This copy is for your personal, non-commercial use only. To order presentation-ready copies for distribution to your colleagues, clients or customers visit https://www.djreprints.com.

https://www.wsj.com/articles/facebook-drug-cartels-human-traffickers-response-is-weak-documents-11631812953

Facebook Employees Flag Drug Cartels and Human Traffickers. The Company's Response Is Weak, Documents Show.

Employees raised alarms about how the site is used in developing countries, where its user base is already huge and expanding

By Justin Scheck, Newley Purnell and Jeff Horwitz
Sept. 16, 2021 1:24 pm ET

In January, a former cop turned Facebook Inc. investigator posted an all-staff memo on the company's internal message board. It began "Happy 2021 to everyone!!" and then proceeded to detail a new set of what he called "learnings." The biggest one: A Mexican drug cartel was using Facebook to recruit, train and pay hit men.

The behavior was shocking and in clear violation of Facebook's rules. But the company didn't stop the cartel from posting on Facebook or Instagram, the company's photo-sharing site.

Scores of internal Facebook documents reviewed by The Wall Street Journal show employees raising alarms about how its platforms are used in some developing countries, where its user base is already huge and expanding. They also show the company's response, which in many instances is inadequate or nothing at all.

A LOOK INSIDE THE FACEBOOK FILES »

Ask your questions now

Employees flagged that human traffickers in the Middle East used the site to lure women into abusive employment situations in which they were treated like slaves or forced to

perform sex work. They warned that armed groups in Ethiopia used the site to incite violence against ethnic minorities. They sent alerts to their bosses on organ selling, pornography and government action against political dissent, according to the documents.

Facebook removes some pages, though many more operate openly, according to the documents.

In some countries where Facebook operates, it has few or no people who speak the dialects needed to identify dangerous or criminal uses of the platform, the documents show.

When problems have surfaced publicly, Facebook has said it addressed them by taking down offending posts. But it hasn't fixed the systems that allowed offenders to repeat the bad behavior. Instead, priority is given to retaining users, helping business partners and at times placating authoritarian governments, whose support Facebook sometimes needs to operate within their borders, the documents show.

Facebook treats harm in developing countries as "simply the cost of doing business" in those places, said Brian Boland, a former Facebook vice president who oversaw partnerships with internet providers in Africa and Asia before resigning at the end of last year. Facebook has focused its safety efforts on wealthier markets with powerful governments and media institutions, he said, even as it has turned to poorer countries for user growth.

THE FACEBOOK FILES »

Facebook Files Sign up here

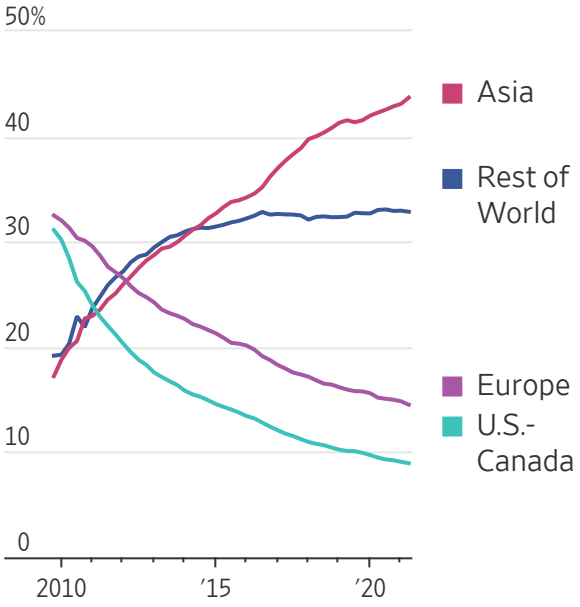
"There is very rarely a significant, concerted effort to invest in fixing those areas," he said.

The developing world already has hundreds of millions more Facebook users than the U.S. —more than 90% of monthly users are now outside the U.S. and Canada. With growth largely stalled there and in Europe, nearly all of Facebook's new users are coming from developing countries, where Facebook is the main online communication channel and source of news. Facebook is rapidly expanding into such countries, planning for

technology such as satellite internet and expanded Wi-Fi to bring users online including in poor areas of Indonesia one document described as "slums."

The documents reviewed by the Journal are reports from employees who are studying the use of Facebook around the world, including human exploitation and other abuses of the platform. They write about their embarrassment and frustration, citing decisions that allow users to post videos of murders, incitements to violence, government threats against pro-democracy campaigners and advertisements for human trafficking.

Percentage of total monthly active Facebook users, by region



Note: Data is quarterly
Source: Facebook

The material is part of extensive company communications reviewed by the Journal that offer unparalleled detail about the company's shortcomings in areas including rules that favor elites, teen mental health and efforts to manage its algorithm.

Some of the most serious issues flagged by the documents are overseas. Activists have complained for years that Facebook does too little to protect overseas users from trouble it knows occurs on its platform. The documents show that many within Facebook agree.

"In countries at risk for conflict and violence, we have a comprehensive strategy, including relying on global teams with native speakers covering over 50 languages, educational resources, and partnerships with local experts and third-party fact checkers to keep people safe," Facebook spokesman Andy Stone said this week.

The employee who identified the Mexican drug cartel is a former police officer and cybercrime expert hired in 2018 as part of a new investigation team focused largely on "at-risk countries," where the rule of law is fragile and violence is common.

That year, hate speech in Myanmar proliferated across Facebook's platforms, and the company has acknowledged it didn't do enough to stop incitements to violence against the minority Rohingya population, which the U.S. said were victims of ethnic cleansing. Executives described the Myanmar violence as a wake-up call to the company's responsibilities in the developing world. Chief Executive Mark Zuckerberg wrote a letter of apology to activists after initially playing down Facebook's role in the violence and pledged to do more.

An internal Facebook report from March said actors including some states were frequently on the platform promoting violence, exacerbating ethnic divides and delegitimizing social institutions. "This is particularly prevalent—and problematic—in At Risk Countries," the report says.

It continues with a header in bold: "Current mitigation strategies are not enough."

The ex-cop and his team untangled the Jalisco New Generation Cartel's online network by examining posts on Facebook and Instagram, as well as private messages on those platforms, according to the documents. (Messages on WhatsApp, another Facebook product, are encrypted by default.)



Mexican police attend a ceremony to honor colleagues killed by the Jalisco New Generation Cartel in 2015.

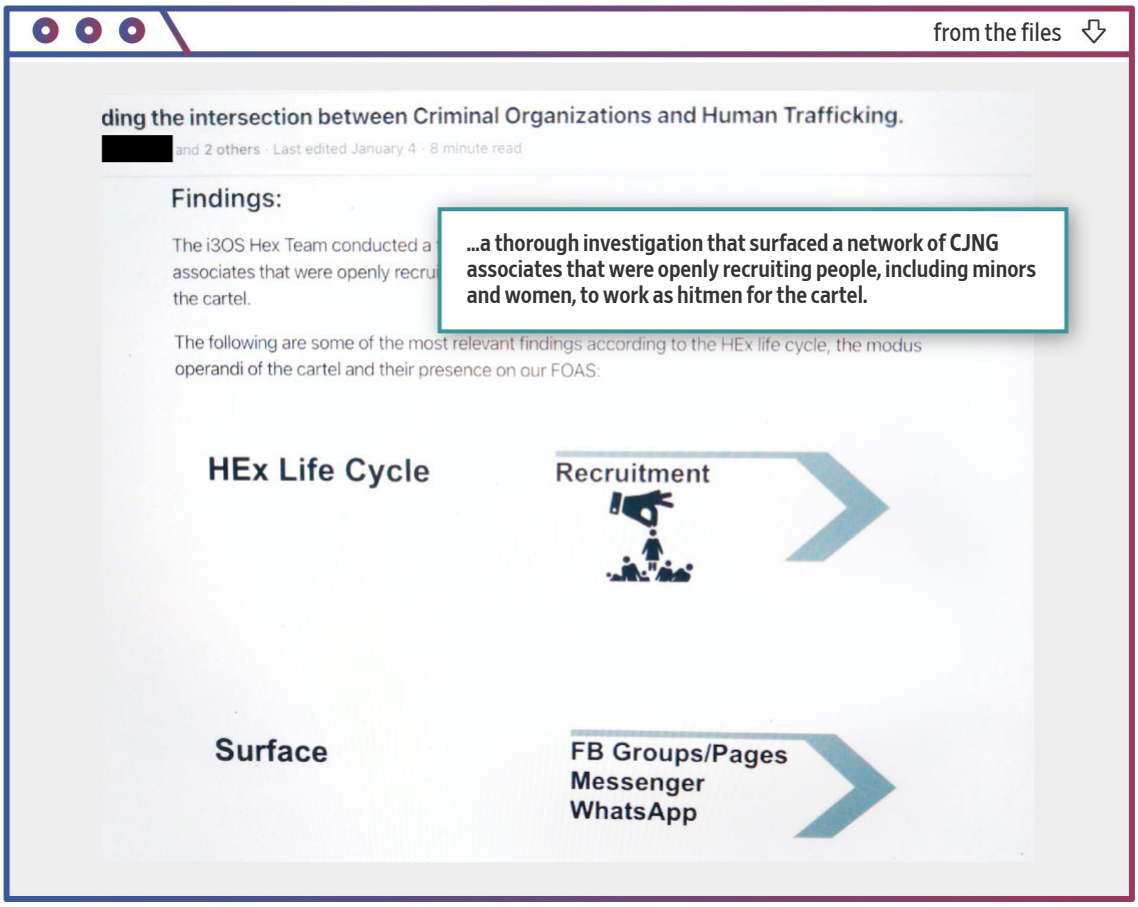
PHOTO: ULISES RUIZ BASURTO/EPA/SHUTTERSTOCK

The team identified key individuals, tracked payments they made to hit men and discovered how they were recruiting poor teenagers to attend hit-man training camps.

Facebook messages showed recruiters warning young would-be hires "about being seriously beaten or killed by the cartel if they try to leave the training camp," the former officer wrote.

The cartel, which law-enforcement officials say is the biggest criminal drug threat to the U.S., didn't hide its activity. It had multiple Facebook pages with photos of gold-plated guns and bloody crime scenes, the documents show.

The Facebook pages were posted under the name "CJNG," widely known as the shorthand for Cartél Jalisco Nueva Generación, even though the company had internally labeled the cartel one of the "Dangerous Individuals and Organizations" whose pages should have been automatically removed from the platform under Facebook policy.



Note: A name has been redacted on this document.
Source: January 2021 internal report titled 'Understanding the Intersection between Criminal Organizations and Human Trafficking'

The former cop recommended the company improve its follow-through to ensure bans on designated groups are enforced and seek to better understand cartel activity.

Facebook didn't fully remove the cartel from its sites. The documents say it took down content tied to the cartel and disrupted the network.

The investigation team asked another Facebook unit tasked with coordinating different divisions to look at ways to make sure a ban on the cartel could be enforced. That wasn't done effectively either, according to the documents, because the team assigned the job didn't follow up.

On Jan. 13, nine days after the report was circulated internally, the first post appeared on a new CJNG Instagram account: A video of a person with a gold pistol shooting a young man in the head while blood spurts from his neck. The next post is a photo of a beaten man tied to a chair; the one after that is a trash bag full of severed hands.

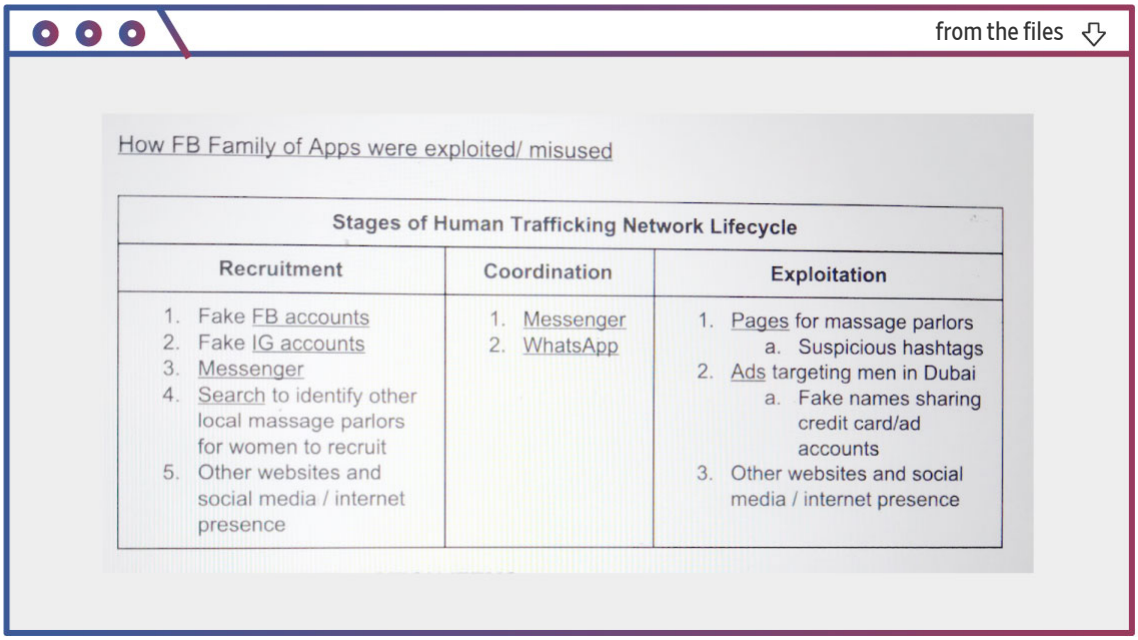
The page, along with other Instagram and Facebook pages advertising the cartel, remained active for at least five months before being taken down. Since then, new pages have appeared under the CJNG name featuring guns and beheadings.

The former officer declined to comment on his findings, and Facebook declined to make him available for an interview.

Facebook said this week its employees know they can improve their anti-cartel efforts, and that the company is investing in artificial intelligence to bolster its enforcement against such groups.

Facebook commits fewer resources to stopping harm overseas than in the U.S., the documents show.

In 2020, Facebook employees and contractors spent more than 3.2 million hours searching out and labeling or, in some cases, taking down information the company concluded was false or misleading, the documents show. Only 13% of those hours were spent working on content from outside the U.S. The company spent almost three times as many hours outside the U.S. working on "brand safety," such as making sure ads don't appear alongside content advertisers may find objectionable.



Source: 2019 'Case Briefs and Insights report' on how human traffickers and criminal networks use Facebook platforms

The investigation team spent more than a year documenting a bustling human-trafficking trade in the Middle East taking place on its services. On Facebook and Instagram,

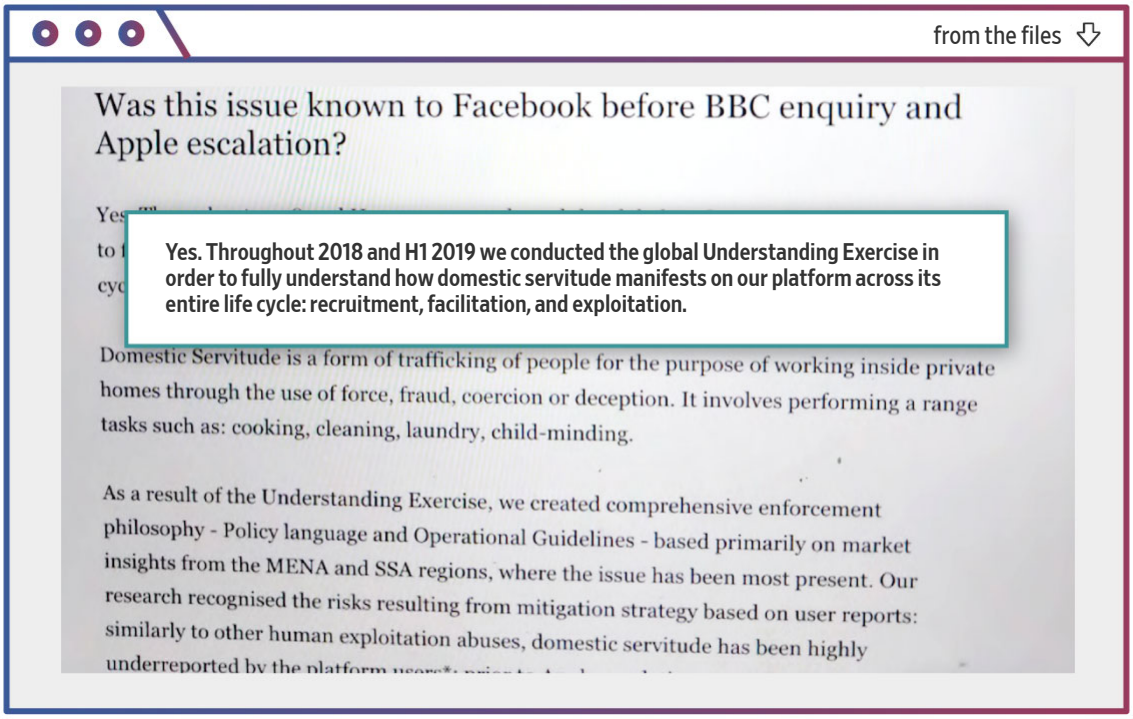
unscrupulous employment agencies advertised workers they could supply under coercive terms, using their photos and describing their skills and personal details.

The practice of signing people to restrictive domestic employment contracts and then selling the contracts is widely abused and has been defined as human trafficking by the U.S. State Department.

The company took down some offending pages, but took only limited action to try to shut down the activity until Apple Inc. threatened to remove Facebook's products from the App Store unless it cracked down on the practice. The threat was in response to a BBC story on maids for sale.

In an internal summary about the episode, a Facebook researcher wrote: "Was this issue known to Facebook before BBC enquiry and Apple escalation?"

The next paragraph begins: "Yes."



Source: 2019 internal report titled 'Apple Escalation on Domestic Servitude - how we made it through this SEV'

One document from earlier this year suggested the company should use a light touch with Arabic-language warnings about human trafficking so as not to "alienate buyers"—meaning Facebook users who buy the domestic laborers' contracts, often in situations akin to slavery.

The Facebook spokesman said the company doesn't follow that guidance. "We prohibit human exploitation in no uncertain terms," Mr. Stone said. "We've been combating human trafficking on our platform since 2015 and our goal remains to prevent anyone who seeks to exploit others from having a home on our platform."

He added: "We have a dedicated team that engages with law enforcement agencies across the globe. In instances of imminent harm, we may also provide relevant information to law enforcement in accordance with applicable law and our terms of service."

In Ethiopia, armed groups have used Facebook to incite violence. The company's internal communications show it doesn't have enough employees who speak some of the relevant languages to help monitor the situation. For some languages, Facebook also failed to build automated systems, called classifiers, that could weed out the worst abuses. Artificial-intelligence systems that form the backbone of Facebook's enforcement don't cover most of the languages used on the site.

Facebook also doesn't publish the "community standards" it requires users to abide by in all of the languages it serves in Ethiopia, so some users may not know the rules they are supposed to follow.

Facebook said this week the standards are available in some Ethiopian languages and that it has started translating them into others.

In a December planning document, a Facebook team wrote that the risk of bad consequences in Ethiopia was dire, and that "most of our great integrity work over the last 2 years doesn't work in much of the world." It said in some high-risk places like Ethiopia, "Our classifiers don't work, and we're largely blind to problems on our site."

Groups associated with the Ethiopian government and state media posted inciting comments on Facebook against the Tigrayan minority, calling them "hyenas" and "a cancer." Posts accusing Tigrayans of crimes such as money laundering were going viral, and some people on the site said the Tigrayans should be wiped out.



Relatives of residents of Togoga, in the Tigray region of Ethiopia, after an airstrike in June.
PHOTO: YASUYOSHI CHIBA/AFP/GETTY IMAGES

Violence escalated toward the end of last year, when the government launched an attack on the Tigray capital, Mekelle.

Secretary of State Antony Blinken said in March that Tigrayans are victims of ethnic cleansing. Ethiopia's government continues to commit violence against Tigrayans, the Journal reported last month.

Facebook said this week it has increased its review capacity in various Ethiopian languages and improved its automated systems to stop harmful content. It said it has a team dedicated to reducing risks in Ethiopia that includes people from the area.

Arabic is spoken by millions of Facebook users across what the company calls a highly sensitive region. Most of Facebook's content reviewers who work in the language speak Moroccan Arabic, and often aren't able to catch abusive or violent content in other dialects or make errors in restricting inoffensive posts, according to a December document. Facebook's enforcement algorithms also weren't capable of handling different dialects.

"It is surely of the highest importance to put more resources to the task of improving Arabic systems," an employee wrote in the document.

When violence broke out between Israel and Palestinians months later, the company erroneously suppressed Arabic-language regional news sources and activists, and began removing posts that included the name "Al Aqsa," an important Jerusalem mosque that

was a focus of the conflict. Al Aqsa is also used in the name of the Al Aqsa Martyrs' Brigade, which the U.S. has designated as a terrorist organization.

"I want to apologize for the frustration these mistakes have caused," one manager wrote in an internal posting.

The issue was previously reported by BuzzFeed.

Facebook publicly apologized and said this week it now has a team focused on preventing similar errors.

India has more than 300 million Facebook users, the most of any country. Company researchers in 2019 set up a test account as a female Indian user and said they encountered a "nightmare" by merely following pages and groups recommended by Facebook's algorithms.

"The test user's News Feed has become a near constant barrage of polarizing nationalist content, misinformation, and violence and gore," they wrote. The video service Facebook Watch "seems to recommend a bunch of softcore porn."

After a suicide bombing killed dozens of Indian paramilitary officers, which India blamed on rival Pakistan, the account displayed drawings depicting beheadings and photos purporting to show a Muslim man's severed torso. "I've seen more images of dead people in the past 3 weeks than I've seen in my entire life total," one researcher wrote.

In a 2017 mission statement, Mr. Zuckerberg said "giving people a voice is a principle our community has been committed to since we began," and that the company would "work on building new tools that encourage thoughtful civic engagement."

In 2018, Facebook Chief Operating Officer Sheryl Sandberg told a Senate committee the company supports democratic principles around the world. When asked about Facebook's operations in Vietnam, she said, "We would only operate in a country when we can do so in keeping with our values."

Facebook restricted the ability of users in Vietnam from seeing the posts of Bui Van Thuan, a prominent critic of Vietnam's authoritarian government, for nine months beginning last year. Mr. Thuan said Facebook acted after a group organized by the government sent the company thousands of complaints about his posts.

Facebook documents show the company's staff agreed the government organized efforts against Mr. Thuan, and used his case and a picture of him and his Facebook profile as an example of what they called systematic harassment.

Facebook tallied 153,000 such reporting incidents over three months via 36 private groups, likely "commissioned and directed by government/military entities." They said the efforts worked, with a "good success % in suppressing the target FB presence."

— Facebook internal document

Facebook last year said it agreed to curtail access to dissident political content deemed illegal in exchange for the Vietnamese government ending its practice of slowing Facebook's local servers to pressure the company.

A former Facebook employee who worked in Asia said Facebook is aware the Vietnamese government is using the platform to silence dissidents, but that it tolerates the abuse because Vietnam is a fast-growing advertising market.

"Our goal is to keep our services running in Vietnam so we can provide a space for as many people as possible to express themselves, connect with friends and run their business," Mr. Stone, the Facebook spokesman, said. "As we shared last year, we do restrict some content in Vietnam to ensure our services remain available for millions of people who rely on them every day."

Restrictions on Mr. Thuan's account were lifted last year, but he said he continues to face chronic harassment on Facebook.

Facebook said this week his profile was restricted in error and the mistake has been corrected.

Facebook's team of human-exploitation investigators, which in addition to the former police officer included a Polish financial expert who previously investigated trafficking finances at HSBC bank and a Moroccan refugee expert who formerly worked at the United Nations High Commissioner for Refugees, gathered evidence of human trafficking.

By looking across Facebook products, they found criminal networks recruiting people from poor countries, coordinating their travel and putting them into domestic servitude or into forced sex work in the United Arab Emirates and other Persian Gulf countries. Facebook products facilitated each step, and the investigators followed communications across platforms to identify perpetrators and victims.

Facebook in 2018 didn't have a protocol for dealing with recruiting posts for domestic servitude. In March 2018, employees found Instagram profiles dedicated to trafficking domestic servants in Saudi Arabia. An internal memo says they were allowed to remain on the site because the company's policies "did not acknowledge the violation."

The investigation team identified multiple trafficking groups in operation, including one with at least 20 victims, and organizers who spent at least $152,000 on Facebook ads for massage parlors.

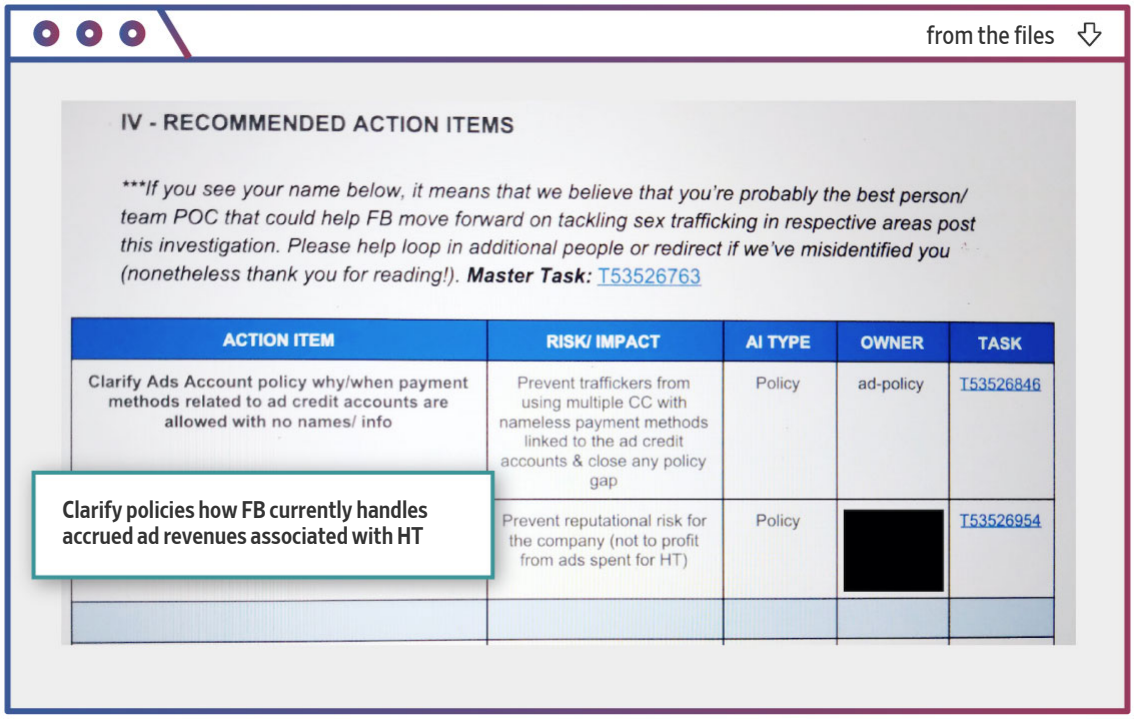
The former police officer recommended that Facebook disable WhatsApp numbers associated with the rings, put in new policies about ads purchased anonymously and improve its artificial intelligence to better root out posts related to human trafficking, according to the documents. He added that Facebook should develop a network to prevent trafficking by sharing findings with other tech companies.

In another memo, the Polish trafficking expert wrote that 18 months after it first identified the problem, Facebook hadn't implemented systems to find and remove the trafficking posts.

The BBC and Apple flagged concerns in 2019. With the threat posing "potentially severe consequences to the business," the trafficking expert wrote, Facebook began moving faster. A proactive sweep using the investigation team's prior research found more than 300,000 instances of potential violations and disabled more than 1,000 accounts.

The team continued finding posts of human trafficking, and Facebook struggled to put effective policies in place. One document says Facebook delayed a project meant to improve understanding of human trafficking.

Another memo notes: "We know we don't want to accept/profit from human exploitation. How do we want to calculate these numbers and what do we want to do with this money?"



Note: Names have been redacted on this document.
Source: 2019 'Case Briefs and Insights report' on how human traffickers and criminal networks use Facebook platforms

At the end of 2020, following three months in which Facebook investigated a dozen networks suspected of human trafficking, a system for detecting it was deactivated. The trafficking investigators said that hurt their efforts, according to the documents.

"We found content violating our domestic servitude policy that should have been detected automatically" by a software tool called the Civic Integrity Detection pipeline, wrote an employee in a document titled "Domestic Servitude: This Shouldn't Happen on FB and How We Can Fix It." She recommended the company reactivate that pipeline.

Facebook said this week similar screening systems are in operation.

The investigation team also struggled to curb sex trafficking. In 2019, they discovered a prostitution ring operating out of massage parlors in the U.S. Facebook gave the information to police, who made arrests.

Facebook discovered a much larger ring that used the site to recruit women from Thailand and other countries. They were held captive, denied access to food and forced to perform sex acts in Dubai massage parlors, according to an internal investigation report.

Facebook removed the posts but didn't alert local law enforcement. The investigation found traffickers bribed the local police to look away, according to the report.

Facebook said this week it launched new programs this year that make it harder for users to find content related to sex trafficking.

Over the past year, Facebook hired an outside consultant to advise it on the risks of the continuing trade in people on its sites. The consultant recommended that if revenue came in from trafficking advertisements, Facebook should develop a policy, such as giving it away, to avoid adding it to Facebook's coffers, according to the documents.

In January, Patricia Wanja Kimani, a 28-year-old tutor and freelance writer in Nairobi, saw a recruitment post on Facebook that promised free airfare and visas—even though Facebook has banned employment ads touting free travel and visa expenses, according to the documents.

"Most of the posts were saying cleaners needed in Saudi Arabia," she said in an interview. She said she was promised $300 a month to work for a cleaning service in Riyadh.

At the Nairobi airport, the recruiter gave her a contract to sign. It said she would receive 10% less pay than she was promised, and that only the employer could terminate the contract. If Ms. Kimani wanted to quit, she would lose her visa and be in Saudi Arabia illegally. Ms. Kimani told the recruiter that she was backing out.

The recruiter responded that since Ms. Kimani's contract had already been sold to an employer, the agency would have to reimburse the employer if she backed out. Ms. Kimani would have to pay the agency to make up for that, she said the recruiter told her. She didn't have any money, so she flew to Riyadh. The agency kept her passport.

She worked in a home where a woman called her a dog. She slept in a storage room without air conditioning. The house's locked courtyard and high walls made leaving impossible. She worked from 5 a.m. until dusk cleaning while "completely detached from the rest of the world," she said.

Ms. Kimani said she got sick and wasn't allowed treatment, and that she wasn't paid.

After two months, she told the agency she wanted to return to Kenya. They said she could pay them $2,000 to buy herself out of the contract. Ms. Kimani didn't have the money, and she posted about her plight on Facebook. She named the employment agency, which pulled her from the job and left Ms. Kimani at a deportation center.

She said there were other Kenyan women there and that one had marks from chains on her wrists and ankles. Eventually, her Facebook posts were forwarded to an official at the International Organization for Migration, a U.N. body, which helped negotiate her release and return to Kenya in July.

Ms. Kimani said Facebook helped her get into and out of the mess. She said she has been warning other people about the risks of getting trafficked, and she would like to see Facebook work harder. "I think something should be done about that so that nobody just goes in blindly," she said.



Ms. Kimani with her daughter at their home in Nairobi last week.

PHOTO: NICHOLE SOBECKI/VII FOR THE WALL STREET JOURNAL

—Neha Wadekar contributed to this article.

—Design by Andrew Levinson. A color filter has been used on some photos.

A Wall Street Journal investigation

A series offering an unparalleled look inside the social-media giant's failings—and its unwillingness or inability to address them.

VIEW THE FULL SERIES →

Write to Justin Scheck at justin.scheck@wsj.com, Newley Purnell at newley.purnell@wsj.com and Jeff Horwitz at Jeff.Horwitz@wsj.com

Appeared in the September 17, 2021, print edition as 'Facebook's Staff Flags Criminals, But Company Often Fails to Act.'

Copyright © 2021 Dow Jones & Company, Inc. All Rights Reserved

This copy is for your personal, non-commercial use only. To order presentation-ready copies for distribution to your colleagues, clients or customers visit https://www.djreprints.com.