National Endowment for Democracy
Supporting freedom around the world

forum
International Forum for Democratic Studies

FORUM Q&A: JONATHON MORGAN ON TRACKING DIGITAL DISINFORMATION

Jonathon Morgan (@jonathonmorgan) is the CEO of NewKnowledge.io, a data science firm dedicated to defending the integrity of public discourse by working with governments to protect citizens from computational propaganda and by helping brands identify and counter information attacks. Jonathon has spent the past twelve years building new technologies and digital products. Previously, he has published research with the Brookings Institution and served as a Special Advisor to the State Department; he also leads Data for Democracy, a volunteer collective of 1,800 data scientists and technologists.

As the challenge of authoritarian disinformation on the internet has become more salient, a community of researchers and technologists has formed to assess the problem and craft solutions. Various techniques for social media analysis have shed light on the issue from various angles. Learning from the fight against online extremism, policymakers and civil society are beginning to apply past lessons to digital disinformation.

Dean Jackson of the International Forum for Democratic Studies spoke with Jonathon Morgan about his work on understanding ISIS's Twitter following, the best ways to study the spread of Russian disinformation on social media, and how civil society and technologists can combine with other stakeholders to preserve the online public sphere. (This interview has been condensed and edited for clarity. The views and opinions expressed here do not necessarily reflect those of the National Endowment for Democracy.)

Dean Jackson: I wanted to start out by asking you about a March 2017 piece you wrote for Data for Democracy, an online community of civic tech activists you created in late 2016. In that piece, you discuss Russian amplification of online purveyors of disinformation, extremist far-right internet communities, and even secessionist movements. How would you describe, in broad strokes, the relationship between Russian information operations and extremist far-right online activity?

Jonathon Morgan: I think the interests of the extremist far-right and the Russian state can be opportunistically aligned. From the Russian perspective, their overarching strategy seems to be selling discord, finding vulnerabilities, finding existing discord, finding small pockets of real polarization, amplifying them, and pushing on pressure points and then seeing what breaks. We see a lot of ideas in extremist far-right communities in less moderated parts of the internet that bubble up in these little meme factories before transitioning into more mainstream spaces where they're ultimately amplified by likely state-driven networks with a high degree of automation. That level of automation and that level of sophistication is usually the thing that signals that there's somebody grown-up in the room and somebody with more resources to bring to bear, which is more in state-actor territory.

How are you able to identify if the actor you suspect is behind more sophisticated attempts? Could you give me a very quick crash course?

There are a few different types of things that are important to identify. One is actually injecting new language into a community by forcing some conversation. And that manifests in the ways that you would expect: the community actually starts to use different sorts of language over time or uses words in a different way. Once you find the set of accounts that are forcing the issue and you scratch the surface a little bit, it turns out that most of them have all the hallmarks of persona accounts—accounts that are sock puppets or that may be a single human operator who might manage many hundreds or sometimes thousands of these accounts. They operate them like a human,

but they are trying to orchestrate and manufacture social consensus by manually spreading messages into a space. So that's one type of actor that we identify.

The second [factor] is sort of more pure automation. These are basically spam bots that, depending on the platform dynamics, are there to amplify the content of the personas and sock puppets. Then, they ultimately cast a wider net and make it seem as if the social consensus is actually quite broad, so that every time you're in a particular type of conversation, you hear a certain point of view from a certain perspective and even the same language. So it starts to become absorbed into the larger conversation. They move from seeding it and amplifying it to ultimately trying to infect the broader population with the ideas that they want to spread. Those are the big things that we look for. And that type of orchestration requires software systems that can operate at platform scale, like at Facebook-size scale, which makes it difficult to imagine the proverbial "guy in his mom's basement" who could be behind attacks of that sophistication.

You've been involved in creating Hamilton 68, which is a German Marshall Fund-supported project devoted to tracking Russian influence over social media. Can you tell us a little bit more about that and the techniques it uses to eliminate the role that Russian social media plays in online discourse in the communities you've analyzed? What has the project found so far?

That project is in conjunction with a few really smart domain experts who understand both the larger context of Russian influence operations and also the dynamics of social media and how different online platforms are manipulated, Twitter in particular. My colleagues on that are Clint Watts, J.M. Berger, and Andrew Weisberd, and it involves a combination of multiple techniques. Previous research conducted by J.M. and myself, for instance, focused on ISIS and other extremist groups trying to manipulate public discourse on Twitter. We developed some techniques for determining who matters in a conversation... Using some of those techniques, we've identified a subset of accounts that we're very confident are core to furthering the Russian narrative in response to mainstream events. We've been monitoring those accounts, analyzing their language, breaking down the main ideas, and representing them in a way that's hopefully consumable by the public.

What we saw is that there's occasionally a lot of focus on what Russia is trying to accomplish with this larger operation that they ran during 2016 and are still running today. But we didn't see anybody focusing on what was happening every day and the daily output of these types of accounts. It's sometimes just having a quick resource for understanding what that conversation is or what that perspective is, which is really valuable. That was Clint's big idea and he brought the right folks to the table, and we

had some domain experts with the technical chops to make sure that we put together a package in a way that was easy for somebody to drive by and see what we've talked about.

I understand there are three tiers of accounts that the product tracks?

Right, we were also trying to understand the relationship between the overt, state-sponsored media narrative and the larger media narrative that they are trying to drive. That's through official accounts, state-sponsored publications like RT and Sputnik, and then what we call covert influence, which are accounts that walk and talk like every day Americans, but are in fact coordinating around a message that's consistent with what the official state-sponsored accounts are publishing. So there's this attributed/unattributed, covert/overt approach that I think really speaks to the sophistication of the actor, in that they can operate on multiple platforms using different tactics that are appropriate for driving a narrative across the entire country.

Going back in time a little bit, in 2015, you co-authored a report for Brookings called the "ISIS Twitter Census." Today, a lot of your work focuses on the extremist far-right and not Islamist extremism. What is it about these two groups that lends itself to similar forms of analysis?

What's always been difficult about the way that people want to fight with groups like ISIS or other kinds of violent extremist groups is that it's really difficult to know when you're winning a war of ideas. And what that led us to do was look at the mechanics of online dialogue. What does it actually mean to change the conversation? How can you measure the difference in conversation today versus yesterday versus last week versus last month? Can you measure the amount of ideology that's present in a community? Could you directly observe the process of radicalization and what that would look like? So we started to explore a lot of techniques and building technologies that might allow us to do that.

Once we got down to that mechanical level, it started to become clear that ideology was pretty much, from a technical perspective, a strong deviation from the mean. So you can, from our point of view, be an extremist about anything: we like to joke that there are extremist Philadelphia Eagles fans, extremist Justin Bieber fans, and extremist fans of *House of Cards*. I think that none of those things are problems. We expect people to have an especially passionate, almost uncomfortably passionate, attachment to things that operate in a cultural space. That's not abnormal, but when those things are associated with ideas in groups that are prone to political violence, then that's the type of extremism that we're interested in observing and ultimately combating. And because

we're looking at it more on a mechanical level and a foundational level, that applies to anybody. We see the same tactics, same movement, same shifts in ideology over time so that you can say, here's an ideology that we're concerned about. How is that ideology infecting a larger population? And how is dynamic changing over time? Who are the primary actors? Who's influencing the conversation? What's the likelihood that it'll lead to violence? These are all questions that we can answer in theory, about any group, but it was really spurred by this interest in having a more objective, outcome-focused approach to countering extremist groups like ISIS.

The world we live in now looks much different than what we were promised in the 1990s when internet access was first expanding rapidly. I want to ask: Can we salvage the internet? Is the relationship between the internet and political freedom always going to be increasingly tenuous? Can we get back to where we thought we were going to be?

The answer is yes, I think we can salvage this, but the first step is accepting that we have a different internet than we were expecting. The internet that we were expecting was an almost purely democratic system, where the wisdom of the crowd pushed the best ideas to the top of the pile.

I don't think that's the reality that we live in. The distribution of information is now governed by effectively a cartel. We have two or three media organizations controlling the flow of information, and not just disseminating it freely person to person, but interjecting their own ethics and their own editorial perspective through the algorithms that they write that choose what information they surface to us. We need to think about how the public, how civil society, and how the government interacts with the system that operates in that way. And I think it's a difficult conversation, it's one the tech industry doesn't want to have, and it's one that government and civil society don't really know how to have. And we only focus on it in times of crisis, like when it has been manipulated and exploited, such as in 2016. As much as I'd like to cling to the idea that information should flow freely point-to-point, person-to-person, and that the best ideas evolve organically from all members participating equally in fluid discussion, that's just not how it turned out.

I think we have to accept the reality and accept the situation at hand, and then think about how we want to share information as a society in the same way that we did when we started broadcasting news directly into people's homes via television, or the same way that we hold other forms of information dissemination to a higher standard than being able to publish whatever they want. I think that's a real sea change. Technologists need to accept that the internet's not what they wanted, and the public may need to be educated about the way that they consume information. We're really the first generation

that has had to deal with this and grapple with these questions. We'll get better over the next five or ten years, but the conversation is starting and I hope that civil society, government, and the tech industry can have it in good faith.

Your website has an intriguing tagline, "AI for Cognitive Security." That term, cognitive security, is one that I see more and more and I wondered if maybe you could explain what cognitive security is, what you're securing, what you're securing us from, and how AI can help with that?

Cognitive security emerged to encompass this idea that we need to protect our information systems in the same way that we protect our networking systems. There's this idea of cybersecurity that's effectively, don't break into my house, or I'm going to try prevent you from breaking into my house. I think cognitive security, in this context, is a lot more about "don't manipulate my community." And as a society, we accept certain types of manipulation. We're more than comfortable with public debate, we're more than comfortable with advertising and marketing. We understand that sometimes that's subtle and subversive, but we've come to a consensus that it's okay. But what the reaction to the election cycle in 2016 in the U.S. shows is that we're not comfortable with actors pretending to be human beings and pretending that there is a ground swell of support for ideas. We don't want systems imitating humans in a way that is unattributed.

Especially when the systems are controlled by actors outside of our own political system.

Jonathon Morgan: Right, particularly in that case, that's a bridge too far. So we decided that that's not okay. And ultimately, that's what we're protecting. We're protecting citizens and we're protecting institutions and corporations from being manipulated by this kind of coordinated, automated, and orchestrated activity that is difficult to detect, unattributed, and usually counter to the public interest. The way that AI supports that is that these are fairly sophisticated operations, they operate on multiple platforms, and the dynamic of these campaigns is always changing. You almost need to be in a thousand different places at once to observe it happening. AI can directly observe small communities, can observe dozens of different communities at the same time, can understand the dynamics of the relationships between individuals, and can understand the language that those communities are using to describe the world in that perspective. It helps see manipulation at a scale that's too large for any one person to observe directly. That's why AI techniques are so effective: because it is difficult for any one person to observe directly.

So it takes the 10,000 foot view?

Jonathon Morgan: It's as if you could rise above and look down on the Facebook and Twitter ecosystems, or the global social ecosystem, and observe how it's being shifted and influenced, then manipulated in ways that wouldn't happen if it were just patterns of human beings communicating with each other in a reasonable and normal social way.