

Researchers Claim AI Can Identify Gang Members on Twitter

 motherboard.vice.com/en_us/article/researchers-claim-ai-can-identify-gang-members-on-twitter

Social media feeds contain a wealth of personal information: daily gripes, tastes in music and movies, and plans for nights out. It's no wonder that [police are interested in mining that data](#) for insights into where crime might spring up.

But can these digital artifacts, taken together, say anything deeper about who you really are? A number of experts believe so: In the near future, algorithms trained on this sort of information may make important decisions about individuals.

Here's a recent example. Researchers from the [Ohio Center of Excellence in Knowledge-enabled Computing](#) (Kno.e.sis) at Wright State University, in [a paper posted to the arXiv preprint server](#), say they've devised a deep learning AI algorithm that can identify street gang members based solely on their Twitter posts, and with 77 percent accuracy.

But, according to one expert contacted by Motherboard, this technology has serious shortcomings that might end up doing more harm than good, especially if a computer pegs someone as a gang member just because they use certain words, enjoy rap, or frequently use certain emojis—all criteria employed by this experimental AI.

The Kno.e.sis team's paper says that their algorithm is trained on a database of gang member Twitter profiles that were previously identified by the researchers. They discovered that the defining characteristics of these "gang member" profiles were: tough talk, use of "RIP," "free," and the n-word, using emojis like the gas pump—which can mean weed—and an affinity for rap. These characteristics, all taken together, are what their algorithm looks for in tweets.

The algorithm didn't take into account location information or any city or neighbourhood-specific terms or hashtags, their reasoning being that such local terms change too quickly to be useful. So, basically, if you post YouTube links to rap songs, use the n-word, employ certain emojis, and publish photos of money or other supposedly gang-related items, then the algorithm may tag you as a gang member.

"No one of these 'features' will lead us to assert one as a gang member," wrote [Amit Sheth](#), study author and executive director of Kno.e.sis, in an email. "But we exploited observations such as these: doing rap does not imply you are a gang member, but listening to gangster rap music increases a possibility that one has a gang association."

To back up this claim, the researchers [cite a paper](#) from another group of authors arguing that social media continues rap's legacy of "keeping it real," and acknowledging that rap helped form "the rebellious, assertive voice of predominantly urban youth." However, that paper does not say that listening to rap or posting rap links increases your likelihood of belonging to a gang.

"No single 'feature' such as this would identify one as a gang member, however a combination of many of these features each using known affinity to gang activity leads to the good result we report," Sheth continued.



Gakirah Barnes (right) was a 17-year-old alleged shooter for a Chicago gang who was killed this year. Her death [has been connected to threats made on Twitter](#) . Photo: [Twitter](#)

[Gang members do use social media](#) , just like everybody else—and [police know it](#) . However, any attempt to peg someone as a gang member based on social media activity should be read in the context of highly controversial initiatives like [California's CalGang database](#) . After years of adding people based on superficial indicators like tattoos and clothing, the database [contains many people who are not gang members](#) , but may be treated as such by police anyway.

Kno.e.sis's algorithm has several problems that may result in a similar situation to CalGang. In the end, the algorithm looks at a set of assumptions about what kinds of people enjoy certain music, or use certain words.

That people who listen to rap or speak a certain way must be gang members or "thugs" is a common assumption by racists, online and off. Indeed, an emerging issue in the development of artificial intelligence is that [machines trained on prejudiced data](#) tend to [reproduce those same, very human, prejudices](#) .

"It's important we think less about labels and trying to identify groups and instead focus on things that shape behaviour"

"Social media is a useful tool for understanding the environmental and situational factors that influence how and why people engage in aggressive communication," said [Desmond Patton](#), a professor at Columbia University's School of Social Work who [uses algorithms to identify at-risk youth](#), in an interview. (Patton was lead author on the paper cited by Sheth, above.) "And that might be an individual who is gang-involved, or individuals who mimic gang-like behavior because they live in neighbourhoods where gang violence has a large presence and they need to engage in survival strategies to stay safe," Patton said.

"So they may talk tough or use language that is similar to gang-involved youth to present themselves as being tough online," he continued. Of course, he also noted, while certain forms of rap like "drill" are often associated with gangs, liking or posting rap music is something that millions of people from every walk of life do every single day.

It's also problematic that the Kno.e.sis team didn't take geographic location or region-specific terms into account in their analysis, Patton said. It's more important to analyze the conditions that encourage young people to act tough online than to label them as gang members, he said, and local knowledge is key to that mission. His work at Columbia involves people who engage with the communities that they look at online.

"Youth are not inherently violent and often times the things communicated online are a function of broader issues are unfolding," Patton said, "and it's important we think less about labels and trying to identify groups and instead focus on things that shape behaviour."

CORRECTION: *An earlier version of this article stated that the Kno.e.sis team's paper was not peer reviewed, when in fact it was.*

CORRECTION: *An earlier version of this article confused the criteria for gang member self-identification in the Kno.e.sis training dataset with the classifiers pulled out by the algorithm. The Twitter accounts in the training set did in fact self-identify as being gang members in text. A section of this article dealing with issues arising from the criteria for inclusion in the dataset has been removed.*

For the Kno.e.sis team's full response to this article, [click here](#).

M