to take effective action to stem these abuses, which has resulted in a series of negative impacts including:

- **Posts by the Myanmar (Burmese) military junta that incited genocide;**
- Cambridge Analytica's misappropriation and abuse of millions of Facebook users' data;
- Russian hackers influencing the outcome of the 2016 U.S. Presidential election;
- Over 45 million images of child pornography and torture made public;
- A proliferation of political advertisements that contain deliberate lies and disinformation;
- **Hate speech linked to anti-immigrant violence;**
- **Libyan Facebook users buying arms, locating foes, and killing them**

29. As such, the shareholders requested a report to assess benefits etc. of:

"the type of enhanced actions put in place during the 2020 election cycle to reduce the platform's amplification of false and divisive information."

30. Facebook rejected the shareholders' proposal, stating:

"We agree that the amplification of false, divisive, hateful, and inciting content is harmful to our community, and we continue to take steps to address this issue . . . We believe that implementing this proposal is unnecessary due to our transparency efforts to date, the significant progress we continue to make to address these issues . . ."

ORIGINAL EVIDENCE CONFIRMING STATEMENTS WERE FALSE

31. **FIRST, Facebook Lacks Adequate Resources for International Issues.**

32. For example, documents show that Facebook prioritizes certain countries, stating:

2020 "Top 3 Policy Priorities"

"1. All in for Tier 1 elections of 2020 with primary focus on Brazil municipal elections and US presidential elections and census.

2. Expand proactive election work to smaller countries (Tier 2) . . ."

"Tier 0" includes Brazil, India, United States

"Tier 1" includes Germany, Indonesia, Iran, Israel, Italy

"Tier 2" includes Azerbaijan, Burundi, Cameroon, ,Chad, Chile, Cote d'Ivoire, Egypt, Ethiopia, France, Hong Kong, Ireland, Kazakhstan, Kuwait, Kyrgyzstan, Mali, Niger, Singapore, Somalia, Sri Lanka, Syria, Togo, and Venezuela

"Tier 3" includes "All others"

Only "Tier 1" has selections for "Investment" in all of the outlined areas:

Proactive technical enforcement; EOC staffing; Regular proactive manual fan-outs; Dashboard and alerting during electoral period; Review on escalations; Ongoing alerting outside electoral period.³⁵

33. In fact, a 2020 FY summary showed:

*"Misinformation Summary" with an "Objective" to "Remove, reduce, inform/measure misinformation on FB Apps, Includes Community Review and Matching" had **Global Remit [n.b. budget] "US - 87%, ROW [Rest of World] (India, France, Italy) - 13%."**³⁶ (This is despite the US and Canada comprising only about 10% of "Daily Active Users" as outlined above.)*

34. As just one example of Facebook's lack of adequate resources in Iraq:

*"Arabic Hatespeech detection . . . [is] 6% only for IG [Instagram]. . . **IG does not detect hate speech for Arabic effectively** . . . Iraqi representation was close to non-existent . . . Iraq is a proxy for cyber armies working on reporting content . . . examples of retroactively reported hate speech comments on IG [Instagram] . . . **Iraq ranks second when comparing countries' reporting volumes on Instagram!** [approximately 10 million] . . . according to . . . recent research **we have close to no reviewers who understand the Iraqi dialect in our reviewers.**"³⁷*

35. SECOND, Documents Show that Facebook's Language Capabilities are Inadequate, Leading to Global Misinformation and Ethnic Violence.

36. For example, as explained in a January 2021 study to address gaps in the "Hate Speech space" in the Afghanistan market, Facebook documentation outlines:

"[N]either the questions that directs a user to the Community Standards page nor the Community Standards are translated into local languages. This is particularly concerning against the backdrop of low literacy rates, low general awareness regarding what constitutes Hate Speech and low specific

³⁵ [REDACTED] Civic Summit Q1 2020, p. 13, 15-16; see also [REDACTED] At Risk Country Prioritization.

³⁶ [REDACTED] RTB Opex Review, p. 37.

³⁷ [REDACTED] MENA Integrity, p. 3, 13-15.

awareness regarding what Facebook's policies about Hate Speech are . . . only 0.2 per cent [of reported Hate Speech] is taken down by automation. While Hate Speech is consistently ranked as one of the top abuse categories **in the Afghanistan market, the action rate for Hate Speech is worryingly low at 0.23 per cent** . . . sub-standard Hate Speech reporting process in local languages, low levels of literacy, low level of awareness regarding Hate Speech, and absence of adequate and precise classifiers . . . only two per cent [of Hate Speech is] caught proactively."³⁸

37. In particular, Facebook's written translations (in limited languages) do not account for regions where significant users cannot read. For example:

"About two-thirds of participants thought each Voice concept would make their overall FB experience better . . . and be particularly useful in opening up FB for those with disabilities, or those who could not read or write. . . Cambodia appears to make up almost 50% of the TOTAL voice clip traffic . . . **[in Cambodia] 30% of daily active messenger users use voice.**"³⁹

38. Nor do they appropriately manage different dialects:

"Arabic is not one language, truly, rather it is better to consider it a family of languages - many of which are mutually incomprehensible . . . [in other dialects] they will still misunderstand cultural or contextual content, which is key to problem areas such as Hate Speech and even Terrorism. . . with the size of the Arabic user base and potential severity of offline harm in almost every Arabic country--as every Arabic nation save Western Sahara is on the At-Risk Countries list and deals with such severe issues as terrorism and sex trafficking--it is surely of the highest importance to put more resources to the task of improving Arabic systems."⁴⁰

39. Internal records show how this problem leads to violence and inciting content:

"RSS [Indian nationalist organization Rashtriya Swayamsevak Sangh] Users, Groups, and Pages promote fear-mongering, anti-Muslim narratives targeted pro-Hindu populations with V&I [violence and inciting] intent. . . There were a number of dehumanizing posts comparing Muslims to 'pigs' and 'dogs' and misinformation claiming the Quran calls for men to rape their female family

³⁸ [REDACTED] Afghanistan Hate Speech, p. 2, 8-10. See also Disclosure re. Hate Speech.

³⁹ [REDACTED] Cambodia Examples, p. 1, 6-7.

⁴⁰ [REDACTED] Proposal to fix arabic language work, p. 6.

members. **Our lack of Hindi and Bengali classifiers means much of this content is never flagged or actioned**, and we have yet to put forth a nomination for designation of this group given political sensitivities.”⁴¹

“Examples of high-risk misinformation shared by politicians [includes] India VPs [Value Per Visitor]: 734,634 Societal Impact: Out-of-context video stirring up anti-Pakistan and anti-Muslim sentiment.”⁴²

“found an instance in Sri Lanka where people were able to (seemingly) **auto-add hundreds of thousands of people to a group, likely to spread violence-inducing & hateful content** in the midst of the crisis there.”⁴³

“Are we considering flagging Benjamin Netanyahu’s [Israeli politician] posts for violence similarly to how twitter flagged Trump’s posts around ‘When the looting starts the shooting starts’? He’s . . . miscategorizing Palestinian civilians as terrorists, he even uses the term ‘disinfect’ when speaking about this population. . .” [Another individual commented in response] “You are reading a faulty translation made by Facebook’s engine . . . The word vandilists was wrongly translated as infections - not sure why, it’s not even a similar word in Hebrew.”⁴⁴

40. **THIRD, Documents Confirm That Facebook’s Actions and Choices Facilitated Harmful Content and Misinformation Around the World.**

41. For instance, Facebook’s algorithms determine what content is displayed to users:

“We have evidence from a variety of sources that hate speech, divisive political speech, and misinformation on Facebook and the family of apps are affecting societies around the world. We also have compelling evidence that our core product mechanics, such as virality, recommendations, and optimizing for engagement, are a significant part of why these types of speech flourish on the platform . . . the net result is that Facebook, taken as a whole, will be actively (if not necessarily consciously) promoting these types of activities. The mechanics of our platform are not neutral.”⁴⁵

⁴¹ [REDACTED] Adversarial Harmful Networks - India Case study, p. 5.

⁴² [REDACTED] Effects of Politician Shared Misinformation, p. 5.

⁴³ [REDACTED] Sri Lanka - Giant Groups Ballooning, p. 1.

⁴⁴ [REDACTED] Request to label netanyahu’s comments, p. 1-2.

⁴⁵ [REDACTED] What is Collateral damage? p. 35.

"Facebook's decision-making on content policy is routinely influenced by political considerations."⁴⁶

"Harmful communities form echo-chambers that normalize harmful attitudes and behaviors, polarize communities in a more extreme direction, and radicalize individual members towards the justification of harmful acts or violence."⁴⁷

"Social media allows misinformation to gain 'unobstructed virality.'"⁴⁸

42. And Facebook's algorithms recommend content and "groups" to individuals:

"40% of Sampled Top VPV [View Port Views] Civic Posters in West Bengal Were Fake/Inauthentic. *The highest-VPV user to be assessed as inauthentic had more than 30M accrued in the L28 . . . Coordinated authentic actors seed and spread civic content to propagate political narratives . . . The message comes to dominate the ecosystem with **over 35% of members having been recommended a cell group by our algorithms.**"⁴⁹*

43. In particular, there is a known issue with "deep reshares," which is explained as:

"reshare depth . . . it's the number of 'hops from the original Facebook post in the reshare chain."⁵⁰

44. Records confirm "deep reshares" further misinformation and violence:

"70% of VPVs [View Port Views, i.e. impressions] on known link misinfo come from reshares . . . 50% to 80% of VPVs on known link misinfo come from the top 1K most seen links per day . . . **Reshare depth appears to be an especially good signal for targeting link misinfo in India and the Philippines.** . . . misinfo prevalence is 15-25X higher among VPVs of depth 2+ reshares . . . "⁵¹

"1-1.5 million predicted misinfo VPVs [View Port Views] per hour in India, Indonesia, and the Philippines at peak hours."⁵²

⁴⁶ [REDACTED] Political Influence Content Policy, p. 1.

⁴⁷ [REDACTED] Harmful Topic Communities, p. 3.

⁴⁸ [REDACTED] Harmful Conspiracies, p. 13; [REDACTED] Lotus Mahal, p. 1; see also [REDACTED] Adversarial Harmful Networks - India Case study.

⁴⁹ [REDACTED] Adversarial Harmful Networks - India Case study, p. 3-4.

⁵⁰ [REDACTED] Reshare ranking exp in India: Indonesia, p. 1, 4; see also [REDACTED] Reshare depth india and indonesia.

⁵¹ [REDACTED] Further reading from Reshare Depth Article, p. 2.

⁵² [REDACTED] Reshare ranking exp in India: Indonesia, p. 1, 4; see also [REDACTED] Reshare depth india and indonesia.

45. Moreover, documents outline how “demoting” these “deep reshares” would mitigate harmful content, as it did in Myanmar until this feature was rolled back:

“An effective, content-agnostic approach to mitigate the harms posted by high-harm misinfo (e.g. civic or health) would be to dampen virality within these topics by hard demoting all deep reshares where the viewer is not a friend or follower of the original poster.”⁵³

*“[In Myanmar] Misinformation, misrepresentation and account issues are a few known high-risk abuse areas that will be exacerbated by a second COVID outbreak . . . The Reshare Depth demotion reduces the distribution of highly-viral content to give more distribution to content produced by friends or connections of friends . . . [although measures enacted to demote reshare depth to reduce the distribution of highly-viral content in fact **reduced Viral Inflammatory Prevalence by 25.1% and Photo Misinformation by 48.5% in Myanmar. Facebook said:] We plan to roll back this intervention after the Myanmar election** in November.”⁵⁴*

46. Regarding another lack of implementation of features in Myanmar as of 2020:

*“**Hate Speech Classifier[s] for Myanmar/Burmese** . . . hate speech text classifier . . . currently being used in production / being maintained? . . . **it doesn’t look like it’s currently in use?**”⁵⁵*

47. Further, records show multiple examples⁵⁶ of harmful content in other countries:

*“**Coordinated Social Harm**” (i.e., coordinated **attempts to mass-spread harmful narratives**, including violating content that is likely to lead to online and offline severe social harm) through a “Diaspora cluster spreading incitement to violence & hate speech to **Ethiopian audiences** [] Mostly based in **Egypt**” and in a “**Sudan-based** cluster of . . . accounts [] Connections with the **Fano armed militia**.”*

*“**Mass Reporting**” (i.e., misuse of reporting systems to harass others) of “activists, journalists, and HRDs in **Vietnam** . . . Good success % in suppressing the target FB presence . . . **Activity commissioned and directed by government/military entities.**”*

⁵³ [REDACTED] Fighting high harm misinfo with deep reshare damping, p. 1; see also [REDACTED] Reshare depth by country.

⁵⁴ [REDACTED] [I’m redoing] Reshare depth exp in Myanmar, p. 1.

⁵⁵ [REDACTED] Burmese hate Speech Classifier.pdf, p. 1.

⁵⁶ [REDACTED] Adversarial Harmful Networks, p. 11, 14, 19, 22-27.

“Brigading” (i.e., mass harassment and coordinated intimidation operations that can manifest in real-world harm) through “Coordinated harassment of critics of the Egyptian government . . .”

48. **Role for the SEC.** The SEC is charged with enforcing the laws that protect investors in public companies like Facebook. Facebook’s investors care about misrepresentations and omissions by Mark Zuckerberg and other Facebook executives on the topic of lack of global support for two reasons. First, to the extent that users become aware of the dangers that Facebook platforms present, they are likely to use the platforms less, leading to lower advertising revenue and lower profits. Second, some investors simply will not want to invest in a company that facilitates harmful content around the world and then engages in misstatements and omissions on the topic.
-
49. Whistleblower Aid is a non-profit legal organization that helps workers report their concerns about violations of the law safely, lawfully, and responsibly. We respectfully request the SEC’s assistance ensuring that our client never faces retaliation.
50. On information and belief, none of the documents enclosed here constitute attorney-client communications, were obtained during a meeting with an attorney, or otherwise indicate that they are in any way privileged.
51. We plan to continue supplementing this disclosure with additional information and evidence. Our client would be happy to meet with investigators at your convenience. Please feel free to contact us using the information below.
52. We are representing an anonymous whistleblower who is making the above disclosures solely for reporting the suspected violation of laws as outlined.

Sincerely,



John N. Tye, Attorney at Law
Chief Disclosure Officer



--REDACTED FOR CONGRESS--

Report government and corporate lawbreaking.
Without breaking the law.



Andrew Bakaj, Attorney at Law
Of Counsel

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

Enclosures:

Internal Facebook documents including —

[REDACTED] Coordinated Social Harm
[REDACTED] - big post
[REDACTED] - exit post and team debate
Civic Summit Q1 2020
At Risk Country Prioritization
RTB Opex Review
MENA Integrity
Afghanistan Hate Speech
Cambodia Examples
Proposal to fix arabic language work
Adversarial Harmful Networks - India Case study
Effects of Politician Shared Misinformation
Sri Lanka - Giant Groups Ballooning
Request to label netanyahu's comments
What is Collateral damage?
Political Influence Content Policy
Harmful Topic Communities

--REDACTED FOR CONGRESS--

*Report government and corporate lawbreaking.
Without breaking the law.*

	Harmful Conspiracies
	Lotus Mahal
	Reshare ranking exp in India: Indonesia
	Reshare depth india and indonesia.
	Further reading from Reshare Depth Article
	Reshare depth india and indonesia
	Fighting high harm misinfo with deep reshare damping
	Reshare depth by country
	Reshare depth exp in Myanmar
	Burmese hate Speech Classifier.pdf, Adversarial Harmful Networks