Facebook Lied – It's Reading Your Private WhatsApp Messages

By Peter Elkind, Jack Gillum and Craig Silverman | ProPublica | The Defender

When Mark Zuckerberg unveiled a new "privacy-focused vision" for Facebook in March 2019, he cited the company's global messaging service, WhatsApp, as a model.

Acknowledging that "we don't currently have a strong reputation for building privacy-protective services," the Facebook CEO wrote that "I believe the future of communication will increasingly shift to private, encrypted services where people can be confident what they say to each other stays secure and their messages and content won't stick around forever. This is the future I hope we will help bring about. We plan to build this the way we've developed WhatsApp."

Zuckerberg's vision centered on WhatsApp's signature feature,

which he said the company was planning to apply to Instagram and Facebook Messenger: end-to-end encryption, which converts all messages into an unreadable format that is only unlocked when they reach their intended destinations.

WhatsApp messages are so secure, he said, that nobody else — not even the company — can read a word. As Zuckerberg had put it earlier, in testimony to the U.S. Senate in 2018, "We don't see any of the content in WhatsApp."

WhatsApp emphasizes this point so consistently that a flag with a similar assurance automatically appears on-screen before users send messages: "No one outside of this chat, not even WhatsApp, can read or listen to them."

Those assurances are not true. WhatsApp has more than 1,000 contract workers filling floors of office buildings in Austin, Texas, Dublin, and Singapore, where they examine millions of pieces of users' content. Seated at computers in pods organized by work assignments, these hourly workers use special Facebook software to sift through streams of private messages, images, and videos that have been reported by WhatsApp users as improper and then screened by the company's artificial intelligence systems.

These contractors pass judgment on whatever flashes on their screen — claims of everything from fraud or spam to child porn and potential terrorist plotting — typically in less than a minute.

Policing users while assuring them that their privacy is sacrosanct makes for an awkward mission at WhatsApp. A 49-slide internal company marketing presentation from December, obtained by ProPublica, emphasizes the "fierce" promotion of WhatsApp's "privacy narrative."

It compares its "brand character" to "the Immigrant Mother" and displays a photo of Malala □□Yousafzai, who survived a shooting by the Taliban and became a Nobel Peace Prize winner,

in a slide titled “Brand tone parameters.” The presentation does not mention the company’s content moderation efforts.

WhatsApp’s director of communications, Carl Woog, acknowledged that teams of contractors in Austin and elsewhere review WhatsApp messages to identify and remove “the worst” abusers. But Woog told ProPublica that the company does not consider this work to be content moderation, saying: “We actually don’t typically use the term for WhatsApp.” The company declined to make executives available for interviews for this article but responded to questions with written comments.

“WhatsApp is a lifeline for millions of people around the world,” the company said. “The decisions we make around how we build our app are focused around the privacy of our users, maintaining a high degree of reliability and preventing abuse.”

WhatsApp’s denial that it moderates content is noticeably different from what Facebook Inc. says about WhatsApp’s corporate siblings, Instagram and Facebook. The company has said that some 15,000 moderators examine content on Facebook and Instagram, neither of which is encrypted. It releases quarterly transparency reports that detail how many accounts Facebook and Instagram have “actioned” for various categories of abusive content. There is no such report for WhatsApp.

Deploying an army of content reviewers is just one of the ways that Facebook Inc. has compromised the privacy of WhatsApp users. Together, the company’s actions have left WhatsApp — the largest messaging app in the world, with two billion users — far less private than its users likely understand or expect.

A ProPublica investigation, drawing on data, documents, and dozens of interviews with current and former employees and contractors, reveals how, since purchasing WhatsApp in 2014, Facebook has quietly undermined its sweeping security assurances in multiple ways. (Two articles this summer noted

the existence of WhatsApp's moderators but focused on their working conditions and pay rather than their effect on users' privacy. This article is the first to reveal the details and extent of the company's ability to scrutinize messages and user data — and to examine what the company does with that information.)

Many of the assertions by content moderators working for WhatsApp are echoed by a confidential whistleblower complaint filed last year with the U.S. Securities and Exchange Commission. The complaint, which ProPublica obtained, details WhatsApp's extensive use of outside contractors, artificial intelligence systems, and account information to examine user messages, images, and videos. It alleges that the company's claims of protecting users' privacy are false. "We haven't seen this complaint," the company spokesperson said. The SEC has taken no public action on it; an agency spokesperson declined to comment.

Facebook Inc. has also downplayed how much data it collects from WhatsApp users, what it does with it and how much it shares with law enforcement authorities. For example, WhatsApp shares metadata, unencrypted records that can reveal a lot about a user's activity, with law enforcement agencies such as the Department of Justice.

Some rivals, such as Signal, intentionally gather much less metadata to avoid incursions on its users' privacy and thus share far less with law enforcement. ("WhatsApp responds to valid legal requests," the company spokesperson said, "including orders that require us to provide on a real-time going forward basis who a specific person is messaging.")

WhatsApp user data, ProPublica has learned, helped prosecutors build a high-profile case against a Treasury Department employee who leaked confidential documents to BuzzFeed News that exposed how dirty money flows through U.S. banks.

Like other social media and communications platforms, WhatsApp is caught between users who expect privacy and law enforcement entities that effectively demand the opposite: that WhatsApp turns over information that will help combat crime and online abuse.

WhatsApp has responded to this dilemma by asserting that it's no dilemma at all. "I think we absolutely can have security and safety for people through end-to-end encryption and work with law enforcement to solve crimes," said Will Cathcart, whose title is Head of WhatsApp, in a YouTube interview with an Australian think tank in July.

The tension between privacy and disseminating information to law enforcement is exacerbated by a second pressure: Facebook's need to make money from WhatsApp. Since paying $22 billion to buy WhatsApp in 2014, Facebook has been trying to figure out how to generate profits from a service that doesn't charge its users a penny.

That conundrum has periodically led to moves that anger users, regulators, or both. The goal of monetizing the app was part of the company's 2016 decision to start sharing WhatsApp user data with Facebook, something the company had told EU regulators was technologically impossible.

The same impulse spurred a controversial plan, abandoned in late 2019, to sell advertising on WhatsApp. And the profit-seeking mandate was behind another botched initiative in January: the introduction of a new privacy policy for user interactions with businesses on WhatsApp, allowing businesses to use customer data in new ways. That announcement triggered a user exodus to competing apps.

WhatsApp's increasingly aggressive business plan is focused on charging companies for an array of services — letting users make payments via WhatsApp and managing customer service chats — that offer convenience but fewer privacy protections. The

result is a confusing two-tiered privacy system within the same app where the protections of end-to-end encryption are further eroded when WhatsApp users employ the service to communicate with businesses.

The company's December marketing presentation captures WhatsApp's diverging imperatives. It states that "privacy will remain important." But it also conveys what seems to be a more urgent mission: the need to "open the aperture of the brand to encompass our future business objectives."

I. "Content moderation associates"

In many ways, the experience of being a content moderator for WhatsApp in Austin is identical to being a moderator for Facebook or Instagram, according to interviews with 29 current and former moderators. Mostly in their 20s and 30s, many with past experience as store clerks, grocery checkers and baristas, the moderators are hired and employed by Accenture, a huge corporate contractor that works for Facebook and other Fortune 500 behemoths.

The job listings advertise "Content Review" positions and make no mention of Facebook or WhatsApp. Employment documents list the workers' initial title as "content moderation associate." Pay starts at around $16.50 an hour. Moderators are instructed to tell anyone who asks that they work for Accenture, and are required to sign sweeping non-disclosure agreements.

Citing the NDAs, almost all the current and former moderators interviewed by ProPublica insisted on anonymity. (An Accenture spokesperson declined to comment, referring all questions about content moderation to WhatsApp.)

When the WhatsApp team was assembled in Austin in 2019, Facebook moderators already occupied the fourth floor of an office tower on Sixth Street, adjacent to the city's famous bar-and-music scene. The WhatsApp team was installed on the floor above, with new glass-enclosed work pods and nicer

bathrooms that sparked a tinge of envy in a few members of the Facebook team.

Most of the WhatsApp team scattered to work from home during the pandemic. Whether in the office or at home, they spend their days in front of screens, using a Facebook software tool to examine a stream of “tickets,” organized by subject into “reactive” and “proactive” queues.

Collectively, the workers scrutinize millions of pieces of WhatsApp content each week. Each reviewer handles upwards of 600 tickets a day, which gives them less than a minute per ticket. WhatsApp declined to reveal how many contract workers are employed for content review, but a partial staffing list reviewed by ProPublica suggests that, at Accenture alone, it’s more than 1,000. WhatsApp moderators, like their Facebook and Instagram counterparts, are expected to meet performance metrics for speed and accuracy, which are audited by Accenture.

Their jobs differ in other ways. Because WhatsApp’s content is encrypted, artificial intelligence systems can’t automatically scan all chats, images, and videos, as they do on Facebook and Instagram. Instead, WhatsApp reviewers gain access to private content when users hit the “report” button on the app, identifying a message as allegedly violating the platform’s terms of service.

This forwards five messages — the allegedly offending one along with the four previous ones in the exchange, including any images or videos — to WhatsApp in unscrambled form, according to former WhatsApp engineers and moderators. Automated systems then feed these tickets into “reactive” queues for contract workers to assess.

Artificial intelligence initiates the second set of queues — so-called proactive ones — by scanning unencrypted data that WhatsApp collects about its users and comparing it against

suspicious account information and messaging patterns (a new account rapidly sending out a high volume of chats is evidence of spam), as well as terms and images that have previously been deemed abusive.

The unencrypted data available for scrutiny is extensive. It includes the names and profiles images of a user's WhatsApp groups as well as their phone number, profile photo, status message, phone battery level, language and time zone, unique mobile phone ID and IP address, wireless signal strength, and phone operating system, as a list of their electronic devices, any related Facebook and Instagram accounts, the last time they used the app and any previous history of violations.

The WhatsApp reviewers have three choices when presented with a ticket for either type of queue: Do nothing, place the user on "watch" for further scrutiny, or ban the account. (Facebook and Instagram content moderators have more options, including removing individual postings. It's that distinction – the fact that WhatsApp reviewers can't delete individual items – that the company cites as its basis for asserting that WhatsApp reviewers are not "content moderators.")

WhatsApp moderators must make subjective, sensitive, and subtle judgments, interviews, and documents examined by ProPublica show. They examine a wide range of categories, including "Spam Report", "Civic Bad Actor" (political hate speech and disinformation), "Terrorism Global Credible Threat", "CEI" (child exploitative imagery), and "CP" (child pornography).

Another set of categories addresses the messaging and conduct of millions of small and large businesses that use WhatsApp to chat with customers and sell their wares. These queues have such titles as "business impersonation prevalence," "commerce policy probable violators" and "business verification."

Moderators say the guidance they get from WhatsApp and

Accenture relies on standards that can be simultaneously arcane and disturbingly graphic. Decisions about abusive sexual imagery, for example, can rest on an assessment of whether a naked child in an image appears adolescent or prepubescent, based on a comparison of hip bones and pubic hair to a medical index chart.

One reviewer recalled a grainy video in a political-speech queue that depicted a machete-wielding man holding up what appeared to be a severed head: “We had to watch and say, ‘Is this a real dead body or a fake dead body?’”

In late 2020, moderators were informed of a new queue for alleged “sextortion.” It was defined in an explanatory memo as “a form of sexual exploitation where people are blackmailed with a nude image of themselves which have been shared by them or someone else on the Internet.” The memo said workers would review messages reported by users that “include predefined keywords typically used in sextortion/blackmail messages.”

WhatsApp’s review system is hampered by impediments, including buggy language translation. The service has users in 180 countries, with the vast majority located outside the U.S. Even though Accenture hires workers who speak a variety of languages, for messages in some languages there’s often no native speaker on-site to assess abuse complaints.

That means using Facebook’s language-translation tool, which reviewers said could be so inaccurate that it sometimes labeled messages in Arabic as being in Spanish. The tool also offered little guidance on local slang, political context, or sexual innuendo. “In the three years I’ve been there,” one moderator said, “it’s always been horrible.”

The process can be rife with errors and misunderstandings. Companies have been flagged for offering weapons for sale when they’re selling straight shaving razors. Bras can be sold, but if the marketing language registers as “adult,” the seller can

be labeled a forbidden “sexually oriented business.” And a flawed translation toolset off an alarm when it detected kids for sale and slaughter, which, upon closer scrutiny, turned out to involve young goats intended to be cooked and eaten in halal meals.

The system is also undercut by the human failings of the people who instigate reports. Complaints are frequently filed to punish, harass or prank someone, according to moderators. In messages from Brazil and Mexico, one moderator explained, “we had a couple of months where AI was banning groups left and right because people were messing with their friends by changing their group names” and then reporting them. “At the worst of it, we were probably getting tens of thousands of those. They figured out some words the algorithm did not like.”

Other reports fail to meet WhatsApp standards for an account ban. “Most of it is not violating,” one of the moderators said. “It’s content that is already on the internet, and it’s just people trying to mess with users.” Still, each case can reveal up to five unencrypted messages, which are then examined by moderators.

The judgment of WhatsApp’s AI is less than perfect, moderators say. “There were a lot of innocent photos on there that were not allowed to be on there,” said Carlos Sauceda, who left Accenture last year after nine months. “It might have been a photo of a child taking a bath, and there was nothing wrong with it.” As another WhatsApp moderator put it, “A lot of the time, the artificial intelligence is not that intelligent.”

Facebook’s written guidance to WhatsApp moderators acknowledges many problems, noting “we have made mistakes and our policies have been weaponized by bad actors to get good actors banned. When users write inquiries pertaining to abusive matters like these, it is up to WhatsApp to respond and act (if necessary) accordingly in a timely and pleasant

manner." Of course, if a user appeals a ban that was prompted by a user report, according to one moderator, it entails having a second moderator examine the user's content.

II. "Industry leaders" in detecting bad behavior

In public statements and on the company's websites, Facebook Inc. is noticeably vague about WhatsApp's monitoring process. The company does not provide a regular accounting of how WhatsApp polices the platform. WhatsApp's FAQ page and online complaint form note that it will receive "the most recent messages" from a user who has been flagged.

They do not, however, disclose how many unencrypted messages are revealed when a report is filed, or that those messages are examined by outside contractors. (WhatsApp told ProPublica it limits that disclosure to keep violators from "gaming" the system.)

By contrast, both Facebook and Instagram post lengthy "Community Standards" documents detailing the criteria its moderators use to police content, along with articles and videos about "the unrecognized heroes who keep Facebook safe" and announcements on new content-review sites. Facebook's transparency reports detail how many pieces of content are "actioned" for each type of violation. WhatsApp is not included in this report.

When dealing with legislators, Facebook Inc. officials also offer few details – but are eager to assure them that they don't let encryption stand in the way of protecting users from images of child sexual abuse and exploitation. For example, when members of the Senate Judiciary Committee grilled Facebook about the impact of encrypting its platforms, the company, in written follow-up questions in January 2020, cited WhatsApp in boasting that it would remain responsive to law enforcement.

"Even within an encrypted system," one respondent noted, "we

will still be able to respond to lawful requests for metadata, including the potentially critical location or account information… We already have an encrypted messaging service, WhatsApp, that – in contrast to some other encrypted services – provides a simple way for people to report abuse or safety concerns."

Sure enough, WhatsApp reported 400,000 instances of possible child-exploitation imagery to the National Center for Missing and Exploited Children in 2020, according to its head, Cathcart. That was ten times as many as in 2019. "We are by far the industry leaders in finding and detecting that behavior in an end-to-end encrypted service," he said.

During his YouTube interview with the Australian think tank, Cathcart also described WhatsApp's reliance on user reporting and its AI systems' ability to examine account information that isn't subject to encryption. Asked how many staffers WhatsApp employed to investigate abuse complaints from an app with more than two billion users, Cathcart didn't mention content moderators or their access to encrypted content.

"There's a lot of people across Facebook who help with WhatsApp," he explained. "If you look at people who work full time on WhatsApp, it's above a thousand. I won't get into the full breakdown of customer service, user reports, engineering, etc. But it's a lot of that."

In written responses for this article, the company spokesperson said: "We build WhatsApp in a manner that limits the data we collect while providing us tools to prevent spam, investigate threats, and ban those engaged in abuse, including based on user reports we receive. This work takes extraordinary effort from security experts and a valued trust and safety team that works tirelessly to help provide the world with private communication."

The spokesperson noted that WhatsApp has released new privacy

features, including “more controls about how people’s messages can disappear” or be viewed only once. He added, “Based on the feedback we’ve received from users, we’re confident people understand when they make reports to WhatsApp we receive the content they send us.”

III. “Deceiving users” about personal privacy

Since the moment Facebook announced plans to buy WhatsApp in 2014, observers wondered how the service, known for its fervent commitment to privacy, would fare inside a corporation known for the opposite.

Zuckerberg had become one of the wealthiest people on the planet by using a “surveillance capitalism” approach: collecting and exploiting reams of user data to sell targeted digital ads. Facebook’s relentless pursuit of growth and profits has generated a series of privacy scandals in which it was accused of deceiving customers and regulators.

By contrast, WhatsApp knew little about its users apart from their phone numbers and shared none of that information with third parties. WhatsApp ran no ads, and its co-founders, Jan Koum and Brian Acton, both former Yahoo engineers, were hostile to them.

“At every company that sells ads,” they wrote in 2012, “a significant portion of their engineering team spends their day tuning data mining, writing better code to collect all your personal data, upgrading the servers that hold all the data, and making sure it’s all being logged and collated and sliced and packed and shipped out,” adding: “Remember when advertising is involved you the user are the product.” At WhatsApp, they noted, “your data isn’t even in the picture. We are simply not interested in any of it.”

Zuckerberg publicly vowed in a 2014 keynote speech that he would keep WhatsApp “exactly the same.” He declared, “We are absolutely not going to change plans around WhatsApp and the

way it uses user data. WhatsApp is going to operate completely autonomously."

In April 2016, WhatsApp completed its long-planned adoption of end-to-end encryption, which helped establish the app as a prized communications platform in 180 countries, including many where text messages and phone calls are cost-prohibitive. International dissidents, whistleblowers, and journalists also turned to WhatsApp to escape government eavesdropping.

Four months later, however, WhatsApp disclosed it would begin sharing user data with Facebook — precisely what Zuckerberg had said would not happen — a move that cleared the way for an array of future revenue-generating plans.

The new WhatsApp terms of service said the app would share information such as users' phone numbers, profile photos, status messages, and IP addresses for the purposes of ad targeting, fighting spam and abuse, and gathering metrics. "By connecting your phone number with Facebook's systems," WhatsApp explained, "Facebook can offer better friend suggestions and show you more relevant ads if you have an account with them."

Such actions were increasingly bringing Facebook into the crosshairs of regulators. In May 2017, EU antitrust regulators fined the company 110 million euros (about $122 million) for falsely claiming three years earlier that it would be impossible to link the user information between WhatsApp and the Facebook family of apps. The EU concluded that Facebook had "intentionally or negligently" deceived regulators. Facebook insisted its false statements in 2014 were not intentional but didn't contest the fine.

By the spring of 2018, the WhatsApp co-founders, now both billionaires, were gone. Acton, in what he later described as an act of "penance" for the "crime" of selling WhatsApp to Facebook, gave $50 million to a foundation backing Signal, a

free encrypted messaging app that would emerge as a WhatsApp rival. (Acton's donor-advised fund has also given money to ProPublica.)

Meanwhile, Facebook was under fire for its security and privacy failures as never before. The pressure culminated in a landmark $5 billion fine by the Federal Trade Commission in July 2019 for violating a previous agreement to protect user privacy. The fine was almost 20 times greater than any previous privacy-related penalty, according to the FTC, and Facebook's transgressions included "deceiving users about their ability to control the privacy of their personal information."

The FTC announced that it was ordering Facebook to take steps to protect privacy going forward, including for WhatsApp users: "As part of Facebook's order-mandated privacy program, which covers WhatsApp and Instagram, Facebook must conduct a privacy review of every new or modified product, service, or practice before it is implemented, and document its decisions about user privacy." Compliance officers would be required to generate a "quarterly privacy review report" and share it with the company and, upon request, the FTC.

Facebook agreed to the FTC's fine and order. Indeed, the negotiations for that agreement were the backdrop, just four months before that, for Zuckerberg's announcement of his new commitment to privacy.

By that point, WhatsApp had begun using Accenture and other outside contractors to hire hundreds of content reviewers. But the company was eager not to step on its larger privacy message — or spook its global user base. It said nothing publicly about its hiring of contractors to review content.

IV. "We kill people based on metadata"

Even as Zuckerberg was touting Facebook Inc.'s new commitment to privacy in 2019, he didn't mention that his company was

apparently sharing more of its WhatsApp users' metadata than ever with the parent company — and with law enforcement.

To the lay ear, the term "metadata" can sound abstract, a word that evokes the intersection of literary criticism and statistics. To use an old, pre-digital analogy, metadata is the equivalent of what's written on the outside of an envelope — the names and addresses of the sender and recipient and the postmark reflecting where and when it was mailed — while the "content" is what's written on the letter sealed inside the envelope. So it is with WhatsApp messages: The content is protected, but the envelope reveals a multitude of telling details (as noted: timestamps, phone numbers, and much more).

Those in the information and intelligence fields understand how crucial this information can be. It was metadata, after all, that the National Security Agency was gathering about millions of Americans not suspected of a crime, prompting a global outcry when it was exposed in 2013 by former NSA contractor Edward Snowden.

"Metadata absolutely tells you everything about somebody's life," former NSA general counsel Stewart Baker once said. "If you have enough metadata, you don't really need content." In a symposium at Johns Hopkins University in 2014, Gen. Michael Hayden, former director of both the CIA and NSA, went even further: "We kill people based on metadata."

U.S. law enforcement has used WhatsApp metadata to help put people in jail. ProPublica found more than a dozen instances in which the Justice Department sought court orders for the platform's metadata since 2017. These represent a fraction of overall requests, known as pen register orders (a phrase borrowed from the technology used to track numbers dialed by landline telephones), as many more are kept from public view by court order.

U.S. government requests for data on outgoing and incoming

messages from all Facebook platforms increased by 276% from the first half of 2017 to the second half of 2020, according to Facebook Inc. statistics (which don't break out the numbers by platform). The company's rate of handing over at least some data in response to such requests has risen from 84% to 95% during that period.

It's not clear exactly what government investigators have been able to gather from WhatsApp, as the results of those orders, too, are often kept from public view. Internally, WhatsApp calls such requests for information about users "prospective message pairs," or PMPs.

These provide data on a user's messaging patterns in response to requests from U.S. law enforcement agencies, as well as those in at least three other countries — the UK, Brazil, and India — according to a person familiar with the matter who shared this information on the condition of anonymity. Law enforcement requests from other countries might only receive basic subscriber profile information.

WhatsApp metadata was pivotal in the arrest and conviction of Natalie "May" Edwards, a former Treasury Department official with the Financial Crimes Enforcement Network, for leaking confidential banking reports about suspicious transactions to BuzzFeed News. The FBI's criminal complaint detailed hundreds of messages between Edwards and a BuzzFeed reporter using an "encrypted application," which interviews and court records confirmed was WhatsApp.

"On or about August 1, 2018, within approximately six hours of the Edwards pen becoming operative — and the day after the July 2018 Buzzfeed article was published — the Edwards cellphone exchanged approximately 70 messages via the encrypted application with the Reporter-1 cellphone during an approximately 20-minute time span between 12:33 a.m. and 12:54 a.m.,"

FBI Special Agent Emily Eckstut wrote in her October 2018 complaint. Edwards and the reporter used WhatsApp because Edwards believed the platform to be secure, according to a person familiar with the matter.

Edwards was sentenced on June 3 to six months in prison after pleading guilty to a conspiracy charge and reported to prison last week. Edwards' attorney declined to comment, as did representatives from the FBI and the Justice Department.

WhatsApp has for years downplayed how much-unencrypted information it shares with law enforcement, largely limiting mentions of the practice to boilerplate language buried deep in its terms of service. It does not routinely keep permanent logs of who users are communicating with and how often, but company officials confirmed they do turn on such tracking at their own discretion – even for internal Facebook leak investigations – or in response to law enforcement requests. The company declined to tell ProPublica how frequently it does so.

The privacy page for WhatsApp assures users that they have total control over their own metadata. It says users can "decide if only contacts, everyone, or nobody can see your profile photo" or when they last opened their status updates or when they last opened the app. Regardless of the settings a user chooses, WhatsApp collects and analyzes all of that data – a fact not mentioned anywhere on the page.

V. "Opening the aperture to encompass business objectives"

The conflict between privacy and security on encrypted platforms seems to be only intensifying. Law enforcement and child safety advocates have urged Zuckerberg to abandon his plan to encrypt all of Facebook's messaging platforms.

In June 2020, three Republican senators introduced the "Lawful Access to Encrypted Data Act," which would require tech companies to assist in providing access to even encrypted

content in response to law enforcement warrants. For its part, WhatsApp recently sued the Indian government to block its requirement that encrypted apps provide "traceability" – a method to identify the sender of any message deemed relevant to law enforcement. WhatsApp has fought similar demands in other countries.

Other encrypted platforms take a vastly different approach to monitoring their users than WhatsApp. Signal employs no content moderators, collects far less user and group data, allows no cloud backups, and generally rejects the notion that it should be policing user activities. It submits no child exploitation reports to NCMEC.

Apple has touted its commitment to privacy as a selling point. It has no "report" button on its iMessage system, and the company has made just a few hundred annual reports to NCMEC, all of them originating from scanning outgoing email, which is unencrypted.

But Apple recently took a new tack and appeared to stumble along the way. Amid intensifying pressure from Congress, in August the company announced a complex new system for identifying child-exploitative imagery on users' iCloud backups.

Apple insisted the new system poses no threat to private content, but privacy advocates accused the company of creating a backdoor that potentially allows authoritarian governments to demand broader content searches, which could result in the targeting of dissidents, journalists, or other critics of the state. On Sept. 3, Apple announced it would delay the implementation of the new system.

Still, it's Facebook that seems to face the most constant skepticism among major tech platforms. It is using encryption to market itself as privacy-friendly while saying little about the other ways it collects data, according to Lloyd

Richardson, the director of IT at the Canadian Centre for Child Protection.

“This whole idea that they’re doing it for personal protection of people is completely ludicrous,” Richardson said. “You’re trusting an app owned and written by Facebook to do exactly what they’re saying. Do you trust that entity to do that?” (On Sept. 2, Irish authorities announced that they are fining WhatsApp 225 million euros, about $267 million, for failing to properly disclose how the company shares user information with other Facebook platforms. WhatsApp is contesting the finding.)

Facebook’s emphasis on promoting WhatsApp as a paragon of privacy is evident in the December marketing document obtained by ProPublica. The “Brand Foundations” presentation says it was the product of a 21-member global team across all of Facebook, involving a half-dozen workshops, quantitative research, “stakeholder interviews” and “endless brainstorms.”

Its aim: to offer “an emotional articulation” of WhatsApp’s benefits, “an inspirational toolkit that helps us tell our story,” and a “brand purpose to champion the deep human connection that leads to progress.” The marketing deck identifies a feeling of “closeness” as WhatsApp’s “ownable emotional territory,” saying the app delivers “the closest thing to an in-person conversation.”

WhatsApp should portray itself as “courageous,” according to another slide because it’s “taking a strong, public stance that is not financially motivated on things we care about,” such as defending encryption and fighting misinformation. But the presentation also speaks of the need to “open the aperture of the brand to encompass our future business objectives. While privacy will remain important, we must accommodate for future innovations.”

WhatsApp is now in the midst of a major drive to make money. It has experienced a rocky start, in part because of broad

suspicions of how WhatsApp will balance privacy and profits. An announced plan to begin running ads inside the app didn't help — it was abandoned in late 2019, just days before it was set to launch.

Early this January, WhatsApp unveiled a change in its privacy policy — accompanied by a one-month deadline to accept the policy or get cut off from the app. The move sparked a revolt, impelling tens of millions of users to flee to rivals such as Signal and Telegram.

> *Facebook is "pestering" WhatsApp users to accept a new policy that allows Facebook to collect more data, while gradually removing key features of the app for those who don't, a coalition of 28 groups say.#TheDefender: SIGN UP –> https://t.co/UTTfyiKViIhttps://t.co/ba9cMY3OxB*
>
> *— Children's Health Defense (@ChildrensHD) May 19, 2021*

The policy change focused on how messages and data would be handled when users communicate with a business in the ever-expanding array of WhatsApp Business offerings. Companies now could store their chats with users and use information about users for marketing purposes, including targeting them with ads on Facebook or Instagram.

Elon Musk tweeted "Use Signal," and WhatsApp users rebelled. Facebook delayed for three months the requirement for users to approve the policy update. In the meantime, it struggled to convince users that the change would have no effect on the privacy protections for their personal communications, with a slightly modified version of its usual assurance: "WhatsApp cannot see your personal messages or hear your calls and neither can Facebook." Just as when the company first bought WhatsApp years before, the message was the same: Trust us.

Originally published by ProPublica.

The views and opinions expressed in this article are those of the authors and do not necessarily reflect the views of Children's Health Defense.

ProPublica