

## Module 2 - Benefits & Limitations of Automated Content

Hello and welcome back to Week 2.

In the last video I showed you some examples of how automated content is being used by various news organizations. Today I want to talk to you a bit more specifically about some of the benefits and also limitations of automated content so that you're better prepared to decide whether or not it makes sense to integrate these technologies and approaches in your own practices and your in your own newsrooms.

So let's jump right in here. These are the four main benefits of automated content: speed, scale accuracy and personalization. Now in terms of speed it's hard to argue with the fact that computers are just faster than people at a lot of tasks. In some domains like in finance or in breaking news scenarios speed really is the name of the game. Particularly in cases where access to financial information and where a lot of money is at stake, we've seen automated systems deployed for the speed benefits that they provide.

Automation also offers the benefit of scale. Content can be scaled up across different geographies or over time so that the same type of event can be covered over and over again with minimal additional effort. Scale across time also creates new opportunities because it allows for a consistency of coverage that might not otherwise have been possible.

Accuracy is another benefit of automation because once a system is set up and debugged it will be consistent in its application of a template and its application of rules. That doesn't mean that automation can't make mistakes though just that it won't make the same types of mistakes and errors as people do.

Finally personalization is about adapting content to a user based on some property of the user like their age, their gender, their location or topical interests and automation creates an entirely new opportunity to be able to adapt content to individuals based on these various background properties of the user. So let's talk about each of these in a little bit more detail and I'll give you some specific examples.

So in terms of speed I wanted to show you this example from the LA Times. The LA Times QuakeBot actually creates a story stub with a map within minutes after an earthquake of magnitude greater than 3.0 is reported by the U.S. Geological Survey. The speed advantage here is essentially for safety reasons since people may want to know as soon as possible if there's a quake near themselves or near a loved one.

Another advantage of automation is scale. In the last video, I showed you this example of the Washington Post's coverage during the 2016 U.S. elections. The Washington Post was able to cover all 4 and 35 house races, all 34 senate races and all 12 gubernatorial races in in their 2016 coverage. In the past, in 2012, they'd only covered 15 percent of congressional races, but in 2016 they were able to cover 100 percent of congressional races. So given that they had data available for all of these different races they were able to scale up their coverage to really just cover everything.

So in general accuracy is also a benefit of automation. At the Associated Press for instance they've seen their error rates go down after introducing automation in financial earnings stories. Automation doesn't

make typos, it won't add up some numbers incorrectly, but that doesn't mean that automation can't still make mistakes just that they make different kinds of mistakes than people do.

Now I wanted to show you this example to point out one potential area where automation can still make errors. In this example we see the implications of an algorithmic writer not understanding context. So it says: "Move Inc. established a new 52 week low yesterday and could be a company to watch at the opening. After opening at \$0.00, Move Inc. dropped to \$0.00 for a new 52 week low."

So clearly this text doesn't make any sense. The system has missed an important bit of context which is that Move Inc. was actually delisted from the stock exchange and when a stock is delisted its share price is shown as zero, but the system didn't understand the concept of "delisting" and so it still generated a story that ended up being a nonsensical.

So finally I want to talk about personalization. Automation really opens up entirely new possibilities for creating personalized content where the texture of the visualizations produced in an article are adapted based on the characteristics of a user like their age their location or maybe their topical interest.

So this example that I'm showing here is an early example of a personalized article coming from the New York Times. At the top of the article, it inserts a localized map and some localized text to connect the article more to the area where you're visiting from.

So if you go to this page it will try to detect your location based on your Internet protocol, address your IP address, and then it will try to automatically adapt the visualization and the text of the article based on that location.

The main benefit of this type of presentation is increase relevance to an individual user and in my opinion, I think there are a lot of possibilities waiting to be tried out with respect to personalized article generation.

OK, let's turn our attention now to some of the limitations of automated content. I want to talk about four main limitations here.

Data access is really probably the biggest limitation of automated content. You need to have structured data and knowledge bases in order to plug into data driven templates that you might write. The quality, the breadth, the richness of available data will really impact whether or not the automated content you're able to produce is compelling or just kind of boring and bland.

Data can also be a competitive differentiator here. Exclusive data means exclusive content, and maybe news organizations actually need to start thinking more closely about acquiring some unique and well differentiated data assets. Perhaps they could digitize documents or submit public records requests or even engage in things like sensor journalism to collect data.

Another limitation is what I would call "Why and how." Explaining why or how something happens often requires higher level cognition and ability to interpret, to think about causal reasoning, you know, why did

one thing cause another thing. And these explanations also often require some degree of social understanding.

People generally tend to be better at these tasks because they can draw on context, they can draw on common sense, and develop rational explanations that way. Now because algorithms and automated content systems lack social understanding, areas related to cultural interpretation and journalism, things like writing a theater review may be particularly difficult for automation to undertake.

The lack of legal knowledge by automated rating systems also means that automation could make mistakes that run afoul of legal regulations like defamation for instance. So defamation involves making a false statement about a person that causes a injury or reputational harm and the potential is there for algorithmic missteps that could lead to legal liability if the automated writing system were to generate some text that hurt someone's reputation.

That's yet another reason perhaps to have humans integrated into the editorial loop, so that people are reviewing the content before it's published. A final limitation that I want to mention here, particularly with respect to template based approaches, is the writing quality of these systems.

Typically template based approaches lack variability and if we if a given user or a person is expected to consume multiple pieces of content from an automated system they could start to see the reputation and get bored with the content that they're reading.

On the other hand in the last video I talked about more sophisticated statistical techniques and those have the capability to create more variability in the text that they output, but they also then also create more potential for introducing errors in the text.

One way around some of these limitations particularly writing quality as well as social legal reasoning and for why and how explanations is to get people more involved in the process. Again this is the idea of blending or hybridizing people into the automation process. So what I'm showing you in this slide kind of gives you a hint at what that hybridization might look like in practice. It's essentially a more complex word processing interface.

You can sort of think of it as an almost alien word processor that lets the author write fragments of text that are controlled by data driven If/Then/Else rules. This is an interface that allows you to write templates with data and it offloads some of the most complex decisions in automated content related to document planning and micro planning. It offload those to a human writer who then authors a template of rules and text fragments.

So, in the next video I'm actually going to demo this interface it's called Arria Studio, so that you get a feel for how to start writing your own data driven templates using this approach.