9/11
MEMORIAL
& MUSEUM

Combatting Extremist Content on the Web (10/9/18)

00:00:25 Noah Rauch: Good evening, welcome. My name is Noah Rauch. I'm the senior vice president of education and public programs here at the 9/11 Memorial & Museum. As always, I'd like to extend a special welcome to our museum members and those tuning in to our live broadcast at 911memorial.org/live.

Tonight's program will explore two overlapping ideas. On the one hand, the rise of the internet and social media allows for ideas to spread globally and instantly. It allows for people to connect and feel heard in ways previously unimagined. And on the other, the platforms upon which these ideas spread and people connect can't or don't inherently discriminate between good ideas and bad ones, between positive connections and dangerous ones.

00:01:10 And so we see online radicalization and the dissemination of extremist content becoming central recruiting strategies for neo-Nazis and groups like ISIS and Al Qaeda. In the face of this reality, new technologies are being developed to prevent these groups from weaponizing the internet. This is a subject we've touched on in several of our previous public programs, and tonight we are very fortunate to have with us two experts who have extensive knowledge of these threats and the methods used to counter them. They bring different experiences and perspectives, and we have been looking forward to this program for quite some time.

00:01:41 Vidhya Ramalingam is founder of Moonshot CVE, a company using technology to disrupt and counter violent extremism globally. Vidhya's work is underpinned by a decade of experience engaging directly with extremists, building new partnerships with activists, and advancing policy design with international governments. She's held various positions, including commissioning panelist for the U.K. security and intelligence agencies and ESRC, associate at the University of Oxford, and board member of Life After Hate.

00:02:11 Prior to founding Moonshot, Vidhya was senior fellow on far right extremism and intolerance at the Institute for Strategic Dialogue. In 2012, she was appointed to lead the E.U.'s capacity-building program on far right violence, working with over 300 practitioners across ten European countries. She holds a Master of Philosophy in migration studies from Oxford University.

00:02:32 Hany Farid was most recently the Albert Bradley 1915 Third Century Professor and chair of computer science at Dartmouth. However, he has just started a new position-- congratulations-- at the Berkeley School of Information. His research focuses on digital forensics, image analysis, and human perception. He received his PhD in computer science from the University of Pennsylvania. He was a recipient of an Alfred P. Sloan fellowship, a John Simon Guggenheim fellowship, and is a fellow of the I.E.E.E. and National Academy of Inventors. He's also the chief technology officer and co-founder of Fourandsix Technologies and a senior adviser to the Counter Extremism Project.

With that, please join me in welcoming Vidhya Ramalingam and Hany Farid in conversation with executive vice president and deputy director of museum programs Clifford Chanin.

(applause)

00:03:27 Clifford Chanin: Thank you, Noah, and welcome, everybody. And I have to say, based on the conversation we had before the program started, you're in for a real treat tonight. So put on your seat belts and let's go. Let me ask each of you to help us define the problem that you're dealing with. What is the scale of it, who are the people who represent the problem, and how does the online culture in which they operate facilitate their access to these problematic communities that they become part of? Let me start with Vidhya. >>

00:04:00 Vidhya Ramalingam: Sure. When talking about this problem, I always caution those I'm talking with to make sure that we realize we're not talking about mass movements here. We're not talking about audiences in the millions. We're usually talking about audiences in the thousands. In some cases, we're really only talking about hundreds, and in some cases, we're talking about the dozens. So we're not talking about large-scale movements that are taking over our planet. We're talking about really, really tiny groups of individuals that are engaging in destructive behavior.

00:04:30 Now, the challenge for us, as organizations that are working in the preventative or intervention space, is, in the big, bad world of the internet, how it is that we identify these really tiny niche communities. And, you know, oftentimes, people talk about big data. It's a bit of a buzzword these days. I usually like to talk about small data.

For us, we're finding... We're looking to find individuals that are really just engaging in very tiny ways that indicate that they are getting involved in these movements. So they leave behind... In many cases, they leave behind what we call a digital footprint. It's a kind of telltale set of clues that lets us know that they're getting involved in this problem.

00:05:09 And the challenge for us as organizations that are bridging the technology space and the social work space is, how we can automate the process of identifying those individuals, and then reach them in very personalized ways in the online space. And I firmly believe that... You know, I'm very optimistic, actually, about what's possible in the online space. I do believe that we can find ways with new technology, and the advancements of technology, to scale up what is incredibly personalized work that needs to happen with individuals at risk.

00:05:36 Clifford Chanin: Wonderful. Hany, could you address the same question?

Hany Farid: You know, I think the-- and I don't think this is a very rewarding answer-- but I think part of the answer is, I don't think we entirely know. And I think it's important to understand, too, that while we're here at the 9/11 Memorial & Museum, we are talking about lots of different bad behavior online, from child exploitation to terrorism to the opioid crisis-- that is largely being fueled online-- to illegal guns.

00:06:02 And so the amount of criminal and dangerous behavior online, I think-- and by the way, if you just look at the surface web, the kind of web that you and I search in, you know, that's not a small number. But then there's also the dark web, and there are things that are happening behind the scenes with P2P, peer-to-peer connections, and encrypted chat lines. So there's a lot of activity happening that I think is very dangerous.

00:06:26 So here's, here's what we know, is that the internet has been weaponized. And it's been weaponized to great effect in many ways. So, for example, if you look at child exploitation, the child pornography problem, the child sexual exploitation, is largely fueled by the internet. I completely agree with you that when we are talking the terrorism space, it is a relatively small number of people. But if you look at the attacks over the last six years, they have been, the vast majority of them, rooted in radicalization online. And so while it may be a small number, the impact is tremendous for our society. So what I've seen-- and having lived through the beginning and the rise, and where we are today on the internet-- is, the internet has gotten weaponized.

00:07:09 And I think things are getting worse in how we are exploiting online communities. It is polarizing our society. It is being fueled by a relatively small number of people, but then being magnified by the way social media works in terms of the echo chamber it creates. And the direction that I see us going in is problematic. And I think a lot of things that we think about is, how do you leverage all the great things of technology-- which, we should acknowledge, there are many of them-- while mitigating what we are seeing, a very real harm from the types of behavior that are being fueled, magnified, and executed online?

00:07:46 Clifford Chanin: Let me ask about this question of numbers, because it is, as you say, a relatively small number of people. Yet, of course, they cause enormous problems when they act on some of these ideas or out of some of these impulses. Do we get the sense from the media that this is a small problem? Or is the media also part of this idea that we are facing this enormous crisis with enormous numbers of people who, whatever their number, seem to be manifesting themselves in societies across the world, having, you know, nothing to do with the terrorism problems of the United States?

00:08:24 But these kind of extremist movements are making themselves felt in the public space across European countries, certainly in the United States, the election just held in Brazil... I mean, again and again and again, these kinds of voices seem to have come into the public sphere in a way that was not accessible to them before. Is this simply a magnifying of their numbers, or does technology empower this in a way that we're only beginning to understand?

00:08:53 Vidhya Ramalingam: Well, media has the power to be incredibly helpful for those of us that are working on violent extremism. It has the power... You know, journalists who are investigating these movements and, again, applying pressure to social media companies and to governments to act, that can be a very important... a very important advocacy tool for those of us that are working on this issue. But it also has the power to completely undermine the efforts of some of us who are working on more hidden groups, groups that don't get the public attention that they should.

00:09:22 The greatest example of this is the underplayed challenge of neo-Nazism and white supremacy across America. It took Charlottesville for the media to really start to cover this issue in a substantive way. And for those of us that have been working on terrorism in the online space in the U.S. for many years, we've known that the numbers of individuals that are engaging in violent neo-Nazi content vastly outweigh the numbers of individuals that are engaging in violent jihadist content. And the media, if you look solely at media representation, you would never get that, that picture. So media has the power to both incredibly help us and also to potentially completely distort the picture on violent extremism across the country.

00:10:02 Hany Farid: I think that's absolutely right. My sense is, the media's good right after the last attack. And there's a lot of talk and a lot of hand-wringing and, "What do we do?" But in between, I mean, you're right. Prior to Charlottesville, there was not a lot of talk about neo-Nazism online. They were certainly there. You would have thought, by the way, that after Charlottesville, all these tech companies that booted some of these groups would have just woken up and realized there were bad people on their networks. They knew they were on the networks, and they tolerated it-- and they tolerated it, I think, inexcusably so.

00:10:33 So I think, sadly, it takes these huge events with deaths and destruction for the media to wake up. And then, of course, there's a ripple effect. Right, so the event happens, the media wakes up. The public wakes up. And then the PR machines start kicking in at the tech companies and they start responding to it in fits and bursts. Look at the privacy issue that we're dealing with in this country. What did it take? It took Cambridge Analytica, right? And then that was the start of an honest discussion. These issues have been there for the last decade, but I don't think we have been talking about them honestly enough. We have been running headfirst into this technology, hoping that everything would work out, and, well, guess what? It doesn't. You have to be a little more thoughtful.

00:11:16 Clifford Chanin: You've each mentioned a number of different problems and, you know, certainly in the conversations that we have here, for obvious reasons, the recruitment of people to jihadist movements-- whether going to the Islamic State or acting as lone-wolf terrorists-- has been the focus of what we do.

00:11:34 But from what you're saying, it seems like the problem itself-- and you've alluded to opioids, and you've alluded to right-wing and neo-Nazi extremists-- the problem itself, then, doesn't really seem to be a particular set of beliefs. The problem itself is, rather, the use of technology to allow people to join up with whatever this cause or movement is, or illegal activity is, because they, they exploit a vulnerability in the system. Is that fair to say?

00:12:08 Hany Farid: Yeah. I would say-- I would just add one more thing to that, which is, the technology is not in and of itself to blame. It's the way the technology has been developed and deployed. So if you look at the Googles, the Facebooks, the Twitters, these large platforms. First of all, they have a speed and a reach around the world that is unprecedented, which in many ways is fantastic and part of the issue. That in and of itself is not the issue. The issue is that these platforms have decided that they are going to operate on the philosophy that they are a platform, that they are neutral in this space.

00:12:47 So whether you're sharing kitten videos or beheading videos or child pornography issues, they remain agnostic, okay? And therein lies part of the problem, because those materials are not on the same caliber. And the reason they do this, by the way, and you should understand this, is that here in the U.S., we have something called the Communications Decency Act, and part of that act, there's a section 230 of that act, that says that platforms-- like Facebook, like Twitter, like Google, like YouTube-- cannot be held liable for what their users do on their platform.

00:13:20 That's not a terrible idea, by the way, because if they were held liable for everything that happened on their platforms, they would have been out of business decades ago. But they have taken section 230 of CDA to an extreme, and they are turning a blind eye to children being trafficked on their platforms, to terror groups and neo-Nazi groups recruiting and radicalizing and executing attacks on their platforms, and they have turned a blind eye to all of that.

00:13:43 And I think that is part of the problem, is this lack of responsibility for the particularly heinous acts that are happening on their platform. And the difficulty, the thing you have to wrestle with is, where do you draw that line? The easiest thing to do in the world, the easiest thing to do, is to say, "We're a neutral platform. We don't draw any lines." That would be nice, but the reality now, we know, is, we can't take that position anymore.

00:14:06 And now we have to start walking back is, where do we pull off content? Where do we decide that this is too dangerous to be online? And that requires really thoughtful discussion, it's hard decisions. But we are talking about, literally, trillion-dollar companies. Surely they have the resources to do this. And I would remind people, by the way, that in this country, we have, for over a decade now, been protecting the rights of the music and the movie industry through the Copyright Millennial Act, that says platforms can be held responsible for infringing on copyright information.

00:14:39 We do not hold them to the same level of responsibility for child pornography, child sex trade, ISIS videos, and violent videos. And so I find that sort of an odd juxtaposition, where we are happy to protect the financial interests of the music and the movie industry, but not the well-being of our children and our society. And I find that sort of an inexcusable position to be in.

00:14:59 Vidhya Ramalingam: Just to add to that, you know, we see striking similarities in the way that violent extremists are using the internet and those that are using the internet for human trafficking purposes, for substance abuse issues, people who are, who are... communicating with others encouraging suicide. There are a range of destructive behaviors that are being... are being advanced in the online space.

Now, the beauty of that... of that... of that kind of complementarity across these different destructive behaviors is that, for those of us that are trying to respond, it presents huge opportunity for us to learn from what's been... what's been done in the child sexual exploitation space, the suicide prevention space.

00:15:41 You know, for us, for us working in counterterrorism in the online space, we've taken huge inspiration from the suicide prevention space, which has for many years now, since the birth of the internet, been encouraging personalized messaging to reach out to individuals that are dropping signs and signals to us that they're considering killing themselves. Now, there's a lot of learning there to take place. We just need to be able to work effectively with social media and tech companies, and that's where some of the challenges arise.

00:16:06 Clifford Chanin: Can you describe, Vidhya, what you do? Because you're really starting at the personal end. There's an interaction that you detect online that signals to you in some way-- and I'd like to know what way that is-- but signals to you that someone is at risk of crossing a line. How do you detect that and what do you do about it?

00:16:27 Vidhya Ramalingam: So I mentioned earlier this digital footprint that individuals leave behind. And I realize that that might actually sound like a scary term in a post-Cambridge Analytica era. Now, the reality is that so much of that digital footprint is data that's put in the public domain... Users consent, you know, whether it's on their Facebook profiles, whether it's on Twitter-- whatever platform it is. There's a lot of public data out there that can be used to identify individuals at risk. And that's without infiltrating groups, without going in, without breaching any privacy regulations. Now, one step beyond that, a data source that we've been using in the last few years through a partnership with Google, is search data.

00:17:04 Now, we don't look at the personal information of anyone who's searching, but we do use a very simple mechanism, and that's actually targeted advertising. It's the same tool that's used by Nike, by Coca-Cola, to advertise, you know, shoes and soda to us. And what we do is, we look at data around how many individuals across the country are searching for violent extremist content. And actually, maybe this is a good time just to show a couple of the visuals. Just a warning to those in the audience, there may be some disturbing words on the screen, so apologies in advance for that.

00:17:34 If we go to slide two... This just gives you a sense, from seven months last year, we were looking at the kinds of search traffic related to right-wing extremism, violent right-wing extremism, that was happening across the U.S. These are the top search terms that we detected. Now, this is just a small sample. We tracked hundreds of search terms related to violent neo-Nazism, violent white supremacy, across the states. All in all, in that seven-month period, we saw 180,000 instances of searches for this sort of content.

00:18:08 If we move on to the next slide, slide three, this is an example of some of the top search terms on the violent jihadist side of the spectrum, again, during the same seven-month period across the United States last year. Now, the numbers for jihadist search traffic on Google were just around 40,000. So 180,000 for violent neo-Nazi search traffic, and 40,000 searches on the violent jihadist side.

Clifford Chanin: Just parenthetically, we were speaking before, it seems that these numbers spike when there's an incident that raises people's interest in reaching out to these groups.

00:18:40 Vidhya Ramalingam: Absolutely, exactly. And, actually, if we move to slide five, this is an example of that spike. So post-Charlottesville, so we've been... we were... We were tracking search data across the U.S. before and after Charlottesville took place, in August of 2017. This gives you an indication as to the impact of an event like that on individuals that are trying to access violent content on Google.

00:19:03 So we saw a 400% increase in searches... in searches related to violent far right. A 493% increase in searches specifically related to joining the KKK. And then, most shockingly, a 200% increase in searches relating to a desire to kill ethnic minorities. Now, this is pretty shocking. If we go on to the next slide, slide six, this is data from after the New York attack that took place almost exactly a year ago, at the end of October of 2017.

00:19:32 Clifford Chanin: So this is the car attack on...

Vidhya Ramalingam: Exactly, the truck attack.

Clifford Chanin: Right down the street from the museum.

Vidhya Ramalingam: Exactly. In the week following that attack, we saw a 414% increase in searches related to killing apostates or infidels. And overall, a 142% increase in searches attempting... attempting to download ISIS products. Now, this for us, this is not publicly available data, in that it's not information that somebody has put onto their public Facebook profile, but it is information... It's information that is available to those of us that are using targeted advertising. And this is where some of our efforts to reach individuals that are searching for this material comes into play.

00:20:10 Clifford Chanin: So what happens? What is the next step after you've found people who are doing this outreach?

Vidhya Ramalingam: So the next step... So, I mentioned the way that, you know, advertisers like Coca-Cola and, you know, other commercial entities use advertising. When you're searching for a product, you'll oftentimes see an ad that pops up. It's the very first result when you're, when you're looking for something on Google. And if I... if I give you an example-- so I think it's slide seven. If I give you an example of a violent extremist search term.

00:20:39 So "Hitler is my hero"—which is a very real search term that we tracked hundreds of instances of across the U.S.-- if you enter "Hitler is my hero" into Google, if we go to the next slide... The results that you get, the top four results, are all neo-Nazi, white supremacist content, including videos that have been put together. It might be tiny writing. That first video that pops up at the top is uploaded by White Aryan Woman, and it is a neo-Nazi account that's uploaded very violent video content.

00:21:11 The first four results are all neo-Nazi content. With the approach that we developed in partnership with Google, which is called the redirect method, our advertisement becomes the first piece of content that an individual sees. And if we go to the next slide, you can see what that looks like. That's one of our ads. Now, the ads that we use don't explicitly say, "Don't believe neo-Nazis, don't join a neo-Nazi group," because of course, if you say that to someone who is so entrenched in their beliefs, that will turn them off right away.

00:21:40 What we've tried to do is meet them with ad text that looks like it's offering them the information that they're, that they're interested in. Now, it's a very fine line there, because you don't want to be inciting violence through your own ads. You don't want to be confirming their beliefs. But we do want to make them think that they will be coming to something which will offer them information that's related to what they searched for. And then that's exactly what we offer them.

00:22:00 If we go to the next slide, what we offer them is information that is directly related to what they were searching for. It's directly linked to the question or the query that they've entered into Google. And the approach that we've taken is not to create single pieces of content that try and debunk their narratives, but instead make use of the range of existing content that's already been uploaded to YouTube by various NGOs and individuals across the, across the country. What you can see here, the video playlist that we're using in this instance, it has testimonies from former white nationalists, former, former terrorists. It has video content that offers more nuanced perspectives around cultural diversity in America.

00:22:40 And what we've seen in terms of results from this approach is, you know, if you look at, actually for noncommercial advertising-- so advertising that's used in the advocacy space-- generally, we're looking at a 1.7% click-through rate. It's very low-- a 1.7 click-through rate for individuals that will click on an ad if they're looking for something that's political or related to an ideology.

00:23:03 For our ads, we get anywhere between three and four percent click-through rates. So nearly double... nearly double the average for campaigns that are in a related space. And then in some places, in some countries where we've delivered this work, we get up to seven percent click-through rate, which is basically unheard-of for advertising.

Where we find our ads are most effective is actually not even ad texts that I showed you there, but when we offer ad texts that deal with the underlying vulnerabilities, the underlying emotions of the individual who might be searching. These are ads that were designed in partnership with social workers, with psychologists, and they deal directly with an appetite or a need for mental health or social health content.

00:23:45 So some of our most effective ads, ads where we've had up to ten percent click-through rate, is ads where we're asking, "Do you feel anxious? Click here for support." "Do you feel hopeless? Click here for... to talk to someone." And we'll offer them resources that are genuinely offering them some form of either self-help or referring them to a real-world service that they can access and that they can engage with.

00:24:08 Clifford Chanin: So this allows people to identify-- and you assume they have the capacity to identify-- this need within themselves as somehow connected to the impulse to glorify Hitler, or something like that.

Vidhya Ramalingam: Exactly, and we've seen... So we've tested mental health, social health ads with audiences at risk of violent extremism. And then we've also tested them with control groups. In every instance, whether it's with neo-Nazis or jihadists, we find that the extremist audience is disproportionately likely to engage with those mental health, social health ads.

00:24:41 There is, in some cases... So for the jihadist audience across the U.S., for those that were searching for religious jihadist content, they were three times more likely than a control group to click on that ad when it was offered to them. So we know there's an appetite there for individuals who are engaging in hateful content to get help, to get some form of support, and to talk about their feelings, their emotions. And that's a real area that we need to exploit when we're trying to reach people online.

00:25:05 Clifford Chanin: Okay. Now, Hany, your work comes from the other side of this. You're not offering services to individuals who are looking around in these troubling areas on the internet. And your work goes back a decade or more to issues of child pornography and child exploitation. And just to sort of set you up, if I could, it really is about finding ways and algorithms that allow these materials to be identified, and essentially disabled or disarmed, so that the people who are looking for them may still have this interest or need, but they're not able to find these materials that are removed quickly from the sites that are hosting them. But tell us more about that.

00:25:46 Hany Farid: That's exactly right. Yeah, so we, we started-- and I think we've talked about this already, that, you know, the issues of terrorism, child exploitation, opioid crisis-- they're all very different in terms of social and economic impact, but they have this common thread.

And a lot of the solutions that we see, as you mentioned, you know, in one space sort of map into the other spaces. And so back in 2008 or so, I started working with the National Center for Missing and Exploited Children, NCMEC-- this is a U.S.-based agency that deals with child protective services in this country and abroad-- to deal with what was then an explosion in online child sexual abuse material.

00:26:25 And at the time-- and even today-- what we would have liked to have done is to develop technology that can look at every single image, every single video, and determine if it's child pornography. So is there a person in this? Is the person underage? Is the content sexually explicit? And back in '08, and even today, we couldn't solve that problem. It's an incredibly hard algorithmic problem to solve, and we particularly couldn't solve it at internet scale.

00:26:52 At a scale of billions of uploads a day, you have to run at an accuracy that is simply not available to us today. So what we did instead, we didn't want to just give up, so what we did instead is, we know that the National Center for Missing and Exploited Children is home-- today it's home to 80 million known child pornographic content, okay?

00:27:14 We also know that that same content-- week in, week out, month in, month out, and year in and year out, and decade in and decade out-- continually gets redistributed. So the idea is the following: If you have identified illegal content, all right? This contains a child, it's sexually explicit... What we want to do now is, we're not going to try to find all the child pornography. We're going to try to find that one piece of content. And the way we do that-- and this is the algorithmic part-- is, we reach into that content and we extract from it a signature. That signature has two important properties. It is distinct, so whatever signature we extract for that image is not shared by any other image in the universe, and it is stable as that content is modified as it makes its way around the internet.

00:28:03 Clifford Chanin: So the signature is always visible in some way to the algorithm.

Hany Farid: That's right. And so think about the human DNA. This is basically human DNA, right? My human DNA is distinct from everybody in the room, and as far as we know, seven some-odd billion people in the world. And as I age, as I change my clothes, as I move around, my DNA stays stable. So that's why we call the technology "photo DNA." It had those two properties.

00:28:25 So what it allowed us to then do is to sit at the pipe of Facebook-- and this, by the way, has been true since 2010-- every single image that you have uploaded to Facebook has been scanned by this software for the last eight years. And all it does is, every image that gets uploaded, we extract the signature from that-- and here's the really nice thing, by the way, since we've been talking about privacy-- once you have the signature, you can tell me nothing about the image. There's no... there's no identifying information. We're not doing face recognition. We're not doing anything that would violate the... Well, Facebook is doing all kinds of things that would violate your privacy.

(laughter)

00:28:57 We're not doing anything to violate your privacy. And we scan that signature against all bad content, whether that's child pornography, terrorism, opioid ads. And anything that is a match simply never makes it online.

Clifford Chanin: So it just blocks it from appearing.

Hany Farid: It just blocks it from upload. It never makes it online. And, by the way, because it's child pornography, it's actually reported to NCMEC, which is then reported to law enforcement, and that's how people go to jail. Now, in the counterterrorism space, we don't report those, because that content is not inherently illegal. We believe it's dangerous. Facebook believes that it violates their terms of service, so they can also block based on that. So this technology says there is all this content out there that we don't want on our network. Either it's illegal, it's a violation of terms of service, it is dangerous. We extract from that content the signature, scan, scan, scan, scan, scan-- and, by the way, that also works for copyright material.

00:29:48 So that's the same technology that these companies have been using for many years to... When you go to YouTube and you'll see a video that's blocked because it has copyright infringement, it's the same idea. When the movies release these things, they actually extract signatures, give those signatures to the YouTubes of the world, and then that's how they block copyrighted material from showing up. And so what we say is, there is-- and this is... We have to be very careful here.

You said something, too, is that you have to be really careful when you write those ads. There's a fine line here. And there's also a line here. In the child exploitation space, it is not always clear what child pornography is. The federal statute here says it has to be sexually explicit. That's not particularly well defined. You have to be under the age of 18.

00:30:28 In the terrorism space, too, there is a line between, this is sort of hurtful speech, maybe even hateful speech, but it doesn't elevate to the point of, "Here is how you make a suicide vest and go and blow people up." So what we try to do is, be very thoughtful on what constitutes hate speech, what constitute dangerous speech, what constitutes violations of terms of service. We try to make those decisions in the most thoughtful way possible. But then once we made that decision, it's over. We extract the signatures from the content, and we can sit it at Facebook, and we can be completely consistent with how we stop that material from showing up online.

00:31:08 Clifford Chanin: Now, your story of this technology runs into the resistance of the tech companies, even though we're talking now about child pornography and child exploitation, which would seem kind of a gimme in this case.

Hany Farid: Yeah, you would think so.

Clifford Chanin: Yeah, so tell us, you know, what happened there and how that was overcome. And then I'll ask you, Vidhya, about your relationship with the tech companies in exactly the same context.

00:31:30 Hany Farid: So you would think, by the way... And again, I don't want to be graphic or say things that are disturbing, but I will tell you, I was very naïve when I started working in this space. I thought child pornography was 16-, 17-year-olds flirting with their sexuality. I'm not excusing that. But I didn't realize these were four-year-olds and six-month-olds and infants. This is what we're talking about. It's really... it's, it's... and you would think... You would really think that this is something we could all get behind. We can all agree that content that shows four-year-olds being sexually abused has no place in our society. We can all agree on this, can't we? It turns out we can't-- it turns out we can't.

00:32:09 So here's the story. 2003, the internet is really in its early days. It's on the rise. And we see an explosion of child sexual abuse online, I mean, just an explosion. And then-Attorney General Ashcroft is briefed on this problem. Of course, it's all happening online. Now, as the A.O.Ls. and the Yahoos! and the early platforms, it was happening there. And in '03, he convened the leaders of the tech industry to Washington, and said, "Guys, your platform is being used to do some pretty awful things. You got to get a handle on this." And they said, "We feel terrible about this. We feel really terrible. But there's nothing we can do. It's too hard of a problem."

00:32:49 And for five years-- this is, by the way, an eternity in technology-- from '03 to '08, nothing happened. So think about what happens in five years in technology. That is insane. And the tech companies got bigger and bigger, they got wealthier and wealthier, and they kept turning a blind eye. And in '08 is when I got involved.

00:33:09 And with leadership primarily from Microsoft and myself, and working with the National Center, we developed this technology. It took five years to finally get a technology that worked. Google adopted this technology in 2016. We deployed it at Microsoft in 2009. So it's stunning to me the... With clear evidence not only that the problem exists, not only that it's illegal, not only that it is incredibly harmful, not only that we have technology that... By the way, we gave that to them for free. We weren't even asking them to pay for it. We just gave it to them for free. "Just use this. Just use it to get rid of this material."

00:33:46 Still, we met against resistance. And it wasn't until really steady, beating them up in the press, basically. It was really just day in and day out, doing the interviews, talking to the press, talking to legislators, and putting pressure on the company. Here's another example of this. You'd think they would have learned their lesson, by the way, at this point, because they deployed the technology, and guess what? It didn't break the internet, and they're doing just fine.

00:34:09 Just this last year, the Senate passed a bill called the Stop Enabling Sex Traffickers Act. And it was designed to allow victims of the sex trade to sue the platforms that knowingly-- and that word is really important-- knowingly allowed them to be trafficked on their platform. There's one particularly bad actor in this space called Backpage, that knowingly trafficked children. And they were responsible for the vast majority of children trafficked in this country, which, by the way, is in the hundreds of thousands.

00:34:37 The tech industry lobbied to kill the bill. This was a bill that was meant to give children the ability to have legal protection. And, by the way, we had been protecting copyright infringement for almost two decades at this point, but are still resisting protecting the most vulnerable among us. So this is not new. The fight on the child exploitation... I'm sure I'm going to hear similar things on counterterrorism. And by the way, you're hearing the same thing on the opioid crisis.

00:35:04 We know that the internet is being used to traffic illicit material, and the companies are turning a blind eye to it. And, you know, we can have an interesting conversation as to why that is. But I think that is the reality. And despite pressure from here, despite pressure from the E.U., despite pressure from other parts of the world, it is in some ways almost business as usual.

00:35:27 There are still-- look, you just saw it. If you didn't read the news yesterday, Google was found covering up a huge security breach where user data was captured. Why did they not release it? The internal memo says they were worried about legislative pressure. They knowingly hid this. And so this-- you have to understand this stuff. You can't believe the hype about Silicon Valley. I am not an anti-technologist. I'm a computer scientist by training.

00:35:53 But they are not here to connect the world. They are not here to make the world a better place. They're here to make a lot of money. That's fine-- I don't have a problem with that. I have zero problem with that. But let's not pretend that they are different than any other industry, and that they somehow should be above regulation and above oversight. Because I think the day of the wild, wild west of the internet is over, and we have to start realizing the real harm that we are seeing and-- and this is important-- and the absolute intransience of the tech companies to come to grips with the reality of how their platforms are being abused.

00:36:25 Clifford Chanin: Vidhya, are you encountering the wild, wild west in the tech world still, in relation to these kinds of extremist movements?

Vidhya Ramalingam: So my company, Moonshot CVE, works very closely with tech companies. And I always say that we are critical friends of tech companies. And, you know, we will work with them, but we also challenge them where challenge is due. The issue of terrorism on platforms like Facebook and on Google as a search engine have existed for years. It's been in the press for years. We've known it for years.

00:36:55 These companies have really only started to invest in this problem, in tackling this problem, in the last two years, I'd say. Facebook made its first counterterrorism hire two years ago. And Google approached us in 2016, two years ago, to develop the redirect method, which is the method I just showed you. It's been, they've been very new to come to the table for those of us that have actually been responding proactively in the online space.

00:37:17 Now, one of the major challenges is, they don't necessarily have the in-house expertise on these issues. And they know that. They're very aware of that, and that's why they outsource quite a lot of their work on these issues-- they hire out. That poses challenges, because those that they're working with who sit outside of their platform staff don't necessarily have the access to all the information that they would require in order to work effectively.

00:37:39 Now, these are not... I'm not spilling any trade secrets here. These are problems that are acknowledged and known by those that work at these companies. Facebook has 200 employees that work exclusively on counterterrorism now. And most of those employees are not experts on counterterrorism at all. In many cases, they're contractors. They're individuals working remotely who are reviewing content and helping to flag and take down content. So there are lots of challenges.

00:38:03 One thing I would say is that Facebook and Google and tech companies, they operate like governments. They move very slowly at times, and then sometimes, they move very quickly out of reactions that are based on advocacy, pressure points, media, and other... and other kind of pushes. So for those of us that are trying to work with tech companies, it's, it's a very challenging environment. Because it can be... it can be very much dependent on the time and place.

00:38:29 Immediately after attacks like Charlottesville or the New York attack, or any attack that happens in the U.S., we see an immediate influx of questions from tech companies to us as to what we can do with them to solve the problem. Many of you will have seen post-Charlottesville, due to huge pressure which came from the private sector, there was a mass takedown across, you know, platforms like GoDaddy that do website hosting, dating sites like OkCupid, eBay, et cetera, to crack down on the way that neo-Nazis were using their platform.

00:38:59 We needed Charlottesville for that move to take place. So it's a very challenging environment to operate in. What I would say is that it's very similar to working with governments in terms of the pressure points, and that's why we do need pressure groups. We need those that are doing investigative work who are revealing the problems on these platforms, because that will help those of us who are trying to do good work make it happen.

00:39:19 Clifford Chanin: So I'd ask each of you, I mean, is this simply a matter of finding ways to affect the bottom line, whether it's through threatened regulation, whether it's media exposure. We were talking before, the actual user doesn't have much power, because as a user, you're not the source of income. You're simply a pair of eyes that's looking...

Hany Farid: Well, you are, but not in the way you'd like. (laughing)

Clifford Chanin: Right. So is it really a matter of, the bottom line has got to be at least threatened or hit in some way for this to change?

00:39:51 Hany Farid: Yeah, I think you're asking the right question. So in my experience, yes, that where the change comes... First, it's reactionary to specific events. Although, I would add that there's often a rebound effect. That often there's a lot of activity because they know that there's some bad press coming, and then they sort of return back to normal behavior.

So I'll give you an example of that. Facebook got in trouble a little while ago for allowing people to advertise housing and jobs to whites only. That was a ProPublica, I believe, expose. They said, "Oh, we feel really bad about this. We're really sorry-- we're going to fix it." A year later, same problem, right?

00:40:25 So there is this reaction, but often we slip back. So I do think that we have seen the needle move. So if you look at the GDPR, the General Data Privacy Regulation that came out of the E.U. You've all gotten a hundred emails about this over the last year. That has definitely moved the needle a little bit, although we just saw a massive security breach with Facebook just this week. I think where the pressure will really come is fines.

00:40:51 So Germany now can fine tech companies up to 50 million euros for every failure to take down content once they are notified. The E.U. is considering similar legislation that would fine tech companies millions of euros, and, if the behavior persists, up to four percent of their revenues. So these are big, big penalties, even for the Googles and the Facebooks of the world. I think where the pressure comes from, the other two pressure points are the media-- you've already mentioned this, and I agree-- is, bad press is bad for these companies and they do respond to it.

00:41:24 And then the third is where the revenue really comes from, which is advertisers. So a couple of years back, there was an expose about how mainstream products were being advertised against beheading videos. And those companies got really mad. And they pulled their advertising from YouTube. And, well, guess what? YouTube suddenly got really smart really fast. And so the advertisers-- I like to say there's probably 12 C.E.Os. in the world that could at the snap of a finger change online tech companies. Those 12 C.E.Os., from the Unilevers, the GMs, the G.Es.-- the biggest advertisers online-- if they said, "We are going to withhold advertising for the next six months until you get your act together," I guarantee you, that will change the game.

00:42:08 Clifford Chanin: Yeah, I think you were mentioning before that Tiffany ads showed up in neo-Nazi searches.

Hany Farid: It was something like this, yeah. And people were, you know, look, they have brands to protect. They don't want-- and, by the way, it was also showing up in child predator, child pornography and child predatory ads, as well. And that's because, you know, again, what you have to understand about these platforms is, to operate at the scale that they operate at, it has to be fully autonomous.

00:42:33 That's the only way the system works. So they have made a killing by having the smallest number of bodies on the ground, and then letting the algorithms do everything. And the algorithms, it turns out, are pretty stupid. They don't know when somebody advertises for, "I want to advertise to people who like 'kill the Jews.'" They don't-- they don't... Nobody is looking over their shoulder. And we don't accept that in our mainstream media. We don't accept that on television, for the most part, and I don't think we should accept it on the platforms.

00:43:01 And in some ways, Facebook and YouTube want to have it both ways. They want to be both a center for news, but they don't want to be a news publisher. And I don't think you can have it both ways. Pick which one you want, but Facebook, whether they like it or not, they are a news aggregator and a source of news. More than 50% of Americans get their news primarily from Facebook. You now have a responsibility, I'm sorry.

Clifford Chanin: Vidhya, your... Is the financial pressure point the one that makes most sense in terms of effecting change?

00:43:29 Vidhya Ramalingam: Oh, I think it's hugely important. And the example I gave of post-Charlottesville, the private sector applying pressure, is the greatest example of that. I would say, in addition to the pressure points that were just discussed—in the media, in particular-- government pressure on social media companies does have an impact. I mean, most recently, so, a mixture of European and North American governments and major tech companies have recently set up what's been called the Global Counterterrorism Internet Forum, which is a space where, essentially, they're meeting very regularly to work through options collaboratively on what can be done on terrorism online.

00:44:03 We've seen a very real impact of some of those kind of for a for that sort of pressure to be placed on those companies. And in particular, what I would say is, we mentioned a lot about Facebook and about Google, and Twitter, as well. We're increasingly seeing a diversification of platforms that are used. I mean, we were just speaking before the event about how Facebook is becoming very passe. I mean, people are using Instagram now, which is owned by Facebook, so you might ask, "Well, is it really... Is it really leaving Facebook if you're on Instagram?" But in addition to those platforms, we're seeing more and more users, especially at-risk users, users at risk of getting involved in violence, engaging on platforms like Telegram, Discord gaming apps, a range of other platforms.

00:44:42 Now, these sorts of for a where government pressure is applied to tech companies also has an impact on those smaller companies as they're on the rise. And we've now seen that pressure pushing those companies to engage with organizations like ours that have expertise on extremism on those platforms so that they can more effectively respond. There are ways to effect change.

00:45:01 Clifford Chanin: You mentioned Telegram, which is an encrypted site, and I wonder if you could each comment on... You know, the next step in this is encryption. So even if everything squares away with the publicly available sites Facebook and Google, and so on and so forth, people now have options for complete, uncrackable—at this point-- privacy. How does that figure into your thinking about this issue?

00:45:24 Hany Farid: Well, so, I think it's important to understand that... Whether it's the solution you've been hearing on stage and whether it's what we're working on, or what you're working on, the problem just doesn't disappear. We just move the problem. And the hope is, we keep moving and moving and moving and narrowing and narrowing and narrowing, and we eventually get control over it.

So, yes, we know, for example, in the child pornography space, that people have moved from the public forums to P2P-- peer-to-peer sharing-- to encrypted channels, to the dark web. So we have to keep innovating and keep moving and keep thinking.

00:45:59 So I'll tell you one thing that we are working on. So I talked a little bit about the signature of digital content, that we can grab a content and extract a signature and tell you, "Ah, this is an image or a video I've seen before." We now have a solution that works in the encrypted domain. So if encrypted data is coming through a pipe, without having to break encryption-- you are allowed to have the privacy you want, and we should... that's good. I mean, encryption is not inherently bad, but it is complex. We can actually analyze the encrypted bits and tell you this is child pornography, this is a beheading video.

00:46:32 And I really like this work, because in some ways, it finds that balance between privacy and security. And that's the balance we have to find. I think we've gone too far into the privacy space and not enough in security. And we need to sort of refigure that a little bit. So this is one technology that we're working on to try to deal with that. I think, also, we have to think about, as we deploy these technologies, not reactively, but proactively, how are they going to be abused?

00:46:59 We have to think out of the gate, not once it's too late and the infrastructure is in place and everything's being abused, and years have gone by, but right out of the gate, you got to start thinking, "All right, we're deploying technology." You can't stop reading about artificial intelligence. This is a good example. Artificial intelligence, we seem to be going through another wave of A.I. We know there's going to be lots of great things that come out of that. But there's going to be some dangers. We have to think about them now and how to mitigate them before they get too far ahead of us.

00:47:27 But you're absolutely right. This is going to be a moving target. It's like everything else. It's a moving target. Think everything else in the cyber world-- spam, malware, viruses-- it's this constant war between the adversary and the defender. And it's hard, and it's tedious. But not doing anything doesn't work, either. So this is, this is going to be the new reality going forward, I think.

Clifford Chanin: Vidhya, in your world, if someone goes into an encrypted space, have you lost them or do you have other ways of...

00:47:54 Vidhya Ramalingam: Oh, absolutely not. There are... there are ways. And what I would say is that in addition to the challenges of identifying content on... on encrypted platforms, which is the space where Hany is one of the greatest experts, for those of us that are trying to do proactive intervention work-- so social work in the online space-- it is... they're far from lost when they enter an encrypted zone.

00:48:14 For us, what we need to do is find clever ways of making sure we are in those spaces with them, so that regardless of encryption, we can still find individuals that are sharing content and engage with them one-on-one. And that's what we've done. We've had to create software which automates the process of joining Telegram groups where jihadist content's being shared, that automates the process of joining a WhatsApp group where hate speech is being shared.

And, you know, let's say a context, like in India, or in Myanmar, on a platform like Viber. You know, we need to be there. And if we're there, then we can engage in one-on-one conversations. So there are clever ways that we need to come up with to do it, but they're far from lost once they're there.

00:48:50 Clifford Chanin: Good to hear. Let's see if we have some questions from our audience here. We'll ask you before you speak to wait for the microphone to get to you. So... anyone would like... Gentleman in the middle there.

Man: Thank you—quick question for you. When you get to the context—or when you get into these Facebook groups or these WhatsApp groups, or what have you, what is the next step after that? Who takes over and who is communicating from that point on, and what happens?

00:49:26 Vidhya Ramalingam: Actually, I do have a slide for that which we haven't shown yet. Slide 11, actually, can give you an example. So what we, what we tend to do when we're trying to do social work in the online space, we will have people who are trained psychologists, who are trained social workers, who will be the ones in those spaces, in those closed spaces, whether it's a closed Facebook group or a Telegram group. They will then send a message to individuals. They will send direct messages or they'll send messages in kind of the group chat.

00:49:53 The kinds of messages that we send vary based on the environment that we're in and the country that we're working in. These are some examples of the kinds of messages that we've sent in a U.S. context. Sometimes we're working with social workers. Other times, we will sometimes bring in a former extremist, someone who has been through these movements, who understands it, and who can reach out on an individual level and say, "Hey, man, I've been there before. I know what you're dealing with. Let me know if you want to talk."

00:50:20 Now, in terms of the responses that we get from these messages, the most effective message text that we've sent is one that reaches someone on a really personal level. So either, you know, it acknowledges something that they've said about themselves, and it kind of meets them on... It acknowledges some sort of, kind of shared background or shared interest. Or offers of help, kind of reaching out with an olive branch. Sometimes we won't get a response for months. Sometimes on Facebook messaging, in particular, where you can wait for months for a message, we will sometimes wait four months before we get a response. But least the individual knows that someone's there, and that they can reach out to someone if they are... if they are actually concerned about their position in a movement and they want to leave. In terms of what we do next, once that conversation actually gets started, our intent is to build a relationship with that individual.

00:51:08 We know that terrorists are using platforms like Facebook, like Skype, like WhatsApp, to build personal connections with individuals. The cases of girls in London that got on planes to travel to Syria-- or tried to get on planes, excuse me, to travel to Syria-- they were speaking on Skype for months with people who were based in Syria and Iraq, and they were building personal relationships. They felt like they knew them. For those of us that are trying to respond through this sort of method, we want to build that same relationship. We want to mimic that kind of person-to-person, peer-to-peer relationship.

00:51:39 Once we build up that relationship, the ultimate intent is to take that message from the online space into an offline meeting. And that's where a lot of this work is still very experimental. We're currently running pilots across the globe of this. We will have results next year, but at the moment, we are currently working on that transition from the online to the offline, which has to be a done in a very managed, carefully managed and safe and secure way. But it is possible to do. It's just about, you know, testing whether you can build personal relationships online.

00:52:09 Clifford Chanin: You know, it strikes me that each of you, working in two very different ways, but each of you have expressed the question or the concern about whether you can simply keep up, that whether you're counting individuals and trying to find ways of intervening on their behalf, or if you're simply trying to keep up with these masses of data--whether it's human or digital-- there's more, perhaps, than you can actually deal with at any given moment. Is that... is that really the worry here?

00:52:37 Hany Farid: Yeah, it depends on the day of the week that you ask. I think there's days where we feel like we're losing, and there are some days where we feel like we're barely keeping up. I think that's a good day, is when we're keeping up. But, yeah, I mean it's... it's, you know, it's a big world. It's a big internet.

And I think you had already said it earlier, without the cooperation of the tech companies... That's maybe part of the frustration, is that, you know, they're the gatekeepers of this content at the end of the day, and it's, you know, two, three, four of them. And, you know, you can only put so much pressure at some point.

00:53:05 And so part of the frustration is, I feel like... You know, if the issue was, we simply don't have the technology or the personnel or the skills to solve the problem, I would say, "Okay, let's get there, but we're not there." But there are days where I feel like we could do so much more than we're doing now, were it not for just for a lack of will. And that to me is very frustrating, is when I feel like there are good people in the world who are trying to do good things, and we are being blocked from trying to make the world a better place because of interests elsewhere. That to me is very frustrating.

Clifford Chanin: Are you keeping up?

00:53:38 Vidhya Ramalingam: I'm hugely optimistic about what's possible in the online space. I'm hugely optimistic. I think one of the challenges for our sector, for the counterterrorism sector, the counter- violent-extremism sector, is that we're working on a very risky... a very risky issue. It's inherently risky. But so many of the actors in the space are risk-averse. They're unwilling to take risks at all. They're unwilling to actually reach out to an individual and start a conversation.

00:54:04 And this is where, I think, actually, the... You know, technology has the power for us to scale up what is incredibly personalized work. It's essentially the equivalent of, you know, the youth worker in the street who just happens to be present there and is able to break up gangs. We want to replicate that in the online space. If we're there, we know we can do something. I'm hugely optimistic.

There are lots of challenges, and we've touched on a lot of those in the conversation, but I do think we're making good progress. We just, yeah, we just need to keep pushing forward and being a bit bolder.

00:54:32 Clifford Chanin: Let me see if we have another question from the floor. The gentleman there.

Man: Well, this talk has made me optimistic, anyway, so thank you. To hear that you are out there and addressing this problem is incredible. I'd just like to know a little bit more about how you get your resources to do this and if there's an opportunity for the public to help with your work.

Clifford Chanin: That's a good question.

00:55:05 Vidhya Ramalingam: That's a great question. So in terms of resources, we work with both tech companies and with governments, sometimes foundations, to deliver this work. We haven't taken public donations, and I... You know, for organizations like ours, I don't... I don't see that in the immediate future, as nice it is to even have someone ask that question.

00:55:24 However, there are lots of organizations that are doing this personalized work, both offline, and also working with us, in some instances, online. Organizations like... so I am on the board of an organization called Life After Hate, which is an NGO in the U.S., that... It was set up by former white supremacists, and post-Charlottesville, in particular, they are working to get people out of neo-Nazi movements. They're starting conversations with people online. We work with them to facilitate their having conversations with individuals online. Organizations like that do accept public donations.

00:55:53 There are a lot of places you can turn if you're interested in being a part of this. I would look at NGOs that are operating across the U.S. For us, you know, we're forging ahead doing what we can with governments and with tech companies, with all of the challenges that come with that. But it allows us... You know, it allows us to sustain this work for a longer period.

Hany Farid: Yeah, I work with a great group called the Counter Extremism Project. They're based here in New York City. It was founded by Ambassador Wallace, who is the former U.S. ambassador to the U.N. It is an NGO. It is, I think, almost entirely private-funded. We get some government grants. And then I use a lot of my funding from Dartmouth College to sort of help support the back end, as well.

00:56:36 So it's, you know, it's... This is sort of what's amazing about these, these are relatively small organizations that we're talking about. And so one of the good things is, we are not necessarily resource-poor. It's really more about the advocacy, sort of getting the media and the people to sort of understand what the problems are and sort of advocating on our behalf in some ways. But I think you're right. There's a lot of organizations out there doing really good work and I'm sure would benefit tremendously from public support.

00:57:06 Clifford Chanin: You know, it's nice, I'm glad you... Did you want to come back?

Man: You know, as you spoke, I just kept thinking to myself, "Boy, I'd love to share that on Facebook."

(laughter)

Hany Farid: Don't.

(laughter)

Hany Farid: In fact, just delete your Facebook account. You'll be so much happier.

(laughter)

00:57:21 Clifford Chanin: You know, it's... And you made the point, Michael, I mean, it's... We have a lot of conversations here about a lot of interesting but complicated and tough issues. And it's nice, I think, to find one where we end on an upbeat note. And so with that, I'm going to get out while the getting's good, and ask you to thank our guests, Vidhya Ramalingam, and Hany Farid.

(applause)