AI4media

ARTIFICIAL INTELLIGENCE FOR THE MEDIA AND SOCIETY

ROADMAP ON AI TECHNOLOGIES & APPLICATIONS FOR THE MEDIA INDUSTRY

SECTION: "ROBOT JOURNALISM"

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 951911

info@ai4media.eu www.ai4media.eu

Author	**Mike Matton** (De Vlaamse Radio- en Televisieomroeporganisatie nv)

This report is part of the deliverable D2.3 - *"AI technologies and applications in media: State of Play, Foresight, and Research Directions"* of the AI4Media project.

You can site this report as follows:

F. Tsalakanidou et al., Deliverable 2.3 - AI technologies and applications in media: State of play, foresight, and research directions, AI4Media Project (Grant Agreement No 951911), 4 March 2022

This report was supported by European Union´s Horizon 2020 research and innovation programme under grant number 951911 - AI4Media (A European Excellence Centre for Media, Society and Democracy).

The information and views set out in this report are those of the author(s) and do not necessarily reflect the official opinion of the European Union. Neither the European Union institutions and bodies nor any person acting on their behalf.

Robot journalism

Current status

Newsrooms are heavily under pressure due to several reasons. First of all, the amount of news and information generated worldwide is exploding. Journalists have to be selective on the information they process. The rise of social media and other platforms has also contributed to this information explosion of the past decade. Secondly, news needs to be produced and distributed faster than ever. The competition to be the first to publish a news item is higher than ever before and sometimes depends on seconds. Finally, the pressure to produce more content is always there. In order to retain consumers, the offering needs to be large and pretty complete.

Over the past few years, advances in AI have caused AI technologies to make their way into the newsrooms. In the narrow sense of the word, "robot journalism" a.k.a. "algorithmic journalism" refers to stories automatically generated by AI algorithms, typically derived from one or several data sets. Lewis et al. define it as "*any process or system of news production under the control of media or electronic devices, with little or no external influence*"[1]. Several organisations worldwide have been using robot journalism to produce part of their content offering. Examples include: the Associated Press, which uses the Wordsmith platform to write compelling narratives on financial recaps; PA Media[2], which developed the RADAR platform to generate data driven news articles on sports, financial data and elections; Forbes, LA Times, etc. Robot journalism is not only limited to large media organisations; smaller organisations such as MittMedia in Sweden[3] have also been using it in practice for the automatic generation of Real Estate articles.

Initially, robot journalism was limited to extracting data from databases and using this data to automatically fill in the spaces in pre-defined template articles. Over the course of the past years however, those technologies have become smarter, with deeper data analysis taking place and the creation of more complex narratives in a story. However, applications are still mostly limited to domains with a lot of structured data available such as sports or financial news. There are currently promising systems available that can automatically generate narrative summaries of sports games. Producing more complex stories or narratives is however still a challenge. Although there are some nice examples[4] around, the creation of those stories remains in the hands of journalists.

The main idea behind robot journalism is that through the automated creation of articles, journalists in the newsroom will have more time to invest in developing more complex news stories, currently out of scope for automated journalism systems. In this way, some of the challenges mentioned earlier can be tackled. In practice however, several newsrooms have fired

[1] Seth C. Lewis, Andrea L. Guzman, and Thomas R. Schmidt. 2019. Automation, Journalism, and Human–Machine Communication: Rethinking Roles and Relationships of Humans and Machines in News. Digital Journalism7, 4 (April 2019), 409–427.

[2] Radar- Combining the latest in AI with skilled writers to dynamically create high-quality content at massive scale: https://pa.media/radar/

[3] L. Southern, Robot writers drove 1,000 paying subscribers for Swedish publisher MittMedia (2019): https://digiday.com/media/robot-writers-drove-1000-paying-subscribers-swedish-publisher-mittmedia/

[4] Guardian, A robot wrote this entire article. Are you scared yet, human? (2020): https://www.theguardian.com/commentisfree/2020/sep/08/robot-wrote-this-article-gpt-3

journalists because their work was now done by robot journalists. The most known example is Microsoft replacing 27 journalists by an AI system in 2020[5].

In newsrooms, "robot journalism" usually has a broader interpretation. It not only stands for completely automated news articles but also for AI and journalist working together in hybrid ways, where AI systems help journalists to navigate information, write parts of articles, assess veracity of information, create summaries etc. In this broader definition of the term, many more newsrooms worldwide are planning to or in the process of integrating such technologies into their workflows.

Research challenges

Most – if not all – of the robot journalism systems heavily rely on the availability of structured data. Unless this structured data is available, the systems will fail or produce unsatisfactory results. Also, the transformation of unstructured data into useful semantic structured data is not a simple task. Fundamental as well as applied research into technologies capable of doing this remains necessary. It requires analytical processing of unstructured information. Humans are much better at analytical thinking and reasoning with unstructured data compared to computers.

In news production, the elements of surprise and creativity are also important. As of today, human journalists are much better in creative thinking and surprising the reader compared to computers. Although advances in artificial creativity are being made, they still are not convincing enough to replace journalistic work.

Furthermore, there is the language challenge. As robot journalism heavily relies on natural language processing/generation systems, their performance is also language dependent. Those systems work much better in major languages such as English compared to smaller languages.

Finally, assessing the veracity of information, and making conclusions on whether or not to publish a story is still a challenge in itself. Disinformation is in fact one of the major hurdles to overcome. AI systems that automatically generate trustworthy news can often be misused to generate disinformation. This is in fact already happening in some fake news farms. Furthermore, tampering with the data sources themselves can remain undetected and generate fake articles without anyone noticing before it is in place. There is still a lot of work on rigid editorial processes and trustworthiness assessment of information before such systems can be used at mass scale.

A good overview of current research challenges can also be found in more detail in a recent article by Kotenidis et al[6].

[5] J. Waterson, Microsoft sacks journalists to replace them with robots (2020): https://www.theguardian.com/technology/2020/may/30/microsoft-sacks-journalists-to-replace-them-with-robots

[6] Kotenidis, Efthimis, and Andreas Veglis. 2021. Algorithmic Journalism—Current Applications and Future Perspectives. Journalism and Media 2: 244–257. https://doi.org/10.3390/journalmedia2020014

Societal and media industry drivers

Vignette: Robot journalism for Olympic Games reporting

Chantal is editor in chief of a large newspaper. Thanks to the digitisation of the past two decades, audiences are more and more reached online with respect to the traditional newspapers. This also brings challenges, however. In the digital world, more content is needed compared to the regular newspaper, and people want to get updates regularly. The traditional collection of articles in a typical newspaper has been replaced by a continuous feed of articles. Readers browse and scroll for interesting content throughout the day, similar to how social media feeds work.

There is a major challenge for Chantal ahead, the Olympic Games. Her newspaper has always been the reference for Olympic Games reporting and she wants to keep this reputation. She has a decent team of journalists and reporters, who will work both on premise as well as in the newsroom. Although her team has a decent size, the team is unable to report on every discipline and every game. Furthermore, reporting is not limited to the Olympic Games period itself. Even in the months before, a lot of stories and previews need to be published in order to raise the attention of the audience and raise the excitement. Even though her team is working at maximum capacity, there is too much to report on, and Chantal knows some readers will be disappointed if their favourite sports is not reported and as a result might move away to other media channels. Moreover, the audience follows the newspaper not only through their website, but also, more importantly, through various social media channels. This creates an additional stress on the team, as they now not only have to cater for the newspaper, their own news website, but also keep the social media channels active and filled with interesting content.

Luckily, the reporters are not the only resource Chantal has available. Her newspaper has bought access to several data feeds on the Olympic Games, with statistics about the athletes, loads of historical information, and live data feeds with events from every single competition in the Olympic Games. Publishing this data as such is not compelling to the audience. Therefore, the newspaper bought a license for a robot journalism system. One year ago, when Chantal was first introduced to the system, she was impressed with its ability to produce live reports on sports games in many different formats and publish them on a variety of channels, including social media feeds. With the right configuration in place, hundreds of short reports, live game reports and highlight overviews are generated automatically. Thanks to the powerful linguistic engine of the system, the articles read smoothly and fluently. Chantal is keen on the fact that the decision to publish is still in the hands of her editorial team. However, thanks to the powerful robot journalism system, these decisions are made efficiently, allowing her and her team of reporters and journalists to focus more on the major highlights of the event.

Although Chantal is happy with the current situation, which keeps her newspaper on the competitive edge, she knows even more will be necessary to keep her newspaper relevant and to retain its consumer base. She hopes that in the near future AI systems will be able to offload more work of the editors such that they can focus on the real cutting edge quality content.

Future trends for the media sector

We foresee that in the near future, every newsroom worldwide will make use of AI technologies in one or another way. Advances in AI technologies will almost certainly help newsrooms to be more efficient in their work, with repetitive and less complex jobs fully automated as discussed before.

However, the replacement of a human journalist by a computer is still a long way ahead. It is even doubtful this will ever happen. The main trend we foresee for the future is the hybrid collaboration between journalist and AI system. The places where AI technologies can augment the work in the newsroom are numerous. Figure 1 presents a (non-exhaustive) overview of processes in the newsroom that can be supported by AI technologies. There is potential throughout the workflows, from ideation over research, production, publication, feedback and interaction, and archival.

For this purpose however, journalists need to adapt their way of working, and learn how to rely on the available AI systems in the newsroom and beyond. More specifically, they need to understand the strengths and limitations of the AI systems. A huge challenge ahead in this is not only the integration of the technology into the newsroom, but also the whole process of change management to adhere to new ways of working. If working well however, these trends will make the newsroom much more efficient, both in terms of quantity as well as in quality of the news stories provided.

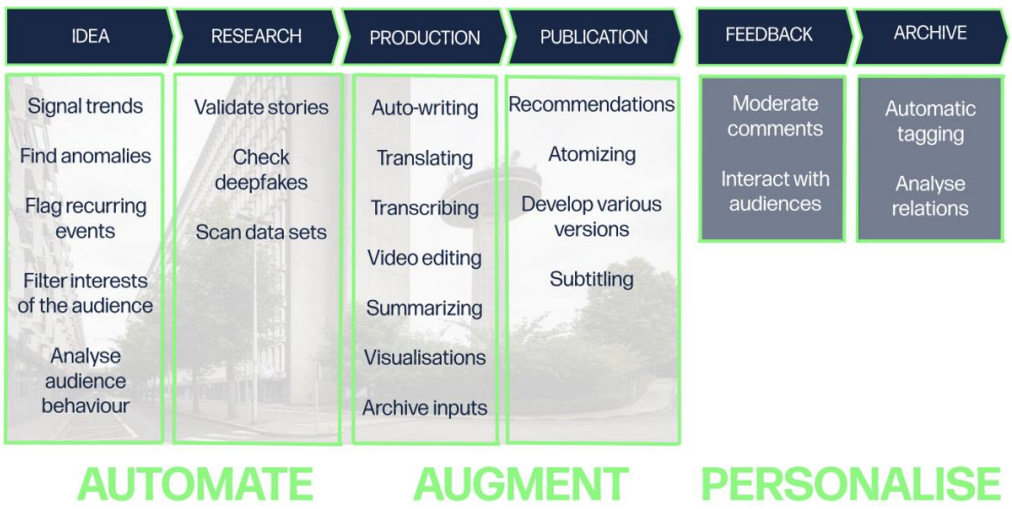

Figure 1: AI assisted newsrooms.

Furthermore, the consumption of news is also diversifying and becoming more personalised. For this purpose, newsrooms need to publish stories in many different formats and on many different platforms. Also for this purpose, AI technologies can certainly assist to generate the many different versions of the content to publish.

Goals for next 10 or 20 years

In the next 10 or 20 years, robot journalism will gradually be introduced in newsrooms. It is not expected however that robot journalism will completely replace journalists due to the complex nature of news production, information assessment and reasoning. Journalists will gradually adapt to work together with AI systems that help them in their daily job.

The successful newsroom of the future will seamlessly integrate fully automated robot journalism and hybrid (AI-assisted) journalism, able to produce a rich content offering on many different platforms, and personalised to the needs of the news consumer.

AI4media

ARTIFICIAL INTELLIGENCE FOR THE MEDIA AND SOCIETY

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 951911

info@ai4media.eu www.ai4media.eu