

UCLA Newsroom

SCIENCE + TECHNOLOGY

UCLA researchers teach computer to read the internet

Sifting through nearly 2 million posts, their algorithm could identify a underlying narrative

Jessica Wolf | December 02, 2016

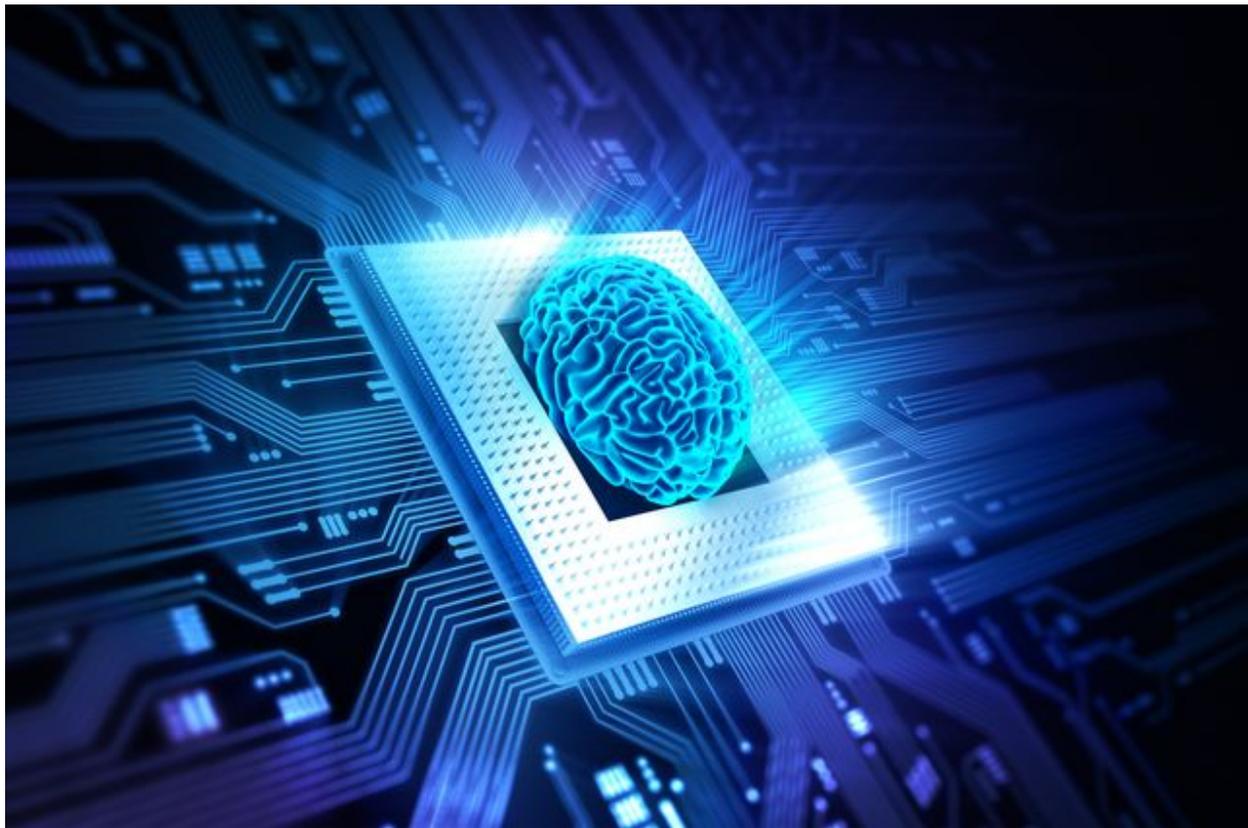
SHARE 285

TWEET

SHARE 8

EMAIL

PRINT



iStock.com/Henrik5000

UCLA researchers say machines could be taught to decipher false narratives as they proliferate.

Teaching computers to read is one thing. But by designing an algorithm that examined nearly 2 million posts from two popular parenting websites, a multidisciplinary team of UCLA researchers has built an elegant computational model that reflects how humans think and communicate, thereby teaching computers to understand structured narratives within the flow

of posts on the internet.

The researchers said their success at managing large-scale data in this way highlights the overarching potential of machine learning, and demonstrates the capability to introduce counter-narratives into internet interactions, break up echo chambers and one day potentially help root out fact from fiction for social media users.

“Our question was, could we devise computational methods to discover an emerging narrative framework underlying internet conversations that was possibly influencing the decision making of many people throughout the country or possibly world?” said Timothy Tangherlini, lead author and a self-described “computational folklorist” who teaches folklore, literature and cultural studies in the Scandinavian section of the UCLA College.

In the [study](#), published in the Journal of Medical Internet Research, Tangherlini and other researchers used sophisticated language modeling to review 1.99 million posts from two parenting sites with active user forums. They examined posts on Mothering.com — a site known to be a hub of anti-vaccine sentiment — and another parenting site (unnamed due to site privacy rules) where opinions on vaccinations were more varied. Those posts came from 40,056 users and were viewed 20.12 million times over a period of nearly nine years ending in 2012. Most users on both sites identified themselves as a mother.

“The anti-vaccine movement was a clear candidate for this type of study,” Tangherlini said. “Tens of thousands of parents were exchanging ideas about child-rearing online and, through those interactions, creating virtual communities where they could share concerns, propose methods to allay those concerns, and share their own experiences.”

The project was partially funded by a grant from the National Institutes of Health. Collaborating with Tangherlini were machine-learning expert Vwani Roychowdhury, UCLA professor of electrical engineering, and Dr. Roshan Bastani, a professor of health policy and management in the Fielding School of Public Health, and director of the UCLA Center for Prevention Research.

For the study, and based on his past scholarship in Danish folklore, Tangherlini and his colleagues came up with a broadly defined model of narrative, making that model a key part of the computational framework.

In this four-part narrative model, a story begins with an orientation, which details the type of event and the major actors in the story, such as family with a newborn infant. The second part, referred to as the complicating action, presents a threat, such as the perceived threat to the infant’s health posed by vaccination. The third part suggests a strategy to counteract that threat, such as a parent’s attempt to figure out how to avoid vaccinating. The resolution of the story evaluates the

success of the strategy in dealing with the threat.

They aligned this narrative model with nearly two million pieces of aggregated content from the parenting sites and, using natural language processing methods, were able to identify characters and the relationships between those characters, discovering the core of the underlying narratives.

On the basis of this work, they discovered that a large number of parents were not only going online to talk about vaccines, their distrust of institutions requiring them, or the perceived health risks of vaccinations, but also to seek out ways to acquire vaccination exemptions for their children.

“Stories often emerge through conversation,” Tangherlini said. “The framework of the underlying narrative emerges through time as more and more stories are circulated, negotiated, aligned and reconfigured.”

Added Roychowdhury: “It’s especially impressive, when you take into consideration the fact that all the machine was fed with, were just web pages, nothing else; and it found all the vaccine related concepts all on its own.”

While this study specifically applied to parents’ discussions about vaccination, the methods could be applied to any topic, said the researchers, who are pursuing follow up projects like incorporating a sequencing mechanism, which would track story plot.

Roychowdhury says the way we learn about how stories take shape around any given topic can be applied to targeted messaging like advertising or fighting misinformation by allowing machine learning to automatically decipher false narratives as they proliferate. For example, users exposed to particular anti-vaccination narrative could be presented with alternate narratives, based on well-tested public health paradigms, using the same extensive online advertising infrastructure currently used by the likes of Google, Facebook and Amazon.

“In public health, we have hundreds of studies trying to understand the facilitators and barriers to getting vaccinated,” Bastani said. “Our data is generally obtained through tools such as questionnaires and electronic medical records. What these tools fail to capture are the very interesting conversations that individuals are having with one another that profoundly shape their views and actions related to vaccinating their children.”

Bastani said this project was one of the most interesting she has participated in, and one that has real implications for those working in the public health field to educate parents about vaccinations.

“We hope to utilize findings from this work to design and test interventions that may positively influence vaccination rates because they are more likely to address some of the key drivers of

resistance,” she said.

SHARE ◀ 285

TWEET

SHARE ◀ 8

EMAIL

PRINT

Tags: [technology](#) [computer science](#) [behavior](#) [media](#)

Media Contact

Jessica Wolf

310-825-1046

jwolf@support.ucla.edu