OPINION
OTHER VOICES

In this May 13, 2021, file photo, a girl is tossed into the air as people gather for Eid al-Fitr prayers at the Dome of the Rock Mosque in the Al-Aqsa Mosque compound in the Old City of Jerusalem. Eid al-Fitr, festival of breaking of the fast, marks the end of the holy month of Ramadan. In May, as the Gaza war raged and tensions surged across the Middle East, Instagram briefly banned the hashtag #AlAqsa, a reference to the Al-Aqsa Mosque in Jerusalem's Old City, a flash point in the conflict. Facebook, which owns Instagram, later apologized, explaining its algorithms had mistaken the third-holiest site in Islam for the militant group Al-Aqsa Martyrs Brigade, an armed offshoot of the secular Fatah party. Inset: In this Dec. 20, 2018, file photo, a Bangladeshi reads a news report that makes mention of Facebook along with other social networking service, on his mobile phone in Dhaka, Bangladesh. (AP)

Internal company documents reveal problems

Language gaps weaken screening of hate, terrorism

Saving face … will Facebook face up?

As the Gaza war raged and tensions surged across the Middle East last May, Instagram briefly banned the hashtag #AlAqsa, a reference to the Al-Aqsa Mosque in Jerusalem's Old City, a flash point in the conflict.

Facebook, which owns Instagram, later apologized, explaining its algorithms had mistaken the third-holiest site in Islam for the militant group Al-Aqsa Martyrs Brigade, an armed offshoot of the secular Fatah party.

For many Arabic-speaking users, it was just the latest potent example of how the social media giant muzzles political speech in the region. Arabic is among the most common languages on Facebook's platforms, and the company issues frequent public apologies after similar botched content removals.

Now, internal company documents from the former Facebook product manager-turned-whistleblower Frances Haugen show the problems are far more systemic than just a few innocent mistakes, and that Facebook has understood the depth of these failings for years while doing little about it.

Such errors are not limited to Arabic. An examination of the files reveals that in some of the world's most volatile regions, terrorist content and hate speech proliferate because the company remains short on moderators who speak local languages and understand cultural contexts. And its platforms have failed to develop artificial-intelligence solutions that can catch harmful content in different languages.

In countries like Afghanistan and Myanmar, these loopholes have allowed inflammatory language to flourish on the platform, while in Syria and the Palestinian territories, Facebook suppresses ordinary speech, imposing blanket bans on common words.

"The root problem is that the platform was never built with the intention it would one day mediate the political speech of everyone in the world," said Eliza Campbell, director of the Middle East Institute's Cyber Program. "But for the amount of political importance and resources that Facebook has, moderation is a bafflingly under-resourced project."

This story, along with others published Monday, is based on Haugen's disclosures to the Securities and Exchange Commission, which were also provided to Congress in redacted form by her legal team. The redacted versions received by Congress were reviewed by a consortium of news organizations, including The Associated Press.

In a statement to the AP, a Facebook spokesperson said that over the last two years the company has invested in recruiting more staff with local dialect and topic expertise to bolster its review capacity around the world.

But when it comes to Arabic content moderation, the company said, "We still have more work to do. ... We conduct research to better understand this complexity and identify how we can improve."

In Myanmar, where Facebook-based misinformation has been linked repeatedly to ethnic and religious violence, the company acknowledged in its internal reports that it had failed to stop the spread of hate speech targeting the minority Rohingya Muslim population.

The Rohingya's persecution, which the U.S. has described as ethnic cleansing, led Facebook to publicly pledge in 2018 that it would recruit 100 native Myanmar language speakers to police its platforms. But the company never disclosed how many content moderators it ultimately hired or revealed which of the nation's many dialects they covered.

Despite Facebook's public promises and many internal reports on the problems, the rights group Global Witness said the company's recommendation algorithm continued to amplify army propaganda and other content that breaches the company's Myanmar policies following a military coup in February.

In India, the documents show Facebook employees debating last March whether it could clamp down on the "fear mongering, anti-Muslim narratives" that Prime Minister Narendra Modi's far-right Hindu nationalist group, Rashtriya Swayamsevak Sangh, broadcasts on its platform.

In one document, the company notes that users linked to Modi's party had created multiple accounts to supercharge the spread of Islamophobic content. Much of this content was "never flagged or actioned," the research found, because Facebook lacked moderators and automated filters with knowledge of Hindi and Bengali.

Arabic poses particular challenges to Facebook's automated systems and human moderators, each of which struggles to understand spoken dialects unique to each country and region, their vocabularies salted with different historical influences and cultural contexts.

The Moroccan colloquial Arabic, for instance, includes French and Berber words, and is spoken with short vowels. Egyptian Arabic, on the other hand, includes some Turkish from the Ottoman conquest. Other dialects are closer to the "official" version found in the Quran. In some cases, these dialects are not mutually comprehensible, and there is no standard way of transcribing colloquial Arabic.

Facebook first developed a massive following in the Middle East during the 2011 Arab Spring uprisings, and users credited the platform with providing a rare opportunity for free expression and a critical source of news in a region where autocratic governments exert tight controls over both. But in recent years, that reputation has changed.

Scores of Palestinian journalists and activists have had their accounts deleted. Archives of the Syrian civil war have disappeared. And a vast vocabulary of everyday words have become off-limits to speakers of Arabic, Facebook's third-most common language with millions of users worldwide.

For Hassan Slaieh, a prominent journalist in the blockaded Gaza Strip, the first message felt like a punch to the gut. "Your account has been permanently disabled for violating Facebook's Community Standards," the company's notification read. That was at the peak of the bloody 2014 Gaza war, following years of his news posts on violence between Israel and Hamas being flagged as content violations.

Within moments, he lost everything he'd collected over six years: personal memories, stories of people's lives in Gaza, photos of Israeli airstrikes pounding the enclave, not to mention 200,000 followers. The most recent Facebook takedown of his page last year came as less of a shock. It was the 17th time that he had to start from scratch.

He had tried to be clever. Like many Palestinians, he'd learned to avoid the typical Arabic words for "martyr" and "prisoner," along with references to Israel's military occupation. If he mentioned militant groups, he'd add symbols or spaces between each letter.

Other users in the region have taken an increasingly savvy approach to tricking Facebook's algorithms, employing a centuries-old Arabic script that lacks the dots and marks that help readers differentiate between otherwise identical letters. The writing style, common before Arabic learning exploded with the spread of Islam, has circumvented hate speech censors on Facebook's Instagram app, according to the internal documents.

But Slaieh's tactics didn't make the cut. He believes Facebook banned him simply for doing his job. As a reporter in Gaza, he posts photos of Palestinian protesters wounded at the Israeli border, mothers weeping over their sons' coffins, statements from the Gaza Strip's militant Hamas rulers.

Criticism, satire and even simple mentions of groups on the company's Dangerous Individuals and Organizations list - a docket modeled on the U.S. government equivalent - are grounds for a takedown.

"We were incorrectly enforcing counterterrorism content in Arabic," one document reads, noting the current system "limits users from participating in political speech, impeding their right to freedom of expression."

The Facebook blacklist includes Gaza's ruling Hamas party, as well as Hezbollah, the militant group that holds seats in Lebanon's Parliament, along with many other groups representing wide swaths of people and territory across the Middle East, the internal documents show, resulting in what Facebook employees describe in the documents as widespread perceptions of censorship.

"If you posted about militant activity without clearly condemning what's happening, we treated you like you supported it," said Mai el-Mahdy, a former Facebook employee who worked on Arabic content moderation until 2017.

In response to questions from the AP, Facebook said it consults independent experts to develop its moderation policies and goes "to great lengths to ensure they are agnostic to religion, region, political outlook or ideology."

"We know our systems are not perfect," it added.

The company's language gaps and biases have led to the widespread perception that its reviewers skew in favor of governments and against minority groups.

Former Facebook employees also say that various governments exert pressure on the company, threatening regulation and fines. Israel, a lucrative source of advertising revenue for Facebook, is the only country in the Mideast where Facebook operates a national office. Its public policy director previously advised former right-wing Prime Minister Benjamin Netanyahu.

Israeli security agencies and watchdogs monitor Facebook and bombard it with thousands of orders to take down Palestinian accounts and posts as they try to crack down on incitement.

"They flood our system, completely overpowering it," said Ashraf Zeitoon, Facebook's former head of policy for the Middle East and North Africa region, who left in 2017. "That forces the system to make mistakes in Israel's favor. Nowhere else in the region had such a deep understanding of how Facebook works."

Continued on Page 3

editor's choice