

How algorithms and human journalists will need to work together

September 1 2017, by Andreas Graefe



Credit: AI-generated image ([disclaimer](#))

Ever since the [Associated Press automated](#) the production and publication of quarterly earnings reports in 2014, algorithms that automatically generate news stories from structured, machine-readable data have been shaking up the news industry. The promises of this technology – often referred to as automated (or robot) journalism – are

enticing: Once developed, such algorithms could create an unlimited number of news stories on a specific topic at little cost. And they could do it faster, cheaper, with fewer errors and in more languages than any human journalist ever could.

This technology provides an opportunity to make money creating content for very small audiences – even, perhaps, customized news feeds for an audience of just one person. And when it works well, readers [perceive the quality](#) of [automated news as on par](#) with news written by human journalists.

As a researcher and creator of automated journalism, I've found that computerized news reporting can offer key strengths. I've also identified important weaknesses that highlight the importance of humans in journalism.

Identifying automation's abilities

In January 2016, I published the "[Guide to Automated Journalism](#)," which reviewed the state of the technology at the time. It also raised key questions for future research, and discussed potential implications for journalists, news consumers, media outlets and society at large. I found that, despite its potential, automated journalism is still in an early phase.

Right now, automated journalism systems are serving specialized audiences, large and small, with very particular information, producing recaps of [lower-league sports events](#), [financial news](#), [crime reports](#) and [earthquake alerts](#). The technology is constrained to these types of tasks because there are limits to what sorts of information it can take in and process into text that humans can easily read and understand.

It works best when handling structured data that is accurate like stock prices. In addition, algorithms can only describe what happened – not

why, making it best for routine stories based solely on facts that have little room for uncertainty and interpretation, such as when and where an earthquake happened. And because the major benefit of computerized reporting is that it can do repetitive work quickly and easily, it is best used to cover repetitive topics that require producing a large number of similar stories, such as sporting event reports.

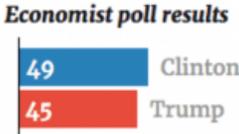
New Economist poll: Clinton holds slim lead



PollyBot
November 8, 2016
Uncategorized
2016NationalEconomist,
Polls (vote intentions)

Results of a new national poll administered by Economist were released. The poll asked respondents for whom they will vote: Republican Donald Trump or Democrat Hillary Clinton.

Of those who responded, 49.0% said that they plan to vote for former Secretary of State Hillary Clinton, whereas 45.0% said that they would give their vote to real estate developer Donald Trump.



The poll was conducted between November 4 and November 7. The sample size was 3669 participants. There is a sampling error of +/-1.7 percentage points. Considering this error margin, the gap between both candidates is statistically significant.

A news article written by an algorithm. Credit: PollyVote, CC BY-ND

Covering elections

Another useful area for automated news reporting is election coverage – specifically regarding results of the numerous polls that come out almost

daily during major campaigns. In late 2016, I teamed up with fellow researchers and the German company [AX Semantics](#) to develop automated news based on forecasts for that year's U.S. presidential election.

The forecasting data were provided by the [PollyVote](#) research project, which also hosted the platform for publishing the resulting texts. We established a completely automated process, from collecting and aggregating the raw forecasting data, to exchanging the data with AX Semantics and generating the texts, to publishing those texts.

Over the course of the election season, we published nearly 22,000 automated news articles in [English](#) and [German](#). Because they came from a fully automated process, the final texts often had errors, such as typos or missing words. We also had to spend much more time than we had expected troubleshooting problems. Most of the issues came from errors in the source data, rather than the [algorithm](#) – highlighting another key challenge of automated journalism.

Finding the limits

The process of developing our own text-generating algorithms taught us firsthand about the [potential and limits of automated journalism](#). It's crucial to make sure the data is as accurate as possible. And it is easy to automate the process of creating text from a single set of facts, such as the results of a single poll. But adding insights, like comparing that poll to others in the past, is much harder.

Perhaps the most important lesson we learned was how quickly we reached the limits of automation. When developing the rules governing how the algorithm would turn data into text, we had to make decisions that might seem easy for people to make – such as whether a candidate's lead should be described as "large" or "small," and what signals could

suggest a candidate had momentum in the polls.

Those sorts of subjective decisions are very hard to formulate into predefined rules that should apply to any situation that has occurred historically – much less to any situation that might occur in future data. One reason is that context matters: A four-point lead for Clinton in the run-up to the election, for example, was normal, whereas a four-point lead for Trump would have been big [news](#). The ability to understand that difference and interpret the numbers accordingly is crucial for readers. It remains a barrier that algorithms will have a hard time overcoming.

But human journalists will have a hard time outcompeting automation when covering routine and repetitive fact-based stories that merely require a conversion of raw data into standard writing, such as sports recaps or company earnings reports. Algorithms will be faster at identifying anomalies in the [data](#) and generating at least first drafts of many stories.

All is not lost for the people, though. Journalists have plenty of opportunities to take on tasks algorithms cannot perform, like putting those numbers in proper context – as well as providing in-depth analyses, behind-the-scenes reporting and interviews with key people. The two types of coverage will likely become closely integrated, with computers using their strengths and the humans focusing on ours.

This article was originally published on [The Conversation](#). Read the [original article](#).

Provided by The Conversation

Citation: How algorithms and human journalists will need to work together (2017, September 1) retrieved 19 April 2024 from

<https://techxplore.com/news/2017-09-algorithms-human-journalists.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.