THE WALL STREET JOURNAL.

How Algorithms Can Help Beat Islamic State

He 'changed the world' by combating child porn. Now his software could suppress terrorists online.

By JOSEPH RAGO

Updated March 10, 2017 6:18 p.m. ET

Hanover, N.H.

You can't blame the message on the medium, not exactly. But maybe, all things considered, arming everyone with pocket supercomputers, and then filtering most of human experience through social-media feedback loops, wasn't the greatest idea.

America recently endured the most electronic and media-saturated presidential campaign in memory, with its hacks, private servers, secret videotapes, fake news, troll armies and hour-by-hour internet outrage across all platforms. And however glorious modern communications may be, they've also empowered a cast of goons, crooks and jihadists to build audiences and influence world-wide.

A technological solution, at least to that last problem, may lie 2,600 miles east of Silicon Valley, in a computer-science laboratory at Dartmouth College. Prof. Hany Farid, chairman of the department, creates algorithms that can sweep digital networks and automatically purge extremist content—if only the tech companies will adopt them.

“If you look at recent attacks, from Orlando to San Bernardino to Nice to Paris to Brussels,” Mr. Farid says, “all of those attackers had been radicalized online. They weren’t going to Syria. They watched YouTube videos.”

He continues: “The dark side of the open internet is that truly fringe and harmful ideas now are mainstream, or at least accessible to 7½ billion people.” Yet “whenever we have one of these attacks, we just wring our hands for a few weeks and then wait for the next one to happen.”

Social networks have created “a new environment for radicalization and recruitment,” says David Ibsen, executive director of the Counter Extremism Project, a nonprofit research and advocacy organization to which Mr. Farid is a senior adviser. Terror groups weaponize Twitter, Google, Facebook a nd other forums to plan or encourage violence; to discover the vulnerable or disaffected; and to publish professional, sophisticated and carefully presented propaganda.

Islamic State is basically a digital-first media startup. (By comparison, al Qaeda was MySpace.) ISIS content is beamed out globally and becomes refractory across the viral web. Some videos show vignettes of ISIS bureaucrats delivering social services or its fighters talking about the battle between belief and unbelief. Others are more savage—beheadings, stonings, drownings, other torture and combat operations.

Mr. Farid slipped into this world slantwise. He’s a founder of the computer-science field known as digital forensics. In the late 1990s as a postdoctoral researcher, he was among the first to recognize that mathematical and computational techniques to authenticate digital images and other media would be useful to society.

Because images so powerfully change what we are willing to believe, the modern era requires a scientific method to ensure we can trust them. How can we prove, for example, that digital photographs aren’t forgeries so they are admissible as evidence in court? Images are increasingly important in cellular, molecular and neurological medicine, Mr. Farid notes, and tampering has led to more than one research-and-retraction scandal. Unscrupulous stringers sometimes file doctored photos with news organizations, and unscrupulous motorists sometimes photoshop pictures to exaggerate fender-benders for insurance claims.

Mr. Farid explains how image authentication works: “We think about how light interacts in the physical world; what happens when that light hits the front of the lens and gets focused and goes through an optical train; what happens when that light hits an electronic sensor and gets converted from an analog to a digital and then goes through a postprocessing and gets saved as .jpeg and then gets posted on Facebook.” By identifying “statistical and geometrical and physical regularities” in this life cycle, software can search for inconsistencies to expose manipulation.

In 2008 this research pulled Mr. Farid

into another underworld—child pornography. In 2002 the U.S. Supreme Court struck down a ban on "virtual" child porn—computer-generated images that "appear to depict minors but were produced without using any real children." Mr. Farid is sometimes brought in as an outside expert when a defendant claims the material at issue is virtual.

The child-porn industry was nearly defunct by the 1990s, because negatives and videotapes can be confiscated and destroyed. "Then the internet came," Mr. Farid says, "and all hell broke loose."

Supply can create its own demand. Much like jihadists, deviants formed a global community, finding each other online and sharing what are really crime-scene photos. Like ISIS agitprop, material is continuously copied, cut, spliced, resized, recompressed and otherwise changed, in part to evade detection as it is retransmitted again and again.

Mr. Farid worked with Microsoft to solve both problems—detection and replication. He coded a tool called Photo DNA that uses "robust hashing" to sweep for child porn. "The hashing part is that you reach into a digital image and extract a unique signature. The robust part is if that image undergoes simple changes, the fingerprint shouldn't change. When you change your clothes, cut your hair, as you age, your DNA stays constant," he says. "That's what you want from this distinct fingerprint."

The algorithm matches against a registry of known illegal signatures, or hashes, to find and delete photographs, audio and video. Photo DNA is engineered to work at "internet scale," says Mr. Farid, meaning it can process billions of uploads a day in microseconds with a low false-positive rate and little human intervention.

Monitoring by Photo DNA, which is licensed by Microsoft at no cost and now used in most networks, revealed that the nature of the problem was "not what we thought it was," says Ernie Allen, the retired head of the National Center for Missing and Exploited Children. Child pornography was far more widely circulated than law enforcement believed. "Hany Farid changed the world," Mr. Allen adds. "His innovation rescued or touched the lives of thousand of kids, and uncovered perpetrators, and prevented terrible revictimization as content was constantly redistributed."

Mr. Farid linked up with the Counter Extremism Project to apply the same robust-hashing method to extremist propaganda. But this effort has encountered resistance. "The pushback from the tech companies has been pretty strong," the project's Mr. Ibsen says dryly.

U.S. law immunizes internet companies from criminal and civil liability for content that travels over their transoms. Their terms of service forbid abusive content, but they rely on users instead of algorithms to police violations. "It's a very slow and tedious process: You wait for it to get reported, somebody has to review it, they make mistakes," Mr. Farid says. "They take down the Vietnam napalm girl on Facebook."

Liability aside, what about their moral obligations to help prevent death, injury and destruction? “In my mind, we’re not asking them even to do something that they haven’t said they want to do already. We’re saying, hey, would you please do the thing that you promised you would do?” he explains. “I am simply saying, look, for free, you can automate this and make it really efficient and really fast and save you money on the side.”

But the “ethos” of Silicon Valley doesn’t include becoming the censors of the internet, and tech firms fear a slippery slope. “The concern they have is, OK, first they came for the child porn, then they came for the extremism, next they’re going to take the kitten videos,” Mr. Farid says. “I think that’s a bit of a hysterical leap. We are talking about content with very clear and well-defined harm. These are not abstract notions—‘I don’t want people to be mean to me.’ We’re not talking about bullying. We are talking about things with very immediate consequences and very real harm.”

One question is how to distinguish support for terrorism from the merely inappropriate or objectionable. What about Islamic State’s black-flag brand, or a declaration of a caliphate, or the sermons of Anwar al-Awlaki? Maybe you know it when you see it.

“Is an ISIS fighter saying ‘Death to the West’ extremism? I don’t know. I don’t want to have that conversation,” Mr. Farid replies. “I’m talking about explicit acts of violence, explicit calls to violence, explicit glorification of violence, depravity, the worst of the worst of the worst.”

His point is that tech companies can make judgment calls about the middle ground, wherever it might be, for themselves: “You decide: Yes, no, yes, no, yes, no, and then we’ll build a cache and eliminate that content from your networks.”

Mr. Farid concedes that there are dangers: “This type of technology is agnostic in what it’s looking for. It can be used in ways we would not approve of, such as stifling speech. You can’t deny that. This is what we’ve learned about technology over the years—it can be used for good and for bad. Social media platforms can be good and bad.”

There has been some progress. Twitter has deleted hundreds of thousands of handles associated with terrorism since 2015, and late last year Twitter, Facebook, Microsoft and YouTube announced an industry antiterror consortium. But Mr. Farid’s robust hashing remains a hard sell.

The irony is that algorithms increasingly govern the world. Networks are perpetually scanned for spam, malware, viruses; Google reads your email to target ads; credit-card companies monitor your financial transactions to prevent fraud. Facebook’s Mark Zuckerberg even promises to use algorithms to distinguish truth from falsehood. As a scholar of the differences between the two, Mr. Farid has a few thoughts.

In the backwash of 2016, Mr. Zuckerberg published a 5,800-

word manifesto that promised Facebook's artificial intelligence would soon learn to sort real news from hoaxes and misinformation, break up "filter bubbles," and draw a line between free speech and suborning terror. The goal, he wrote, is to preserve "our shared sense of reality."

Mr. Farid is a skeptic: "As somebody who worked for a long time in this space, I think he's underestimating what a hard problem this is." Mr. Zuckerberg "paints this picture like machine learning is going to be fully automatic—basically you'll be able to set criteria on your page, 'I don't want to see violence, I don't want to see bad words,' and it'll just work.

"Even as a technologist, and despite all the advance of technology, the human brain is astonishing at what it does. Our ability to make these types of assessments that are really hard for these AI algorithms is humbling. I don't think we'll get it in the next five or 10 years."

Meantime, Mr. Farid has developed a technology that could work today to contain a growing threat. While we await the Facebook utopia, perhaps our digital lives—and our real lives—would be healthier if it were widely deployed.

Mr. Rago is a member of The Wall Street Journal's editorial board.