Google’s Director of Regime Change For The Democrats Reveals Google’s Secret For Mass Mood Manipulation To Steer Votes and Elections

- Google’s Jared Cohen is Sid Blumenthal 2.0
- Cohen can “kill any of Google’s political enemies reputation in two clicks”

Author: Andy Greenberg. Andy Greenberg Security

Inside Google’s “Internet Justice League”

Around midnight one Saturday in January, Sarah Jeong was on her couch, browsing Twitter, when she spontaneously wrote what she now bitterly refers to as “the tweet that launched a thousand ships.” The 28-year-old journalist and author of *The Internet of Garbage*, a book on spam and online harassment, had been watching Bernie Sanders boosters attacking feminists and supporters of the Black Lives Matter movement. In what was meant to be a hyperbolic joke, she tweeted out a list of political caricatures, one of which called the typical Sanders fan a “vitriolic cryptoracist who spends 20 hours a day on the Internet yelling at women.”

Related Stories

Andy Greenberg

Google’s Clever Plan to Stop Aspiring ISIS Recruits

- 

  Andy Greenberg

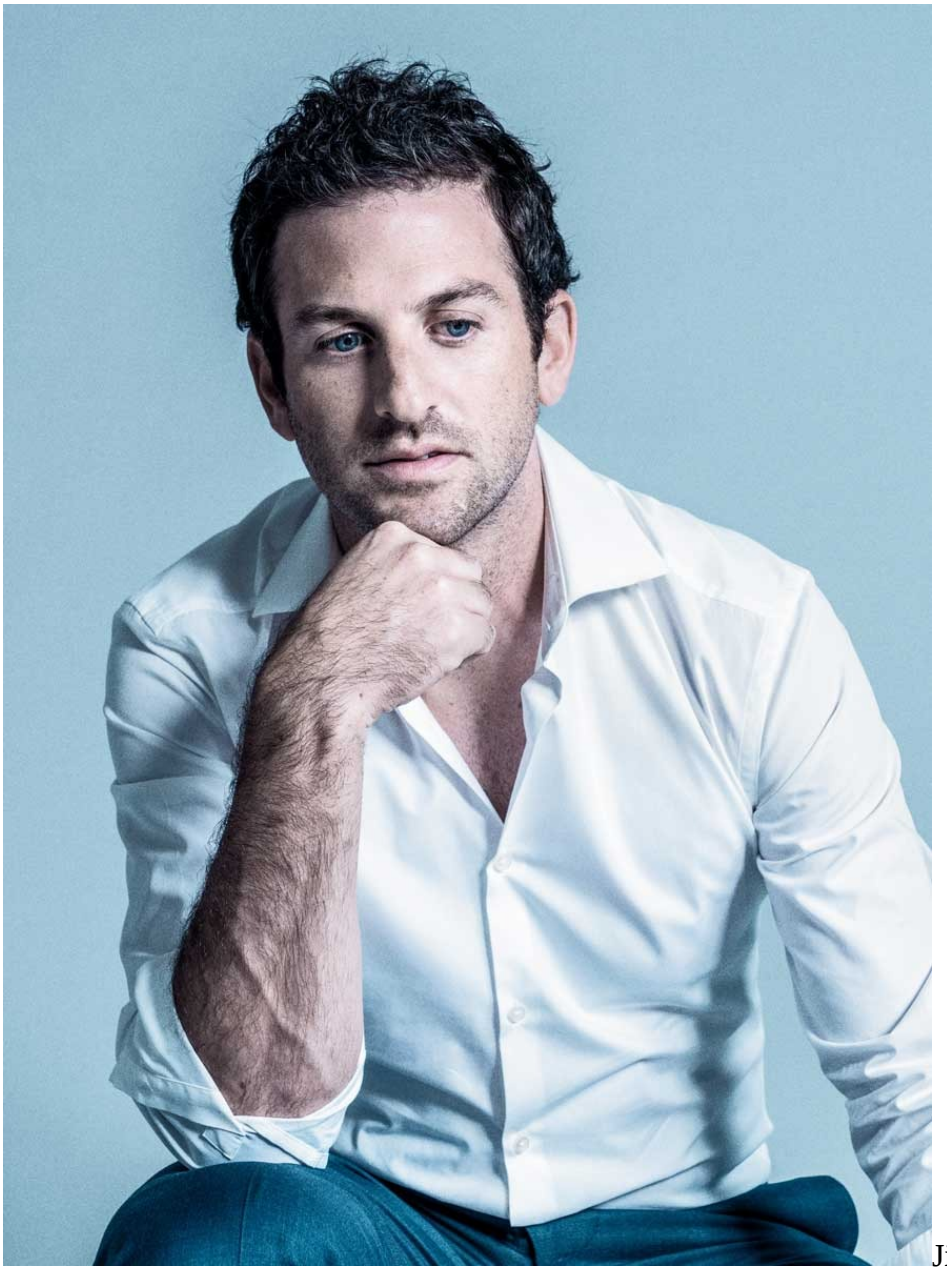
  Google Wants to Control ALL News Sites From Cyberattacks—For Free

---

The ill-advised late-night tweet was, Jeong admits, provocative and absurd—she even supported Sanders. But what happened next was the kind of backlash that's all too familiar to women, minorities, and anyone who has a strong opinion online. By the time Jeong went to sleep, a swarm of Sanders supporters were calling her a neoliberal shill. By sunrise, a broader, darker wave of abuse had begun. She received nude photos and links to disturbing videos. One troll promised to "rip each one of [her] hairs out" and "twist her tits clear off."

The attacks continued for weeks. "I was in crisis mode," she recalls. So she did what many victims of mass harassment do: She gave up and let her abusers have the last word. Jeong made her tweets private, removing herself from the public conversation for a month. And she took a two-week unpaid leave from her job as a contributor to the tech news site *Motherboard*.

For years now, on Twitter and practically any other freewheeling public forum, the trolls have been out in force. Just in recent months: Trump's anti-Semitic supporters mobbed Jewish public figures with menacing Holocaust "jokes." Anonymous racists bullied African American comedian Leslie Jones off Twitter temporarily with pictures of apes and Photoshopped images of semen on her face. *Guardian* columnist Jessica Valenti quit the service after a horde of misogynist attackers resorted to rape threats against her 5-year-old daughter. "It's too much," she signed off. "I can't live like this." Feminist writer Sady Doyle says her experience of mass harassment has induced a kind of permanent self-censorship. "There are things I won't allow myself to talk about," she says. "Names I won't allow myself to say."


Jigsaw’s Jared Cohen: “I want us to feel the responsibility of the burden we’re shouldering.”Benedict Evans

Mass harassment online has proved so effective that it’s emerging as a weapon of repressive governments. In late 2014, Finnish journalist Jessikka Aro reported on Russia’s troll farms, where day laborers regurgitate messages that promote the government’s interests and inundate opponents with vitriol on every possible outlet, including Twitter and Facebook. In turn, she’s been barraged daily by bullies on social media, in the comments of news stories, and via email. They call her a liar, a “NATO skank,” even a drug dealer, after digging up a fine she received 12 years ago for possessing amphetamines. “They want to normalize hate speech, to create chaos and mistrust,” Aro says. “It’s just a way of making people disillusioned.”

All this abuse, in other words, has evolved into a form of censorship, driving people offline, silencing their voices. For years, victims have been calling on—clamoring for—the companies that created these

platforms to help slay the monster they brought to life. But their solutions generally have amounted to a Sisyphean game of whack-a-troll.

Now a small subsidiary of Google named Jigsaw is about to release an entirely new type of response: a set of tools called Conversation AI. The software is designed to use machine learning to automatically spot the language of abuse and harassment—with, Jigsaw engineers say, an accuracy far better than any keyword filter and far faster than any team of human moderators. “I want to use the best technology we have at our disposal to begin to take on trolling and other nefarious tactics that give hostile voices disproportionate weight,” says Jigsaw founder and president Jared Cohen. “To do everything we can to level the playing field.”

Jigsaw is applying artificial intelligence to solve the very human problem of making people be nicer on the Internet.

Conversation AI represents just one of Jigsaw’s wildly ambitious projects. The New York–based think tank and tech incubator aims to build products that use Google’s massive infrastructure and engineering muscle not to advance the best possibilities of the Internet but to fix the worst of it: surveillance, extremist indoctrination, censorship. The group sees its work, in part, as taking on the most intractable jobs in Google’s larger mission to make the world’s information “universally accessible and useful.”

Cohen founded Jigsaw, which now has about 50 staffers (almost half are engineers), after a brief high-profile and controversial career in the US State Department, where he worked to focus American diplomacy on the Internet like never before. One of the moon-shot goals he’s set for Jigsaw is to end censorship within a decade, whether it comes in the form of politically motivated cyberattacks on opposition websites or government strangleholds on Internet service providers. And if that task isn’t daunting enough, Jigsaw is about to unleash Conversation AI on the murky challenge of harassment, where the only way to protect some of the web’s most repressed voices may be to selectively shut up others. If it can find a path through that free-speech paradox, Jigsaw will have pulled off an unlikely coup: applying artificial intelligence to solve the very human problem of making people be nicer on the Internet.

-

Merjin Hos

Jigsaw is the outgrowth of an earlier effort called Google Ideas, which Google's then-CEO Eric Schmidt and Cohen launched in 2010 as a "think/do tank." But aside from organizing conferences and creating fancy data visualizations, Ideas didn't actually *do* much at first. "People would come around and talk a bunch of bullshit for a couple days," one Google Ideas conference attendee remembers. "Nothing came out of it."

But slowly, the group's lofty challenges began to attract engineers, some joining from other parts of Google after volunteering for Cohen's team. One of their first creations was a tool called uProxy that allows anyone whose Internet access is censored to bounce their traffic through a friend's connection outside the firewall; it's now used in more than 100 countries. Another tool, a Chrome add-on called Password Alert, aims to block phishing by warning people when they're retyping their Gmail password into a malicious look-alike site; the company developed it for Syrian activists targeted by government-friendly hackers, but when it proved effective, it was rolled out to all of Google's users.

"We are not going to be one of those groups that just *imagines* what vulnerable populations are experiencing. We're going to get to know our users."

In February, the group was renamed Jigsaw to reflect its focus on building practical products. A program called Montage lets war correspondents and nonprofits crowdsource the analysis of YouTube videos to track conflicts and gather evidence of human rights violations. Another free service called Project Shield uses Google's servers to absorb government-sponsored cyberattacks intended to take down the websites of media, election-monitoring, and human rights organizations. And an initiative, aimed at deradicalizing ISIS recruits, identifies would-be jihadis based on their search terms, then shows them ads redirecting them to videos by former extremists who explain the downsides of joining an ultraviolent, apocalyptic cult. In a pilot project, the anti-ISIS ads were so effective that they were in some cases two to three times more likely to be clicked than typical search advertising campaigns.

The common thread that binds these projects, Cohen says, is a focus on what he calls "vulnerable populations." To that end, he gives new hires an assignment: Draw a scrap of paper from a baseball cap filled with the names of the world's most troubled or repressive countries; track down someone under threat there and talk to them about their life online. Then present their stories to other Jigsaw employees.

At one recent meeting, Cohen leans over a conference table as 15 or so Jigsaw recruits—engineers, designers, and foreign policy wonks—prepare to report back from the dark corners of the Internet. "We are not going to be one of those groups that sits in our offices and imagines what vulnerable populations around the world are experiencing," Cohen says. "We're going to get to know our users." He speaks in a fast-forward, geeky patter that contrasts with his blue-eyed, broad-shouldered good looks, like a politician disguised as a Silicon Valley executive or vice versa. "Every single day, I want us to feel the burden of the responsibility we're shouldering."

"Jigsaw recruits will hear stories about people being tortured for their passwords or of state-sponsored cyberbullying."

We hear about an Albanian LGBT activist who tries to hide his identity on Facebook despite its real-names-only policy, an administrator for a Libyan youth group wary of government infiltrators, a defector's memories from the digital black hole of North Korea. Many of the T-shirt-and-sandal--

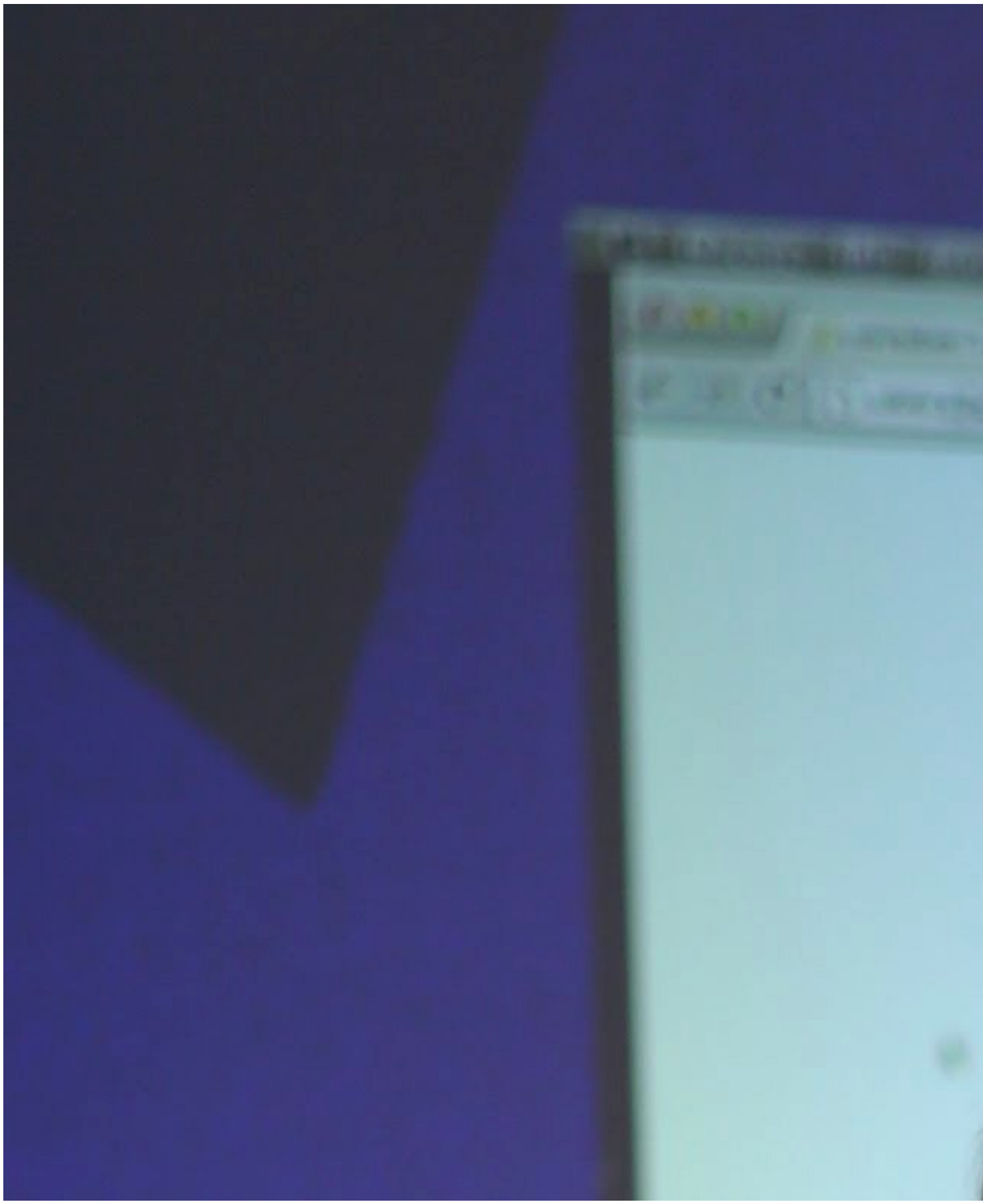
wearing Googlers in the room will later be sent to some of those far-flung places to meet their contacts face-to-face.

“They’ll hear stories about people being tortured for their passwords or of state-sponsored cyberbullying,” Cohen tells me later. The purpose of these field trips isn’t simply to get feedback for future products, he says. They’re about creating personal investment in otherwise distant, invisible problems—a sense of investment Cohen says he himself gained in his twenties during his four-year stint in the State Department, and before that during extensive travel in the Middle East and Africa as a student.

Cohen reports directly to Alphabet’s top execs, but in practice, Jigsaw functions as Google’s blue-sky, human-rights-focused skunkworks. At the group’s launch, Schmidt declared its audacious mission to be “tackling the world’s toughest geopolitical problems” and listed some of the challenges within its remit: “money laundering, organized crime, police brutality, human trafficking, and terrorism.” In an interview in Google’s New York office, Schmidt (now chair of Alphabet) summarized them to me as the “problems that bedevil humanity involving information.”

Jigsaw, in other words, has become Google’s Internet justice league, and it represents the notion that the company is no longer content with merely not being evil. It wants—as difficult and even ethically fraught as the impulse may be—to do good.

-



Yasmin Green, Jigsaw's head of research and development.Jigsaw

In September of 2015, Yasmin Green, then head of operations and strategy for Google Ideas, the working group that would become Jigsaw, invited 10 women who had been harassment victims to come to the office and discuss their experiences. Some of them had been targeted by members of the antifeminist Gamergate movement. Game developer Zoë Quinn had been threatened repeatedly with rape, and her attackers had dug up and distributed old nude photos of her. Another visitor, Anita Sarkeesian, had moved out of her home temporarily because of numerous death threats.

At the end of the session, Green and a few other Google employees took a photo with the women and posted it to the company's Twitter account. Almost immediately, the Gamergate trolls turned their ire against Google itself. Over the next 48 hours, tens of thousands of comments on Reddit and Twitter demanded the Googlers be fired for enabling "feminazis."

"It's like you walk into Madison Square Garden and you have 50,000 people saying you suck, you're horrible, die," Green says. "If you really believe that's what the universe thinks about you, you certainly shut up. And you might just take your own life."

To combat trolling, services like Reddit, YouTube, and Facebook have for years depended on users to flag abuse for review by overworked staffers or an offshore workforce of content moderators in countries like the Philippines. The task is expensive and can be scarring for the employees who spend days on end reviewing loathsome content—yet often it's still not enough to keep up with the real-time flood of filth. Twitter recently introduced new filters designed to keep users from seeing unwanted tweets, but it's not yet clear whether the move will tame determined trolls.

The meeting with the Gamergate victims was the genesis for another approach. Lucas Dixon, a wide-eyed Scot with a doctorate in machine learning, and product manager CJ Adams wondered: Could an abuse-detecting AI clean up online conversations by detecting toxic language—with all its idioms and ambiguities—as reliably as humans?

Show millions of vile Internet comments to Google's self-improving artificial intelligence engine and it can recognize a troll.

To create a viable tool, Jigsaw first needed to teach its algorithm to tell the difference between harmless banter and harassment. For that, it would need a massive number of examples. So the group partnered with *The New York Times*, which gave Jigsaw's engineers 17 million comments from *Times* stories, along with data about which of those comments were flagged as inappropriate by moderators. Jigsaw also worked with the Wikimedia Foundation to parse 130,000 snippets of discussion around Wikipedia pages. It showed those text strings to panels of 10 people recruited randomly from the CrowdFlower crowdsourcing service and asked whether they found each snippet to represent a "personal attack" or "harassment." Jigsaw then fed the massive corpus of online conversation and human evaluations into Google's open source machine learning software, TensorFlow.

Machine learning, a branch of computer science that Google uses to continually improve everything from Google Translate to its core search engine, works something like human learning. Instead of programming an algorithm, you teach it with examples. Show a toddler enough shapes identified as a cat and eventually she can recognize a cat. Show millions of vile Internet comments to Google's self-improving artificial intelligence engine and it can recognize a troll.

In fact, by some measures Jigsaw has now trained Conversation AI to spot toxic language with impressive accuracy. Feed a string of text into its Wikipedia harassment-detection engine and it can, with what Google describes as more than 92 percent certainty and a 10 percent false-positive rate, come up with a judgment that matches a human test panel as to whether that line represents an attack. For now the tool looks only at the content of that single string of text. But Green says Jigsaw has also looked into detecting methods of mass harassment based on the volume of messages and other long-term patterns.

Wikipedia and the *Times* will be the first to try out Google's automated harassment detector on comment threads and article discussion pages. Wikimedia is still considering exactly how it will use the tool, while the *Times* plans to make Conversation AI the first pass of its website's comments, blocking any abuse it detects until it can be moderated by a human. Jigsaw will also make its work open source, letting any web forum or social media platform adopt it to automatically flag insults, scold harassers, or even auto-delete toxic language, preventing an intended harassment victim from ever seeing the offending comment. The hope is that "anyone can take these models and run with them," says Adams, who helped lead the machine learning project.

Adams types in "What's up, bitch?" and clicks Score. Conversation AI instantly rates it a 63 out of 100 on the attack scale.

What's more, some limited evidence suggests that this kind of quick detection can actually help to tame trolling. Conversation AI was inspired in part by an experiment undertaken by Riot Games, the video-game company that runs the world's biggest multiplayer world, known as *League of Legends*, with 67 million players. Starting in late 2012, Riot began using machine learning to try to analyze the results of in-game conversations that led to players being banned. It used the resulting algorithm to show players in real time when they had made sexist or abusive remarks. When players saw immediate automated warnings, 92 percent of them changed their behavior for the better, according to a report in the science journal *Nature*.

My own hands-on test of Conversation AI comes one summer afternoon in Jigsaw's office, when the group's engineers show me a prototype and invite me to come up with a sample of verbal filth for it to analyze. Wincing, I suggest the first ambiguously abusive and misogynist phrase that comes to mind: "What's up, bitch?" Adams types in the sentence and clicks Score. Conversation AI instantly rates it a 63 out of 100 on the attack scale. Then, for contrast, Adams shows me the results of a more clearly vicious phrase: "You are such a bitch." It rates a 96.

In fact, Conversation AI's algorithm goes on to make impressively subtle distinctions. Pluralizing my trashy greeting to "What's up bitches?" drops the attack score to 45. Add a smiling emoji and it falls to 39. So far, so good.

But later, after I've left Google's office, I open the Conversation AI prototype in the privacy of my apartment and try out the worst phrase that had haunted Sarah Jeong: "I'm going to rip each one of her hairs out and twist her tits clear off." It rates an attack score of 10, a glaring oversight. Swapping out "her" for "your" boosts it to a 62. Conversation AI likely hasn't yet been taught that threats don't have to be addressed directly at a victim to have their intended effect. The algorithm, it seems, still has some lessons to learn.

For a tech executive taking on would-be terrorists, state-sponsored trolls, and tyrannical surveillance regimes, Jigsaw's creator has a surprisingly sunny outlook on the battle between the people who use the Internet and the authorities that seek to control them. "I have a fundamental belief that technology empowers people," Jared Cohen says. Between us sits a coffee table covered in souvenirs from his travels: a clay prayer coin from Iraq, a plastic-wrapped nut bar from Syria, a packet of North Korean cigarettes. "It's hard for me to imagine a world where there's not a continued cat-and-mouse game. But over time, the mouse might just become bigger than the cat."

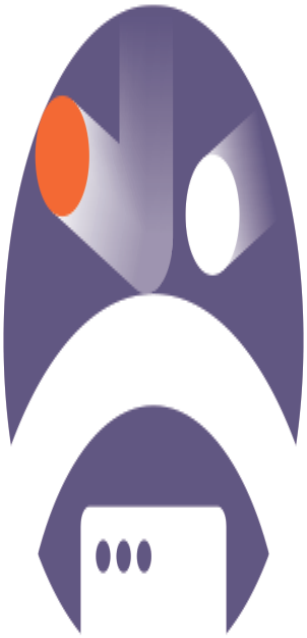
Jigsaw's Projects

The incubator is dedicated to geopolitical moon shots, tackling issues from online censorship to violent extremism. Here are a few of its efforts. —Gregory Barber

uProxy

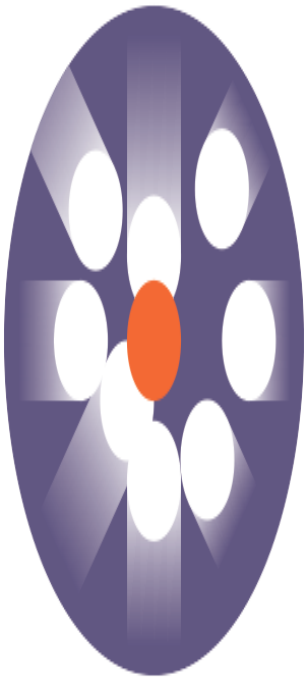
A Chrome browser buddy system that lets any censored Internet user route around the firewall by using a friend's unblocked connection.

Project Shield



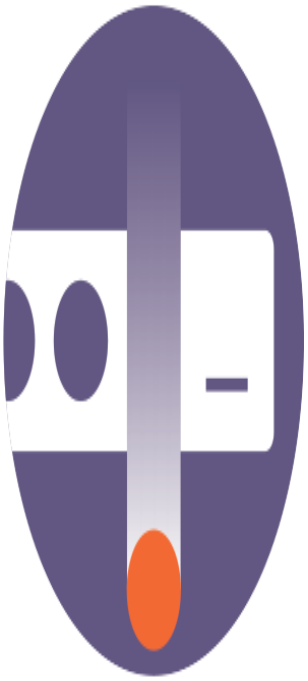
Free protection for media, election monitors, and human rights groups to defend themselves against cyberattacks aimed at taking down websites.

Montage



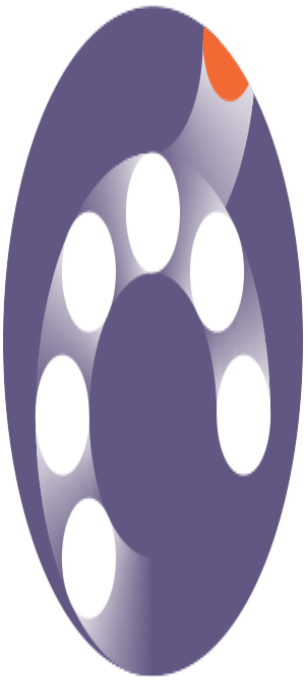
Crowdsourced analysis of YouTube videos to help journalists and humanitarian groups document conflict and human rights violations.

Password Alert



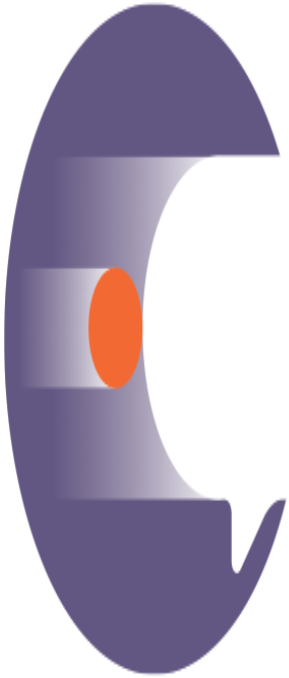
Warns people when they type a Gmail password into a phishing website mocked up to look like one of Google’s.

The Redirect Method



Identifies would-be jihadis based on search terms and redirects them to anti-ISIS videos featuring former extremists.

Conversation AI



A filter for online discussion that uses machine learning to automatically detect insults or hate speech.

Digital Attack Map

A real-time visualization of DDoS cyberattacks around the world, including those where freedom of expression is being limited.

That sense of digital populism, as Cohen tells it, was instilled in him during his travels through Iran, Syria, Lebanon, and Iraq in the early 2000s as a Rhodes scholar. His most formative memories from that time are of watching young people use technology— cell phones everywhere, gay-nightclub promoters in Iran sending text messages to strangers via Bluetooth, and satellite TV blanketing the

region with otherwise-censored Western culture. He was particularly struck by the time he spent with two Internet-savvy, cell-phone-obsessed young Syrian women in Homs who acted as his hosts, walked in public with him—an American man—and wore makeup and short-sleeved shirts amid the burkas and disapproving stares surrounding them. “Unlike their mothers, these girls know what they’re missing out on,” he’d write in a book about his travels, *Children of Jihad*. “Society has changed, and technology has opened their eyes in ways that their parents cannot begin to understand.”

When Cohen became the youngest person ever to join the State Department’s Policy Planning Staff in 2006, he brought with him a notion that he’d formed from seeing digitally shrewd Middle Eastern youths flout systems of control: that the Internet could be a force for political empowerment and even upheaval. And as Facebook, then YouTube and Twitter, started to evolve into tools of protest and even revolution, that theory earned him access to officials far above his pay grade—all the way up to secretaries of state Condoleezza Rice and later Hillary Clinton. Rice would describe Cohen in her memoirs as an “inspired” appointment. Former Policy Planning director Anne-Marie Slaughter, his boss under Clinton, remembers him as “ferociously intelligent.”

Many of his ideas had a digital twist. After visiting Afghanistan, Cohen helped create a cell-phone-based payment system for local police, a move that allowed officers to speed up cash transfers to remote family members. And in June of 2009, when Twitter had scheduled downtime for maintenance during a massive Iranian protest against hardliner president Mahmoud Ahmadinejad, Cohen emailed founder Jack Dorsey and asked him to keep the service online. The unauthorized move, which violated the Obama administration’s noninterference policy with Iran, nearly cost Cohen his job. But when Clinton backed Cohen, it signaled a shift in the State Department’s relationship with both Iran and Silicon Valley.

Around the same time, Cohen began calling up tech CEOs and inviting them on tech delegation trips, or “techdels”—conceived to somehow inspire them to build products that could help people in repressed corners of the world. He asked Google’s Schmidt to visit Iraq, a trip that sparked the relationship that a year later would result in Schmidt’s invitation to Cohen to create Google Ideas. But it was Cohen’s email to Twitter during the Iran protests that most impressed Schmidt. “He wasn’t following a playbook,” Schmidt tells me. “He was inventing the playbook.”

The story Cohen’s critics focus on, however, is his involvement in a notorious piece of software called Haystack, intended to provide online anonymity and circumvent censorship. They say Cohen helped to hype the tool in early 2010 as a potential boon to Iranian dissidents. After the US government fast-tracked it for approval, however, a security researcher revealed it had egregious vulnerabilities that put any dissident who used it in grave danger of detection. Today, Cohen disclaims any responsibility for Haystack, but two former colleagues say he championed the project. His former boss Slaughter describes his time in government more diplomatically: “At State there was a mismatch between the scale of Jared’s ideas and the tools the department had to deliver on them,” she says. “Jigsaw is a much better match.”

But inserting Google into thorny geopolitical problems has led to new questions about the role of a multinational corporation. Some have accused the group of trying to monetize the sensitive issues they’re taking on; the Electronic Frontier Foundation’s director of international free expression, Jillian York, calls its work “a little bit imperialistic.” For all its altruistic talk, she points out, Jigsaw is part of

a for-profit entity. And on that point, Schmidt is clear: Alphabet hopes to someday make money from Jigsaw's work. "The easiest way to understand it is, better connectivity, better information access, we make more money," he explains to me. He draws an analogy to the company's efforts to lay fiber in some developing countries. "Why would we try to wire up Africa?" he asks. "Because eventually there will be advertising markets there."

"We're not a government," Eric Schmidt says slowly and carefully. "We're not engaged in regime change. We don't do that stuff."

Wikileaks founder Julian Assange has accused Cohen of continuing to work as a de facto State Department employee, quietly advancing the government's foreign policy goals from within Google, and labeled him the company's "director of regime change." When I raise that quote with Schmidt, he visibly tenses, then vehemently rejects the notion. "We're not a government," he says slowly and carefully. "We're not engaged in regime change. We don't do that stuff. But if it turns out that empowering citizens with smartphones and information causes changes in their country … you know, that's probably a good thing, don't you think?"

Beyond the issue of Jigsaw's profit motives or imagined government ties, however, another point nags at Cohen's optimistic digital interventionism: Technology has unintended consequences. A tool like Haystack that was intended to help Iranians could have put them in danger. Twitter, with all its revolutionary potential, enabled new forms of abuse. And Conversation AI, meant to curb that abuse, could take down its own share of legitimate speech in the process.

During her worst days of being targeted by a gang of misogynists last year, feminist writer Sady Doyle would look down at her phone after an hour and find a hundred new Twitter notifications, many of them crude sexual comments and attacks on her history of mental health issues. But when I present her with the notion of Conversation AI as a solution, she hesitates. "People need to be able to talk in whatever register they talk," she says. "Imagine what the Internet would be like if you couldn't say 'Donald Trump is a moron.'" In fact, when I run the phrase though the Conversation AI prototype, I find that calling someone a moron scores a full 99 out of 100 on its personal attack scale.

The example highlights Conversation AI's potential for false positives or suppressing the gray areas of speech. After all, even without automated flagging, Twitter and Facebook have been criticized for blocking legitimate, even politically powerful, content: Last year Twitter banned Politwoops, a feed that collected the deleted tweets of political figures to catch damning off-the-cuff statements. Facebook blocked photos of drowned migrant children intended to make Americans more aware of the tragedy of Syria's refugee crisis.

Sarah Jeong, the *Motherboard* writer who was silenced by Bernie bros, says she supports the notion of Conversation AI, in theory. "The Internet needs moderation," she says. But she warns that no one should be foolish enough to let Conversation AI run wild with automated comment deletion: "These are human interactions." Any fix for the worst of those interactions, she says, will need to be human too.

"An automated detection system can open the door to the delete-it-all option," adds Emma Llansó, director of the Free Expression Project at the nonprofit Center for Democracy and Technology, "rather than spending the time and resources to identify false positives."

My tests of Conversation AI do in fact produce outright false positives. "I shit you not" somehow got an attack score of 98 out of 100, the same as the far more offensive "you are shit." The rather harmless

phrase “you suck all the fun out of life” scored a 98, just a point shy of “you suck.” And most problematic of all, perhaps: “You are a troll”—the go-to response for any troll victim—was flagged with an attack score of 93.

“When you’re looking at curbing online harassment and at free expression, there’s a tension between the two. We don’t claim to have all the answers.”

Throwing out well-intentioned speech that *resembles* harassment could be a blow to exactly the open civil society Jigsaw has vowed to protect. When I ask Conversation AI’s inventors about its potential for collateral damage, the engineers argue that its false positive rate will improve over time as the software continues to train itself. But on the question of how its judgments will be enforced, they say that’s up to whoever uses the tool. “We want to let communities have the discussions they want to have,” says Conversation AI cocreator Lucas Dixon. And if that favors a sanitized Internet over a freewheeling one? Better to err on the side of civility. “There are already plenty of nasty places on the Internet. What we can do is create places where people can have better conversations.”

On a muggy morning in June, I join Jared Cohen at one of his favorite spots in New York: the Soldiers’ and Sailors’ Monument, an empty, expansive, tomblike dome of worn marble in sleepy Riverside Park. When Cohen arrives, he tells me the place reminds him of the quiet ruins he liked to roam during his travels in rural Syria.

Our meeting is in part to air the criticisms I’ve heard of Conversation AI. But when I mention the possibility of false positives actually censoring speech, he answers with surprising humility. “We’ve been asking these exact questions,” he says. And they apply not just to Conversation AI but to everything Jigsaw builds, he says. “What’s the most dangerous use case for this? Are there risks we haven’t sufficiently stress-tested?”

Jigsaw runs all of its projects by groups of beta testers and asks for input from the same groups it intends to recruit as users, he says. But Cohen admits he never knows if they’re getting enough feedback, or the right kind. Conversation AI in particular, he says, remains an experiment. “When you’re looking at curbing online harassment and at free expression, there’s a tension between the two,” he acknowledges, a far more measured response than what I’d heard from Conversation AI’s developers. “We don’t claim to have all the answers.”

And if that experiment fails, and the tool ends up harming the exact free speech it’s trying to protect, would Jigsaw kill it? “Could be,” Cohen answers without hesitation.

I start to ask another question, but Cohen interrupts, unwilling to drop the notion that Jigsaw’s tools may have unintended consequences. He wants to talk about the people he met while wandering through the Middle East’s most repressive countries, the friends who hosted him and served as his guide, seemingly out of sheer curiosity and hospitality.

It wasn’t until after Cohen returned to the US that he realized how dangerous it had been for them to help him or even to be seen with him, a Jewish American during a peak of anti-Americanism. “My very presence could have put them at risk,” he says, with what sounds like genuine throat-tightening emotion. “To the extent I have a guilt I act on, it’s that. I never want to make that mistake again.”

Cohen still sends some of those friends, particularly ones in the war-torn orbit of Syria and ISIS, an encrypted message almost daily, simply to confirm that they’re alive and well. It’s an exercise, like the

one he assigns to new Jigsaw hires but designed as maintenance for his own conscience: a daily check-in to assure himself his interventions in the world have left it better than it was before.

“Ten years from now I’ll look back at where my head is at today too,” he says. “What I got right and what I got wrong.” He hopes he’ll have done good.

Andy Greenberg (@a_greenberg) wrote about cryptographer Moxie Marlinspike in issue 24.08.

This article appears in the October issue.

Grooming by Veronica Velez / Aubri Balk

Go Back to Top. Skip To: Start of Article.

- #magazine-24.10

Skip Social. Skip to: Latest News.

- **Share**
- Share 56
- Tweet
- Pin 3
- Comment 2
- Email

Skip Comments. Skip to: Footer. View comments

- Comments
- **Community**
- Login
- 1
- Recommend
- Sort by Best

Join the discussion…

- –
-

hammr25 • 30 minutes ago

The link on your front page doesn't work although I was able to use google to find the article anyway. Oh and I laugh when a company says their computer algorithm will be better than humans at detecting human behavior.

- △▽
- •
- Reply
- •
- Share ›
    -
-
- –
- ⚑

Bri • an hour ago

> “We’re not a government,” he says slowly and carefully. “We’re not engaged in regime change. We don’t do that stuff. But if it turns out that empowering citizens with smartphones and information causes changes in their country … you know, that’s probably a good thing, don’t you think?”

Horse fud! Cohen should be on trial for aiding and abetting the US government's illegal attempts to overthrow Assad in Syria! The guy violated international law and basically fomented the US led resistance in Syria.

- △▽
- •
- Reply
- •
- Share ›
    -
-

- Powered by Disqus
- ✉Subscribe
- d Add
- 🔒 Privacy