



REUTERS
INSTITUTE for the
STUDY of
JOURNALISM

24

Reuters Institute Fellowship Paper
University of Oxford

**Putting Europe's Robots on the Map:
Automated journalism in news agencies**

by Alexander Fanta

Hilary and Trinity Term 2017

Sponsor: Google Digital News Initiative

Acknowledgements

I dedicate this work to all those who helped and inspired me to come to Oxford and pursue my research here.

I am indebted to the contribution of my Austrian friends Julian Ausserhofer, who gave me the idea for writing about automated journalism, and Pascal Lebersorger, who created the robot illustration on the next page.

Many thanks go to the Reuters Institute and in particular to the Director of the Fellowship Programme, James Painter and my supervisor Chris Westcott, who patiently guided me in researching and writing. A further thank you goes to Richard Fletcher, who never tired of debating politics and the media with me.

I also want to thank my journalist fellows from around the world. Their help and companionship made my experience in Oxford truly great.

And I am ever grateful to Hanna, who is always there for me.



Table of Contents

Introduction	4
Literature Review	6
Methodology	8
Key findings on automation in news agencies	10
Discussion: The limits of automation	13
Data access is key	13
Missing appetite for innovation	14
Bottom-up automation	15
The news robots of the future	16
Journalism and AI	17
Conclusion	19
Bibliography	20
Appendix: list of interviewees	23

Introduction

Automation is changing journalism. At major news organizations, algorithms create thousands of stories every day without much human interference. Robot-written stories have vastly expanded coverage in certain fields, especially in financial news and sports.

The Associated Press says it now produces automated stories on quarterly earnings reports for around 3,700 companies, up from 400 companies covered previously¹. AP has also automated some of its sports coverage. AP's closest competitors Reuters and AFP likewise produce thousands of automated stories each year.²

Smaller news agencies, too, have started to catch on. Norwegian agency NTB is currently working on an algorithm to produce stories for all 20,000 league football games played in the country each year.³ The stories are based on data from the Football Association of Norway.

News agencies have different reasons for automating news writing. Some, such as Reuters, focus on financial news and are devoted to speed. Others use automation to expand coverage in particular fields, specialize their coverage, provide more regional variations in reporting or customize their content to the reader. Industry leaders hope that automation will provide additional value to their customers and open up new streams of revenue.

Stories written by robots often still read formulaic, closely hewing to a certain pattern. Developers are trying to change this and make Natural Language Generation (NLG) more complex by improving linguistic variety. The advances have been sufficient that, in surveys, readers see stories produced by robots as hard to discern from human stories⁴.

Using robots to generate text is just the high-point of what has been, as researchers such as Carl-Gustav Linden⁵ have pointed out, decades of automation in the newsroom. Much as PCs have made typists and clerks redundant, automated text generation and news-finding algorithms could put reporters and monitoring teams out of work in the future. This could lead to less diversity and complexity in news, some journalists fear.

Not only news writing is being automated, but also news detection and video production. Recent algorithmic developments include a tool, the Reuters News Tracer, which alerts reporters of breaking news on social media. Reuters is also trying text-to-video technology by the software firm Wibbitz to produce videos. Meanwhile, the BBC is developing automated video translation software.

¹ Marconi, Francesco/Siegman, Alex (2017): The Future of Augmented Journalism: A guide for newsrooms in the age of smart machines, Associated Press

² Talk by Reuters' Reg Chua at the Newsrewired conference in London on February 8, 2017 and email from AFP's Sophie Huet-Trupheme on March 3, 2017

³ Telephone interview with Helen Vogt on February 15, 2017

⁴ Christer Clerwall (2014), Enter the Robot Journalist, *Journalism Practice*, 8:5, 519-531

⁵ Linden, Carl-Gustav (2017), Decades of Automation in the Newsroom, *Digital Journalism*, Vol. 5, Iss. 2, 2017, page 12

So far, there has been no comprehensive survey of automation projects at different news agencies. My research attempts to close that gap.

In this paper, I provide an overview of news agencies' efforts in Natural Language Generation. My research draws on interviews and email exchanges with news agencies in the US and across Europe. The responses allow a perspective on how senior journalists and developers at different news organisations evaluate the use of automation and its further potential.

My paper addresses the question of what kind of content news agencies automate and how they do it. The research highlights the sophistication of the technology in use, and what technical and organisational challenges news agencies face in introducing automated journalism. My work also interrogates the pivotal role of data for automation.

I first became interested in the subject while working as reporter at the Austrian Press Agency. I tinkered with improving the algorithms for distribution of stories from other news agencies to different news desks, and thought about ways to automate the production of metadata. The experience gathered from these experiments motivated me to research other ways to automate journalism, especially news writing.

To get a better sense of how the field will develop further, my paper samples opinions from leading academic and industry experts in the field. Their expertise helps me to discuss the constraints that data access puts on further automation and it draws attention to the reliance on third-party providers as a possible pitfall for news organisations.

In my concluding chapter, I outline the technical and organisational limits for the present-day form of automation and discuss possible ways in which the technology could develop next. My intention is to give researchers and practitioners a balanced overview of the state of automated journalism in news agencies, and offer ideas on the path forward.

Literature Review

There is a small but fast-growing body of academic literature on automated journalism. Research has addressed topics such as the technical possibilities of current technology, ethical challenges of algorithmic news creation and the perception of NLG among the public and journalists. In this chapter, I highlight some notable work in the field.

Researchers have attempted to provide insight into how news organisations use automation technology. Andreas Graefe explains in his *Guide to Automated Journalism*⁶ that algorithms can ‘create thousands of news stories for a particular topic, they also do it more quickly, cheaply, and potentially with fewer errors than any human journalist.’ Graefe aptly describes the most common form of NGL technology as ‘simple code that extracts numbers from a database, which are then used to fill in the blanks in pre-written template stories’.

Graefe and earlier his colleague Konstantin Dörr note that the availability of data limits automation to fields with an abundance of pre-structured information, such as weather, traffic, sports and finance. ‘Data for text generation has to available in a scalable amount as programming and individualization of journalistic products are cost-intensive’, Dörr writes⁷. This observation is mirrored in my finding that news agencies only consider automation on a large scale, eschewing smaller journalist-driven innovation.

Wal-Mart tops 1Q profit forecasts.

BENTONVILLE, Ark. (AP) — Wal-Mart Stores Inc. is reporting fiscal first-quarter earnings of \$3.04 billion.

On a per-share basis, the Bentonville, Arkansas-based company said it had profit of \$1.

The results surpassed Wall Street expectations. The average estimate of 14 analysts surveyed by Zacks Investment Research was for earnings of 96 cents per share.

The world's largest retailer posted revenue of \$117.54 billion in the period, missing Street forecasts. Eight analysts surveyed by Zacks expected \$117.63 billion.

Wal-Mart expects its per-share earnings to range from \$1 to \$1.08 for the current quarter.

Analysts forecast adjusted earnings per share of \$1.07.

This is a recent example of an automated AP story.⁸

Research has addressed the ethical challenges of automated journalism. The question posed is whether code should reflect journalistic values of media organizations such as objectivity, responsibility, and accuracy. Diakopoulos and Koliska⁹ debate the need for

⁶ Andreas Graefe (2016) *Guide to Automated Journalism*, Tow Center for Digital Journalism, <http://towcenter.org/research/guide-to-automated-journalism/>

⁷ Konstantin Nicholas Dörr (2016) *Mapping the field of Algorithmic Journalism*, *Digital Journalism*, 4:6, 700-722

⁸ As taken from the Business Insider website, May 18, 2017

⁹ Diakopoulos, Nicholas & Koliska, Michael (2016): *Algorithmic Transparency in the News Media*, *Digital Journalism*, DOI: 10.1080/21670811.2016.1208053

algorithmic transparency to reflect those values. Dörr and Hollnbuchner¹⁰ raise the issue of data integrity; they propose that missing items can lead to bias in content generation.

Journalism researchers Montal and Reich note that there is no consistent policy across news organizations on whether to attribute stories to algorithms. This assertion is mirrored in my findings. According to Montal and Reich, a consistent attribution policy would help to 'minimize the dangers of ambiguous authorship and demystify the naive perceptions shared by audiences and journalists regarding the seemingly unbiased nature of algorithms and the possible decline in readers' judgement'.¹¹

Interesting contributions to the study of automation in journalism come from legal and business studies. Lin Weeks notes in his paper Media Law and the Copyright Implications of Automated Journalism that content generated by algorithm raises difficult questions about authorship. Weeks states that it is conceivable that rights could be assigned to the computer programme itself¹².

A study of the impact of AP's robot-generated earnings stories on capital markets found that coverage of firms that previously got little media attention led to a significant increase in trading volume.¹³ Thus automated journalism is moving, if not people, then at least markets.

Researchers have also scrutinized the views of the audience and human journalists on their robot colleagues. Christer Clerwall compared reader reactions to human and robot-written stories and found that there was no significant differences in how texts were perceived, with the notable exception of the human story being seen as more pleasant to read.¹⁴

A study by Neil Thurman et. al. deals with the perception of journalists after having first-hand experience working with state-of-the-art automation software¹⁵. The journalists acknowledge the potential of automated journalism for reducing cost, increasing speed and expanding coverage, but note the limited potential of the technology at the moment. In their findings, the authors of the study assert that limitations of automated news generation become evident 'by contrasting it with the multi-sourced, interrogated, and contextualized journalism humans can produce'.

¹⁰ Konstantin Nicholas Dörr & Katharina Hollnbuchner (2017) Ethical Challenges of Algorithmic Journalism, *Digital Journalism*, 5:4, 404-419, page 9

¹¹ Tal Montal & Zvi Reich (2016): I, Robot. You, Journalist. Who is the Author? *Digital Journalism*

¹² Lin Weeks (2014) Media Law and Copyright Implications of Automated Journalism, *Jipel No. 4 - Vol. 1*, <http://jipel.law.nyu.edu/vol-4-no-1-3-weeks/>

¹³ Elizabeth Blankespoor, Ed deHaan and Christina Zhu (2017) Capital Market Effects of Media Synthesis and Dissemination: Evidence from Robo-Journalism, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2872784

¹⁴ Christer Clerwall (2014), Enter the Robot Journalist, *Journalism Practice*, 8:5, 519-531, page 526

¹⁵ Neil Thurman, Konstantin Dörr & Jessica Kunert (2017) When Reporters Get Hands-on with Robo-Writing, *Digital Journalism*

Methodology

The objective of my research is to find out how news agencies all over Europe and the US use automation technology. To this end, I decided to survey 15 news agencies to get a sense of their experience and understanding of the technology.

My assumption is that news agencies are more interested in broad coverage than other types of media organizations. Therefore, I hope my survey of news agencies offers an insightful perspective on automation efforts in other parts of the news industry.

I spoke with the people responsible for automation in their agencies about the capabilities and cost of automation. Topics of discussion included the use of data, the complexity of the algorithm and possibilities for creating new revenue streams. My intention was to get a broad picture of the technical and organizational aspects of using automation technology.

In total, I conducted semi-structured interviews with representatives of 5 news agencies via the telephone and had at least partial answers by email from another 10 news organizations. 8 news agencies did not respond to repeated requests.

I transcribed all the interviews and supplemented the responses with public information from press releases and news stories to gain a fuller picture. I structured the replies into a few broad categories to be able to compare them directly. These categories included the type of automation used, data sources and the mode of publication. I then fed the information into an Excel spreadsheet. The key findings chapter below gives a summary of my findings.

The table below is part of the attempt to structure the information I gathered. I try to provide a somewhat comprehensive overview over developments. Some questions were left unanswered by respondents, so my data is incomplete in these areas. But I tried to account for these gaps and give approximate answers where full ones can't be given.

News agency	Response	Interview	In use? In development?
Reuters	y	y	In use
Associated Press	y	y	In use
NTB	y	y	In use
STT	y	y	Prep-stage
dpa	y	y	In use
Ritzau	y	n	In use
CTK	y	n	None
Efe	y	n	None
PA	y	n	In development
AFP	y	n	In use

TT	y	n	In use
ANP	y	n	In use
Ansa	y	n	None
Lusa	y	n	Former use
APA	y	n	In use

To supplement the data with further insight on the subject, I conducted interviews with three leading academic researchers on automated journalism and an industry expert. I discuss their input in the chapters on data access and constraints for automation projects.

My research addresses the question of automation in news agencies. It does not take note of similar projects done by other media organisations, such as public broadcasters and local newspapers. I therefore can only make assumptions about the state of automation across the wider publishing industry. Due to time constraints, I was only able to speak with news agencies in Europe and the USA and I did not get full answers from all of them.

A further limitation is posed by my research design. My focus was on statements from the news agencies and I did not take a closer look at the automated output itself and the audience reception of it. This restricts my paper to a survey of opinion from within the industry and academia. Future studies might try to assess audience reaction and create a comprehensive register of all automation activity in newsrooms.

Key findings on automation in news agencies

News agencies across Europe are adopting automation in their newsrooms. My survey includes 14 organisations based in Europe and the Associated Press (AP) headquartered in the US. Among these, nine say they use some form of automated text creation. Two others say they are working on automation projects.

One news agency, Lusa of Portugal, experimented with automated reporting for the 2015 Portuguese parliamentary election. Lusa has not used the technology since, but 'the agency is aware that this technology will be part of the future of the journalism', a senior figure at the news agency says¹⁶.

Three news agencies, CTK of Czechia, Efe of Spain and Ansa of Italy, say they do not use any form of automation.

The most common fields of use for robot-journalists are finance and sports. 7 out of 9 news agencies use automation in one or both of these fields. A reason why many news agencies focus on automating sports is the availability of data, says Maija Lappalainen of the Finnish news agency STT. She notes that *'in sports you have the rules, you know this is the winner, this is the loser and this is what happened. It is kind of an easy way to start with.'*¹⁷

Robot journalists also write about election results as well as official figures on education, unemployment and real estate. These are usually dependent on official sources providing figures in an easily usable format.

The most common application are templates that put tabularized data into words, also known as Natural Language Generation (NLG). 8 out of 9 agencies use automation technology in that way. But there are other forms of text-automation as well. A versioning tool developed by Dutch news agency ANP is a bot that rewrites human stories into more simple language for the wire services' Kids Feed.

Most robot-journalists aren't very sophisticated. Only two news organizations surveyed, Reuters and the Norwegian agency NTB, say that their algorithms are able to compare new information to historical data, and interpret that data into statements. A widget launched by the Swedish news agency TT compares local figures for school performance and real estate prices with the national average.

Other news agencies use robots only to report new data, providing basic snapshots rather than bigger-picture stories that put events into a historical frame. One reason for this is caution against possible mistakes when comparing new data with old. *'We have really stayed away from that, because obviously the more sophisticated you try to get with it, the more you increase the chance of error'*, says Lisa Gibbs of the AP¹⁸.

¹⁶ Email from Teresa Gomes of Lusa on February 20, 2017

¹⁷ Interview with Maija Lappalainen on February 17, 2017

¹⁸ Telephone interview by the author with Lisa Gibbs on February 23, 2017

Not all use of automation is made transparent to customers and readers. Reuters, AP and NTB usually tag their robot stories. However, this does not apply to single-line alerts, so-called snaps, which Reuters sends out. At least two news agencies produce partial stories from templates without mentioning the robot as a co-author.

Much automated content is published without being looked at by human editors. Reg Chua, an innovation manager at Reuters, says his news agency generates 950 alerts and 400 stories a day that get published without human interference¹⁹. At least four other news agencies publish at least some automated material without it being looked at by humans.

The responsible editors argue that the value of automation was severely limited if stories would have to go through human review. Another factor mentioned is that robots, when trained well, are very accurate. *'One of the great things about algorithms is that they don't make the same mistake twice. Once you corrected the mistake, they won't make it again'*, Helen Vogt of NTB says.²⁰

An important motivation to automate content is the wish to create new products. At least 6 news agencies in this survey automated content to create a new product or greatly expand the existing service in certain fields.

On the contrary, cutting costs of human labour was not mentioned by any news agency as factor in the decision to automate. However, it does play a role in allocating resources in the newsroom. After automating earnings reports, the AP said that saved them three full-time staff, who were deployed to other tasks.²¹

One possible reason why automation might not replace much human staff is cost. Several respondents to the survey mentioned development expenses as a barrier to entry and to further expansion. The necessary technical expertise is usually not found in-house, and therefore can create considerable expense for outside development.

Some news agencies also noted the high running costs for automated products. *'People think automation is cheap, but automation is in fact not that cheap. If you automate, it costs you money. You have to maintain it, you have to track it, you have to manage it. It's actually not [there] to save a lot of money'*, said Reg Chua of Reuters.²²

Automation has created the new task of data maintenance. Data sets for earnings reports need to be constantly updated to account for companies changing names and merging, or they might produce incorrect or misleading stories. *'It is not one of these things where you create the process and then just forget about it. You have to constantly be managing that data set'*, said Lisa Gibbs of the AP.²³

¹⁹ Talk by Reg Chua at the Newsrewired conference in London on February 8, 2017

²⁰ Telephone interview with Helen Vogt on February 15, 2017

²¹ Telephone interview with Lisa Gibbs on February 23, 2017

²² Telephone interview with Reg Chua on March 10, 2017

²³ Telephone interview with Lisa Gibbs on February 23, 2017

Access to data is frequently reliant on third-party providers and official sources. My data is somewhat incomplete, but at least four news agencies depend on data from external sources for their automation. At least two news agencies pay for that data. While this might not represent an overall picture, it is a strong indicator that news agencies have a tendency to source their data externally rather than developing own resources. A notable exception to this is Reuters, which finds and structures its own financial dataset.

Discussion: The limits of automation

The availability of data is a crucial factor limiting the use of more far-ranging applications of automation in journalism. This has been acknowledged in earlier assessments of the potential of automated journalism, as noted by Graefe²⁴, and is mirrored in my findings.

Data access is key

Most robot-written stories in news agencies are done in areas where data is readily available, mainly sports and finance. Structured data in these fields is in many cases provided by third-party organizations such as sports federations and financial analytics firms.

In areas where no external data is available, news agencies appear reluctant to find and structure their own data. With the notable exception of financial news providers Reuters and Bloomberg, I find that news agencies do not think of themselves as data hosts and providers. Their reluctance to build up data handling capability limits the fields in which automation can be used. It also potentially restricts the complexity of the output - after all, commercial providers might ask higher prices when asked to provide more rich data.

Using third-party data saves news agencies money in data maintenance. However, in the long run it can constrain growth and raise the issue of autonomy. News organizations should avoid making themselves reliant on data providers, warns researcher Konstantin Dörr.

*'This dependence is absolutely unnecessary, if they would resolve to look at certain subjects, and mine and structure the data themselves.'*²⁵

Another academic expert on automation in journalism, Nick Diakopoulos of the University of Maryland, says news organisations should look to build their own capabilities to make their offering distinguishable from competitors.

*'[I]n order to be competitive and make a unique offering, you can't just use the same data set that everyone else has.'*²⁶

With control of the data as key point in using automation, the role of the media in providing news might itself be at stake. After all, what would happen if the media wasn't the first source for news, asks Neil Thurman²⁷, who researches computational journalism at LMU Munich. He warns that news organisations could soon have unexpected competitors.

'If [news organizations] don't own the data and this technology advances to the point where anybody can take a data source and turn it into a story, then what's to stop Manchester United from producing their own stories and cutting out the middleman?'

²⁴ Grafe 2016

²⁵ Telephone interview with Konstantin Dörr on May 25, 2017

²⁶ Telephone interview with Nick Diakopoulos on May 22, 2017

²⁷ Telephone interview with Neil Thurman on April 28, 2017

News organisations have geared much development effort towards producing text from pre-structured data. Less has been done on data mining and structuring, says Dörr. These fields, however, are in his view crucial for the further expansion of automation technology.

*'If this potential is used, these processes will consequently boost other developments in automation.'*²⁸

Another important path for news agencies to accessing data is government regulation. Some countries in Europe and the US have laws providing for open data from official sources, and some also try to standardize the way in which this data is accessible. The Global Open Data Index 2017 of the Open Knowledge foundation provides a good overview over the field.²⁹

News agencies such as TT in Sweden use data feeds from governmental sources for automated stories. The more official data is available in machine-readable format and an open licence, the easier it must be assumed to be usable as source material for automation.

Requiring governments to report data in an accessible, regular and consistent way would facilitate the use of automated journalism as a public service. The same applies if companies and other entities were required to make mandatory publications such as earning reports in structured formats to allow easier use for journalism.

Missing appetite for innovation

The use of automated journalism in news agencies has so far been limited in its complexity and the fields in which it is applied. The previous chapter discussed the role of data in limiting the scope of automation, but there are further contributing factors.

As mentioned above, some news organisations cite the risks associated with complexity. Others mention the lack of technical ability and resources.

In itself, the technical barrier to automation in its current form is not deemed particularly high, at least compared to other technology used in newsrooms. The type of NLG software news agencies use was technologically feasible already in the 1990s, says Neil Thurman.

*'The automated journalism that you see at the moment is a sort of a glorified mail merge, essentially. It's something that we were doing with Microsoft Word 20 years ago, but obviously a bit more sophisticated.'*³⁰

So why has automation not caught on earlier? One explanation lies with the publishing industry that controls news businesses. The industry has been slow to take up innovation in the digital field, says Paul Maidment, who was editor-in-chief of Forbes.com and now is Director of Analysis and Managing Editor at the consulting firm Oxford Analytica.³¹

²⁸ Telephone interview with Konstantin Dörr on May 25, 2017

²⁹ Global Open Data Index 2017, Open Knowledge Foundation

³⁰ Telephone interview with Neil Thurman on April 28, 2017

³¹ In-person interview with Paul Maidment on May 18, 2017

In Maidment's time at Forbes.com, from 2001 to 2010, the company tried out automated journalism, yet it was unable to fund further development to make it into a new product. 'We were experimented with it, and it was very clunky, and then 2008 came along and there wasn't money to do anything', Maidment recounts. Automated content was eventually introduced after he left.

Maidment says the publishing industry has missed opportunities to develop digital innovation projects. This includes, in his view, archives of news organizations offering a trove of data that can be structured and used to create new products. The industry is focussed too much on creating new content in the traditional way instead of thinking about how to spin new business out of the old, Maidment says.

*'There is a huge lack of entrepreneurial appetite and skill in publishing companies. (...) Breaking up content, packaging it, that is outside of the experience of anyone in the publishing industry. (...) The whole innovation culture just wasn't there. And the journalists had no background in that either.'*³²

Media management researcher Lucy Küng comes to a similar assessment about innovation in the news business. Traditional media companies, including news agencies, in her view struggle to build new digital businesses on top of their old ones because they lack organisational capacity and know-how. Companies are 'not only under-financed but in a broader sense under-resourced in terms of their digital capabilities', Küng says.³³

Bottom-up automation

Another factor that limits use of template-based automation is the top-down nature of newsroom culture. Innovation typically is initiated by leading figures and conceptualized on an organizational scale. Thus Lisa Gibbs of the AP says that a script automating 50 stories on employment a month is not worth their time. She remarks, 'we have yet to discover that next big thing that is going to enable us to drive thousands of stories.'³⁴ From the management perspective, the low-hanging fruit in automation have been picked.

However, it is possible that reporters themselves automate their own work, rather than marshalling the resources of development teams. An example for this can be found at the news agency Efe, a leading news provider in Spain and the Spanish-speaking world.

Efe's former Washington Correspondent Jairo Mejia says he became interested in digital innovation after being a journalist fellow at the Reuters Institute for the Study of Journalism in Oxford in 2014. On his own initiative, Mejia created software to automate simple stories on monthly official figures for US unemployment and oil supply. The technology is now used by

³² In-person interview with Paul Maidment on May 18, 2017

³³ Tokbaeva, D. (2016): When Media and Management Collide: An Interview with Lucy Küng, Westminster Papers in Communication and Culture, 11(1), 26–30

³⁴ Telephone interview with Lisa Gibbs on February 23, 2017

journalists at the Washington bureau of Efe. *'The robot is not super-sophisticated, but it works flawlessly so far and it saves precious time'*, Mejia says.³⁵

Meanwhile, the management of Efe says it is not working on any new automation projects. Mejia has since left the company to form a startup called Superwire.

The example of Efe shows the possibility of journalist-driven automation. But if news organisations want to develop new technology in collaboration with their journalists, this requires commitment by the management to providing resources and training to their staff.

One step towards bottom-up automation is to design user-friendly software for everyday work around the newsroom. Automation software companies have started to pick up on this. The German firm AX Semantics offers their own markup language, ATML3, to make it easier to write automation code. *'Because we removed the "programming" part of Natural Language Generation, you don't need to be a computational linguist anymore'*, they claim³⁶.

However, journalist-driven innovation likely will still require improving journalists' digital skills, empowering them to take charge of more aspects of digital innovation. *'Do journalists need to have computational literacy, mathematical literacy, data literacy? These skills will be, in my opinion, essential in the long term'*, says researcher Konstantin Dörr.³⁷

This opinion is mirrored in a recent AP report on Artificial Intelligence in journalism. *'We've put a lot of effort into putting more journalists who have programming skills in the newsrooms'*, says The New York Times' chief technical officer, Nick Rockwell³⁸.

The news robots of the future

So far, automation is still quite limited. In a recent survey, journalists voiced concern about, among other things, the inability of current news robots to interrogate data. They noted that current automation is reliant on single, isolated data streams and is one-dimensional due to the quantitative data feeds it relies on³⁹.

It is yet beyond the scope of currently used automation technology to critically examine figures, let alone find causal links between data points or integrate external events. However, more sophisticated news robots are being thought about. Future developments are meant to address deficiencies in the current generation of automated journalism.

One point is variety. The next generation of algorithms will be able to handle multiple sources, says Neil Thurman. *'[T]echnology providers are working in this area try to provide*

³⁵ Email by Mejia to the author on May 4, 2017

³⁶ AX Semantics, Facebook post on August 7, 2015

³⁷ Telephone interview with Konstantin Dörr on May 25, 2017

³⁸ Marconi, Francesco/Siegman, Alex (2017): The Future of Augmented Journalism: A guide for newsrooms in the age of smart machines, Associated Press, page 5

³⁹ See Thurmann 2017, page 15

*additional data sources that can be referenced in the template that produces the automated story, in order to provide some kind of context.*⁴⁰

Thus an automated story on unemployment figures might provide a spotlight on the overall economic situation, or a company earnings story might feature figures on the situation of the wider industry.

Automation will be used to identify stories, too. Algorithms will find interesting patterns or outlier events in the data and flag them up for journalists to look at - a dimension of the technology Reg Chua of Reuters refers to as 'automating for insight'. This might take the form of automatically generated statements on the data, to be followed up on by journalists looking for explanations and reactions.

An example of 'automating for insight' is the US polling data algorithm trialled by Reuters. The robot looks at the polling data and generates statements on the most statistically significant results.⁴¹ This allows human reporters to make a quick assessment of the most interesting events within their data.

Another approach suggests moving beyond numbers. David Caswell and Konstantin Dörr have argued that new data models will have to be developed to allow for event-driven storytelling, rather than the static description of figures found in current automated stories. In a recent paper⁴², Caswell and Dörr outline how encoding text information in the right way allows stories about a sequences of events to be told.

The researchers describe a case study where they coded events based on Los Angeles police data to narrate a car chase. Their 'structured narratives' represent a novel approach to automated journalism that could allow for more complex robot storytelling in the future, at least for those who do not shy away from building sophisticated story databases.

Journalism and AI

Ultimately, the development of automated journalism will also depend on advances in the wider field of Artificial Intelligence (AI). Self-learning software could eventually lead to breakthroughs in data mining and make it much easier to work with unstructured data. User feedback could help to improve Natural Language Generation and enhance the linguistic diversity of stories.

Meanwhile, technologies such as speech-to-text and face recognition from images will vastly improve the workflow speed of journalists and add value to their existing archives by making them machine-searchable in new ways.

⁴⁰ Telephone interview with Neil Thurman on April 28, 2017

⁴¹ Talk by Reg Chua at the Newsrewired conference in London on February 8, 2017

⁴² Dörr/Caswell 2017

The use of AI in automation could not only mass-produce stories for news agencies, but also facilitate the work of investigative journalists and other forms of data-driven journalism. AI is a daunting prospect when looking at large data sets such as the 13.5 million documents related to the Panama Papers investigation, says Matthew Caruana Galizia of the International Consortium of Investigative Journalists (ICIJ). He thinks AI can greatly speed up investigations done by human journalists⁴³.

AI could help to find crucial information in vast data sets. In the case of the ICIJ's Offshore Leaks investigation, the software could learn to *'extract references to real-world entities, like corporations and people, and start looking for relationships between them, essentially building up context around each entity'*, Nick Diakopoulos speculated in 2013⁴⁴. It thereby could identify people such as the business partners of politicians hiding money offshore and accountants that set up thousands of shell companies.

The discussion above hints at some paths toward future automation. It remains to be seen what approaches will be adopted by news agencies, and how widespread the application of the technology will be. Reviewing some of the technology currently used and developed, it seems clear that automation technology in journalism is still in its infancy.

⁴³ Marconi/Siegman 2017, page 2

⁴⁴ Diakopoulos, Nick (2013): Can Artificial Intelligence Like IBM's Watson Do Investigative Journalism?, Fast Company

Conclusion

Automation is an increasingly important tool for content production in news agencies. The largest news agencies in the world, AP, Reuters and AFP, now produce thousands of stories each month with the help of algorithms. Robot journalism is also being tested and used by several smaller news organisations in Europe.

However, some news agencies are sitting on the fence, including large organisations such as the Spanish agency Efe and Ansa in Italy. This indicates that automated journalism cannot yet be considered ubiquitous in news production. News agencies appear reluctant to invest in automation projects that do not promise great economies or new business.

In my research, I took stock of automation projects in 15 news agencies across Europe and the US. I looked at what kind of content agencies automate and how sophisticated robot journalism software is. Furthermore, I investigated the organisational challenges news agencies face in automating content, and I highlighted the decisive role that data access plays in expanding automated journalism further.

I find that so far, automation is limited in its scope and complexity. Natural Language Generation is used mainly for reporting events in finance and sports, where structured data is most commonly available. Only two news agencies I surveyed use algorithms to compare new to historical data, thereby adding analytical value to their stories.

My findings suggest that the focus of news agencies is still very much on the low-hanging fruit: areas that are easy to automate with large numbers of stories as a reward. There has been far less small-scale automation or more complex storytelling based on historical data.

In the assessment of experts I surveyed for this paper, news agencies face organisational challenges that limit the progress of automation. I discussed the top-down mentality that, in some organisations, prevents easy-to-use-automation technology from being handed over to the newsroom. I also noted the lack of innovation culture in legacy media companies that plays a role in inhibiting investment in more far-reaching automation technology.

As my paper highlights, a key point for further expansion of automated text generation is access to structured data. Rich data sets allow for interpretation of patterns and comparison with historical figures. This is crucial for adding value to automated journalism beyond merely narrating figures in a financial earnings report or sports result.

Yet, I find that news agencies mostly rely on third parties for access to data. Few agencies are willing to create and maintain their own databases and invest resources in structuring historical data they already have.

This means automated journalism in news agencies is reliant on data providers that will, in some cases, take a share of the profit. Data providers such as sports associations and government agencies might also have non-monetary motives in providing access to data, and could restrict it when no longer expedient.

In my view, further development in the use and scope of automation will rely on building data handling capacity in organisations. Also, it can be advanced by handing over the necessary tools and skills to the newsroom.

Bottom-up automation requires making tools easier to use for non-programmers, and spreading knowledge on data literacy and computational skills. Journalists can thus be put in the position of automating parts of their work they deem unrewarding or repetitive, and focus on more high-value tasks.

Automation, however, is not just a help-meet for tired journalists. It likely will be crucial to breaking the link between the amount of content produced and the number of journalists there to produce it. News organizations in financially precarious situations thus will be able to use automation to shift workload away from their decreasing or stagnating human workforce, and allow their journalists to focus on adding higher-value human input into stories.

To conclude, we are far from being able to say automated journalism will become universal practice. Applications are limited by data access, technological limits and newsroom culture.

Yet news organizations, including leading news agencies, are willing to invest resources into developing the expansion of automation. This suggests that automation will be expanded in the future, and more sophisticated robots can take over tasks from human journalists. I thus suggest that those news agencies who advance their capabilities in data handling and Natural Language Generation will likely be the ones defining the future of journalism.

Bibliography

AX Semantics (2015), Facebook post on August 7, 2015, <https://de-de.facebook.com/AXsemantics/posts/499957900167973>

Blankespoor, Elisabeth/deHaan, Ed/Zhu, Christina (2017) Capital Market Effects of Media Synthesis and Dissemination: Evidence from Robo-Journalism, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2872784

Clerwall, Christer (2014): Enter the Robot Journalist, *Journalism Practice* 8:5, 519-531

Diakopoulos, Nicholas & Koliska, Michael (2016): Algorithmic Transparency in the News Media, *Digital Journalism*, DOI: 10.1080/21670811.2016.1208053

Diakopoulos, Nicholas (2013): Can Artificial Intelligence Like IBM's Watson Do Investigative Journalism?, *Fast Company*
<https://www.fastcompany.com/3021545/can-artificial-intelligence-like-ibms-watson-do-investigative-journalism>

Dörr, Konstantin/Caswell, (2017): Automated Journalism 2.0: Event-driven narratives, *Journalism Practice*

Dörr, Konstantin (2016) Mapping the field of Algorithmic Journalism, *Digital Journalism*, 4:6, 700-722

Dörr, Konstantin & Hollnbuchner, Katharina (2017) Ethical Challenges of Algorithmic Journalism, *Digital Journalism*, 5:4, 404-419

Global Open Data Index (2017), Open Knowledge Foundation, <https://index.okfn.org/>

Graefe, Andreas (2016): Guide to Automated Journalism, Tow Center for Digital Journalism, <http://towcenter.org/research/guide-to-automated-journalism/>

Linden, Carl-Gustav (2016): Decades of Automation in the Newsroom, *Digital Journalism*

Marconi, Francesco/Siegman, Alex (2017): The Future of Augmented Journalism: A guide for newsrooms in the age of smart machines, Associated Press

Monta, Tall & Reich, Zvi (2016): I, Robot. You, Journalist. Who is the Author? *Digital Journalism*

Thurman, Neil/Dörr, Konstantin/Kunert, Jessica (2017): When Reporters Get Hands-on with Robo-Writing, *Digital Journalism*

Tokbaeva, D. (2016): When Media and Management Collide: An Interview with Lucy Küng, *Westminster Papers in Communication and Culture*, 11(1), 26–30

Weeks, Lin (2014) Media Law and Copyright Implications of Automated Journalism, Jipel
No. 4 - Vol. 1, <http://jipel.law.nyu.edu/vol-4-no-1-3-weeks/>

Appendix: List of interviews

Telephone interview with Helen Vogt of NTB on February 15, 2017
Telephone interview with Maija Lappalainen on February 17, 2017
Telephone interview with Lisa Gibbs of the AP on February 23, 2017
Telephone interview with Reg Chua of Reuters on March 10, 2017
Telephone interview with Neil Thurman on April 28, 2017
In-person interview with Paul Maidment on May 18, 2017
Telephone interview with Nick Diakopoulos on May 22, 2017
Telephone interview with Konstantin Dörr on May 25, 2017