JUST SECURITY

Extreme Vetting by Algorithm

Faiza Patel

November 20, 2017

Last week, a group of machine learning and data mining experts wrote to the acting secretary of DHS urging her to reconsider an automated Extreme Vetting Initiative being proposed by Immigration and Customs Enforcement (ICE). Simultaneously, civil society groups (including the Brennan Center where I work) released a letter urging the department to abandon the initiative as discriminatory and a threat to freedom of speech and assembly.

There are fairly obvious fundamental flaws with the proposed program, which came to light when *The Intercept* reported that at a July 2017 "Industry Day" the agency sought input from the private sector about developing a automated process for the "current manual vetting process" that would evaluate whether a visa applicant: 1) would become "positively contributing member of society;" 2) had the ability to "contribute to the national interests;" and 3) "intends to commit criminal or terrorist acts after entering the United States."

The evaluation would be based on information about potential visitors from publicly accessible platforms such as "media, blogs, public hearings, conferences, academic websites, social media websites…radio, television, press, geospatial sources, internet sites" as well as information held in government databases.

The inherent subjectivity of measuring whether someone will contribute positively to society or the national interest is obvious. As the tech experts pointed out:

As a result, "[a]lgorithms designed to predict these undefined qualities could be used to arbitrarily flag groups of immigrants under a veneer of objectivity."

Perhaps less obvious – but equally well-established – is the inadequacy of automated tools to predict whether an individual intends to commit terrorist or criminal acts after entering the United States. As the tech experts point out:

[T]here is a wealth of literature demonstrating that even the 'best" automated decision-making models generate an unacceptable number of errors when predicting rare events. On the scale of the American population and immigration rates, criminal acts are relatively rare, and terrorist acts are extremely rare. … As a result, even the most accurate possible model would generate a very large number of false positives – innocent individuals falsely identified as presenting a risk of crime or terrorism who would face serious repercussions not connected to their real level of risk.

For context, a Cato Institute Study shows that over the past forty years the United States issued 7.38 million for each one issued to a terrorist, amounting to a near-zero statistically insignificant 0.0000136 percent. Indeed, even in the domestic context – where we are dealing with a much larger number of events – predictive tools have inevitably failed to live up their hype.
It is no surprise that automated tools have difficulty making predictive judgments about human behavior. As the civil society letter notes, "[t]he meaning of content posted on social media is highly context-dependent. Errors in *human* judgment about the real meaning of social media posts are common." For example, in a widely-reported incident, two British tourists were detained overnight at Los Angeles airport because agents were concerned about social media postings in which he said he was going to "destroy America" (apparently slang for partying) and was planning to "dig up Marilyn Monroe's grave" (apparently a joke). Algorithms do even worse, often struggling to "make even simple determinations, such as whether a social media post is positive, negative, or neutral."

Closely related to the issue of ineffectiveness is the discriminatory impulse and potential of these types of programs. President Trump's first Muslim ban executive order directed the State Department and security agencies to develop a screening system that included "a process to evaluate the applicant's likelihood of becoming a positively contributing member of society and the applicant's ability to make contributions to the national interest; and a mechanism to assess whether or not the applicant has the intent to commit criminal or terrorist acts after entering the United States." While that language was removed from later versions of the order as part of the administration's effort to make it seem less obviously aimed at Muslims, Trump has made it clear that the original reflects his true intentions. Indeed, he has often juxtaposed "extreme vetting" as an adjunct to outright bans. Malleable concepts such as value to "society" and the "national interest" could easily be used to keep out Muslims on the theory that they present a threat to American values as this president and his inner circle clearly believe. Domestic countering violent extremism programs already use factors such as concerns about U.S. foreign policy as indicators of pre-terrorism among Muslims and would almost certainly be built into any automated vetting system as well.

Then there are the predictable effects on free speech of a system designed to snoop on people's social media and public statements looking for undefined qualities. Anyone seeking to come to the United States – whether to reunite with family, study, or conduct business – would feel pressure to censor themselves online if they thought it would affect their chances. So too would family and friends in the U.S. with whom they are communicating. Indeed, there are strong indications that ICE is considering implementing continuous vetting of this sort inside the United States as well. The program's statement of objectives identifies the failure to continuously vet permanent residents as creating "significant risk in ICE's ability to identify emerging risks, such as radicalization, that may occur after an individual arrives in the United States." Perhaps relatedly, DHS recently issued a systems of records notice indicating that it was planning to include social media information in Alien (A) files, immediately setting off alarm bellsthat the agency is continuously collecting such information on permanent residents and naturalized citizens.

All in all, there can be little doubt that this latest ICE initiative is poorly conceived, unlikely to be effective and likely to be discriminatory and to suppress speech. While these concerns are at

their height with an administration that is both openly xenophobic and contemptuous of constitutional norms, these types of programs didn't start with Trump. Efforts to use social media monitoring as a screening tool for travelers date back to at least 2015 when DHS began piloting several programs in secret. In 2016, over the strong objections of civil society groups, the Obama administration began asking travelers from visa waiver countries to provide their social media handles, presumably so they could be analyzed – although it is not clear for what. These efforts have gone full steam ahead even though there is no evidence that they work. Indeed, as the DHS inspector general recently pointed out, the agency has failed to measure whether its social media monitoring programs are effective. Companies that develop monitoring and predictive technology no doubt have an interest in upselling their products. But they should stop and consider the reputational harm they will incur when their participation in discriminatory schemes becomes known. And, it is past time for government agencies to stop drinking the Kool Aid and take seriously the difficulty of interpreting and making predictions based on what happens in the virtual world, as well as real world civil rights and civil liberties risks.