

Brave new world? Robot reporters take over beats

by Benjamin Lyons

When an earthquake occurred at 6:25 a.m. on March 17, it may have given “robot” journalism its first big break. The early morning tremor allowed an algorithm created by L.A. Times programmer and journalist Ken Schwencke to report the story ahead of other outlets.

The story took only about three minutes to appear online, drawing information such as the quake’s time, magnitude and epicenter from the United States Geological Survey and inserting it into a pre-fabricated template.

This came some 40 years after the first story-writing algorithm was created at Yale University in 1977. Now, the Associated Press says it will distribute financial reports generated by Automated Insights software. These stories will draw from information produced by an investment research company.

AP, which has acquired a small stake in Automated Insights, expects the new method will allow it to publish 10 times as many earnings reports as in the past. AP Managing Editor Lou Ferrara says rather than replace human reporters, automation will free them up for more meaningful work, such as identifying trends and locating exclusives.

Algorithmic copy has been seeping into newspapers for a few years now. Ferrara said AP had already used the technology to generate sports stories from box scores. Other boilerplate news may follow.

Narrative Science, another major automated content producer, has been servicing Forbes with similar earning reports since 2011. Its co-founder, Kris Hammond, is bullish about their future. In an interview with WIRED magazine, he estimated 90 percent of what the public thinks of as news-based journalism will be written via algorithm within 15 years.

'Makes everybody's job more interesting'

Schwencke argues the technology is not a replacement, but a means for quicker dissemination for some types of news. "It's supplemental," he told Slate's Will Oremus. In addition to the Quakebot algorithm, Schwencke and the L.A. Times have used similar tools to monitor local homicides, which human editors filtered for newsworthiness.

Comparable algorithms could be used to cover high school sports, city council meetings, police blotter and other data-heavy, narrative-light stories. To do so, they need to draw upon trusted databases, such as the one maintained by the USGS that provided fodder for the L.A. Times earthquake story.

The move to robotic journalism can be seen as a polar opposite to the rise of "explainer" sites like Ezra Klein's Vox and the New York Times' Upshot. Whereas those sites make extensive use of data to probe complex relationships and contextualize current events, robotic reporting conveys simple data on a revolving basis, without tackling any "why" questions.

In its current form, algorithmic journalism poses little danger to human employment. Algorithms are incapable of handling subjective beats, such as those covering the arts, or those requiring legwork, such as investigative reporting, as Angela Washeck of 10,000 Words has noted.

Ferrara suggested the same, saying the change was a "doubling down" on the reporters on the financial beat, not a challenge to them.

"It doesn't eliminate anybody's job as much as it makes everybody's job more interesting," Schwenke said.

In his own defense, Automated Insights' CEO Robbie Allen similarly told Poynter, "We're creating content where it didn't exist before."

But the general public may have a hard time not seeing the shift as part of larger, bleaker trend toward job loss. Bloomberg reported on March 12 that half of existing U.S. occupations could be automated in the next few decades.

Known unknowns

Nieman Journalism Lab's Joshua Benton highlights a current problem with journalism algorithms: They are usually "black boxes, unknowable to the public." How do they work?

Nick Diakopoulos, a Tow Fellow at the Columbia University Journalism School,

suggests that as new, opaque power brokers, algorithms should be subject to the investigative reporting that politicians and business leaders have been in the past. This could be done technically, by "reverse engineering," or through old-fashioned interviews with their creators. After all, they are tools of human design.

Barring this, patent filings provide some basic insight. Looking at Narrative Science's filing, Diakopoulos discovered the basic decision-making process. The first step is simple: ingest large amounts of data. After reading this, the algorithm must compute its interesting features.

This is done through statistical tests — looking for large deviations from the mean, swings in value or other violations of previous predictions.

With these outliers in hand, the software then selects angles from a pre-authored library. These are basic narrative structures. Angles, matched with individual data points, can then be supplemented with additional factual content drawn from internet databases. Finally, natural language is generated using templates.

Diakopoulos argues that while Narrative Science does appear concerned with journalistic values such as newsworthiness, a broader conceptualization of this value is lacking beyond basic "deviance." They could try to mimic in code more sophisticated notions of newsworthiness, such as those found by journalism scholars Tony Harcup and Deidre O'Neill, he says.

Given more "editorial" insight, what kind of content might an algorithm produce? More interesting news, hopefully. But our own biases could lurk within.

Diakopoulos notes that since algorithms are systematic, they tend to be seen as objective. "But in the systematic application of its decision criteria," he says, "the algorithm might be introducing bias that is not obvious given its programming." Systemic biases are all the more insidious.

As news-generating algorithms become more complex, there will be more potential for this to occur.

One example is a system built by IBM Research that can summarize sports events based only on tweets about them. To know which information to include or exclude, the algorithm must prioritize it. To do so, IBM's algorithm focuses on "spiky" events, or those that receive the biggest bursts of attention. But as Diakopoulos says, other stories may "simmer below the threshold."

Logical conclusion?

Although most industry types seem to think not much will change in the wake of automated content, Kate Patrick of the Daily Caller sees one potential unintended side effect. If robots excel at the brief enumeration of facts that frees humans to pursue long form reporting, and readers already trend toward headline-only style reading, then humans' and technology's respective audience shares may become out of whack.

From the perspective of readers, not much research has been done. One study, conducted in Sweden, found that people were "not able to discern automated content from content written by a human." But as Ryan Chittum wrote for the Columbia Journalism Review, news outlets spread this finding credulously, despite its questionable basis.

Chittum points out that the study was conducted in English, despite being most Swedes' second or third language. Even more distressing, the topic was NFL Football. Further, the two writing samples did not cover the same event; one was a game recap and the other a player analysis.

A few outlets — WIRED UK, Vice's Motherboard — admitted the scale of the study was small (45 students participated), but its many other limitations went unnoticed. Still, the Guardian's headline for the story asked "Could robots be the journalists of the future?" Chittum noted this was a self-fulfilling prophecy.

Mathew Ingram, senior writer for tech/media site Gigaom, links the anxiety over robots in journalism to the similar concern over non-professionals — citizen or amateur journalists — saturating the market with free news. But he contends all these changes bring is an increase in the total amount of journalism being done.

"All it means is that as a professional journalist, you now have to make sure that you are better than a robot," Ingram says.

But Ingram overlooks one potential caveat. The shift may have an outsized effect on entry-level reporters, who are usually assigned to many of the same tedious beats algorithms are now beginning to cover. These reporters are already highly expendable in a workforce expected to decline by 7,200 jobs over the next eight years, according to the Bureau of Labor Statistics.

Younger journalists may have less turf on which to prove themselves "better than a robot." ■