

NiemanLab

Want to bring automation to your newsroom? A new AP report details best practices

LINK: [INSIGHTS.AP.ORG](https://insights.ap.org) | **POSTED BY:** JOSEPH LICHTERMAN | APRIL 5, 2017

In 2014, the Associated Press began automating some of its coverage of corporate earnings reports. Instead of having humans cover the basic finance stories, the AP, working with the firm Automated Insights, was able to use algorithms to speed up the process and free up human reporters to pursue more complex stories.

The AP estimates that the automated stories have freed up 20 percent of the time its journalists spent on earnings reports as well as allowed it to cover additional companies that it didn't have the capacity to report on before. The newswire has since started automating some of its minor league baseball coverage, and it told me last year that it has plans to expand its usage of algorithms in the newsroom.

“Through automation, AP is providing customers with 12 times the corporate earnings stories as before (to over 3,700), including for a lot of



RELATED ARTICLE

The AP wants to use machine learning to automate turning print stories into broadcast ones

October 31, 2016

very small companies that never received much attention,” Lisa Gibbs, AP’s global business editor, said in a report the AP released Wednesday.

The AP’s report — written by AP strategy and development manager Francesco Marconi and AP research fellow Alex Siegman, along with help from multiple AI systems — details some of the wire’s efforts toward automating its reporting while also sharing best practices and explaining the technology that’s involved, including machine learning, natural language processing, and more.

The report additionally identifies three particular areas of note that newsrooms should pay attention to as they consider introducing augmented journalism: unchecked algorithms, workflow disruption, and the widening gap in skills needed among human reporters to produce this type of reporting.

To highlight the challenges of using algorithmic journalism, the report constructed a situation where a team of reporters covering oil drilling and deforestation used AI to analyze satellite images to find areas impacted by drilling and deforestation:

Our hypothetical team begins by feeding their AI system a series of satellite images that they know represent deforestation via oil drilling, as well as a series of satellite images that they know do not represent deforestation via oil drilling. Using this training data, the machine should be able to view a novel satellite image and determine whether the land depicted is ultimately of any interest to the journalists.

The system reviews the training data and outputs a list of four locations the machine says are definitely representative of rapid deforestation caused by nearby drilling activity. But later, when the team actually visits each location in pursuit of the story, they find that the deforestation was not caused by drilling. In one case, there was a fire; in another, a timber company was responsible.

It appears that when reviewing the training data, the system taught itself to determine whether an area with rapid deforestation was near a mountainous area — because every image the journalists used as training data had mountains in the photos. Oil drilling wasn’t taken into consideration. Had the

team known how their system was learning, they could have avoided such a mistake.

Algorithms are created by humans, and journalists need to be aware of their biases and cognizant that they can make mistakes. “We need to treat numbers with the same kind of care that we would treat facts in a story,” Dan Keyserling, head of communications at Jigsaw, the technology incubator within Google’s parent company Alphabet. “They need to be checked, they need to be qualified and their context needs to be understood.”

That means the automation systems need maintenance and upkeep, which could change the workflow and processes of editors within the newsroom:

Story templates were built for the automated output by experienced AP editors. Special data feeds were designed by a third-party provider to feed the templates. Continuing maintenance is required on these components as basic company information changes quarter to quarter, and although the stories are generated and sent directly out on the AP wires without human intervention, the journalists have to watch for any errors and correct them.

Automation also changes the type of work journalists do. For instance, when it comes to the AP’s corporate earnings stories, Gibbs, the global business editor, explained that reporters are now pursuing different types of reporting.

“With the freed-up time, AP journalists are able to engage with more user-generated content, develop multimedia reports, pursue investigative work and focus on more complex stories,” Gibbs said.

Still, in order to use this type of automated reporting, newsrooms must employ data scientists, technologists, and others who are able to implement and maintain the algorithms. “We’ve put a lot of effort into putting more journalists who have programming skills in the newsrooms,” said New York Times chief technical officer Nick Rockwell.

The report emphasizes that communication and collaboration are critical, especially while keeping a news organization’s journalistic mission front

and center. The report outlined how it views the role data scientists play:

Data scientists are individuals with the technical capabilities to implement the artificial intelligence systems necessary to augment journalism. They are principally scientists, but they have an understanding as to what makes a good story and what makes good journalism, and they know how to communicate well with journalists.

“It’s important to bring science into newsrooms because the standards of good science — transparency and reproducibility — fit right at home in journalism,” said [Larry Fenn](#), a trained mathematician working as a journalist in AP’s data team.

The full AP study is available [here](#).

SHOW TAGS **NO COMMENTS / LEAVE A COMMENT**