Automation of news production in Norway:

Augmenting newsroom with artificial intelligence.

Ekaterina Pashevich

Master's Thesis in Nordic Media
Department of Media & Communication
Faculty of Humanities

UNIVERSITY OF OSLO

Spring 2018

2018

Automation of news production in Norway: Augmenting newsroom with artificial intelligence.

Ekaterina Pashevich

http://www.duo.uio.no

Trykk: Reprosentralen, Universitetet i Oslo

Abstract

The current master thesis researches the problem of implementing automation in Norwegian newsrooms. The technology of Natural Language Generation, powered by artificial intelligence, allows partial automation of mundane and repetitive tasks, while freeing journalists to do more creative and challenging work, like data analysis and interviews. Along with the benefits in speed and accuracy, automated journalism allows creating innovative news products using personalization algorithms. Looking at the phenomenon from the media innovations perspective, the thesis answers the questions: *(1) How is automated journalism currently being implemented in newsrooms in Norway?* and *(2) What is its potential as an innovation?*

The study is based on empirical data from 11 in-depth interviews with journalists, system developers and scholars working with automated journalism in Norway, Sweden and Germany. Since the topic is so new, it was important to put Norwegian experience into the European context. The findings show that automation is suitable only for certain types of tasks in journalism, and it is beneficial only for specific types of newsrooms. The ethics of algorithms is still an important issue to explore, as computer reasoning is different from that of human and its outcome is difficult to predict. In the thesis I argue that journalism, although transformed by computational tools, still stays strong as a profession (at least for now) and mostly benefits from the introduction of automated text generation software. The possibilities of automation in newsroom along with the ethics of it should be thoroughly discussed by media scholars and made openly available for the news professionals.

Table of contents

Table of figures

Preface

This work started with my honest wish to “learn about the future”. For obvious reasons, it turned out to be impossible. Instead I learned something about “now” that made me think differently about the future: artificial intelligence is gradually diffusing into various spheres of our lives, leaving us with barely any choice but to accept these changes. I took a controversial topic of automation in journalism, probably because I wanted to be proven wrong. I wanted to find the evidence that no, algorithms are not as mighty as the human brain and far less creative. I found it only to realize that I was asking the wrong question. The algorithms do not have to simulate human intelligence to be useful and to transform the media industry as we know it.

I would probably not dare to take a topic like this if it was not for my brilliant supervisor Anders Fagerjord, who convinced me to study something that I “really, really wanted”. He always came up with very creative tricks in times of my deepest emotional struggle to finish this work. I want to thank the journalists, developers and scholars who agreed to reflect on the issue of newsroom automation and dared to bring this innovation to life in their organizations. You are a true inspiration to me and I hope you will continue to inspire your colleagues. During my research I was supported by the Work Research Institute (AFI), Oslo Metropolitan University (OsloMet) and the Norwegian Council of Applied Media Research (RAM), who trusted me with their scholarships in order to accomplish my rather ambitious project to compare the situation with automated journalism in several countries.

During the whole writing process I was lucky to have my co-student Jessica Robinson by my side. Her American work ethic and warm support, as well as her valuable feedback, made the exhausting and lonely writing process more enjoyable. I want to thank my friends who took their time to read and give me their comments on the thesis: Robert Vaagan, Ingrid Thorseth, Anna Sendra, Cormac Lally and Ion Dronic. Finally, I want to thank my family in Russia who always support me, and especially this last year. In particular, this thesis would never be completed without my mother’s vigilance.

I hope you will enjoy the reading!

1 Introduction

The global automation trend occupies many industry leaders nowadays. While technology brings significant advantages for more successful competition in the markets, it is threatening to replace certain kinds of jobs, which may result in severe unemployment and unprecedented income inequality (Allen, 2017). Starting with industrial automation, when human craft was substituted by machines in factories, automation is now able to perform not only manual, but also cognitive tasks, and is gradually spreading through banking and finance to more creative spheres, like journalism. IBM Chief Executive Officer Ginni Rometty argues that the danger lies not so much in the number of jobs, but their changing character and complexity: "when it comes to complete job replacement, it will be a very small percentage; when it comes to changing a job and what you do, it'll be 100 percent" (Cao, 2017). But journalism and creative writing are still thought to be human prerogatives. In addition to the creative side of the profession, it also has a vital societal role of informing the citizens about the current state of events and thus supporting the democracy. Should we let an impersonal algorithm influence our democratic choices? Therefore, automation in journalism is a highly controversial issue, creating heated debates between editorial management, journalists, journalistic trade unions and scholars.

The present master thesis studies the automation tendency in modern newsrooms. The recent progress in the development of artificial intelligence and computation produced the technology capable of automatically writing texts which could be used along with human-written materials. The Natural Language Generation (NLG) technology that lies at the core of automated journalism allows for producing news faster, at larger scale and to automate repetitive and manual journalistic tasks. Kris Hammond, CEO for Narrative Science, the startup developing software for automated journalism, once made a prognosis that more than 90 percent of news would be written by computers by 2030 (Levy, 2012). The sphere of journalism is steadily becoming more technologically saturated, but the adoption of artificial intelligence software for the task of writing texts can be seen as a break-through phenomenon with the potential to shatter the fundamental ethics of journalism and its established practices. In other words, automated journalism might become a disruptive innovation for the news media market. The way automated journalism is being implemented in newsrooms today is a topic worth researching as it touches the issues of the quality of journalism, its ethics and even

employability of journalists. It seems necessary to study this conflict in order to see the clear picture of the possible developments of this technology, as well as its likelihood of being adopted by newsrooms.

Despite its relative novelty, the phenomenon of automated journalism has already attracted significant attention of scholars in various fields. The research to date can be divided into several major problematic sections: the quality of machine-written texts and their perception by the audience, the issues of authorship and regulation, the ethical issues and journalistic practices, last but not least, concerns about further job loss as a consequence of automation. The goal of the present research is to describe the phenomenon of automated journalism and to (1) put it in context with similar newsroom innovations, (2) identify its positive and negative sides and (3) to research its risks. The research questions of the study are *(1) How is automated journalism currently being implemented in newsrooms in Norway?* and *(2) What is its potential as an innovation?* Since global automation seems inevitable (Rotman, 2013; Manyika et al., 2017), it is of high importance to study which parts of the profession should and which must never be automated. The study aims primarily to analyze the development of automated journalism in Norway, and in order to answer the second research question, it is necessary to compare the results of the Norwegian study with the results from other countries developing similar technologies for journalism. The current thesis is targeted at students and researchers interested in the interdisciplinary studies of journalism, media economy and artificial intelligence. The results of the research will also be of use for media professionals pondering the thought of automating certain types of tasks in their newsrooms.

To answer the research questions, 11 qualitative in-depth interviews were conducted with professionals working with automated journalism or interested in this innovation in Norway, Sweden and Germany. The interviews covered the topics of technological functionality, innovative potential and the impact of automated journalism on the journalistic profession. This method allowed creating a coherent picture of the state of this innovation in Norway, as well as to give a broader perspective when it comes to the spread of this technology in the world.

The thesis, taking into consideration the complexity of the phenomenon, is structured thematically. Chapters 2 and 3 explain the theoretical and methodological approaches to answering the research questions, where Chapter 2 gives an overview of the aspects that lie in the background of the studied phenomenon of automated journalism as a media innovation, and presents the Diffusion of Innovations theory (Rogers, 2003), chosen as a basis for the analysis.

The methods chapter discusses the possible approaches to answering the given research questions and explains the design of the chosen methods. The results and the discussion parts of the thesis are divided into three major sections describing technological (Chapter 4), economic (Chapter 5) and practical (Chapter 6) aspects of implementing automated journalism. In the first part of Chapter 4 I show the variety of technological forms that automated journalism can take, the quality benefits which it offers to journalism (its value proposition), then I justify why human journalism will not be harmed or substituted by automation, and discuss the future potential of this innovation. In the second part of the technological chapter, I look at the ethical challenges of automated journalism, while drawing attention to the most common problems with data sources and algorithms. Chapter 5 directly answers the research questions. In the first part, I discuss automated journalism as an innovation and argue why it can be considered disruptive for the news media market. Then, with the help of the Diffusion of Innovations theory (Rogers, 2003), I analyze how the technology is being developed in newsrooms in Norway and whether it has a potential to be successfully adopted by some of them. Here I pay special attention to the choices that newsrooms make based on their available resources and needs. The second part of the chapter 5 discusses automated journalism in connection to another recent journalistic phenomenon – personalization of news, powered by artificial intelligence. Here I explain why news personalization arguably presents the main revenue source for newsrooms deciding to introduce automation and talk about the ethical challenges. Chapter 6 is devoted to the changes in journalistic practices due to the recent computational turn in the profession. Here I discuss the arguments for and against journalists acquiring technical skills and introducing automation. The chapter also touches upon the topic of cooperation between journalists and system developers in newsrooms and different models of developing automation products. Chapter 7 summarizes the results of the study, its limitations and gives recommendations for future research.

2 Theoretical approach and previous research

2.1 "Automated journalism" as a concept

"Automated journalism" is now a widely accepted term for the automation technology used for writing texts that is being implemented in newsrooms (Graefe, 2016; Linden, 2017a). There exist other automated technologies developed in newsrooms for different purposes, while automated text writing is used in other spheres, like e-commerce and finance. In the early studies, the scholars mainly used the term "robot-" or "robotic journalism". Van Dalen (2012) and Clerwall (2014) focused on the advantages and disadvantages of "robots" over human journalists and showed that the algorithms were capable of writing texts of human-level quality, though they were rather boring to read. Clerwall (2014) even raised a question of the need for human journalists: "if journalistic content produced by a piece of software is not (or is barely) discernible from content produced by a journalist, and/or if it is just a bit more boring and less pleasant to read, then why should news organizations allocate resources to human writers?" (pp. 526-527). The heated debates on the issue of robots replacing human journalists followed immediately. Linden (2017a) later answered to these concerns with a definite "no", arguing that journalism, as a creative profession with strong ideology, is at low or no risk of automation. Nevertheless, calling the phenomenon "robotic" triggered incorrect associations with robotics (embodied agents typing news), while in fact at the core of automated journalism is a piece of software. Although robots may become a part of the future newsroom (Latar, 2015; Marconi et al., 2017), the concept discussed in this study is the use of NLG technology in journalism, first defined by Matt Carlson as "algorithmic processes that convert data into narrative news texts with limited to no human intervention beyond the initial programming choices" (2015, p. 416), meaning that the technology allows composing journalistic materials, using provided data, and automatically publishing articles without additional editing.

Currently there are two major scholarly attitudes towards automated journalism. The first one presents technological optimism – the idea of technology gradually improving the lives of people and making the world a better place (Agar, 2015). More profoundly, techno optimists regard automated journalism as an opportunity to cut costs and increase the quantity and quality

of news (Graefe, 2016). The alternative point of view is techno-pessimism – the idea that technology often leads to unpredictable, even dangerous results. Automated journalism invokes concerns about its practical implementation, the need to dismiss even more journalists, as well as the ethical and legal challenges. The recent studies on automated journalism identified several problems with the use of artificial intelligence in journalism. Weeks (2014), Montal & Reich (2017) and Ombelet et al. (2016) discussed the legal aspects of newsroom automation, such as authorship, copyright and legal responsibility for inaccurate content; Dörr & Hollnbuchner (2016) are occupied with the ethical issues around automated journalism, while Van der Kaa & Krahmer (2014), Latar (2015) and Bucher (2016) studied practical concerns around it. Since it became rather obvious that automation was not going to fully replace journalists in newsrooms in the near future, but only some of their tasks (Graefe, 2016; Linden, 2017a; Marconi et al., 2017), which the present study also proves, the latest reports focus on automated journalism as a part of a complex of artificial intelligence tools that are going to "augment" the newsroom (Marconi, 2017).

In Norway, the topic was recently discussed by the journalists Magne Soundjock Otterdal and Geir Terje Ruud in their essay "Autostory" (2017), where they explore the use of automated journalism and similar computational innovations in Norwegian newsrooms and reflect on the future of journalism in the "world of robots". It is the opinion of the author of this thesis that the innovative aspect of automated journalism has not yet been paid enough attention by scholars, considering the general artificial intelligence trend and the obvious benefits that automation can bring to the news industry. The study done by Linden (2017b) looks at automated journalism from the media innovation perspective and touches upon several issues of practical use, technological possibilities and ethical concerns. The study is based on 24 interviews, including some Norwegian experts as well, but the main focus of the paper was mainly American news organizations. The present study, however, focuses on the potential of automated journalism to be adopted by newsrooms in Norway and the obstacles that this process might meet. The situation in Norway is also compared to that of neighboring Sweden and the technological leader of the continent – Germany.

Studying automated journalism as a media innovation requires acquiring a comparative knowledge in various interconnected spheres. Here I present a visualization of connections between the different concepts (see Figure 1), which I will elaborate on further in this chapter.

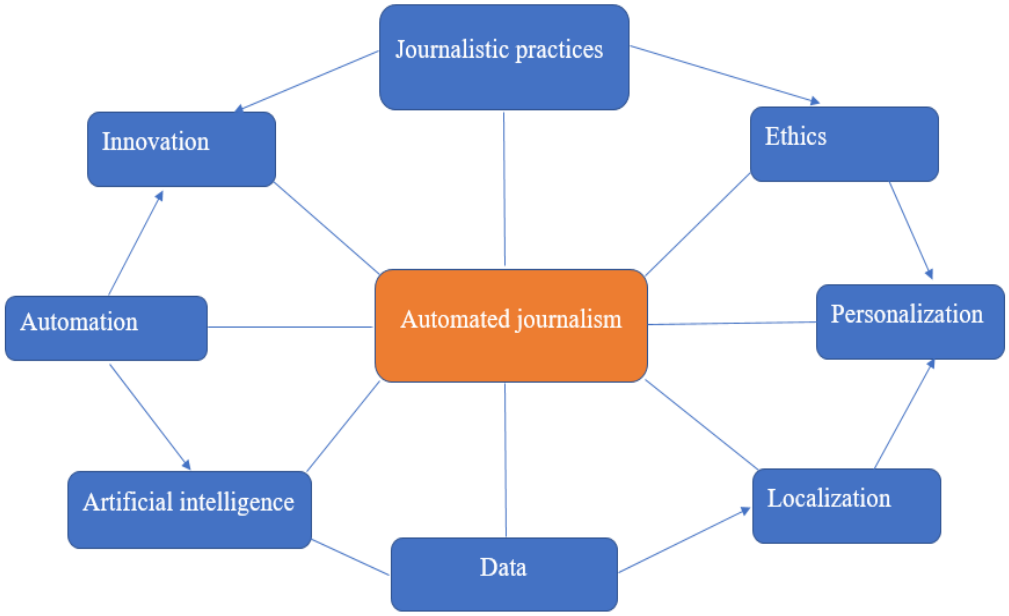

Figure 1. The relations between the areas where automated journalism plays an important role.

Firstly, I will look at the technology that lies at the core of automated journalism – Natural Language Generation (NLG) – and its relation to other types of artificial Intelligence (AI). Secondly, I will discuss the appropriate framework for analyzing the ethics of automated journalism – a sensitive topic due to the strong ethical and professional regulations in the journalistic sphere. Then I will explain and justify the choice of the Diffusion of Innovations theory (Rogers, 2003) that helps to assess the likelihood of automated journalism being successfully adopted in newsrooms. Since the issue of automated journalism is closely connected to the personalization (and customization) of news, which may bring significant economic benefits to media organizations and presents a promising innovation, the personalization theory will also be discussed. Here I will also discuss the history of relationships between journalism and technology, and will explain what is meant by the profession's "quantitative turn". Finally, there will be a summary of the highlights of the existing research on automated journalism.

2.2 NLG technology and its relation to AI

If less than a decade ago "artificial intelligence" as a concept was only discussed among computer engineers and data scientists (and was mainly a science fiction phenomenon), today it becomes steadily more popular to explore the applications of artificial intelligence in various spheres of human life. The ideas about designing intelligent machines were widely discussed in the 1950-60s, followed by two decades of relative stagnation (Tørresen, 2013). Thanks to three modern breakthroughs – inexpensive parallel computation, enormous volume of collected data on the internet and the advances in algorithms which made it possible to optimize the results of neural networks (deep learning) – the field of artificial intelligence in the recent decade experienced a resurgence (Pohl, 2015). And thanks to the recent victory of Google DeepMind's AlphaGo in the ancient Chinese game of Go, such terms as *artificial intelligence, machine learning* and *deep learning* became a part of the daily agenda. However, when we talk about artificial intelligence and the global automation trend, it is important to distinguish between the different types of this technology.

Artificial intelligence is not a solid concept, but rather an umbrella term for different techniques meant to increase the "intelligence" of computers. First of all, in scientific literature, the term *artificial intelligence* refers to artificial general intelligence (AGI) – the scientific attempts to create a machine of human-level intelligence, which is also the main goal of the research in this sphere. This term understandably lacks a clear definition. Goertzel & Pennachin (2007) describe AGI as "systems that possess a reasonable degree of self-understanding and autonomous self-control, and have the ability to solve a variety of complex problems in a variety of contexts, and to learn to solve new problems that they didn't know about at the time of their creation" (VI). There is a need for further explanations: what is "a reasonable degree" and how can "self-understanding" be interpreted for computers? The questions are endless. That is why pioneer of artificial intelligence Alan Turing in 1950 suggested his famous "test", which he called "The Imitation Game". In order to pass the test, the machine, by giving "answers that would naturally be given by a man", without being seen, needs to trick the human interrogator into thinking that there is a human on the receiving end (Turing, 1950). The rules of this game have been changing with progress in the research field of artificial intelligence. The film "Ex Machina", in which a female robot, Ava, manages to outsmart her creator and escape from the laboratory, reflects the fears of people being similarly tricked but in real life circumstances. The Google researcher and famous tech prognosticator, Ray Kurzweil, argues that machines can

reach the level of general intelligence by 2029 and superhuman intelligence (1000 times smarter than human) by 2045 (Barrat, 2013). There exist several theories of how it can be done, but neither the optimists, nor the skeptics can prove or disprove the possibility that such a machine can be created. Kurzweil suggested building a human-level artificial general intelligence by mapping the human brain, or the approach of *reverse-engineering* (Ganapati, 2010), while Goertzel, arguing that the architecture of a plane's wings is not similar to those of birds, suggested creating an artificial brain from scratch, relying solely on algorithms usual for the computer-type of thinking (Goertzel et al., 2014). Tsvelik (2012), however, criticizes these materialistic approaches and denies the possibility of the machines reaching the human-level intelligence, referring to the concept of "soul", which can only exist in living organisms.

Putting aside the aspirations to build a superhuman machine, other types of artificial intelligence are already capable of solving specific tasks much better than humans, which makes our lives easier. Artificial narrow intelligence (ANI), or *weak AI*, is an artificial intelligence method dedicated to automation of a particular task, like playing chess, driving cars, diagnosing diseases (Goertzel, 2007), or recognizing voices and faces. Personal assistants, like Apple's Siri, Microsoft's Cortana and Amazon's Alexa, are the prominent examples of a rather powerful narrow AI. Automated journalism is also a typical example of ANI – the smart assistant in a specific task – creating narratives. To build a narrow AI, the method of *machine learning* is used. In a nutshell, this process means making machines learn concepts and relations between them by "training" the algorithms on various examples, without explicit programming or any specific instructions, so that the machine could recognize different and similar patterns in the data. It is often used for making predictions based on data and building models explaining various processes (Shalev-Shwartz & Ben-David, 2014).

Though today there are numerous examples of weak AI, humanity is not even close to creating an intelligent and self-conscious robot. The reason the question of people being replaced by machines is being discussed today is the previously mentioned breakthrough in the research on artificial intelligence, namely *deep learning* (Goodfellow et al., 2016) – an approach in machine learning, in which algorithms are trained to learn from the experience to solve specific problems, using very large and deep (due to many layers) neural networks. Neural networks are machine learning models that connect many linear and non-linear functions in a layered way to make a prediction about a problem; they are "logical constructions which ask a number of binary questions of every bit of data which passes through them and then classifies the answers received" (Murray-Hundley, 2017). The hierarchy of concepts allows the computer

to learn complicated concepts by building them out of simpler ones: “If we draw a graph showing how these concepts are built on top of each other, the graph is deep, with many layers. For this reason, we call this approach to AI deep learning” (Goodfellow et al., 2016, pp. 1-2). The technology among others helped significantly improve one of the most popular linguistic services, Google Translate (Lewis-Kraus, 2016).

Despite its linguistic complexity, automated journalism does not require the use of deep learning, as for most of the cases machine learning is enough. The technology behind the most advanced forms of automated journalism – Natural Language Generation (NLG) – is a type of Natural Language Processing (NLP), also called Computational Linguistics (CL). Another type of NLP is Natural Language Understanding (NLU) – software transforming raw texts into the structured data sets (Reiter, 2017). NLG does the exact opposite: it transforms well-structured data into a narrative (Wright, 2015). NLG as a concept has existed since the middle of 1980s (McDonald & Bolc, 1988), however, the recent advancements in machine learning allowed the business sphere to widely implement this narrow AI technology: NLG is used for creating automated content in several languages in e-commerce (product descriptions), online advertising, financial reports, personalized customer communications and real estate property descriptions (Automated Insights, 2017). NLG provides the opportunity to communicate properly analyzed and interpreted data in a narrative format, fast and on a large scale, which is crucial in the age of ubiquitous big data. NLG generally allows the reduction of production costs, as well as personalization of content based on customer demands. This software is widely used in programming of personal assistants (Siri, Alexa and Cortana) to produce limited responses. In short, NLG makes it possible to take a high quality well-structured data input, get it through a template-based algorithmic tree – the system of hierarchically organized (from “root” to “leaves”) rules of grammar and syntax – and get a coherent text as an output (see Figure 2).

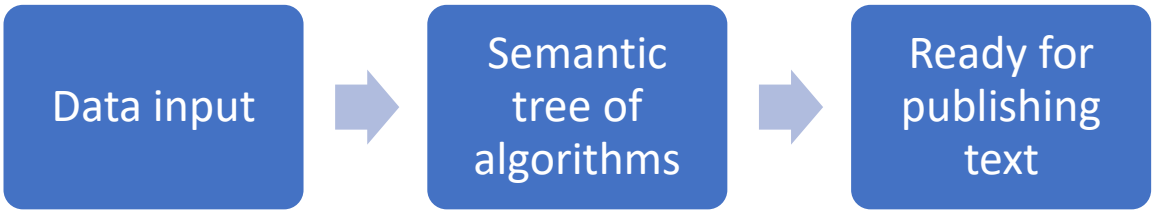

Figure 2. The structure and functionality of automated journalism.

Narrative Science identifies the following 3 stages of NLG complexity (Glascott, 2017, March 14):

1. Basic NLG – automatically translates data into text via Excel-like functions (used in mail merge to restate numbers into language);
2. Templated NLG – the algorithmic tree built based on the user-written templates. The system basically recombines the sentences and phrases from a limited number of texts according to the user-determined fixed rules of connecting ideas and the interpretation of output. Templated NLG systems use a limited number of data sources for their input, cannot perform advanced analytics and should be programmed for each purpose almost from scratch. They are not very "intelligent", but suit perfectly the journalistic needs, as they are relatively easy to control (used in the majority of automated journalistic reports on finance, sports, weather, etc.);
3. Advanced NLG – is meant to communicate like a human. It can identify interesting events in the data, target them to a specific audience and write coherent materials based on this autonomous research. Its language is fluent, easy and pleasant to read. Advanced NLG understands the context and can write suitable texts at an inhuman speed and scale. This level of complexity, however, is full of challenges (used in personal assistants, like Siri, Alexa and Cortana).

The applications of natural language generation go far beyond the journalistic sphere: the products of NLG software companies target mainly the financial industry, advertising and any business where content production is involved. The journalistic sphere, however, appears to be the most controversial area for automation. Here, the automated generation of texts meets certain ethical concerns and, therefore, needs to be well-researched before being widely implemented.

Marconi et al. (2017) introduce the concept of "augmented journalism" and discuss various types of artificial intelligence technologies that will assist journalists in their work. These include: supervised and unsupervised machine learning for finding interesting stories by

comparing the data in the data sets and official, as well as commercial, documents; Natural Language Processing (NLP) divided into Natural Language Understanding (NLU) and Natural Language Generation (NLG = automated journalism); software transforming speech to text and back; image recognition and computer vision software; and also robotics (Marconi et al., 2017, p. 7). Therefore, automated journalism represents only one facet of emerging forms of newsroom automation. In the near future, artificial intelligence technologies will assist journalists in various tasks, including writing materials. But there is a big obstacle looming over this optimistic plan – namely ethics.

2.3 Ethical aspects of automated journalism

When in 2014 the first automated "journalist" entered the newsroom of Associated Press, this decision immediately caused ethical concerns: should the readers know that the author of the article is a machine or should this fact be concealed? Who is the author of the material from the legal perspective (Weeks, 2014; Montal & Reich, 2017)? Who takes responsibility for the correctness of the facts due to the possible errors in data (Ombelet et al., 2016)? How do you teach a robot to obey the journalistic ethics (Diakopoulos & Koliska, 2016)? There seem to be no end to the ethical concerns about the automation of text writing in the absolute human domain – the newsroom.

A possible way of approaching ethics in regard to automated journalism is to perceive them as lying at the intersection of computer ethics, ethics of building intelligent machines (AI) and journalistic ethics. The Institute of Electrical and Electronics Engineers (IEEE, 2017) gives an overview of ethical problems in designing advanced computer systems in order for them to be aligned to our moral values and to prioritize human wellbeing. Among them: consensual use of personal data, avoiding discrimination, universal internet access, legal status of intelligent machines, respect of the internationally recognized human rights, accountability of designers and producers, minimization of risks of their misuse and other issues. Amodei et al. (2016) discuss the research problems connected to the general issue of "accidents" or "unintended and harmful behavior" of machine learning systems depending on where in the process these problems occur; while Hovy and Spruit (2016) discuss the social impact of NLP systems as the technology becomes more wide-spread, apart from the privacy concerns, including social justice (as all data sets carry demographic biases), equal access to resources in society,

representativeness of various social groups and other typical problems with statistics. At the moment, due to the relative simplicity of the technology that serves journalistic purposes and its limited implementation, it is not relevant to discuss the ethics of intelligent machines when we talk about automated journalism. But the ethics of machine learning systems are important to take into account in this case.

Journalistic ethics, however, plays a more significant role in the conflict of introducing automated writers in newsrooms. Since mass media have a great impact on society, the ethics of journalism has always been widely discussed by media scholars (Allern, 2001; Jacquette, 2007). The professional code of journalistic ethics comprises the principles of ethics and of good practice to meet the specific challenges faced by journalists. The modern global code of journalistic ethics, which appeared in the middle of the previous century as a form of self-regulation between the members of the journalistic profession (IFJ, 1954), includes the following principles: the duty of the journalists to provide the public with truthful information, taking full responsibility for the credibility of their sources, as well as protection of these sources, and preventing themselves and other journalists from the information manipulation.

Ess (2009) discusses more broadly "digital media ethics", which studies the moral, legal, and social issues (privacy, property, free speech, etc.) in cyber technologies. Dörr and Hollnbuchner (2016) argue that the ethics of automated journalism should be regarded at the intersection of digital media ethics (from social sciences), with a special focus on journalistic ethics, and cyberethics (from computer science). The scholars identify the following areas of technical ethical concerns in automated journalism: on the levels of data search and origin (Bradshaw 2014), algorithmic authority (Carlson, 2015), algorithmic objectivity (Gillespie, 2014), algorithmic transparency (McBride & Rosenstiel, 2013), data usage and abuse (Zion & Craig, 2014), and on the level of values and reasoning embedded in code (Kraemer et al., 2011; Young & Hermida, 2014). In this study, the ethics of automated journalism will be discussed in two major dimensions: technical (the quality of automated journalism in relation to human journalism) and the impact of automated journalism on the established journalistic practices and the newsroom organization.

Since the technology presupposes automated publishing without pre-editing (Carlson, 2015), the danger that a factual mistake might easily slip into journalistic materials is high enough to seriously discuss this issue. The danger of the automated journalism making misleading statements occupies many researchers. From the technological point of view, Dörr and Hollnbuchner (2016) distinguish the following possible stages where the ethical conflict

can appear in automated journalism: before (input), during (throughput), and after (output) content generation. The researchers from Thomson Reuters in the US and Britain discuss more deeply the ethical aspects of Natural Language Generation systems (Smiley et al., 2017) and raise a question of the correct interpretation of underlying data by algorithms inside the software. The researchers even argue this kind of misleading information might have "a real-world impact" (Smiley et al., 2017, p.103).

Unfortunately, we cannot be sure whether algorithms give a more objective assessment than a human does alone. The reason is the underlying assumptions that the algorithms base their assessments on, which are not often clear even to their designers (Spielkamp, 2017). Not many developers dare to admit this. When trusting a machine algorithm, it is important to be aware that it was programmed by a human and is not neutral by definition. ProPublica, a Pulitzer Prize-winning nonprofit news organization, is studying the biases in algorithms. In 2016, they investigated the accuracy of algorithms in the risk assessment software COMPAS that is used in the US for predicting future criminals, or the likelihood of known criminals of committing a new crime in the future. The journalists detected significant bias against black criminals, who, according to the algorithm, were much more likely to commit a crime than the white criminals, despite the fact that the latter had more serious crime records (Angwin et al., 2016). If the algorithms produce such mistakes now, who will guarantee they work correctly in journalism? If such a mistake is revealed in automated journalism, the consequences for the news organizations might be grave.

Human assessment abilities are not perfect either, however. The study of Danziger et al. (2011) from the Ben-Gurion University of the Negev shows that the decisions of Israeli judges are influenced not only by laws and facts, but also by rest and meal breaks. The researchers discovered that "the percentage of favorable rulings drops gradually from ≈65% to nearly zero within each decision session and returns abruptly to ≈65% after a break" (p. 6889). There are many modern examples that prove the fact that the algorithms must be taken under control because they, although originally considered neutral, are made by people, who all have certain prejudices. Especially when trained using the internet as a data input (as opposed to specific examples) the algorithms become the reflections of the average thinking in society, advocates of the points of view of the majority. The Microsoft Tay-bot, an example of a robot trained freely on (taught on the examples from) the internet, very soon started to produce racist and hateful remarks (Vincent, 2016). Therefore, specific attention should be paid while training the algorithms in NLG systems that later will be used in journalism. The most efficient result will

be achieved by augmenting the machine and human assessment skills with regular checks of the activity of both.

Smiley et al. (2017) urge both the journalistic and research communities to focus on developing automated text-writing systems that could avoid ethical violations. That is why they offered "the list of best practices for building NLG systems", including the sections on "human consequences", "data issues", "generation issues" and "data provenance". This checklist is based on another ethical checklist offered earlier by the standards editor in Associated Press, Tom Kent (2015), who argues it is important to "devote to your news decisions on automated news the same amount of effort you devote to your ethics and objectivity decisions at any other kind of news". In order for NLG technology to become widely implemented in newsrooms, the ethical principles of digital media and journalism must be respected and integrated into these systems. It is, therefore, understandable why the idea of introducing automated writers to newsrooms causes so many doubts among journalists, as making the machine act according to the established journalistic rules might be a difficult task.

Along with the main ethical concerns about automated journalism – namely the quality of data and its interpretation by NLG algorithms – there is another important ethical aspect that occurs during the process of implementation of automated journalism in newsrooms: any innovation in the newsroom affects established journalistic practices, which in turn affect the success of the innovation (Ekdale et al., 2015). The introduction of automated journalism in newsrooms triggers a discussion about the increase in unemployment in journalism due to the automation of journalistic tasks and the cost-saving interest of the newsroom management, as well as additional pressure on journalists when it comes to acquiring digital skills and learning programming. Cohen (2015) expresses concerns about the future development of automated journalism. He worries this labor-saving technology can reorganize the news production in a way that the responsibility for the quality of the content will be taken from journalists almost completely, resulting in, among other things, lowering of labor costs and weakening of trade unions. Linden (2017a), however, underlines that even the most advanced forms of automated journalism will not necessarily directly lead to job losses, as journalists always adapt to new forms of technology. The outcome of the implementation of automated journalism in newsrooms will depend on the characteristics of this innovation, as well as on the way it is introduced and spread.

2.4 Technological media innovation

News production has always been a business, where publishers strive for the reduction of costs to keep their enterprises afloat. Today, due to the digital shift which has weakened the economic power of print, the news media industry is experiencing a never before seen crisis and is on constant alert for the innovative solutions that can reduce its financial pressure (Pavlik, 2013; Vaagan & Barland, 2015). News organizations in North America and Western Europe have started closing their foreign bureaus and reducing newsroom staffing (Boyles, 2016). But the real problem rests in the fact that the news organizations did not come up with the appropriate business models for digital on time. Giving the history of internet, readers of digital content often expect it to be available online for free, though the value of content remains equal both in print and online. Therefore, news organizations, in order to survive, are forced to create new business models, while simultaneously developing new habits in the digital content consumers (Barland, 2015). In this thesis I strive to analyze automated journalism primarily from the media innovations perspective in order to see whether the introduction of automation in the media industry can help relieve the financial burden on print.

An innovation is an idea, a concept, or a technology, perceived as new by a community ready for its adoption (Rogers, 2003). Being new does not mean a completely new phenomenon, as, more often than not, an innovation is a new combination of existing technologies, methods, and business models (Schumpeter, 1934). A new idea becomes an innovation only when it is introduced on a market and has a commercial value (Fagerberg et al., 2004). Innovation is also a process of putting new ideas into practice (Tidd et al., 2005). Kline and Rosenberg (1986) argue that innovations are not well-defined and homogenous and, therefore, can undergo drastic changes during their adoption. In an attempt to explain the innovation process in firms, the researchers introduced the chain-linked model of innovation with “feedback loops”, which allow making subsequent improvements after the new technology was first implemented (Kline and Rosenberg, 1986). When it comes to the process of media innovation, the phases of production/development of an innovation and its diffusion/spread are often closely connected (Dogruel, 2014).

Innovations can have different effects and consequences. Some of them bring small changes within an enterprise, while others can revolutionize their market. Schumpeter (1934) divided innovations into incremental and radical according to the way they are introduced and the scale of their impact. *Incremental innovation* follows logical development and brings small

improvements to an existing product, business model, or a process. It is similar to Christensen's term *sustaining technologies* which aim to improve the performance of a product (Christensen, 1997). Christensen makes a distinction between sustaining and *disruptive technologies*. The latter lead to a temporarily worse performance of the product, but, by bringing a different value proposition (better solution to the customer's problem) to the market, they become more attractive to the customers and eventually change the market. Moreover, such innovations often precipitate the leading firm's failure (Christensen, 1997). Disruptive innovations are often what Schumpeter called "radical innovations", which change the economy and destroy the established order through the process of *creative destruction* (Schumpeter, 1942).

Innovations are considered crucial for meeting the social, economic and technological challenges; the ability to come up with innovative solutions makes it possible for businesses to survive in highly competitive environments (Küng, 2013). Innovations are needed for the sustainable development of the media industry. According to Lucy Küng (2013), there has never been a period in the media history when the industry was not innovative. Dogruel (2014) mentions "continuous need for newness" as one of the major characteristics of media innovations (p. 55). Today, with the rapid technological progress, the necessity to innovate in media is as urgent and challenging as ever, as technology and innovation are "inextricably linked" (Storsul & Krumsvik, 2013). Moreover, today the role of technology in modern newsrooms has changed. Previously, technology was quite stable for long periods of time, being, as Küng (2013, p. 9) puts it, an "enabler of, not a contributor to" the generation of content. Today, in the age of internet, web 2.0 and social media, the barriers between technology and content creation have become virtually non-existent. Technologies steadily integrate with the content, opening new possibilities for storytelling. Dogruel (2014) names "close interaction between creative and technological aspects" as one of the major characteristics of media innovations (p. 56). According to Küng (2013), this leads to the situation where technological skills become equally important in the media industry as the ability to produce high-quality content. This tendency in journalism proves the changing character of the newsroom staff, where programmers are working steadily more closely with journalists, and where journalists become programmers and vice versa.

In order to assess the likelihood of automated journalism being adopted by newsrooms and its future development in Norway, I will use the Diffusion of Innovations theory by Rogers (2003). It is common among the researchers in the field of media innovations, focusing on new technological devices as new media consumer products, to use this theory (Garrison, 2001;

Micó et al., 2013; Ekdale et al., 2015). Automated journalism can be classified as such "device" in newsrooms. Rogers defines "diffusion" as "the process by which an innovation is communicated through certain channels over time among the members of a social system" (Rogers, 2003, p. 11). The theory originally emerged from an attempt to explain the uneven adoption of hybrid seed corn (Ryan & Gross, 1943) by studying non-economic factors in farmers' economic decisions, and later was applied to various spheres from banking to education. The theory found its successful implication in technological adoption, for example, in medicine (Fitzgerald et al., 2002; Greenhalgh et al., 2008). When applied to journalism, the theory helps understand the various innovative processes in newsrooms and explain the struggles of introducing innovations to the journalistic community.

The Diffusion of Innovations theory defines the timeline for the innovation's occurrence: starting with invention, to diffusion (or communication) through the social system, and finally to consequences – these being either adoption or rejection. Rogers argues (2003) that all innovations first go through the innovation-decision process, which consists of 5 sequential stages: knowledge, persuasion, decision, implementation and confirmation (see Figure 3).

- *Knowledge* stage comes when a subject of an innovation implementation (an individual or a company) gets to know about the innovation and starts to explore its functionality;
- *Persuasion* occurs when a subject or a decision-maker forms an opinion about the innovation – whether it is favorable or not;
- *Decision* comes with the actual activities that lead to a choice of either adoption of the innovation or its rejection;
- *Implementation* phase means that this innovation is put to use;
- *Confirmation* is the result of testing the new idea, which leads to its possible abortion or reinforcement of the taken decision.

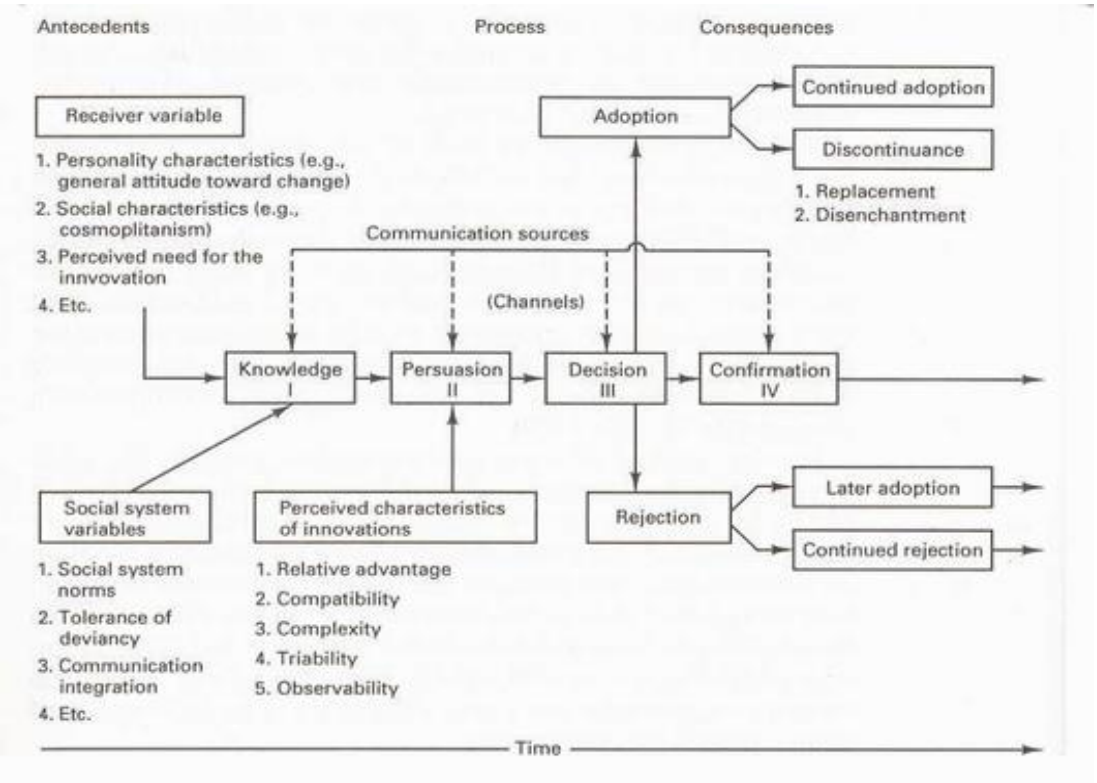

Figure 3. Diffusion of Innovations model. Source: Rogers (2003, p.170)

The essence of the innovation-decision process is to reduce uncertainty about the advantages and disadvantages of an innovation. During the *persuasion* phase, the new technology is assessed by the members of the community, which in the case of automated journalism are journalists and the newsroom management. Whether automated journalism will be adopted by newsrooms depends on the characteristics of this innovation, which can increase or decrease the likelihood of its successful diffusion. Rogers (2003) provides the researchers with the following innovation attributes that aim to clarify why certain ideas diffuse more quickly and easily than others, or to determine their *rate of adoption* – "the relative speed with which an innovation is adopted by members of a social system" (Rogers, 2003, p. 221):

- *Relative advantage*, or "the degree to which an innovation is perceived as being better than the idea it supersedes" (Rogers, 2003, p. 229);
- *Compatibility*, or "the degree to which an innovation is perceived as consistent with the existing values, past experiences and needs of potential adopters" (p. 240);
- *Complexity*, or "the degree to which an innovation is perceived as relatively difficult to understand and use" (p. 257);

- *Trialability*, or "the degree to which an innovation may be experimented with on a limited basis" (p. 258);
- *Observability*, or "the degree to which the results of an innovation are visible to others" (p. 258).

Rogers (2003) argues that the adoption of an innovation does not happen simultaneously by all the members of the community, but rather, gradually. The stages of adoption present a normal distribution curve (as many human traits are normally distributed, see Figure 4), where the "innovators" and "early adopters" represent approximately 15% of the community members; and even after the innovation has been successfully adopted, 15% are still "laggards" who refuse to use the innovation to the last. Rogers (2003) identifies "adopter categories" based on the innovativeness (how early an individual is in adoption of new ideas) of their members.

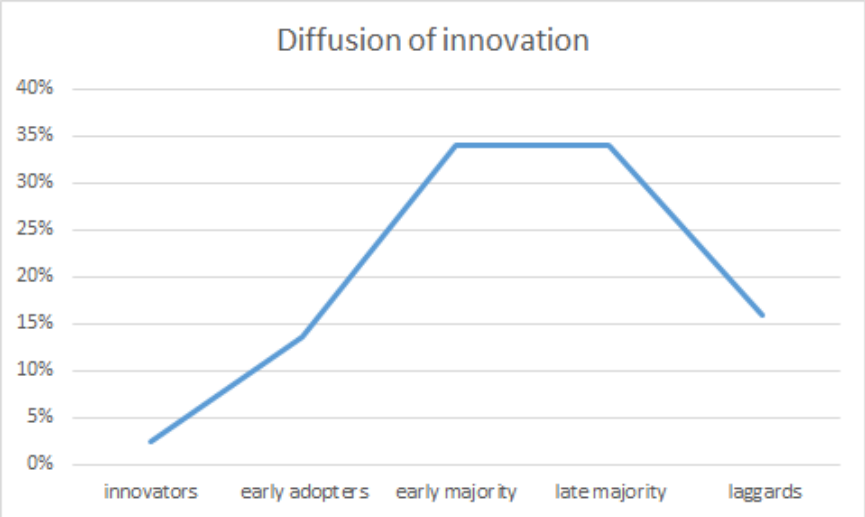

Figure 4. Normal distribution of Rogers' "adopter categories" (2003).

Yet another aspect of the theory that is important to keep in mind is that the diffusion of innovations does not happen exclusively and independently. The innovations overlap and interact with each other, influencing their development and adoption by communities, while forming the clusters of new ideas (Ekdale et al., 2015). Thus, automated journalism, though representing a completely new technology of Natural Language Generation (NLG), being analyzed from the media innovations perspective in newsrooms, is an extension of other computer and data-related journalistic tools, having computational journalism as its closest sibling (Coddington, 2015).

In organizations, the innovation process is rather complex and has, according to Rogers (2003), the following five stages – 2 in *initiation* and 3 in *implementation* phases of innovation diffusion:

1. *agenda-setting* (identifying organizational problems that create a perceived need for innovation);
2. *matching* (fitting an innovation to the identified problem and planning for implementation);
3. *redefinition and restructuring* (as the innovation is tailored to the organization's specific needs and structures, which are typically altered in the process);
4. *clarifying* (occurs as an innovation gains more widespread use throughout the organization, and its contextual meaning becomes clearer to adopting individuals);
5. *routinizing* (when the innovation is incorporated into the normal organizational activities, the major challenge is to secure its *sustainability* – "the degree to which an innovation continues to be used after the initial effort to secure adoption is completed" (Rogers, 2003, p. 435).

The consequences of the innovation diffusion are rarely studied by scholars due to the uncertainty and unpredictability of their nature. Often it is assumed that innovations are generally beneficial for their adopters, but this demonstrates the pro-innovation bias (Rogers, 2003). Innovations often have both positive and negative consequences. Ideally, the potential results and consequences of an innovation should be explored before its implementation, but the changes produced by an innovation usually happen over extended periods of time. Therefore, to study consequences, a researcher should conduct interviews both before and after the innovation is introduced. The current thesis studies automated journalism as an innovation in Norwegian newsrooms at quite an early stage, and, therefore, focuses on its current implementation and its potential. To assess the consequences of this innovation, there will be a need in a new study in several years.

Automated journalism plays an important role in the personalization of news content trend, as it is capable of producing numerous articles in seconds and even to customize them to any type of reader. Furthermore, personalized news is a product which may make the NLG

technology profitable for news organizations and, therefore, directly influence the adoption or rejection of this innovation by newsrooms.

2.5 Personalization and customization of news

Personalization in general means making content personal, relevant and, therefore, more favorable for the consumers. In the modern media context of information inundation, it is crucial to be able to filter the incoming information and to digest it at a reasonable pace in order to make sense of it. The idea of making content personal has always been attractive for the media industry. Personalization allows establishing direct contact with the consumer, learning about her preferences and habits, which presents a great source of income for the industry (Fink et al., 2002).

Hindman (2012) claims similar ideas were discussed back in 1970s. In 1993 the progressive MIT community launched a project "Fishwrap", which was an experimental electronic newspaper that would provide the readers – mainly MIT freshman students – with the news from their hometowns and stories of their interests, together with a selection of general news (Chesnais et al., 1995). Bender et al. (1996) discuss this example in the context of news-as-a-service model, designed for closer cooperation between news provider and news consumer. In 1995, Nicholas Negroponte in his book *Being Digital* described a futuristic way of consuming news – unlike in a traditional newspaper, the news would not be prepared by journalists according to their expertise and taste, but would be delivered tailored specifically to the reader's interests and preferences – a personalized summary of daily news. Negroponte (1995, p. 153) imagines a potential economic solution for the newspaper industry using personalization:

> You might be willing to pay the Boston Globe a lot more for ten pages than for a hundred pages, if you could be confident that it was delivering you the right subset of information. You would consume every bit. Call it *The Daily Me*.

Negroponte's utopia of news personalization becomes steadily more realistic today because of two major factors: improved filtering technologies (resulted in recommendation systems) and the emergence of social media (Hindman, 2012). Filtering algorithms help collect, sort and prioritize data about the users' behavior. The "father" of algorithmic recommendations – online market Amazon – in 1998 introduced the item-based collaborative filtering that has

been providing relevant recommendations to the customers of this online market, finding its later implementation in such services as Netflix, YouTube, etc. (Smith & Linden, 2017). Pariser (2011) claims the "era of personalization" really began in the end of 2009 when Google published an announcement about the tracking of users' online data by 57 "signals", among them the type of browser, the time spent online/offline and search history. According to Google, this would improve their service by providing more relevant search results for an individual person, while the business reason behind this was better opportunities for advertisement targeting – the company's main revenue source (Battelle, 2006). Algorithms can target the ads for specific audiences, making it possible for advertisers to increase the chances of selling their products. This is what the majority of media companies today, including Google and Facebook, are built on: seemingly "free" service always comes with a price, which is, in that case, personal information (Pariser, 2011). Today it is possible to measure and accurately track people's behavior online, and for the media industry this means offering exactly the product that users want to consume. Many news websites today use "cookies", which register different categories of data while a person is active on the given website, so that later they could give the same IP-address a personalized, most suitable offer (Pariser, 2011). Social media are the platforms where people voluntarily share their unique personal information, often unaware of how this information is used by the businesses.

The pioneer of personalized news was the project Google News (Hindman, 2012), which was officially launched in 2006, after 4 years in beta-version (Bharat, 2006). Today Google, being challenged by Facebook as the main source of news for the young audience (Associated Press, NORC Center for Public Affairs Research & American Press Institute, 2015), is launching its "feed", aiming to make it the page people will start their day with (Newton, 2017). The study by Thurman & Schifferes (2012) showed that the news organizations that used personalization on their websites and apps are mostly financial press and the news providers which were preparing to introduce a paywall. The scholars explain the need in personalization for news organizations as a natural way of adjusting to the logic of internet and cost-per-click (CPC) revenue model in advertising that seeks to maximize the relevance of content for individual users.

When talking about news personalization, it is important to distinguish between the concepts of *personalization* and *customization* of news. Often used as synonyms, they differ in terms of agency (Sundar & Marathe, 2010). By personalization we usually mean the process of a product being tailored to each customer by the product designer based on the users' personal

data. Most often personalization exists in the form of recommendations – "you might also like…" (Calandrino et al., 2011) – which suggest new products based on your previous choices. At the same time, customization presupposes an active participation of every consumer: the user herself adjusts the settings of the news product by choosing, for example, the categories of news. This is what Thurman & Schifferes (2012) mean by personalization, when they define it as "a form of user-to-system interactivity that uses a set of technological features to adapt the content, delivery, and arrangement of a communication to individual users' explicitly registered and/or implicitly determined preferences" (p. 776). In other words, it is the opportunity for the users to create their customized news agenda. Thurman & Schifferes (2012) use the terms *explicit* and *implicit* personalization (also, *active* and *passive*) for customization and personalization respectively. The researchers state that news consumption is largely a passive process, and that most people have no habit of news customization, unless it is a part of their work. Therefore, passive forms of personalization are growing faster (Thurman & Schifferes, 2012). Very often personalization and customization approaches are combined in one product.

Automation of news writing introduces a new type of personalization of content, when tailoring news to the customer's interests goes further from personalized news aggregation to *content personalization* – text modification within one article based on the reader's preferences and settings. Adar et al. (2017) describe a system for creating personalized content – PersaLog – which customizes text and media features for a specific article. In this case, a journalist writes an article where the embedded code personalizes certain parts of it for every reader. Marconi (2016) imagines a newsroom where artificial intelligence would write "infinite versions of an article", which, he argues, would result in increasingly relevant information being directly delivered to every reader, leading to "a more engaged audience". The personalization of news, as Otterdal and Ruud put it (2017), signifies the shift from the invention of tabloid "you-journalism" towards "me-journalism".

However, the personalization trend has its opponents. First of all, the collection of personal data creates privacy concerns: users' digital behavior on the internet is being tracked, and this information is stored by media companies. These data allows the media corporations to make personalized products for their users, but at the cost of privacy. Secondly, personalization arguably pleases the news consumers with information relevant for them. And as long as it is local news and weather reports, there is not much danger in personalizing such news. However, already now Facebook and Google algorithms are personalizing advertising and global news stories. Facebook's "trending topics" feature – suggesting the news content to

the readers based on their interests and preferences – received a lot of criticism for luring and trapping the readers in their "filter bubbles" (Greenslade, 2016). Similarly, when Google algorithms began to create "most read" lists for news stories (Otterdal & Ruud, 2017), it led to people reading more stories that are popular, while ignoring other important journalistic materials. Moreover, it sometimes even leads to "fake news" topping the "trending topics", like the one during the US elections (Allcott & Gentzkow, 2017). Lavie et al. (2010) talk about establishing a balance in personalization – where the news interesting and relevant, but does not limit readers too much with extreme personalization. The authors advocate for systems where personalized news is supplied by general topics. Negroponte (1995, p.153) criticizes his own idea of personalized news and warns the developers of personalized content: "Done well, this is likely to be a magnificent news medium. Done badly, it will be hell".

Both automation and news personalization are considered rather controversial by the journalistic community. Along with their quality and economic benefits, these innovations present a considerable danger to established practices and journalistic ethics.

2.6 Journalism and new technologies

Journalism is about curiosity, about digging into the previously unknown. The profession has always welcomed new technologies and, since the middle of the previous century, has been heavily mediatized with the appearance of radio- and TV-broadcasting, and later Web 2.0 and social media. In the digital environment, the professional competencies of journalists went beyond the traditional gathering and processing of information, adding to it such skills as photography and video production, as well as photo and video editing. This happened mainly because of the media convergence (Kammer, 2013), but also due to economic issues, where newsrooms could not afford to send both writers and photographers to events. Surely, this became possible only because of technological progress, which allowed making both photos and videos with light, portable cameras. Recently, the ranks of journalistic skills joined copywriting (since many journalists engaged themselves in content marketing) and social media marketing (promoting materials through various social media). This multitasking of journalists arguably increases the risk of errors and harms the quality of materials (Kammer, 2013).

The increased working speed in modern newsrooms led to less accuracy. If previously the integral part of newsroom presented correctors and proofreaders, today their number is significantly reduced, which leads to the increased frequency of serious mistakes in materials, especially when it comes to numbers. Moreover, with so many different tasks journalists are not able to concentrate on their crucial traditional practices (double-checking of sources, finding contextual information, etc.) (Kammer, 2013). Thus, the de-skilling of journalists manifests itself in the tendency towards the standardization of news, when journalists only slightly edit the news agency stories and publish them quickly on the website. While the technology made newsgathering much easier, journalists nowadays rarely work "in the field" (Kammer, 2013).

Likewise, computational assistance, though being quite difficult to grasp and use, eventually might become a breakthrough in the way people do journalism. Today the journalistic field is not a stranger to various technological innovations, from data journalism to VR-technology. Even automation in journalism has a longer history than many would expect. As early as in 1952 the large UNIVAC computer calculated the results during the US elections and predicted the victory of Dwight Eisenhower (Henn, 2012). Later, computer-assisted reporting developed into data and computational journalism. Coddington (2015) discusses the "quantitative turn" in journalism and identifies its following stages:

Computer-assisted reporting (CAR)

CAR embodies two journalistic practices: data gathering with conducting statistical analysis and computer-based information-gathering skills (online and archival research, email interviews, etc.). In the 1990s, CAR emerged from the interaction between programmers and journalists, with more programmers moving to newsrooms and professional journalists increasingly being drawn to programming's technical capabilities and norms. This tendency was heavily influenced by the *open-source movement*, which arguably first brought together "hacks" (journalists) and "hackers" (technologists) (Coddington, 2015). An important role in the movement towards computer-assisted journalism has been played by the National Institute for Computer-Assisted Reporting (NICAR), which was the main organization for CAR in the 1990s and remains the central organization for those practicing data journalism today. Coddington calls CAR "the foundation for the modern approaches of data journalism and computational journalism" (2015, p. 334).

Data-driven journalism

In the late 2000s a "hybrid" appeared in newsrooms that combined statistical analysis, computer science, visualization, web design, and reporting (Coddington, 2015). This concept closely co-exists with the emergence of data-visualization.

Computational journalism

This journalistic practice emerged from both CAR and data-driven journalism. Coddington (2015) argues that computational journalism lies farther from journalism than data-driven journalism and has tighter connections to programming. It was defined by Hamilton and Turner (2009, p. 2) as "the combination of algorithms, data, and knowledge from the social sciences to supplement the accountability function of journalism". Coddington (2015) explains the term less broadly as "a strand of technologically oriented journalism centered on the application of computing and computational thinking to the practices of information gathering, sense-making, and information presentation" (p. 335). The very profession of journalism underwent critical changes in the late 2000s, as developers started working in newsrooms and became computational journalists (Karlsen & Stavelin, 2014). Another way, chosen by many high-quality newspapers, was to integrate the technical staff into the newsroom (Kammer, 2013), without mixing system developers and journalists.

Automated journalism

Automated journalism, unlike computational or "algorithmic journalism" (Anderson, 2013), aims to assist human journalists in compositional and framing practices (Carlson, 2015) rather than in research. Unlike data visualization, which also serves a presentation purpose, NLG systems present data in the form of a narrative. If previously, computers were the source of information and passive assistants, today, with the introduction of NLG technology, for the first time in history machines are learning to write journalistic texts, playing a role of an active assistant.

2.7 Summary of the previous research results

The present study focuses on the technological possibilities and economic benefits of automated journalism for newsrooms in Norway. The thesis is divided into 5 major areas: NLG

technology, ethics, innovation, personalization and journalistic practices. The research on automated journalism to date has found that technologically automated journalism is capable of producing texts similar to the texts written by journalists. It requires, nevertheless, a decent amount of work and a long testing phase. The present thesis aims to check the status quo of the NLG technology and to explore how far its potential goes. When it comes to the ethics of automated journalism, the studies show a number of serious concerns regarding the quality of data used for input and the extensiveness of the algorithm interpreting the data. Later in the thesis I will show how the ethical challenges can be an obstacle to the introduction of automated journalism in Norwegian newsrooms.

Automated journalism as an innovation has not really been explored yet due to its novelty and a small number of adopters. However, the case of NTB in Norway gives enough information to talk about the character of this innovation. Personalization using automated journalism at the moment exists only as a project and has not yet been implemented. But its potential impact on the news industry makes it important to discuss it in this thesis.

In the final part of the thesis I will talk about how the "quantitative turn" in journalism, through the introduction of automated journalism, influences journalistic practices in Norway.

3 Methods

The main aim of this study is to look at the problem of the implementation of automated journalism in newsrooms in Norway. The four key areas or "research problems" that needed to be addressed in connection with this topic were: (1) the technical capabilities of the automated journalism software, (2) the ethical concerns in relation to the quality of data, programming and maintaining the system, (3) the innovative nature of the technology for the news industry and its economic benefits and (4) the influence of automated writers on the journalistic practices and ethics. Based on these problems, the following research questions were developed: *(1) How is automated journalism currently being implemented in newsrooms in Norway?* and *(2) What is its potential as an innovation?* In order to answer these research questions, the method of in-depth qualitative interviews with news professionals was chosen. This chapter justifies the choice of this research method, and discusses its design. Finally, the chapter discusses how the collected data was analyzed, and how the conclusions about the major issues surrounding automated journalism were drawn from it.

3.1 The choice of the method

The methods of the current study consist of primary and secondary data. The choice of secondary data, presented in the theory chapter, was made according to the main research problems. Since the topics of ethics, journalistic practices and personalization are closely connected to the use of NLG technology in newsrooms and its economic potential, the studies focusing on those issues were included in the analysis. Primarily, books, articles and essays published in Norway were studied; and then the analysis was complemented with research from countries where the NLG technology has been actively implemented – Germany, UK, US, etc. Finally, the studies on newsroom innovations, diffusion of innovations and personalization were included in the secondary data.

The results and findings from the analysis of the secondary data were used in the methodological design of the primary data collection, so that it would be possible to compare, confirm or disprove the results of the previous research. For this study on automated journalism the method of qualitative in-depth interviews was chosen because the aim was to find out how

the pioneers of automated journalism in Norway implement and develop the NLG software, and to learn about the challenges they meet. Additionally, similar interviews with professionals from Sweden and Germany were conducted to compare the adoption of this innovation in different news media markets. The starting point of the research was not only to describe the existing technology and its current implementation, but also to investigate the possibilities of its further development and its more efficient use. Since the method of qualitative interviews allows the researcher to learn about the views, opinions and plans of the informants, as well as to talk about their experiences (Kvale & Brinkmann, 2009), this method was the most suitable for answering the research questions of this explorative study.

Alternatively, there was a possibility to do a case study of NTB and their football "robot", however, given the nascent stage of automated journalism in Norway, it would limit the final results and would not allow creating a bigger picture of the possibilities and challenges of automated journalism, especially in other types of newsrooms. Moreover, the interviews taken in Sweden and Germany added depth to the study and allowed to see the contrasts between the different ways of implementing this technology in Europe.

3.2 Qualitative interviews

The method for conducting qualitative interviews in this paper was designed according to the 7-stage model suggested by Kvale and Brinkmann (2009), from deciding on the topic and the research questions to reporting the results. After researching existing literature on the subject and talking to people in my network about this project, the research problems and the research questions were developed. Potential informants were identified in the beginning of 2016. The interview guide was structured according to the identified research problems.

3.2.1 Interview guide

Rubin and Rubin (2005) suggest an interactive design of the interview guide, which the researcher can later adapt to the changing circumstances and the new knowledge gained on the way. Comparing to other, more conservative designs with predefined plans, this method allows leaving the questions generally open, so that it would be possible to adjust them while pursuing new topics suggested by informants and not being trapped in initial hypotheses. Given the character of the studied phenomenon, it was clear from the start that the information collected

in the interviews could alter the focus of the study. Therefore, the collection of primary data was done in two major stages, where the second stage provided the study with a broader perspective, deeper technological approach and a timeline of the evolution of automated journalism in Norway.

When designing the interview guide for the first round of interviews, I structured it by moving from the most important and specific questions to more general. I always started with asking about the automated journalism projects that the respondent is/was involved in, because it was equally interesting for both of us and constituted the highest level of competency for the respondent. Then, trying to maintain the natural flow of the conversation, I asked follow-up questions about the topics of my interest mentioned in the descriptions of the projects. Thus, going naturally from one aspect of automated journalism to another, in the interviews we managed to cover the major research problems: NLG technology, ethics, innovation, journalistic practices and personalization. Usually, I ended interviews with more general and even philosophical questions about the future of journalism in the age of intelligent machines and digital progress in general, which helped shed light on the potential use of the technology in newsrooms, as well as on the problems connected to the increasing complexity of it, and which also felt like a natural way to end the conversation. Considering the different areas of expertise of the respondents, the interview guide was adjusted every time in order to collect more relevant and valuable information from every respondent. For example, during the interviews with developers, more technical issues of automated journalism and artificial intelligence were discussed, while with journalists the aspects of media innovations, the effect of automation on quality journalism and ethics were discussed more thoroughly.

3.2.2 Conducting interviews

The first round of qualitative interviews was completed in August 2016 in Oslo, Norway. Almost all the potential respondents the author reached out to agreed to the interview and were very helpful. The first four respondents were chosen based on their closeness to the subject, their experience with automated journalism or similar innovations in media industry, as well as their interest in the topic of media innovations. Since the NLG technology is still not widely implemented in Norwegian newsrooms and is currently in the hands of a few pioneers, the journalists and system developers who directly participated in the development of automated journalism in Norway were interviewed first. The sample for analysis was limited to journalists working directly with automated journalism, developing the system in their

newsrooms or simply trying it out as an opportunity. These interviews were supplied by two more, with journalists who were open about their interest in computational media innovations in Norway. The choice of informants in Norway, thus, came down to three journalists, two developers and one scholar/former journalist:

1. Magnus Aabech, then sports editor and now news developer at the Norwegian news agency NTB, who developed a "robot" for the coverage of football matches, based in Oslo, Norway;
2. Ingeborg Volan, at that time director of innovation and newsroom development, Adresseavisen (now – readership development editor at Dagens Næringsliv), and the president of Norwegian Online News Association, at that time based in Trondheim, Norway (Skype interview);
3. Jari Bakken, newsroom developer at Norway's biggest news outlet, Verdens Gang (VG), who published an interactive article about municipalities' debts in Norway, based in Oslo, Norway;
4. Jens Barland, associate professor, Norwegian University of Science and Technology (NTNU), who researches the topic of innovation in media business, based Gjøvik, Norway (the interview was conducted in Oslo, Norway);
5. Gunnar Aastrand Grimnes, CTO Orbit and Bakken & Bæck, the developer of NTB's football robot, based in Bonn, Germany (Skype interview);
6. Geir Terje Ruud, head of the development at NTB, former journalist and editor at VG, based in Oslo, Norway.

All the respondents were first contacted by e-mail with the description of the project and the invitation for an approximately 1-hour talk about their automated journalism projects. The collection of interviews was made during the summer and autumn 2016, right after NTB's football robot was launched in the beginning of summer 2016 (Johansen, 2016). The interviews were transcribed and analyzed, and the preliminary results were presented at the International Symposium on Media Innovations 2016 in Oslo, Norway, which allowed the author to receive a valuable feedback from scholars in the field of media innovations.

Since then the secondary data was constantly updated. New studies on automated journalism published in 2016-2017 and the results taken from the analysis of the interviews showed the need for exploring the topic more deeply. For example, the issue of personalization of news content as a big potential for automated journalism was one of the results of the first round of interviews. An explorative interview taken in Germany with a representative of an IT

company working with NLG technology for journalism showed a different perspective of the use of automated journalism in this country. The participation in NordMedia-2017 conference in Tampere, Finland, with a paper looking at automated journalism in the context of other computational innovations in newsrooms broadened my research network and led to the establishment of new contacts in Finland and Sweden. Therefore, in order to give the study a necessary perspective, the second round of interviews was completed in September 2017, one year after the previous round. The additional interviews were collected in Sweden and Germany in order to compare the development of automated journalism in Norway with other media realities. The last interview was a follow-up of the 2016 interview at NTB, but this time it concerned the latest automation project of NTB – the elections robot.

In Sweden, the choice of the respondents was limited to the creators of the sports robot for the news organization MittMedia due to the similarity of this case to the football robot created at NTB. I interviewed the developers of the robot from both sides, journalistic and technological:

1. Henning Johannesson, chief sports editor, MittMedia (Skype);
2. Søren Karlsson, CEO, United Robots, who developed the software for the robot (Skype).

In Germany, the idea was to follow the initial design and collect data both from the technological and journalistic perspectives, but due to the peculiarities of the German journalistic reality, described later in the thesis, the sample was limited to the three major players on the market of IT-companies developing NLG technologies and offering them for journalism. The companies' representatives were interviewed:

1. Frank Feulner, CVO, AX Semantics, Stuttgart, Germany;
2. Eveline Śliwowska, director of commercial operations, Retresco, Berlin, Germany (Skype);
3. Hermann Bense, CEO, Textomatic, Dortmund, Germany (Skype).

There were conducted 11 interviews in total, 6 of them in Norway. The described above methodological design (Rubin & Rubin, 2005) allowed me to build on my findings from the first round of interviews and literature research and to test the ideas so that a relatively small number of interviewees with diverse backgrounds provided the research with unique and highly valuable results. The sample of respondents fully covered the issue of the implementation of automated journalism in Norwegian newsrooms, both its technical and economic sides, as well as purely journalistic concerns about the subject.

3.2.3 Analysis of the interviews

All 11 interviews were recorded and fully transcribed with the help of the oTranscribe tool, available for free under the MIT license, which allowed speeding up and slowing down the recordings for more correct speech interpretation. The interviewees expressed themselves clearly enough, despite the fact that for all of them English was not their native language, which eliminated the need to contact the respondents afterwards for clarifications.

The analysis of the interview data was done in several stages: after the first round of interviews in 2016, after the additional interviews and after the second round of interviews in 2017. During the coding phase, the quotes were sorted out firstly by the 5 main research problems with automated journalism – technology, data and algorithmic ethics, innovation, personalization and journalistic practices. Then in each section the quotes were logically divided by arguments and put into the thesis template. The interviews were coded manually in an Excel file, which made it easier to search the quotes by topics, respondents and countries.

The thesis was structured primarily based on the analysis of the interviews taken in Norway, while adding information from the interviews taken in Sweden in Germany to the relevant sections in order to show the similarities and contrasts with other media realities. The analysis was done with the help of the results from the previous studies and the Diffusion of Innovations theory by Rogers (2003).

3.2.4 Verification of the interviews

When analyzing the interview data, the concepts and arguments that appeared more often were distinguished from the concepts and arguments unique for each case and discussed in the thesis respectively, which helped ensure the reliability of the study (Kvale & Brinkmann, 2009). Since the inquiry had two dimensions – the current state of implementation of automated journalism and its future potential, the study draws a distinct line between the facts of the present situation and theoretical assumptions that should be studied further. The differences in the development of automated journalism in Norway, Sweden and Germany prove that the process of the implementation of automated journalism in newsrooms depends on various unique factors for every country and every newsroom. However, the results regarding the technological aspect of automated journalism can be applied to any media reality.

3.3 Ethical considerations

In the beginning of the project, in 2016, I had an agreement with the Norwegian Centre for Research Data to anonymize the data collected from the interviews (see Appendix 1), because I thought of such information as rather sensitive (as each of the respondents is a representative of a media company s/he is working for). However, during the data collection process it became clear that the research will significantly benefit from identifying the respondents, as the peculiarities of the introduction of automated journalism highly depend on the character of the newsroom and the person developing the software. Moreover, every respondent could be easily identified, since there are yet not many news organizations that introduce automated journalism in their newsrooms in Norway, Sweden and Germany, and they usually talk about their projects openly in media. Therefore, the agreement with NSD was later changed via email. All the respondents gave me their oral consent to publish their name and title in the company in the beginning of each interview.

To sum up, the discussed qualitative methods were chosen in accordance with the main research questions of the study. In addition to the review of the relevant literature, I conducted qualitative in-depth interviews. The chosen methodology allowed making a comparative analysis of automated journalism in three countries and answering the research questions of the study.

4 Technological aspects of automated journalism

Starting from this chapter, I will present the results of my empirical study and discuss their significance. The study is based on the information from 11 in-depth interviews conducted by me in 2016-2017.

The current chapter is devoted to the technological possibilities and limitations of automated journalism. The first section aims to present the various models of automated text writing in journalism, discuss the advantages and disadvantages of automation in relation to human journalism, and to reflect upon the future development of this technology. In the second section of this chapter, I discuss the ethical issues related to the technological side of automated journalism, mainly focusing on the quality of data and algorithms.

4.1 Technology behind automated journalism

As discussed previously, from a technological point of view automated journalism is a kind of a narrow artificial intelligence – Natural Language Generation or NLG, which is a data-to-text technology. The scheme below (see Figure 5) shows what usually happens inside the NLG system (Graefe, 2016). It consists of the data input, generated text in the output and the central segment – semantic algorithmic tree – that needs to be developed.

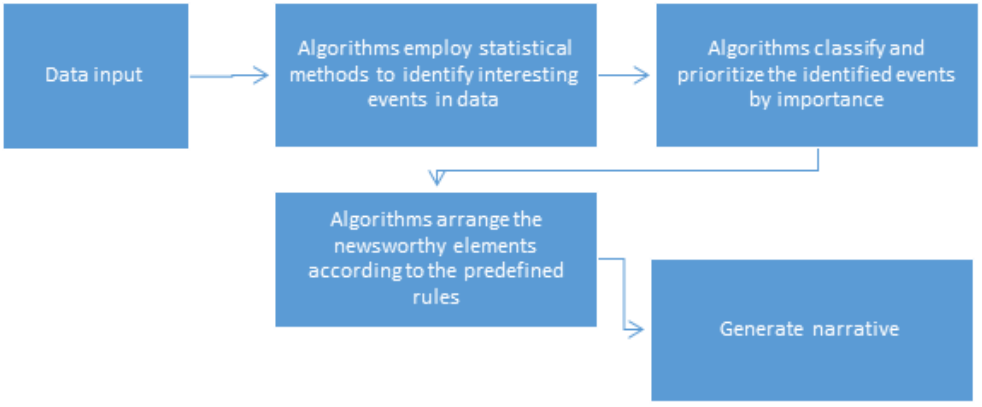

Figure 5. Detailed functionality of automated journalism. Source: Graefe (2016).

4.1.1 The models of automated journalism

There can be various levels of complexity of NLG technology in automated journalism for different journalistic tasks and business models. Grimnes (Germany, interview 22 November 2016) identifies the following 4 major levels of advancement in automated journalism:

1. One or several fully-written templates with drop-down menus.

The basic level of automated journalism is one or several articles with interchangeable parameters – numbers, places, etc. Grimnes argues that this type of automation can be performed even in Microsoft Word. This kind of automation project was developed by Jari Bakken for Norway's biggest news outlet – Verdens Gang (VG). The newspaper published the municipalities' debt report, which consisted of four templates (different for the municipalities with and without debt). Each template was otherwise the same, and the algorithm was able to generate 428 articles out of 4 templates by changing the name of the municipality (Ertzaas & Bakken, 2016). Bakken says in an interview to *Journalisten* that it took more time for him to find the pictures of all 150 mayors than to write this algorithm (Jensen, 2016a).

2. Template-based NLG.

This is the first model in which NLG technology is used to automate newswriting. At the core of the software, there is a tree of several hundred templates, where words and sentences can be interchanged. This is the type of automated journalism that the Norwegian news agency NTB used to develop their football robot. According to Grimnes, the use of this rather advanced technology "contributes to much more variety and richness of texts". A system like this requires that a journalist writes all the texts for the templates. Compared to more advanced NLG systems, this option allows taking full control of the texts and the language, which, Grimnes explains, was the NTB's main wish: "if they didn't like the way something was written, they could change it". At the same time, Grimnes thinks, this level of technology allows for very limited opportunities of the personalization of content: "if you wanted to write a simpler article about football for 8-year-olds, you would have to write new templates".

3. Semantically-advanced NLG.

On this level, the software is trained to "learn" the language (the semantic connections and syntax). The data input is still limited, but the quality of writing becomes more varied. In order to develop this level of NLG technology, the company needs to have linguists onboard to train the system – for example, the meaning of the verb, if it is a reflexive verb and the conjugation, says Grimnes. Then the software generates a text based on that structured logical forms of sentences. The algorithm chooses the language to express the data input. It can, for example, choose to combine sentences. German software company AX Semantics claims to be developing this level of NLG systems; they are also working on translation of the output to different languages, according to Feulner (Germany, interview 23 February 2017). There is a lot of work for the linguists. Such systems, according to Grimnes, are harder to control and correct, as training a robot to learn a language is not an easy task:

> The words often have several meanings, which are hard for the algorithm to distinguish between. A hypothetical system does not learn itself, and the linguistic knowledge is hard to code.

However, this level of software development opens opportunities for content personalization for various target groups, says Grimnes, which will be further discussed in the next chapter.

4. Fully-automated system that learns the language itself and generates texts using the internet as an input. Impossible to control.

The most advanced level of the NLG technology is a system that uses machine learning and can choose freely the data input from the internet and is trained to learn the language. That means that it actually learns the linguistic knowledge itself. This has not been done yet, though many developers worldwide are working on it. Currently, according to Grimnes, it is possible to build models that replicate correct grammar and spelling, but the correction of the output is still the most difficult part of the programming. This means that the articles often do not make any sense and it is next to impossible to predict what the system is going to produce. Such systems require enormous amounts of training data (probably around 10,000 articles) to reach an adequate result. Grimnes thinks it is possible to achieve this level of advancement of NLG technology, but it will never be fully autonomous since there will still be a high chance of producing nonsense. Therefore, this kind of systems may never be implemented in journalism, because, if something goes wrong, there is no way to fix it, according to Grimnes:

> It must be trained in a huge neural network, which might have billions of parameters defining how the things are connected. It is impossible to fix the parameter number 374 and put number 5 instead of number 6, for example. The system needs to be *retrained*.

Such systems can be implemented in other spheres, but not in journalism. Grimnes gives an example of an advanced NLG implemented in computer games: "You can make a character foreign, or a bit stupid, or an alien; or if you are writing spam-bots hassling people on Twitter to get them to support Donald Trump. In these cases, it doesn't matter if the language is slightly broken". In journalism, it is crucial that the language of the automatically written texts is satisfactory.

Since the control over the automatically generated text is so important for news organizations, it is a worldwide trend now that software companies which started from developing complicated AI technologies (Automated Insights, Narrative Science, Arria) are now providing simplified products – user-friendly platforms with editorial interfaces for the autonomous use by journalists, a kind of DIY kit (Wordsmith, Quill and Arria NLG Studio tool respectively). For example, German company, AX Semantics, according to Feulner, is working on a platform that journalists can adjust to the needs of their news outlet almost independently:

> First they need to tell the machine how to segment the text into logical statements and when to apply them – put the rules into the system; then to code the linguistic knowledge, like which idioms to use to construct the phrases in different contexts; finally, in the storytelling part, the machine is taught how to recombine the parts of the text.

Śliwowska (Germany, interview 21, 22 September 2017) from Retresco explains this trend like this: "You have to do the complicated stuff first and to have many common use cases in order to develop simpler products, like automation for non-developers. You engage with many customers and look for what they need most from your service". Bense (Germany, interview 26 September 2017) from Textomatic argues that such simplified platforms cannot serve journalism properly, because they are too simple. To produce a really well-structured, complicated article, you need close collaboration between the newsroom and the software company developing NLG. The different modes of such collaborations I will discuss in chapter 6.

The classification above demonstrates the growth in potential technological complexity of automated journalism. However, due to the ethical considerations, journalism benefits more from the simpler forms of NLG technology, which could be easily controlled and altered when

needed. Moreover, it is even possible to create relatively advanced systems for partial automation of news writing without NLG. In 2017, NTB automated the elections results reporting in Norway without the use of NLG, instead, according to Aabech (Norway, interview 12 October 2017) they used software apps, where they built in the templates with a number of simple rules – the biggest party, the biggest change, the mandate change – using the programming language JavaScript.

4.1.2 The benefits of automated journalism

Graefe (2016) summarizes the following major benefits that automated journalism is going to bring to the news industry:

- **speed of reporting** – the ability to publish short stories immediately (in several seconds) with a more detailed update;
- **increased scale of reporting** – automated journalism allows media organizations to do more extensive reporting of stories, as well as stories that were never reported before due to the resource shortage in newsrooms;
- **good enough quality of machine-written texts** – which is generally not inferior to the human-written (Clerwall, 2014; Graefe et al., 2016). The reason is that the NLG system uses the templates written by journalists. During the development and test phases, journalists and developers make sure that the algorithm makes no significant mistakes and that the language is satisfactory;
- **accuracy and objectivity** in reporting numbers and facts;
- **relieving journalists from routine and repetitive tasks.**

Here I want to find out if this is the case for Norway too.

Speed of reporting

Both Volan (Norway, interview 16 August 2016) and Aabech (Norway, interview 29 August 2016) confirm that the speed of reporting is a crucial issue in the discussion about automated journalism, and also one of the major arguments for developing the technology in newsrooms: “we all try to be first on a story, that's a big thing in news” (Volan). Several months after the launch of the football robot at NTB Aabech commented: “We are now able to publish

short articles immediately; we have a story in 30 seconds after the game. It takes approximately 12 seconds for the robot to publish an article, and then the journalist works on the story. There is a need to publish much quicker, it is very important for us". Feulner agrees that automated journalism offers an opportunity to newsrooms who "want to play the actuality card to the bleeding edge" and gives an example of using it in case of natural catastrophes, like earthquakes: "an editor is probably going to stand in a door frame, while the automation can already start writing about it".

Larger scale

As Aabech argues, every match is important to somebody. Therefore, NTB is now considering automating the coverage of the lower divisions in football. And if previously the newspapers had to prioritize bigger and more significant matches due to the lack of resources, with NLG technology it became possible to produce hundreds of articles simultaneously, covering all the stories that have well-structured data sets, as well as to distribute personalized content on demand. The personalization aspect of automated journalism will be further discussed in chapter 5.

Satisfying quality of texts

The experiment conducted by Clerwall (2014) showed that readers were not able to distinguish between machine and human-written journalistic texts, although the articles written by journalists gained more points in "coherence", "well-written" and "pleasant to read" categories, while machine-produced articles scored more on "objectivity" and "trustworthiness". According to Aabech, NTB received generally positive feedback from their clients and, despite a handful of mistakes during the first weeks after the launch, the automated texts were of good quality: "We have been given a lot of feedback on the content, people say it looks good, it doesn't look like it has been written by a machine, it sounds as if a journalist wrote it. That means we did a good job in preparation".

Accuracy and objectivity

Volan perceives automated journalism as "very accurate", once the data is correct. A study from South Korea showed that the audience (including journalists) tended to trust the journalistic materials written by software more than those written by journalists (Jung et al., 2017). Automated journalism reduces the risk of human error, which is quite often in journalistic work, minding the usual time pressure. Barland (Norway, interview 16 August 2016) even names "mistakes" as a marker for distinguishing the texts written by human journalists. Additionally, automation helps in data-intensive work, when the risk of making errors by humans is increasing. According to Volan, "a lot of journalists do not really enjoy math. That is why you become a journalist and not a mathematician". She also mentions the recent examples of mistakes made by Norway's largest high-quality newspapers in calculating financial results for companies, refugee numbers to the current date and the numbers in a climate report for a major corporation. She says, "algorithms do not make those kinds of mistakes". Aabech adds the "football robot is very objective, since it just reports the facts. It does not have any subjective slants, while a reporter has his opinions in the story". Therefore, automated journalism is mostly suitable in the areas where the correct reporting of pure facts is most favorable.

Automation of journalistic routine tasks and repetitive topics

The most important benefit of automated journalism for newsrooms is the opportunity to automate the repetitive routine tasks that are neither interesting, nor challenging for journalists. Grimnes argues that automation is suitable for repetitive events that are reported based on the well-structured data sets in the same format: "the more of the same you need to do, the more sense it makes to automate it. There is no point in automating a single task to do it once. Only if you need to do it once every 10 minutes". Bakken (Norway, interview 10 August 2016) agrees with this statement and adds that the job of automating can also be rather mundane: "someone needs to write the algorithm for automation, although, only once".

The areas in journalism that are mostly suitable for automation are those mainly relying on well-structured data and which do not require creativity and opinion. Those are the spheres dealing with facts which should be presented accurately and quickly. For example, weather forecasting, sports news, traffic reporting, financial analysis, earthquake warnings, and crime

(Graefe, 2016). Automated journalism has its roots in such data-saturated spheres of journalism as sports and finance. Both two major pioneers developing automated writers, Narrative Science and Automated Insights, started from the recaps of sporting events (Graefe, 2016). According to Grimnes, some sports are easier to automate than others. For example, baseball heavily depends on statistics. Firstly, the statistics in baseball are recorded meticulously. Secondly, journalists reporting about baseball pay a lot of attention to statistical details. Game sports are in general easier to automate than, for example, swimming or ice skating. Barland admits it was only logical for NTB to automate football as "almost all of the events on a football pitch can be described in standard formulations: in every game, there is a ball pass, a score, players".

Financial reports have been another typical area to welcome the news automation: the Associated Press started from automating quarterly earnings reports (Miller, 2015). Another perfect area for automation is weather reporting. With automation, it becomes possible to easily produce weather reports for every city, town or village. Grimnes thinks that Norway can benefit significantly from automating weather reports, because Norway has control over some parts of the international waters and as a state is responsible for providing weather reports for these bits of ocean. Since someone needs to write the weather forecasts for each area of approximately 10x10 kilometers, this would be an excellent task for automation, says Grimnes: "there is not much text; it does not have to be varied or creative. It is a work perfect for algorithms". The speed and accuracy of reporting are paramount in these spheres of journalism.

Bakken states that any manual labor could be replaced by a machine: "You do not need a human-level AI for tasks that do not require human creativity. Why not replace it with machines?" The developer also points out the difference of tasks performed by journalists and machines in newsrooms:

> They [machines] can see things in the data sets that humans would never notice, connections between things that would take years if you had people looking for them. I think it is good if you can free up time to have journalists do difficult tasks that machines cannot do.

Ingeborg Volan distinguishes between the tasks in newsrooms that are suitable for automation and the "real" journalism, meaning that automation is not a competitor of journalists, but a useful assistant:

> What automated journalism does is not what we are usually expecting from real-life journalists. Here we talk about mass-produced scalable journalism. It can be painful to sit and type

> weather reports or stock market reports. It is still important journalism, only labor-intensive and not challenging. We can relieve journalists from these tasks, so that they could use their creativity and critical mindsets more.

In NTB the football robot "works" together with journalists. According to Aabech, the robot publishes a story immediately after the game is finished, in around 10 seconds, so that a journalist could use this time to work on the story, finding better angles and quotes. Automated journalism has the potential of finding new journalistic stories in data, thinks Aabech: "there might be a story here and there, and the robot's report helps you find them".

To sum up, automated journalism is only suitable for particular journalistic tasks, and there is no need to try to automate everything in journalism, as some tasks benefit from automation more than others. However, there are a lot of data-saturated and routine tasks in journalism that could benefit from automation. Some of them could be partly automated, which will still bring a relief for the journalists. Karlsson (Sweden, interview 22 September 2017) agrees here with Marconi et al. (2017), saying that modern newsrooms should be "augmented" with artificial intelligence, because "there are very few jobs that can be done by machines 100%, but almost every job can in some parts be automated. […] Let the machine build a platform of automated content, let it do the ground work and produce a story, and then let the editors work on top of that platform, call the experts, make an analysis, etc. Let both parts do what they can do best".

4.1.3 Automated vs. human journalism

Historically, we tend to compare ourselves with the current most advanced mechanisms, Like René Descartes in his 1633 "Treatise on Man" compared the human body to a machine and explained that pain was a disturbance that passed down along the nerve fibers up to the brain, which in turn sent the signals to the blood and tissues, making them expand. His description, though largely true, is not quite accurate from the modern perspective. The human body appeared to be much more complex. Likewise, when we talk about the similarities between humans and the intelligent machines, we tend to anthropomorphize the latter and to present ourselves as advanced algorithms, forgetting the completely different nature of the two.

Machines demonstrated the ability to write news faster and more accurately than human journalists, but there exist various journalistic genres with different requirements, especially

when it comes to accuracy and creativity. Śliwowska has doubts about the claim made by Narrative Science about an algorithm winning a Pulitzer Prize in the near future (Levy, 2012): "Yes, algorithms already write articles. But if you want an algorithm to write you an opinion piece, I think it would be more challenging". No one would compare a theater play review or an investigative story with a short automatically composed baseball report. Automation is suitable for repetitive newswriting. However, the unique, significant historical moments like the tragic loss of a Brazilian football team to the future world champions – Germany – in 2014, as described by Otterdal and Ruud (2017), are impossible for a machine to cover. Graefe states (2016) that automation is obviously not suitable for the spheres of journalism where creativity and opinions present the biggest value. As Barland puts it, "automated journalism can be used to report the facts, not the opinions or judgements". Similarly, sensitive issues in journalism should not be covered by software. Volan gives an example: "I wouldn't ask a robot to interview someone who has just lost a child. That doesn't mean it is not a good story, but it is a story that should be told by a human". Barland argues that the less organized and standardized the case is, the harder it is to automate. Sports have certain sets of rules – the place, the number of players, the time are known and fixed – but if one attempts to automate reporting from the courtroom or on traffic accidents, the level of uncertainty rises immediately. At some point the automation can become dangerous for the reputation of the news organization. The ethics of implementation of automated journalism will be further discussed in this chapter.

A machine is not a journalist by nature. It is a piece of software, a technology allowing the automation of narratives. Grimnes denies the machine's ability to do journalism: "all it does is writing texts from patterns written by a journalist. It does not write any new texts, talk to people or do any research". Grimnes also adds that machines do not really understand what they write, for them it all is just a set of interchangeable parameters:

> The robot does not know anything about football. It does not know what it means to "score" when it writes that someone "scored". It does not know who the "person" is, what a "ball" is.

This phenomenon is called *symbol grounding problem*. When people grow up and develop, they establish complex connections between the various phenomena and the symbols for them – words in their languages. We may have played football in childhood and we remember the meaning of the words "home team", "player", "score", they trigger memories and emotions in our heads. Whereas for the computer, these are just the parameters that could be replaced with A, B, C, D – random labels, as it has no real understanding of the phenomena behind them. Even the most advanced machines trained with deep learning on big neural

networks can only establish the links between different parameters, but they will not have any emotional connections. For machines, it is equally hard to establish metaphorical connections and correctly use words with several meanings. Human journalists know that "a goal" also means a certain achievement, like in "my goal in life". We understand such metaphorical connections easily, as they are a part of our culture and language.

Van Dalen's (2012) study of the automated news creator Statsheet Network, after analyzing the reaction of the journalistic community to the phenomenon of automated journalism, concluded that analytical skills, creativity and linguistically complex phrases were the major advantages of journalists against the algorithms. Bakken mentions asking difficult questions, interviewing people and contextualizing information among the tasks that machines will not be able to perform (and that journalists are exceptionally good at): "the most interesting journalism is not going to come from a robot, it comes from a human". Volan names creativity as the core element in journalism: "thinking of new stories, new angles, things that no one has done yet, the stories that no one has reported yet – to me that is the heart of journalism. The machines are good at replicating ideas someone has already had and do not have their own ideas. Another important quality of human reporters, according to Volan, is critical thinking: "You cannot trust an algorithm to be critical of sources, it can only processes the data, while approaching the source of data with a critical mindset is one of the main frameworks of journalism. An algorithm cannot look beyond the numbers, which is a huge limitation". Aabech agrees with the above and says: "You need journalists to assess difficult situations, ask difficult questions and write about it in colorful language".

One of the main questions that are being asked in connection to automated journalism is: will the machines ever be independent "journalists"? (Then we would have to discuss ethical, legal and other serious issues.) Or they will only remain the tools assisting journalists? (Then what are the best ways to use them?) One of the pioneers of automated journalism in newsrooms, Ken Schwencke is skeptical of the possibility of machines becoming autonomous in newsrooms:

> The way we use it, it's supplemental. It saves people a lot of time, and for certain types of stories, it gets the information out there in usually about as good a way as anybody else would. The way I see it is, it doesn't eliminate anybody's job as much as it makes everybody's job more interesting (Oremus, 2014).

Volan agrees and expresses hope that better automated journalism in the next few years will augment human journalism: "once we start saying that algorithms can solve all the

problems in the future, we have a huge problem". All the respondents were quite confident in saying that automation is not going to replace journalists or even lead to additional dismissals of journalists. As Śliwowska puts it: "automation is your friend, but it cannot do the work for you". Until, of course, a conscious artificial intelligence (AGI) is developed, which is yet not the case.

Human participation in automation too must not be underestimated. Bainbridge writes (1983, p. 776): "the automatic control system has been put in because it can do the job better than the operator, but yet the operator is being asked to monitor that it is working effectively". Similarly, even a simple automation of journalistic tasks will require monitoring of the activity in order to secure that it works correctly. The NTB's football robot works according to this principle: the first version of the story is published automatically by a robot to be later revised by a journalist. Aabech says: "the story updated by the journalist comes 11 min after the one written by the robot". The role of human journalists is especially evident during the robot's development. As Aabech tells it, "when we were developing the robot, we had football journalists. If we do the automation of the elections, we'll need a political correspondent. We need the journalistic knowledge to develop the robot". Aabech also calls automated journalism a "good supplement" and a "helpful tool" for journalists, because it gives journalists more time to write a good story: "we need to have robots and journalists hand in hand. That is the thing: the robot can do whatever you tell it to do, but you have to tell it everything. That is why we spent months just calculating: "if that happens, how should we write this", and then monitoring if this worked or not".

Finally, journalism is not only about fast and correct reporting, sometimes it is about a certain experience, a pleasure. Barland talks about the meaning of journalism for the readers:

> We ask, why people need journalism, and the usual answers are – freedom of speech, to vote and to have an informed debate. But when you talk to people, they say "to get some time free". For them journalism is an experience.

Thus, Barland reminds us, while automated journalism can write coherent texts, there should still be a journalist with feelings and emotions, who understands them, who knows how to stimulate the feelings of the readers with the media content. Johannesson (Sweden, interview 15 September 2017) goes even further saying that even in sports, "if we have a really good journalist writing a game, it will always be better".

4.1.4 The potential of automated journalism

Another important question in the discussion about automated journalism is its potential, how advanced it can actually get. When we talk about the areas of implementation of automated journalism, we keep in mind the state of the technology to date. However, the studies of artificial intelligence predict that technology has the potential of rapid development. Thus, it is possible that the quality of machine writing will improve to an extent that more areas of journalism could be automated. At the moment, there is no artificial intelligence similar to the human level, and it is still possible to tell whether the text is written by a machine. However, the advancements in the sphere of machine learning can bring the technology closer to this point. Feulner argues that their algorithm already has a "human touch", since it is programmed by a human journalist: "it gives the same idioms, the same phrases that you would use as a journalist".

Volan would like to see the machines looking into data and identifying interesting and irregular events in order to find new stories and better angles: "I am sure in a few years we will be able in a lot more situations than we are in today to have computers do that for us, and to be able to run huge data sets and find patterns and idiosyncrasies". She is also curious about the potential use of the physical robots in journalism, for example, for reporting from the war zones or in a heavily polluted city. On the other hand, Grimnes is not so optimistic about the future of automation in journalism: "the machines are not going to make completely new things. But if it comes to texts, I think it should really be artificial intelligence that you could talk to, which feels like it is conscious".

This study showed that there exist different ways of automating text writing. It is not necessary to have the artificial intelligent element in the software to introduce automated journalism. In fact, artificial intelligence can cause such complications that it should be intentionally limited, especially for the sensitive topics. The ethical issues of data and algorithms will be discussed further in this chapter.

4.2 The importance of good data and the ethics of algorithms

Currently, NLG algorithms are neither creative, nor self-learning, and, therefore, the full responsibility for the correctness of its functioning according to the journalistic standards lies in the structure of its code and the character of the data used for the input. This section discusses the necessary requirements for the data used as an input in automated journalism, the credibility of data sources, as well as the potential problems with data and algorithmic mistakes. Finally, the section discusses the ethics of automated journalism.

4.2.1 Needed qualities of data

Automated journalism heavily relies on the data input. First of all, data must be reliable. As Aabech puts it: "the data has to be correct and it has to be true". The majority of mistakes in automated journalism come from mistakes in the data sets. Therefore, it is important to be sure of the correctness of the data used as an input. As mentioned in the previous section, the respondents agree on the fact that the data for automated journalism must be well-structured, always in the same format and standardized. According to Aabech, "if you do not have structured data, you will not be able to have a good automated product. The data should be of excellent quality". Grimnes thinks that the lack of standardization of financial data in Norway leads to the inability to automate quarterly earnings reports, as it was done by Associated Press:

> In the US, the companies fill in a standard form, which makes it easy to compare their numbers. There is no relative interpretation. While Norwegian earnings reports list the companies with revenues sometimes before and sometimes after tax, sometimes including the revenues of the daughter company, and sometimes not.

Karlsson met a similar format challenge for the automation of earnings reports in Sweden. The leader of United Robots also adds that it is crucial to secure the constant and stable stream of data if one wants to develop automation in the newsroom: "we are providing newsrooms with texts to be published for a monthly fee; therefore, we are really dependent on the stable stream of correct data".

Finally, there should be enough data to cover the whole story. NLG systems should check for missing values in data and do not perform calculations on data if some of the values are missing. They should also be able to explain the absence of these values (Smiley et al., 2017). For example, while programming a football robot, the data needs to cover every aspect of the programmed parts in order to avoid the holes in reporting. If the code requires a certain number, which is missing from the data, the reporting will have blind spots. Therefore, to build a successful journalistic robot, journalists need to imagine all the possible outcomes of the process in all the situations. Aabech shares his concerns about automating the reporting of lower divisions in football:

> Our next project will include more divisions, but then we have the challenge with the data, which might not be as correct or there won't be enough of it. Now we have live reporters, and if we have to rely on other sources, we need to make sure that they are reliable. We might have to simplify the stories a bit.

Similarly, in ranking, in order to state that a parameter has the highest or the lowest values, there should be enough parameters present in data, because stating that something is "the biggest", when there are only two values, is not representative or important for statistics (Smiley et al., 2017).

4.2.2 Sources of data

Data should be accurate, which highly depends on the data sources and their credibility. Among the credible sources of data are: publicly announced numbers from companies, a stock exchange or government (Kent, 2015). The sources must be defined before the development of automated journalism, as the reporters who work with automated journalism take the responsibility for the outcome and, therefore, should know the origin of the data used. Smiley et al. (2017) argue that the origin of data used for NLG must be transparent to the readers as well.

The decision about the software development often comes from the availability of relevant data sets. Sports and finance reporting usually have well-structured and standardized data sets, which makes them the first choice for automation in newsroom. In Sweden, MittMedia with the help of United Robots automated sports reporting, because they had a unique partnership with the largest provider of sports results Everysport, according to Karlsson. According to Aabech, when NTB developed their football robot, they had access to the football

database dating back to 1949, but for reporting from lower divisions, it is more challenging to acquire good data: “for lower divisions now we have live reports from the matches, they also use the data from the database”. Aabech also reflects on the future possibilities of acquiring more correct data, using, instead of live reporting, “sensors on the shoes or on the ball, knowing which of the players hit the ball”. Currently the data is still collected by people, which means that mistakes in data are inevitable. Johannesson explains that the sports results, especially in lower divisions, are still collected via phone calls to the clubs and their coaches, which might not be 100% objective and accurate. Karlsson says the callers often ask, “who scored the goal” and get an answer “I think it was…”, which may later turn out not to be the case. Since the software cannot detect mistakes in data without a human, it is crucial for news organizations to be sure of the accuracy and credibility of data used for automation.

Finally, it is important to be sure that data is legally acquired, and the news organization should have the rights to modify it and publish (for example, the rights on publishing photos). There is also an ethical side to choosing the data sources. Aabech speculates about the possibility of automating police tweets for criminal reporting: “You can always do a robot story from the tweets, but is it the right thing to do? With the Norwegian police it might be ok, but other police departments around the world might not be as decent, they might represent some regimes that would want to use this for their purposes”. Karlsson expresses a concern regarding the availability of data from the public authorities for automation purposes: “the state wants to get paid for providing us the data. I think it is principally wrong: there should be more open policy in providing the data which we as taxpayers already pay for, to let companies and institutions use this data to develop services like ours”. Otterdal and Ruud (2017, p. 59) also write about the “increasing need to make available the public databases in formats which can be used by intelligent software”.

4.2.3 Potential problems and mistakes

There are two sources of problems and mistakes in automated journalism: incorrect data and the limited programming of algorithms. Even if the data is absolutely correct, reliable, structured and standardized, there is still a problem of explicit programming. Machines have a different way of reasoning. Riedl admits (2016), “the AI makes decisions that sometimes can be hard for us to make sense of. Their failures are often due to the fact that they cannot make sense of what we are trying to accomplish or why”. Bakken agrees: “machines are just putting

the data together, and if something is wrong in the algorithm, it could look really stupid. Automated journalism can be pretty accurate, but it can write something that a human would spot immediately". Therefore, the rules and instructions for the code of automated journalism should be meticulously described.

When it comes to temporal scope, while reporting usually starts in the present and then goes sometime in the past, the computer should be given the exact definitions of "current" and the "past" – and precisely how many years back. For example, if the Gross Domestic Product (GDP) for the current year is not available yet, the robot should inform the reader that the data is current as of the latest year available (Smiley et al., 2017). Johannesson gives an example of the mistakes that can occur when the algorithm is not well thought through:

> At the start of the new season, it is stupid to write where the teams are on the league table, like "the team tops the table", when there was only one match. Instead, the robot should write something about the last meeting of the teams. This can also lead to mistakes. Once we wrote "the last time the teams met, the score was 7:0", which was correct, but the teams met last time 12 years ago, which was no longer relevant for our readers.

Grimnes admits that football is not the best sport for automation, because the rules are a bit too complicated for a robot to understand:

> Nothing is well-defined enough to be the facts for the football robot. You can have a match where one team dominated, but they never managed to score a goal. And the other team was consciously defending, but none of them scored, so there will be null-null.

Johannesson says MittMedia's sports robot sometimes would produce such headlines as "The "winner" won against the "winner", because there was a mistake in the algorithm: "the robot was probably programmed to write "the "winner" won against "home", instead of the "loser". For this reason, NTB met several problems with the robot during the testing process before the launch, which they managed to fix quickly:

> Sometimes we saw sentences you'd think would work didn't work. We had to remove them. We had a few mistakes, because there was something wrong with the connections: when a person with the same name was at the same time the player who scored a goal and a referee. Then someone scored a self-goal and the robot thought the team won.

Luckily, the developers saw these mistakes in the development process and could fix the code when that happened. Aabech says that after the launch there were almost no mistakes

made, only sometimes the algorithm takes a bit of a harsh angle, which the journalists can adjust later. Grimnes gives another example of a problem that can happen with the algorithm:

> We had one technical issue, where a game was postponed – sometimes games don't get played when they are meant to be played, because one of the teams had to play another game – and the robot was confused, because it expected the game to happen on a certain day and it didn't.

Thus, NLG systems still need human supervision, but they get better. According to Aabech, "the computer does not know when it makes mistakes. The good thing is that it makes them only once, because, when corrected, it never makes these mistakes again". Although, more complicated cases and areas of journalism increase the scale of possible mistakes.

Yet another concern appears in case of errors in the published materials. Who should be monitoring the automation? As long as the system is in the testing phase and the scale is not that big, journalists can easily spot the errors in the machine-produced texts. As soon as newsrooms will be providing the readers with personalized content, how are they going to solve the problem of content monitoring? The news organizations should come up with an immediate response in case the algorithm makes a mistake, as well as be ready to explain the facts taken by the algorithm from the data (Kent, 2015).

4.2.4 Ethics of automated journalism

The ethical issues should be discussed in the planning stage of automated journalism, and before the start of its programming. News organizations should take into consideration that computer scientists and NLG start-ups that provide the automation services are often not aware of the code of journalistic ethics, especially when news organizations have, in addition, their own guidelines. Therefore, it is important to instruct the programmers of automated journalism about the journalistic ethics and the norms of a particular news organization that this robot will be serving. According to Grimnes, NTB provided him with the book of ethics and linguistic peculiarities, and possible difficult cases in advance. Such guides often include stylistic issues. Kent's guide (2015) warns news organizations ready to welcome automated journalism that "spellings, general writing style and capitalization should match the rest of your content. If they don't, readers may be suspicious of copy that doesn't feel like the rest of your journalism".

It is important to mention that the ethical problems occur in automated journalism not due to the evil nature of technology, but because of the way it is used. Bakken, as pretty much all the other respondents, argued that automated journalism, as any technology, has no ideological color. It becomes what people who use it make of it: "automated journalism is neither a blessing, nor a curse; it is just a tool, like a hammer. A hammer is neutral; it is people who define for what kinds of purposes it will be used. The same with automated journalism". Barland admits there are many possibilities for manipulation in automated journalism, just as they are present in journalism written by humans, because people create and use the software. Therefore, argue Smiley et al. (2017), there should be established measures against code fraud in automated journalism. When programming an NLG system there should always be a code audit to ensure that the written code was not manipulated in the interest of one or another side.

There are indeed certain characteristics of automation that make ethical errors more likely to occur; therefore, journalists, as well as programmers, need to be aware of the "weak spots" in this technology. Still, these failures can be fixed and not repeated, as these errors are not intentional. The computer is not conscious and it creates narratives only based on the data input and the rules it received. The problem with algorithms is that they cannot find errors in data and in the narrative, unless they receive the relevant rules and unless the data is absolutely correct. This is the sphere where automated journalism causes the majority of concerns. When journalistic reputation is at stake, that kind of difficulty might be the obstacle to innovation. That is why the testing phase is so important. Barland gives an example from his practice, when a newspaper reported an incident with a presumably drunk driver, whose state was later confirmed. The journalist didn't write explicitly that the driver was drunk but described the way the driver was handling the car so that the readers understood the driver's condition. The newspaper was punished for reporting the facts and indirectly accusing the driver. If that could happen with a journalist, says Barland, that kind of assumptions and judgements made by a robot can cause serious problems for the editors.

NLG creates narrative based on the embedded rules. In simpler cases of sports and finances with well-structured and standardized data, the rules are relatively easy to imagine. However, when data gets messy, when the topics are non-standardized, ethical issues arise. Ruud (Norway, interview 12 October 2017) argues that the development of automation should adhere to the ethics of journalism. Every new step in automation complexity must involve building a system which will take control over the ethical complications. Thus, by developing the automation in newsroom gradually, it will be possible to automate more tasks in the future:

“we are not jumping from here to there; we will have 50 steps before it gets really complicated. And I think we will have lots of time to think about it”. Aabech states that there will always be a discussion in newsrooms about ethics and possible mistakes, and whether anything should be fully automated.

An interesting ethical issue with the use of robot journalism regards the presentation of the materials written by robots and the author/robot identity problem. Some news outlets, like the Associated Press or NTB, are proudly showing their technological innovation via a disclaimer “generated by Automated Insights robot”, which also solves the problem of authorship of machine-written texts. Others, like in Germany, prefer to hide the authorship of the automated articles. According to Feulner, “there are quite a lot of customers who do not want to admit using automated software or who want to keep it in an experimental form”. Śliwowska argues that “a lot of publishers are reluctant to disclose that they work with a third party, with technology providers”. This happens due to various reasons. Some news outlets are not sure of how the robot materials will be met by their readers and do not want to risk their reputation. Johannesson from MittMedia noticed that the readers in general have a positive attitude towards automated reporting. However, the moment there is a mistake in reporting or another problem, the company gets criticized for using the robot: “when they want to criticize us, how we write about football, they write “ah, they have a robot writing our games”. Other media companies immediately edit the materials written by a robot, and, therefore, the edited materials get their human author and there is no need to write that an algorithm has generated the article. Some news outlets are unwilling to disclose the use of automation due to the resistance from the journalistic trade unions that are protecting the jobs that could potentially be taken by machines. Feulner argues that there are certain agreements between the trade unions and media organizations regarding the use of automation in Germany, saying “ok, you can use automation, but don’t make it a big thing” or “do not make it a part of the company’s identity”. Bense says, however, that his NLG company, Textomatic, meets no resistance from the publishers’ side, and the stock exchange reports they provide for their arguably biggest client, Handelsblatt, all have a disclaimer saying that the text was produced by the Textomatic algorithm. Kent (2015) argues that it is ethically correct to disclose to the readers that the story was automatically produced via providing the link to the source of the data — the company that provides the automation, with an explanation of how the process works.

How can we embed the journalistic values in lines of code? The value sensitive design, proposed by Friedman et al. (2006) – “a theoretically grounded approach to the design of

technology that accounts for human values in a principled and comprehensive manner throughout the design process" – could be of assistance in the process of programming a journalistic robot. Smiley et al. (2017, p. 107) offer the method of presupposition check – "describing pragmatic issues in language by adding semantic and pragmatic formal constraints" that are supposed to make the NLG systems more transparent and less potentially misleading. In that case, the automated check of all the potential areas of mistakes ensures that if an error occurs, the system detects it immediately before generating the narrative. Another suggested method is the output evaluation, i.e. monitoring and evaluating the quality of the generated text, rated by the domain experts.

4.3 Summary of chapter 4

In this chapter I discussed the technological aspect of automated journalism, and closely related to it, ethical concerns. Firstly, I showed that the automation of news writing does not necessarily require the development of complex AI systems, as even simple computational techniques can help in the majority of journalistic tasks. Then I discussed the benefits of automated journalism and argued that there is no doubt that in certain data-saturated journalistic fields automated journalism can play a crucial role and bring the benefits of speed and increased scale, while maintaining the satisfactory level of quality. Still, the possibility of machines replacing human journalists is quite far-off. Machines are getting better at accomplishing specific goals, but they will always lack feelings, compassion, and creative and critical thinking, which are essential in journalism. Therefore, in newsrooms, algorithms can only aspire to the role of useful tools assisting journalists.

The implementation of automated journalism in newsrooms sees a number of ethical obstacles: on the levels of data choice, explicit programming of the algorithms and building the journalistic norms into the systems. Another aspect is the choice of the journalistic tasks for automation, which will be discussed more thoroughly in chapter 5.

5 Automated journalism as an innovation

The media industry today finds itself in a severe economic crisis. The changes in content proliferation, the habits of media consumers and the obsolete business model of selling audience to advertisers trigger innovation in media organizations. Innovation in the news media industry is often caused by a business strategy phenomenon – "first-mover advantage" (Agarwal & Gort, 2001; Berman et al., 2007): an innovative media organization immediately benefits from being a pioneer, while others do not want to be left behind. Thus, it is common in the media industry to test new technologies as they come. Automated journalism is not an exception, but in this case newsroom automation is an extension of the global automation trend. As it was expressed by Reginald Chua, executive editor for editorial operations, data, and innovation at Thomson Reuters: "you can't compete, if you don't automate" (Graefe, 2016).

In this chapter, I look at automated journalism from the media innovations perspective and analyze how the technology is being developed in Norway, what is its potential and value for the media industry, and what kind of benefits in terms of economy it brings. The analysis is based on the empirical material from 11 in-depth interviews conducted in 2016-2017. Firstly, the chapter discusses whether automated journalism can be called an "innovation" and which type of innovation it represents. Then I describe the diffusion process of automated journalism in Norway. Finally, in the second section of the chapter, special attention is paid to the opportunities of automated journalism applied to news personalization, which can arguably become the new revenue source for the news industry.

5.1 Technological innovation

Automated journalism is a technological innovation implemented in news media organizations mainly for the benefits of speed and scale of reporting. This section discusses the characteristics of automated journalism as a media innovation, whether this is an incremental or a radical innovation, and then, with the help of the Diffusion of Innovations theory (Rogers, 2003) analyzes the proliferation of automated journalism in newsrooms in Norway.

To discuss the characteristics of automated journalism as a media innovation, it is important to make sure the phenomenon meets the necessary requirements to be called an

innovation. Taking the traditional definition by Schumpeter (1934), innovations should contain a new idea and have an economic value. When we talk about the novelty of automated journalism, opinions diverge. Barland argues that similar technologies have been here for a long time. Automated journalism is a phenomenon of implementing the technology of Natural Language Generation in journalism, which was not used in newsrooms before 2013 (Marconi, 2017). The NLG technology was first developed in the 1950s as a part of machine translation efforts (McDonald 2010; Reiter 2017), but until recently it was quite primitive and was never used in journalism due to the bad quality of texts. However, there were several attempts of automation in newsrooms before the introduction of NLG technology. Thomson Reuters has been using computer algorithms to generate financial news stories since 2006 (Tran, 2006). Earlier the same year Google officially launched the aggregator of news headlines Google News, which existed in beta version since 2002 (Bharat, 2006). In 2013 the first text generation software was developed by Automated Insights, and in January 2015 the Associated Press published their first automated quarterly earnings report (Miller, 2015).

According to Schumpeter (1934), the innovator can only be the company which was first to implement the technology. In the case of automated journalism, it is the Associated Press. This definition makes all the other companies developing automated journalism worldwide the "imitators" (Nelson & Winter, 1982). Rogers (2003) suggests the term *re-invention* for describing the innovations that undergo significant changes when implemented in the environments other than original. The newsrooms generally tend to modify the technology to customize it and to adapt it to their needs. Grimnes sees the football robot developed at NTB as an innovation for Norway, even though the NLG technology was used for automation of other types of sports before: "I think the concept is new. I cannot claim this was a global innovation. People have done it for other sports years ago. But when we did it, it was something new".

The second part of Schumpeter's (1934) definition of "innovation" is its commercial value: an innovation has to bring profit. Since there is yet no clear evidence that automated journalism brings profit to media organizations, it may not be called an innovation yet. Despite the arguments for its contribution to the economy of the newsroom resources, it does not really sell. Automated journalism allows saving the production costs, while reallocating the resources more efficiently in newsrooms. The Associated Press argues (Marconi, 2017) that the agency managed to save 20% of journalists' time to do better journalism. In Scandinavia, it is not yet clear whether this innovation will bring profit. Bakken says that at the moment he does not see

"the economic angle" in automated journalism: "Yes, you can reduce costs by using more technology, like in any business". But the majority of respondents agree: it can become profitable in the future. The economic potential of automated journalism will be discussed thoroughly later in this chapter.

Innovations do not always radically change the status quo. In fact, the majority of innovations bring small changes and improvements to existing products, processes and services. Radical innovations have strong economic effects and significant improvements, but incremental innovations often have even more important outcomes. Automation in newsrooms, as Bakken and Barland argue, has deeper roots and started with automation of simple tasks. The concept was reflected in proofreading programs, automated sports scores alerts and translation software. According to Barland, automated journalism is the next small step on the road towards technological progress in journalism:

> Nothing new is completely new. Before automated journalism, there was text formatting (though it was not the language generation), there were automated sports results that would publish the name of the player and his team, Google's automated real-time traffic reports, translation programs that are now getting better.

Bakken thinks that automated journalism got so much attention nowadays because it was called "robot" journalism, which sparked the associations with science fiction with its fear of intelligent machines, while the automation concept is not new in journalism:

> Interestingly, we have always used practically the same technology based on large data sets to create different types of output – charts and maps, but as soon as we use the same data to write text articles, the interest is extreme. It happens because it is perceived as something only people can do.

Thus, automated journalism can be considered an incremental innovation – it has grown from the technological development of computational journalism and the advancements in the sphere of artificial intelligence. Barland points out that incremental innovations are very common for the media industry, as newsrooms rarely introduce radical changes.

According to Bakken, the automated project they did at VG followed similar projects, like automated reporting of train delays and the experiments with Natural Language Processing (NLP) or data mining, where the algorithm compares various data sets to find links and hidden patterns in the data:

> Let's say we have a data set of people working in the government, another data set of taxpayers, and the third data set with all

> Norwegian shareholders. We can connect these together and see the crossovers between them: I find this person here, can I find him here and here, what is interesting about it?

At Orbit, Grimnes does other projects using machine learning that do not involve automated text writing. One of them is the project on distributional semantics, where the machine is learning the language, being trained on a lot of texts. Grimnes argues that data mining can be a very powerful tool in journalism, as it can run through data sets finding unusual events extremely fast, while it would take months and years for journalists. The developer reflects upon the possibility of combining the data mining and automated text writing into a single system, although, he admits, it would need a certain degree of control before publishing. Śliwowska says the combination of both NLP (NLU+NLG) technologies (Reiter, 2017) is already being implemented in news chat bots, which, after receiving the user request, search for the relevant articles and publish a little summary about it.

Despite the relatively long history of automation in newsrooms, automated journalism can be regarded as a disruptive innovation as well. In accordance with Christensen's definition (1997), although journalistic texts produced by the algorithm are slightly worse in quality (boring, standardized, uncreative) than those written by journalists, this innovation brings a number of significant benefits that force media management to rethink the organization of the newsroom and the character of content. Carlson (2015) argues that automated journalism is the most disruptive among all the data-centric journalistic practices, as it challenges the status quo of journalistic labor, compositional form, and the question of authority. Karlsson agrees and says that in his opinion, "all kinds of publishing in the future will require some automation". Ruud thinks that, in the future, there might be a need for a clearer distinction between author journalism and journalism produced by automation, as well as between "branded" journalism and "fake news".

5.2 Diffusion of automated journalism

The current thesis explores the diffusion of automated journalism in Norway. Rogers (2004) argues that the character of diffusion does not depend on a particular innovation, but on its adopters, place of implementation and the adopters' culture. Ekdale et al. (2015) point out that technological change in newsrooms itself does not present a difficulty as it is perceived mostly as an opportunity that has benefits for news production. Meanwhile, the audience

relationships and changes of the professional culture of journalism are more challenging. Ekdale et al. (2015) also notice that changes are welcome in the news production industry when they bring the improvement of the quality of journalism and correspond with the existing norms and values. At the same time, the changes that may lead to disruption of journalistic autonomy, damage the news product, or those poorly presented by the leadership, face strong resistance (Ekdale et al., 2015). The present study tests these hypotheses in Norwegian newsrooms using the case of automated journalism.

5.2.1 The Innovation-Decision Process: to automate or not to automate?

Dogruel (2014) argues that media innovations are "high-risk" products and processes, because they usually require significant investments in development and production, but can cost the newsroom a lot in case of failure. The risk can be reduced by studying innovation processes in newsrooms, including the stages of decision-making, implementation and confirmation (Rogers, 2003). The innovation-decision process is a series of choices on the road from the initial knowledge about the innovation towards its implementation, adoption or rejection. At this stage, the actors of the innovation evaluate the suitability of the new idea for its incorporation into the existing practice. The majority of respondents in this study are finding themselves at the stages of knowledge and persuasion, studying the phenomenon and discussing whether it could suit their newsrooms. NTB is the only player on the Norwegian market that reached the stages of decision, implementation and, arguably, confirmation.

The big question is whether automation of text writing should take place at all in newsrooms. Volan thinks automated journalism will soon become a standard innovation in newsrooms, and the main challenge for the journalists will be to find the best ways of augmenting their work with automation. Bakken, however, warns that the decision to automate should be well thought through, because text might not always be the best way to present the data: "You should always think what you want to deliver to the reader and in what form". Aabech agrees that in some cases graphics or simple news flashes suit the purpose better. He also admits that there is no need to automate everything as some areas of journalism are more susceptible to mistakes and can lead to ethical controversies. However, a lot of areas can benefit significantly from automation. Therefore, it is the editorial task to assess the suitability of automated journalism for the various news projects.

When the decision to automate a certain task has been taken, there comes a question of the format: how the automation should be done? After the launch of the football robot, NTB had been pondering the thought of automating the results of the parliamentary elections in Norway in 2017, but they were not sure about the cost-effectiveness of such project. Aabech thought it would simply be too expensive to develop for one night only. However, the project eventually saw the light, although in a simpler form. The NTB's football report has a relatively advanced NLG system at its core, but such complexity was not necessary for one night of the elections. Thus, the company did not develop a new NLG system; instead, journalists programmed the elections robot themselves, which made it slightly less expensive, because the newsroom used its internal resources. Instead of sending the reports automatically, they first went through an editor. If previously there were several journalists working on election night, with the robot, one editing journalist, who also took part in programming the robot, was enough: the quality and the number of texts were the same as the previous election cycle. In this example, NTB showed a creative approach to the problem and found the optimal solution that both satisfied their needs and conserved resources.

Another potential project of NTB in the sphere of automation is making automated reports from Statistics Norway (SSB): first national, then on the levels of counties and municipalities. Aabech explains why it should not be a hard task to do it: "the API is out there, so we can maybe even do it ourselves, without external developers. We could automatically report the unemployment figures, real estate prices, and refugee numbers, possibly as a graph with a smaller text". The agency had plans to automate the results of the Olympics in Rio in August 2016, but they were a bit late. Aabech talks about assessing the potential need in a new service while thinking about new projects involving automated journalism: "if we can deliver a better service by doing it, is it worth the investment and time".

The development and implementation of automated journalism requires different levels of expertise, available budget and human resources. In the US and UK news agencies were the first clients of the software companies that offered automated text generation – the Associated Press and Thomson Reuters. Later, big newspapers followed the lead. The experience of Norway shows that indeed news agencies have the necessary resources to develop automated journalism and that, no matter how beneficial the technology is for the local press, smaller newsrooms cannot afford to develop a text-writing robot in-house. Volan also underlines the necessity of newsrooms developing automated journalism that the demand matches the increased scale of automated reporting:

> NTB has a relevant production scale. For me, at Adresseavisen, developing a robot would not pay off, because the financial investment in the development may exceed the demand for such news among my readers. Smaller newsrooms usually do not have that kind of resources.

Volan is skeptical about the possibility of a Norwegian newspaper developing its own NLG software, but she sees the potential in automated journalism as a business-to-business (B-to-B) service and thinks it is a "brilliant business idea" for the news agencies: "it would be too labor-intensive and costly for tiny newspapers, which will not develop automated journalism until someone will provide it to them as a service. I think there is a whole new market here for offering B-to-B services to news organizations". Barland and Bakken agree with her. Barland adds that the major adopters of automated journalism will be sports and financial newsrooms as they have always been at the forefront of the adaptation of new technologies, especially when it comes to automation. NTB's example clearly shows why sport will be one of the first choices for automation. Barland explains that also the customers of financial news have the necessary resources and the interest in the technological improvement of financial and business news reporting: "they have very high interest in getting the information first, best and accurately". The financial news market is also a market with a lot of resources that could be used to develop such software for automation.

Śliwowska argues that it depends a lot on the value proposition of the news outlet, as the Financial Times did not even ponder introducing automation in their newsroom, because their product is quality opinion journalism; while Reuters and Bloomberg were among the pioneers of automation, because their main product is high speed journalism. Therefore, in order to answer the question of whether your news outlet should introduce automation, it is enough to ask: why do people pay for my journalism?

Bakken says that in VG, the decision about developing new computational tools usually comes from the cooperation between journalists and developers: "we are a small team of programmers. We talk to journalists, asking them if they need some tools, and then we build them. We usually do very small projects, quick development". However, the suggestion about developing an automated system for writing the municipality debt report came from the management, according to Bakken: "I think he had been to a course or a workshop, he had seen some examples, got an idea and asked if we could do something with it. And we did it".

Aabech explains that the initiative to program the football robot at NTB also came from the management, who saw the possibilities of automated journalism in a presentation by the

Associated Press during a meeting with other world news agencies. But the idea became an innovation only because the agency felt the need for a solution to a problem they encountered: "we got a feedback last year that we were too slow for some of the matches. Other customers said they did not need so much information on every article; the basic info would be enough. That is when I started looking at it, trying to figure out what we could do". At MittMedia, Johannesson met a similar problem, which gave him an idea to introduce automation:

> We recognized that we had a lot of lower divisions in football that were covered by freelance journalists who were too late with their reports, sometimes a couple of hours after the match was finished; we had very few readers of this materials because of that. We also had to look through their language before publishing. We needed a cheaper, quicker and more automated way of writing and publishing texts. Plus, we had the necessary data from Everysport.

Rogers (2003) argues that individuals usually do not seek innovations, as they have a tendency for *selective exposure*, which is reflected in the habit of individuals to be willing to expose themselves only to ideas that go in accordance with their existing predispositions. The individuals are not seeking innovations unless they feel the need for them. Therefore, for the message about an innovation to be received, it should be relevant for the individual's needs (this phenomenon is called *selective perception*). Thus, the need for an innovation usually precedes the process of acquiring knowledge about it. However, an innovation can also lead to the realization of needs and even to their creation. Therefore, by the time the decision-makers at NTB got the knowledge about automated journalism, they were already open to this innovation. The challenge, according to Aabech, was to produce the articles faster without the necessity to hire more journalists, as the resources of the newsroom were limited: "it is difficult to add more people to the newsroom in this financial situation, so we needed to find something clever instead".

5.2.2 Persuasion phase

At this stage of the diffusion process the decision-makers assess the innovation's advantages and disadvantages in order to decide whether it should be implemented further (Rogers, 2003). The likelihood of automated journalism being adopted in newsrooms in Norway can be measured by the parameter offered by Rogers – the *rate of adoption*, which

consists of the following attributes: relative advantage, compatibility, complexity, trialability and observability.

Cost-effectiveness as a relative advantage of automated journalism

Innovations are introduced in order to improve the current situation. Therefore, in order to have a relative advantage, an innovation should offer a better idea than its predecessor. Rogers (2003) mentions that the economic profitability often determines the degree of relative advantage. The reason to develop automated journalism in newsrooms (as automation in any business) is mainly its cost-saving potential. The resources in news organizations are usually limited; therefore, the opportunities to save some of the costs and time for doing other tasks are highly welcomed. Automated journalism promises to reduce the content production costs, but it requires significant investments in the beginning, during the development phase, which can be an obstacle to its adoption, for example, in smaller newsrooms and local press, says Volan. Mainly the costs include the external technical knowledge or the license. Aabech argues that the costs of the development of NTB's football robot were reasonable:

> The project itself was not that bad. We paid the developers, then me and another journalist spent a couple of hundred hours – we didn't work full-time on the project. I have a full-time job as a sports editor, but from January until May I worked one day a week or 4 hours a week on the software.

With better promotion of the technology and the development of better NLG products, it will be possible to turn this disadvantage into a benefit. As Rogers (2003) argues, "When the price of a new product decreases so dramatically during its diffusion process, a rapid rate of adoption is encouraged". With more competitors on the market of NLG software, the prices for development may drop significantly, or cheaper versions of the products will be available (like the journalistic platforms discussed in chapter 4).

Aabech and Ruud point out that when investing in the development of automation projects, the newsroom at the same time invests in the training of journalists and developing their technical skills, which arguably leads to a more efficient and varied journalism. Grimnes agrees that automated journalism not only makes it possible for the media industry to save money, but also to do journalism more efficiently: it produces news faster, at a larger scale and with more accuracy (if the data is correct). Aabech argues that the football robot at the moment does not save the money for the company, but allows the organization to shift their resources:

"instead of having three people working on reports – we had one person doing live report, another writing the match report and one editing – now we only have one doing the live report. We hope we will be able to save money in the long run". Johannesson raises a question about the need to pay the journalists for the work that can be done by machines: "we see that in the digital world, people are not paying for the stuff that can be automated". Thus, the automation of repetitive and relatively simple news reports presents a better alternative to time-consuming manual work that journalists currently have to perform in the majority of newsrooms. The issue of changing journalistic practices is further discussed in chapter 6.

Sometimes the advantages of the innovation seem so obvious that the adopters forget other considerations. Many of the respondents in the present study warn against *overadoption* – automation of every journalistic task. In this case the ethical issues and quality considerations, as well as the cost-effectiveness of automation of non-repetitive tasks should be considered.

Consistency with existing norms and values of the journalistic community (compatibility)

By compatibility of an innovation Rogers means its consistency with the existing values, past experiences and needs of potential adopters (2003). Previously, managerial studies of traditional newsrooms have shown that news organizations are often highly resistant to implementing change. Boyles (2016) summarizes three reasons for that:

- the publisher and editor-in-chief controversy: the economic value of the news product often comes in conflict with the civic value of journalism;
- traditionally high degree of autonomy for individual journalists due to the favored informal managerial structures makes it difficult to implement changes across a news organization;
- barely any time for strategic planning due to the nature of the 24-hour news cycle.

The history shows that journalistic managers have always been "too consumed with the day-to-day operations of newswork to think strategically" (Boyles, 2016, p. 234). Ekdale et al. (2015) state that the innovation can succeed in newsrooms only if it:

a). improves the quality of journalism;

b). corresponds with the existent norms and values.

Here the question is not whether automated journalism as an idea is compatible with the existing values and professional practices, but rather how it should be implemented so that it will be compatible. The technology itself presents an opportunity, but the way it is used defines

whether it will succeed. When it comes to past experiences, journalistic community has never been a stranger to the technological challenge, as well as to computation. Therefore, automation of news production is usually quite welcomed by journalists. As potential adopters, they see the need in automation of repetitive and mundane tasks in newsrooms, often performed by interns, in automation of the data-saturated spheres of journalism. This attribute is of special importance, when we talk about the journalistic community, as the new technology should correspond with the existing practices of news production. This aspect will be discussed more thoroughly in Chapter 6.

Technological complexity of automated journalism

This attribute demonstrates whether the given innovation is difficult to understand and use. If the members of community perceive an innovation as complex, it has a negative impact on its rate of adoption (Rogers, 2003). The seemingly complex innovation – that is how automated journalism is usually perceived by non-developers – can easily become a barrier to the adoption in newsrooms, since journalists should be able to learn how to employ the program in a fast and effective way. Moreover, they need to train the software afterwards. One could argue that this opinion only demonstrates the lack of knowledge about the innovation. Bakken, Grimnes and Aabech agree that once a news professional starts dealing with Natural Language Generation technology, s/he realizes how easy it is to develop and use. Though the very system might not require human assistance, the correction of possible errors will require an understanding of how the robot functions. Therefore, it is true that automated journalism is quite a complex innovation and requires the knowledge of relatively advanced computer skills and the willingness to learn basic programming. Again, the availability of more user-friendly platforms for automated journalism may effectively solve this problem in the future.

Trialability

In automated journalism, it is possible to experiment with the technology in its simplest form. Any journalist can put a couple of templates in a .doc-file, adding drop-down menus for varying data, like Bakken did for VG. However, when it comes to more complex NLG systems like the NTB used, the experiment becomes quite costly. Nevertheless, Aabech strongly advises to try developing that kind of software. Moreover, there is always a possibility to divide the

innovations into simpler versions. In the case of automated journalism, it might be trying to automate smaller, simple (in terms of language structure) reports, saturated with data, like weather and traffic reports, slowly shifting to finance and sports. The trial phase is crucial for automated journalism. It is not advisable to use the finished software without a test phase of several months, according to Aabech and Grimnes.

Observability

The Diffusion of Innovations theory defines observability as "the degree to which the results of an innovation are visible to others" (p. 258). Is the idea of automated journalism easy enough to communicate to the people with limited understanding of programming and AI technology? Its popular name "robot journalism" clearly demonstrates that this task is not easy: the misunderstanding that was caused by this inaccurate name prompted the discussion about "robots taking the jobs", before realizing how the technology actually works and at what stage of development it currently is. Other technologies, such as 360-video, drones and VR are easier to observe and communicate, as it is easy to see how they work. Rogers (2003) argues that technological innovations in terms of observability can have *hardware* and *software* aspects. From his point of view, hardware – physical tools, like VR-goggles and drones – score higher in observability than software, like the NLG technology, as it is "not so apparent to observation". Therefore, software might have a generally slower rate of adoption.

To sum up, if we analyze the chances of automated journalism to be adopted by newsrooms using the parameter "rate of adoption", offered by Rogers (2003), it is not the most obvious choice for the majority of news organizations. The development of this technology requires significant investments and faces certain concerns from the journalistic community. However, the relative advantages, namely the benefits in speed, accuracy and the cost-effectiveness of automated journalism can determine the adoption of this innovation by steadily more newsrooms in one form or another in the future.

5.2.3 Rethinking the business models in journalism

At the moment in Norway only NTB reached the implementation phase of the innovation-decision process, while all the other newsrooms are still between the persuasion and decision phases. For several reasons they have decided this innovation is not suitable or cost-

efficient and did not introduce it. At the moment, the NLG technology is in active use only in the news agency NTB with their football robot. Applying the Diffusion of Innovations theory by Rogers (2003), I can state that the sports newsroom at NTB presents the early adopters of automated journalism, while VG's Jari Bakken finds himself currently among the innovators that tried the technology, but did not implement it.

Volan thinks that Norway has been relatively slow in the diffusion of automated journalism. The journalist sees several reasons to it: the quite small size of the country (and its population) leads to the smaller number of news outlets in general and the lack of stored linguistic knowledge, also because a large amount of research in Norway is published in English. Another problem is the lack of scientific labs at the universities that could develop NLG software: the software for NTB's football robot was developed by the students at NTNU. But if there were several similar universities, it could probably be done earlier. Therefore, according to Volan, such a development pace is natural: "it would take a little longer for Norwegian newsrooms to realize there is a market for the scalable output". She mentions that Norwegian media are usually quite innovative, also because the country is highly digitized:

> The Norwegian audience is possibly more advanced than what you find on the American West Coast, around San Francisco. Norway by many is viewed as one of the major laboratories for studying how news consumers will act across the globe, because we consume a lot of news and we are technologically-savvy. Moreover, we are rich enough to have all the necessary equipment.

The study showed that automated journalism is not going to be a sole solution to the crisis in the media industry. As Volan puts it: "it is one of the many little steps". However, put together with other innovations, by creating new products and new ways of delivering the content it can significantly diminish the financial burden on the industry. The respondents agreed that the main problem in the news industry today is the old business model – selling audience to advertisers, which does not bring as much profit today as it did in the course of several centuries. Therefore, there is a need to rethink the existing business model in journalism with the assistance of automated journalism.

Bakken sees the main challenge for the news industry in the digital shift as the unwillingness of consumers to pay for online content: "it is hard to make money from digital advertising. It is not that people are less interested in news, but they have so many free sources available that they simply do not want to pay". Therefore, he means, it is important for the news industry to cultivate the habit of paying for content itself, paying for the quality journalism,

among the news consumers. The recent report of the Reuters Institute for the Study of Journalism (2016) showed that 27% of Norwegians are already paying for news on digital platforms, which is the biggest share of the population compared with other countries. Barland laments the significant decrease in the most stable revenue model for the mass media industry – the subscription:

> Firstly, those who were subscribed to 4 newspapers before are now subscribed to 3; those who had 3, now have 2. Many stopped their subscriptions. Secondly, although the combined paper and digital subscriptions cost the same as the paper subscription before, the pure digital subscriptions are much cheaper. They do it, because they have to sell their product to be able to stay on the market.

There is no point in looking at a particular innovation and making assumptions about whether it has the potential to save the whole situation, but innovations definitely make the processes more efficient, and a combination of innovations may change the situation for the better. As Rogers (2003) argues, innovations do not happen independently and exclusively. Since it is hard to predict whether an innovation will succeed, there are often several innovations that are being implemented simultaneously in newsrooms (Ekdale et al., 2015). As Barland puts it, "in digital development, there is never a single solution – some innovation here, a little better payment model there, success in selling something, and when you've done all this, you see that it works, then comes another problem".

Aabech says that one of the reasons to introduce automation in NTB was the fact that the usual journalistic model of prioritizing the biggest matches as the most important gradually became obsolete. At the moment, NTB does not have a new business model for the football robot, because it is still a new service. Karlsson talks about the need to rethink the business models in journalism: "the way of publishing, the readers' demand, the technology – everything has changed. […] You need to rethink: what is audience, what is content, what is newsworthy? Thinking like that, you can grow a new audience and new revenue streams from new services". Śliwowska thinks that automated journalism is not going to be profitable for the media organizations, but only for technology providers and for the niche areas that will be willing to make significant investments in content personalization. The biggest economic potential automated journalism can bring arguably only through personalization of news. This opportunity will be discussed in the next section of this chapter.

5.3 Personalization, localization and customization

As was shown in the previous section, automated journalism is not here to create profit for media organizations. It is partly cost- and time-saving innovation; partly it helps relocate the human resources for other journalistic tasks, from the mundane and repetitive ones. It seems that this innovation cannot save the news media from the crisis. Or maybe it can? Newsroom professionals, looking at the economic possibilities of automated journalism, argue the technology can bring profit through content personalization, if there is a demand for that. Aabech argues that the technology can turn from the cost-saving to profitable, when the newsrooms start using it to adapt the news stories to the readers' interests. Personalization already made the CEO of Amazon the richest person on Earth (Vinton, 2017). Jeff Bezos offered the ultimate algorithm for book recommendations, which made Amazon's crazy high sales possible (Pariser, 2011). The personalization of news is an extremely popular topic among news professionals today. In this section I will discuss the opportunities and dangers that personalization of news content can bring with the help of automation.

5.3.1 Personalization of news

Today in Norway the Polaris Media group specializes in personalization of news. The company started experimenting with personalization in 2012 (Hagen, 2017). Jørgen Frøland, responsible for the personalization project at the company, admits (Hagen, 2017) that the big problem they face is the limited capacity to produce a sufficient number of articles: local newspapers simply do not possess the resources for the scale necessary for personalization. Automated journalism can solve this problem. Volan from Adresseavisen, which belongs to Polaris Media, argues that automated journalism brings the possibility to cover the stories that were not usually covered due to the lack of resources in newsrooms and because the newspapers had to prioritize the most important cases: "we are a big regional newspaper with around a hundred journalists, but we do not cover sports in the lower divisions in our region, because we do not have the resources to be everywhere. Robots could do that for us". Aabech faces the similar problem at NTB, where a relatively small sports newsroom needs to cover the football matches for the whole country: "since the newspapers do not cover the matches themselves,

they rely on us". The question of giving priority to certain matches is central for the agency, according to Aabech:

> Sometimes we have 4-5 games simultaneously. When we decide what to give the priority, we think just like the newspapers: what is the most important? We usually wrote about the biggest matches first. And for every match we consider as the most important, there will always be a match more important for another newspaper.

Volan thinks that the readers of Adresseavisen would appreciate the personalized product made with the help of automated journalism: "with automated journalism I can serve the news that we think fits best just for you. We are working quite heavily with personalization at Polaris Media. Now you can start adding geography-based filters, age brackets, building in the limits to data sets, narrowing them down". Barland also looks at the problem from the readers' point of view and demonstrates how journalism can benefit from personalization:

> It's interesting to read about the places where you live. If localized, even the boring statistics becomes an interesting journalism. There is a lot of information that could be important for me, which I never receive, because it has not been structured and delivered in a way that would attract me.

The personalized news content can be especially useful for local newspapers, if they can get the national statistics for their regions, argues Volan: "if NTB made an algorithm covering the elections, that would be useful for me; if they sent us this region's numbers and then the top national statistics. The numbers for all the regions are not interesting for my readers". The respondents see an opportunity in the personalization of some parts of articles. Ruud talks about the need for NTB to stay "relevant to each customer", which can be achieved by automating certain parts in news articles: "Almost every national story could have at least a sentence or two that were relevant to each of the 200 customers". Aabech argues that NTB already provides similar services to the customers in various regions. He agrees with Volan about the usefulness of automated journalism for the local press: "the newspapers can get 2 teams they needed, 4 teams or one team instead of just getting everything or nothing. This is also where we can start earning money with automated journalism. Because when you upscale, you have a new product – personalized service for newspapers".

5.3.2 Creating new personalized media products

Aabech assumes that with the help of personalization the Norwegian news agency may switch its football robot from a cost-saving to beneficial innovation by developing new products (and new business models for their distribution). There currently exist several directions where personalization, with the help of automation, can create new media products: localization, increased scale, different languages and news on demand.

Localization

Feulner says that already today a lot of news media use localization as a way of tailoring personalized content to their readers. Especially, it is common for weather forecast services. Wetter.de in Germany uses the software to send individual messages to customers, based on their zip-code. For example, says Feulner, instead of a pictogram with a sun or a cloud, you can get a recommendation: "you'd better put on a warmer coat today" or "you'll probably want to put your car in the garage, it's going to be a hail storm". Another localization format that Feulner mentions is micro grid that allows tracking, down to the exact street, which can also track the movements of the weather. Johannesson thinks that users should be able to choose several location settings, because today people are rather mobile: they were born and raised in one place, studied in another and work in yet another place; therefore, they might be interested in news from not just the place they currently live in.

Larger scale

The most obvious way of creating new products with automated journalism is to increase the scale of reporting: "when you are able to scale, making local stories, you might have a product that you might want to charge a bit extra for […] the scale is the most exciting part of automated journalism". Johannesson agrees that only with automated journalism did it become possible to send to users "a push that says this team won or this team lost" immediately after the game, without having to prioritize certain matches, while sparing the newsroom resources. Karlsson gives the example of real estate market reports: if previously, there were just the reports done by journalists about the most expensive or interesting houses once in a while, today the newspapers can make a whole new product out of this data – daily summaries

or 500 stories a day about this market for interested users. He even calls automated journalism "the economy of scale". Ruud, however, points out the distribution problem with personalized news: "we need to find out how to distribute hundreds of different stories to the customers so that they won't get confused".

Translation to different languages

Another possibility of automated journalism to create new media products is the ability to automatically translate content to different languages. Barland describes a possible implementation of automated journalism in the newsroom, using translation: "you can take an article in Swedish or English and then automatically publish it in Norwegian". The researcher warns that the software might not be so advanced yet, and that the journalist might have to go through the text before publishing it. Still, such automated translation can be of great help in newsrooms: journalists often have to use the materials published in other countries and in various languages when searching for stories. Even when the journalist knows 1-2 foreign languages, it is very inconvenient to read the text in other languages, which limits the possibilities of journalists. With automated article translation, it is possible to see what the article is about and to give it to a professional for translation when necessary. Such automated journalism products could spare significant amount of time and resources in newsrooms. Barland argues that this can be a way to build a new business model for journalism:

> There can be created international text databases, where journalists could publish their texts and buy them. Then you could say "I want to read everything in Norwegian" and to find articles written in New Zealand and Argentina, Israel and Kazakhstan.

In Germany, AX Semantics and Textomatic are working on automating texts in various languages. Feulner argues their technology currently allows localizing the written materials into almost 20 languages. It gives news organizations, among others, an opportunity to compete in other language markets, if this is their aim. Bense thinks that such a service can also be useful for tourists, who can get news and information about the country they are traveling in their language.

Production of news on demand

Another lucrative possibility for personalization with automated journalism is the “on demand” reporting. With the ability to automate short news stories, there is no more need for journalists to make a choice and prioritize which story to write about, as they can be written automatically upon request from the readers. Barland talks about the implementing on demand reporting for weather forecasts in Norwegian weather service YR, in written form instead of only visual icons: “when they relaunch YR, it will be much more text-based. And that text is of course automatically generated, because Norway is divided into a grid of 2x2 km tiny routes, and there should be a text for each. The text will be generated on the screen only when someone asks for it”. Aabech looks at the issue from the B-to-B perspective: “our clients can say “we just want this team”, “only prices for this area”, “the election results on different levels of reporting". We were not able to do this before, because if a person does all this, it will take forever”. Feulner says it can be also possible to receive certain news on a certain time, like the weather at 8 a.m. when one is leaving for work: “automation is there for you 24/7”.

5.3.3 Personalization or customization?

While personalization is a way of delivering individual content to the consumer, customization is an opportunity for the users to adjust the content to their preferences. Usually, says Feulner, it is possible to do via logging in to a news platform (publications that introduced paywalls already require their users to log in) and setting the preferences. A lot of media use cookies on their websites, but Feulner finds it slightly unethical to use cookies for personalization of content, as the user might not be aware of the personalization and it is similar to spying on the user’s browser. Johannesson thinks it is too dangerous to rely on the readers customizing their personalized news content and offers the model suggested by Amazon: “if you buy this thing, you are probably also interested in this” or “we've seen that you buy this a lot of times, so probably you should look at this as well”. Grimnes gives an example of personalization of news content for various social groups that the pioneers of automated journalism are currently occupied with:

> I think Automated Insights, Narrative Science and the UK software company Arria have a lot of linguistic knowledge recorded, and now they work on personalization for different target groups: you can write texts that use simple vocabulary and

> sentences for children and people with learning disabilities, for example.

In the Netherlands, the ANP news agency created a tool that adapts news stories for children by automatically simplifying the language and republishes them in its Kids Feed (Bilton, 2017).

5.3.4 The critics of personalization

The production of content customized to the profiles of individual readers, instead of writing one text and hoping that it will be equally interesting for many, brings opportunities for new business models through automated journalism. However, it is important that the readers are aware of the price they pay for the opportunity to receive their personalized content – their personal data, geo location, and potential imprisonment in their filter bubbles. Now, many readers do not know about personalization (Smiley et al., 2017), while the ethical way to treat this problem would be to underline the fact that their personal data is being tracked. Today, according to Feulner, it is even possible to adjust the text for the reader in real time if the reader is not engaged enough with the material, which for many sounds rather spooky. Bakken expresses his concern that further personalization of news will lead to further fragmentation of society with gradually more radical views and even worse misunderstanding, blocking the way to enlightenment and shared experience of knowledge:

> If all news is personalized for you, then you never get exposed to new ideas. You could argue that newspapers do not necessarily give it to you in Norway either, but at least they give a shared experience: we read the same newspapers; we get a shared understanding of what is happening. If everything is personalized, the society gets split into smaller groups with very strong views that repeat themselves.

Though Norwegians are not particularly happy about being trapped in the personalized bubble, it is arguably the future of journalism (Jensen, 2016b). According to the report prepared by the Reuters Institute for the Study of Journalism (2016), people in Norway are mostly worried about missing out on important news in their personalized news feeds; many also want to be exposed to the challenging points of view unlike their own. Norwegians are concerned with their privacy when it comes to personalization. Therefore, it is important for news media when they develop new personalized products to take these factors into consideration. In any

case, as Karlsson points out, our relationship with information and news becomes more close and personal every day, but “when you personalize too much, it is not journalism anymore, it becomes marketing”.

5.4 Summary of chapter 5

This chapter covered the aspect of automated journalism as an innovation. Firstly, I argued and proved that automated journalism can be considered an innovation in newsrooms, and, although it can be said that the phenomenon has developed naturally, it is more of a disruptive innovation as it has a potential to change the way people consume news and, therefore, the news market.

Secondly, I described the diffusion process of automated journalism in Norway and, with the help of Rogers’ Diffusion of Innovations theory, I assessed the likelihood of automated journalism being adopted in Norwegian newsrooms. The findings showed that automated journalism is most likely to be developed by big news providers, such as news agencies and big national newspapers. But it will only become cost-effective when news organizations develop new suitable business models for the products that automation, together with personalization of news, can offer.

Finally, the personalization of news, the dream of the news industry, can finally become possible with automation. The trend will most likely revolutionize the industry, but the providers of personalized news should take into account the ethical considerations.

6 Journalists going technical

The present chapter describes the challenges that journalists meet today in regard to the recent digital and quantitative turns (Coddington, 2014) in their profession. The possibility of automating some journalistic tasks, the growing need for technical skills in journalism and the changing nature of the profession in the digital world trigger skepticism towards automated journalism. The idea of introducing automation in newsrooms meets a certain resistance from the journalistic staff. Marconi et al. (2017) argue that the impact of artificial intelligence on journalism will not be expressed in how the machines work, but in how journalists adapt to the new technology. The chapter discusses the relations between journalists and machines in newsrooms, and the shift in professional knowledge and skills towards digital and it pays special attention to the relationships between journalistic staff and the programmers in the development of automated journalism.

6.1 Arguments for and against journalists going technical

Technical skills are gaining more and more value in journalism. Being able to find information and present it in innovative ways makes journalism more interesting. Bakken talks about several journalists at VG who are taking a year off to study computation, because they want to be able to use data more efficiently: “they are now able to analyze data sets, to create maps, tools and databases for themselves”. The developer also admits that the acquired skills give journalists a number of advantages over their colleagues, while bringing technical skills to the newsroom allows creating the projects that were not possible to do before, like exploring and analyzing massive data sets in various ways.

The current study showed that automated journalism is capable of increasing journalistic efficiency by reducing the number of tasks that could be automated, while freeing up time for creative and interesting journalism. Volan sees automated journalism as an opportunity for newsrooms to prioritize their resources better, but also emphasizes the new technological possibilities that automated journalism brings for journalists: “I do think that journalists need to know more about computing, not out of the fear of being fired if they don’t. Journalists should

be more enthusiastic about what this technology can bring to journalism as a profession. I see a huge mixture of things that are beneficial for the media industry that are more than 'oh, we can lay off the journalists because there is a robot that is doing what that journalist did'". Aabech argues that automation brings a lot of opportunities for journalists: "they can coach the robot. Some people say, it is not journalism, it is just the data and text, but there are a lot of journalists that have gone into the development, from deciding on the language to what is important/not important. That is a journalistic process". Volan agrees: "I think that in the sort of best future scenario the best journalists are those who are able to work alongside automated journalism, possibly even as the software trainers, if we can see that as a future journalistic profession. They [machines] need to be taught by humans".

Ruud and Aabech advocate for the inclusion of more journalists in the automation projects in the newsroom, arguing that this will lead to more ideas and new projects: "we need sports journalists working on sports robot, political and news journalists working on the elections and national statistics. For an economic robot you need someone with the knowledge of economics". Ruud and Aabech think that many newsroom divisions can benefit from automation and, therefore, they need more journalists to be involved to understand the opportunities with this technology. Additionally, the knowledge and skills acquired while working on automation, according to Ruud, will give these journalists certain benefits over other journalists, as these unique skills may secure them jobs in the future.

There are, however, also opponents of the introduction of automated journalism in the newsrooms. There can be different reasons to this. Bakken sees age as one of the factors that creates the resistance to the automation technology in some journalists: "journalists need more technological skills then they used to have, when we only made the paper issue. There is kind of a generational shift: the older generations of journalists… some of them are eager to learn, but some of them really don't want to touch anything, they simply don't understand. But some people get more out of it". Johannesson also thinks it is hard for experienced journalists to change their practices overnight: "they have been writing for all their lives. Even if you tell them they can do anything because they are talented journalists, it is always tough for someone to realize that their job as they knew it is gone".

Aabech considers this skepticism as normal: "of course, some people would be skeptical". Aabech and Ruud think that the reason some journalists are unwilling to engage in automation in the newsroom is because they think no machine can do a better job than they do, and thus, ignore the benefits that automation of some of their most boring tasks can bring. Such

journalists are also reluctant to accept that a human is simply incapable of producing hundreds of articles simultaneously, while increasing the scale of reporting is one of the current priorities for NTB. Johannesson agrees that some journalists "don't yet realize where we are heading with the digital media", but adds that the concerns often come from their responsibility towards the public for the quality of the media content: "they [sports journalists] think they have pressure from the people to always be present at the games". Karlsson thinks that journalists consider their jobs under threat of automation because "they have an idolized and romantic picture of their profession. It is not just any work, it needs to be protected, it is important, and it serves the democracy". Therefore, it can be hard to introduce change in a newsroom.

Aabech thinks that the key to the problem is to get the journalists acquainted with this technology: "when they understand how it works, why you do it, when you show them 25 different ways of saying the same sentence with automation, they are actually pretty impressed". Johannesson agrees that the knowledge about this technology and the company's development plans towards further digital will help journalists accept and embrace this innovation. The journalists need to be a part of the decision-making process in the news organization.

6.2 Common models of automated journalism development

The development of automated journalism requires cooperation between the developers and the journalistic staff. In different countries and in different newsrooms there are different ways of developing automated journalism. There are several common ways to do it, among which most of the newsrooms choose the most suitable for their needs, based on the available resources. In the simplest cases of automated journalism, the journalist and the programmer is the same person, while for more technologically complicated systems, newsrooms often use external knowledge, sometimes even hiring the IT-companies to develop the algorithm. The relationships between the journalists and programmers in those cases are therefore important to look at while researching the phenomenon of automated journalism as a media innovation.

Individual developers

The first group of developers of automated journalism was made up of the computational journalists and the developers in newsrooms that were trying to automate simple journalistic tasks. One of the first journalists to develop a simple news robot was Ken Schwencke from the Los Angeles Times. In 2014, he developed Quakebot – an automated program that took data from the US Geological Survey and published short reports about earthquakes in Los Angeles a couple of minutes after they occurred (Oremus, 2014). In this case, the software was programmed to wait for the approval of a human journalist. Later, the L.A. Times used their invention to automate reports on homicides in the city.

In Norway, Jarri Bakken was the individual developer behind a relatively simple municipality debt report, which created more than 400 articles out of 4 written templates. Bakken says that in VG they are working as a small team of programmers, helping journalists with computational tasks, such as analyzing data sets, creating maps and visualizations and other computational tools.

Developing automated journalism in newsroom with external expertise

The second way of developing automated journalism is to do it in house with the help of external developers. This model is usually used by newsrooms for developing more complicated systems that use the NLG technology. In Sweden, MittMedia developed a sports robot with the help of the developers from United Robots. Henning Johannesson worked on the front-end, while the programmers from United Robots developed the back-end. Here, automated journalism was applied to sports, similar to NTB in Norway. Even the reason behind it was almost the same – to improve the speed and quality of the lower division reports.

NTB has developed their football robot in cooperation with the IT-company Bakken & Bæck, whose project Orbit is focusing on artificial intelligence. Aabech mentions that the technology was in fact first developed by the students of the Norwegian University of Science and Technology (NTNU), from whom Orbit bought the license.

Buying license to use a ready-made product, developed by IT-companies

Two software companies, Automated Insights and Narrative Science, specializing in the technology of Natural Language Generation (NLG), have launched self-service platforms for news outlets and thus opened a new market for editorial content (Wright, 2015). Today their products are used by, respectively, the Associated Press and Forbes. Similar products are currently being developed in at least 9 companies worldwide – 5 in Germany, 2 in France, and 1 each in the UK, China and Russia (Graefe, 2016). The NLG technology is not always directed towards journalism – it has various implementations in the business sphere or any field that is dealing with data and narrative generation – for example, in advertising or finances (Wright, 2015).

According to Grimnes, such software companies have worked with journalists on the development of automation for news writing on several projects before they started working on their ready-made products: "they had a lot of projects, similar to our project with NTB – big consultancy projects, where you work closely with your customers. Wordsmith is one of the tools they are trying to develop to be exclusively used by customers. This allows them to reach the bigger market". Grimnes thinks such ready-made products must have relatively simple systems:

> The higher levels of technology are still so complicated that they cannot be used by customers independently. That is why you need to have someone writing the templates or encoding the linguistic knowledge, choosing the data you want to write about and the logic form of its presentation. Wordsmith is simpler than what we developed for NTB.

Karlsson thinks such simplified services are only suitable if simple texts based on small data sets are needed: "that is not what we are selling to media companies. We see ourselves [United Robots] more like a news agency without reporters: we have algorithms".

6.3 Coding like a journalist

The respondents working with automated journalism admit that coding will soon be among the normal journalistic skills. Bakken thinks that if the programmers create user-friendly interfaces journalists can learn to program automation software for their needs, teaching the

machine how to write journalistic texts. Aabech explains that at NTB journalists already started to code inside the program for the football robot: “first we just worked in Google Docs. Eventually we learned how to code in the program as well, so we could do it faster. Once you understood it, it was quite easy. There were different rules inserted there, but inside those rules I could add my code”. Johannesson also admits that he can be considered both a journalist and a programmer now, since he doesn’t “just write text and then give it to someone”. He started from this, but eventually realized it was easier to do everything himself: “If I don’t remember some code, I have a really good connection with the programmers and can get my answer in 5 minutes. It wasn’t a big challenge for me to start programming. I’ve worked a lot with computers and HTML code previously. But, probably, an average journalist wouldn’t have set in as quickly as I did”.

Volan underlines the necessity of involving journalists in the development of automated journalism, as they possess the necessary knowledge for building a successful system: “journalists should be definitely involved in the training of the robot to see, what sort of data is used as an input and how to communicate it in a way that feels natural to a reader or a TV-consumer”. She underlines storytelling as one of the main aspects of reporting that journalists should teach the machine: “a good Norwegian language for football is not the same for weather or finance. You need to teach them to speak a language that an ordinary news reader or a consumer will understand and appreciate”. Volan gives an example of the difference between a programmer’s and a journalist’s thinking when teaching a machine the language:

> The programmer generates a list of surnames of all soccer players within a league. Then the journalist says: “but you wouldn't want to use his surname too often in a story” or “you might want to change his first/middle/last name”. Making the automated language fluent is a new journalistic task associated with automated journalism.

Aabech emphasizes that journalists’ coaching of the robot is not limited to teaching it the proper language. It is journalists who work with the developers to build in the proper rules and priorities in the software: “we told the developers, ‘when this happens, this is the priority, when this happens, this is more important, etc. Sometimes you can use both’”. After NTB launched its first football robot, the journalists and developers were monitoring the activity of the software during matches. Aabech says: “we were sitting on Skype with the developers and chatting, looking if we could see some mistakes”.

The experience from the development of the football robot made it easier for the newsroom to program a new robot. For the elections in Norway in 2017, journalists at NTB with the help of an internal developer programmed the system themselves. Aabech explains that it was a relatively simple task: “we used software apps for the front-end, wrote the code for the templates in JavaScript. The developer told us how to write the code already inside the program. It made the working process a lot smoother”. Since the system this time was simpler, NTB decided to put a journalist, who participated in the development, in charge of publishing the automatically produced texts in case there were repetitions or mistakes, as a “safety net”. Aabech wants to attract more journalists and developers at NTB to develop various automation projects, while Karlsson thinks that journalists won’t have to acquire too complicated technical skills: “They will need an understanding of how their automatically written draft was produced”.

6.4 Consequences of automation for journalistic jobs

Bakken warns about the close connection between saving resources in newsrooms and the dismissal of journalists: “the obvious solution would be to get rid of the journalists, it will save the media economy. But the really important tasks in journalism still need humans”. Aabech does not agree with him, saying that getting rid of journalists was never the goal for automation at NTB. Barland notes that journalistic layoffs will happen with or without automation, because the revenues from the obsolete business model of selling advertisements continue falling. On the contrary, automated news production can help media organizations continue producing as much content as before with fewer journalists, keeping the business afloat.

Aabech, however, does not agree with this statement, because the experience of NTB showed that journalists are needed as much with the presence of the automation as before: “we do not see how we can have fewer journalists on duty. I do not think we can say we are going to fire two people. Automation of news can help save some costs, but it is never going to replace the journalists”. According to Aabech, there are currently 8 full-time journalists in the sports section at NTB, plus some part-time reporters. On Sundays, there are usually 3-4 journalists in the newsroom. They are reporting on all the sports and are rather busy. Aabech thinks the most

valuable outcome of the development of a robot in the newsroom is currently not its cost-effectiveness, but the training of journalists in such projects. The purpose of the elections robot, according to Aabech, was to free several journalists during the elections night to do other kinds of reporting, but also to get more experience in developing and working with automation in newsroom, so that later, with similar projects, it could become cost-effective and profitable.

Karlsson thinks here like a manager: “I know that we had put some journalists out of work, but productivity is a factor even in our business. I used to say that if journalists don’t think they can compete with an algorithm, then maybe they should look for other jobs”. He, however, is sure that there will be plenty demand for high quality human-produced journalism in the future.

6.5 Summary of chapter 6

The automation trend is indeed changing the way journalism is produced today. If the changes are currently not that obvious, there is really no way to stop the progress: because in any business, including the media industry, if a process can be improved by automation, it will be done. In the chapter 6 I discussed the different points of view on the introduction of automated journalism in newsrooms and argued that many spheres of journalism can benefit from automation and that it is important for journalists to get to know their opportunities without the irrational fear of the coming changes. There are certain concerns, however, that management should convey to the journalistic staff. But journalists need to be a part of the discussion about automation in the newsroom. Then I described the several common models of developing automated journalism in newsrooms and underlined once again that each newsroom needs to choose one based on their needs and existing resources. Finally, I discussed whether it is important for journalists to learn coding to develop a journalistic robot and presented different views on the question of the loss of journalistic jobs in the future. I want to finish this section by the quote of Peter Thiel from his book “Zero to One”: “The most valuable businesses of coming decades will be built by entrepreneurs who seek to empower people rather than try to make them obsolete”. This thought is true also for journalism in the age of ubiquitous automation.

7 Summary and conclusion

The present thesis is an attempt to describe the phenomenon of automated journalism from the media innovations perspective on its rather early stage of implementation. The automation technology for writing articles has been developed in several newsrooms around the world since 2014 and is mainly used for the data-saturated spheres of journalism, like sports, finances and weather reports. But how is it implemented in various newsrooms and what is its future potential? The thesis aims to answer these two questions using the method of in-depth qualitative interviews with journalists and developers interested in newsroom automation.

Chapter 1 outlined the context of the phenomenon and presented the research questions of the study: *(1) How is automated journalism currently being implemented in newsrooms in Norway?* and *(2) What is its potential as an innovation?* In chapter 2 I defined and discussed the terms I used, described the Diffusion of Innovations theory and made a summary of the existing research on automated journalism to date. The methodological approach of the study – qualitative in-depth interviews – was justified and described in details in chapter 3. The results and discussion parts were organized by topics and formed 3 chapters: chapter 4 analyzed the technical aspects of automated journalism and its ethical challenges; automated journalism as an innovation and its role in the personalization of news was discussed in chapter 5; chapter 6 is devoted to the potential changes in journalistic practices due to the introduction of automated journalism in newsrooms.

Firstly, I described the technological functionality of automated journalism and found out that automation has a number of benefits for the media industry, among them speed, accuracy and the scale of reporting. Moreover, it became evident that is possible to introduce automation using various levels of complexity, from a Word's drop-down menu to the algorithms powered by artificial intelligence. The quality of text is rather satisfactory and not that distant from an article written manually by a journalist, since journalists write the templates for the algorithm. However, the talks about machines fully replacing human labor in journalism lack the necessary understanding of the distinctions between the human and algorithmic capacity: they are simply good at different tasks. Therefore, the two are meant to complement each other rather than substitute one another. Then, in the same chapter I covered the ethical issues with automated journalism, arguing that the developers of the automation technologies in newsrooms should pay significant attention to the choice of data for the input and the explicit

programming of the algorithm. There are many ethical considerations that each newsroom should discuss before they launch their journalistic robot, so that it would be compatible with the norms and values of a particular news outlet.

Secondly, I discussed automated journalism's potential as a newsroom innovation. I argued that, although automation of certain journalistic tasks has a rather long history in newsrooms, the algorithmic text writing for autonomous publishing is a new phenomenon and has a potential to become a disruptive innovation for the media industry, as it may completely change the way people consume news. Currently in Norway, automated journalism is spreading rather slowly. The news agency NTB is the only news organization that has implemented this technology, while some other actors evaluate their needs and availability of resources to develop the automation algorithm. The study showed that automated journalism can be suitable for specific data-saturated areas of journalism, where automation does not contradict the ethical considerations. Some of the areas are perfectly suitable for partial automation. Arguably the main reason why automated journalism is spreading so slowly is the lack of an appropriate business model – how to make rather expensive technology cost-effective? One of the answers to this problem lies in the opportunities that automated journalism brings together with the news personalization, namely in creating new personalized journalistic products.

Finally, in the last chapter I discussed the compatibility of automation with the existing journalistic practices. Automated journalism brings several undeniable benefits to journalism, which should be properly communicated to the journalistic staff in newsrooms where the technology is implemented. The majority of the concerns come from the fear of the unknown: Will automation affect the quality of journalism? Will it reduce the number of journalistic jobs? How will the readers react to the news written by algorithms? Furthermore, it is not a requirement for journalists to learn advanced programming in order to work with automation. In the majority of cases, simply to understand how the article is produced would be enough. Generally, journalists tend to welcome technological innovations; they only want to understand clearly their benefits and risks (Ekdale et al., 2015). There are several ways in which it is possible to help the newsrooms to embrace the opportunities that automated journalism brings:

- To develop and introduce user-friendly interfaced for programming journalistic robots in-house;
- To introduce the relevant courses for journalists in semi-coding automation assistants;

- To give the information and explanation of the possibilities of automated journalism for every journalistic task from the management of the news organizations (Ekdale et al., 2015).

Creating fiction and storytelling, according to historian Harari (2015), is the major cause of evolutionary advancement of Homo Sapiens over other human species and animals. Therefore, it is rightfully considered a human prerogative. Today, when artificial intelligence becomes significantly more advanced than anyone expected, the ability of machines to write narratives shocks society. Naturally, automated journalism is perceived as a threat to human journalism. Bainbridge (1983) writes that the nature of automation implies that the human operator is unreliable, or inefficient. This might be why automated journalism causes such heated debates. Journalists are writing the diary of humanity, and their journalistic pride is wounded: how can a robot be a better journalist? But the answer is, it can't and won't ever be a better journalist, it can hardly be called journalist at all.

To stress this point even more, I argue that machines will soon reach a very high level of competency in journalism, but there is no reason for newsrooms to dismiss journalists in favor of the algorithms, because the human touch constitutes a very big part of the profession, and it can never be substituted. Automated journalism can help redistribute the resources in newsrooms in order to give journalists more time for doing interesting journalistic work, like taking interviews, elaborating on stories, doing deeper research and more in-depth analysis, as well as fact-checking. Automated journalism can even create more journalistic jobs in media industries, like robot coaches and robot developers, because a satisfying level of automation requires a significant input of human knowledge and journalistic, as well as linguistic, experience. Therefore, automated journalism should be regarded as a helping tool rather than an autonomous actor in a newsroom, despite its seemingly autonomous nature. The newsroom of the future will see more such collaborations between journalists and machines (Marconi et al., 2017).

More studies should be done on the possibilities of automating specific tasks in newsroom.

Limitations and recommendations for future research

As a Master thesis, the study was limited in resources. The time constraints did not allow following the diffusion of this innovation in Norwegian newsrooms over a long period of time.

The study was also limited in terms of chosen methodology: it would have benefited from an observation done at NTB and a field study in the US. But a researcher has to make choices.

The ethical issues with automated journalism deserve further and deeper research as they present the main obstacle to the innovation's diffusion, according to the results of present study.

The study would benefit from input regarding the history of journalism and media business in Norway, and better knowledge of the particulars of Norwegian society.

Bibliography

Adar, E., Gearig, C., Balasubramanian, A., Hullman, J. (2017). PersaLog: Personalization of News Article Content. *Proceedings of the 2017 CHI conference on Human Factors in Computing Systems, Denver, Colorado, US, May 06-11, 2017*. http://dx.doi.org/10.1145/3025453.3025631

Agar, N. (2015, July 31). Why Techno-Optimism Is Dangerous. Huffingtonpost.com. Retrieved from http://www.huffingtonpost.com/nicholas-agar/techno-optimism-dangerous_b_7746682.html

Agarwal, R., Gort, M. (2001). First Mover Advantage and the Speed of Competitive Entry, 1887–1986. *The Journal of Law and Economics, 44, 1,* 161-177. https://doi.org/10.1086/320279

Allcott, H., Gentzkow, M. (2017). Social Media and Fake News in the 2016 Election. *Journal of Economic Perspectives, 31, 2*, 211–236. https://doi.org/10.1257/jep.31.2.211

Allen, R.C. (2017, October 18). Lessons from history for the future of work. *Nature, 550, 7676,* 321-324. Retrieved from http://www.nature.com/news/lessons-from-history-for-the-future-of-work-1.22825

Allern, S. (2001). Nyhetsverdier. *Om markedsføring og journalistikk i ti norske aviser*. Kristiansand: IJ-forlaget.

Amodei, D., Olah, C., Steinhardt, J., Christiano, P., Schulman, J., Mane, D. (2016). Concrete problems in AI safety. arXiv preprint arXiv:1606.06565.

Anderson, C.W. (2013). Towards a sociology of computational and algorithmic journalism. *New Media & Society, 15, 7,* 1005 –1021. http://dx.doi.org/10.1177/1461444812465137

Angwin, J, Larson, J., Mattu, S., Kirchner, L. (2016, May 23). Machine Bias. There's software used across the country to predict future criminals. And it's biased against blacks. *ProPublica*. Retrieved from https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Associated Press, NORC Center for Public Affairs Research, American Press Institute. (2015, March). *How Millennials Get News: Inside the Habits of America's First Digital Generation*. Chicago: Rosenstiel, T., Sonderman, J., Loker, K., Tran, M., Worthington, L., Swanson, E., Tompson, T., Benz, J., Reimer, R., Alvarez, E., Malato, D., Sterrett, D., Willcoxon, N., Zeng, W. Retrieved from http://www.mediainsight.org/PDFs/Millennials/Millennials%20Report%20FINAL.pdf

Automated Insights. (2017, November 10). Natural Language Generation. Use Cases. *Automatedinsights.com*. Retrieved from https://automatedinsights.com/use-cases

Bainbridge, L. (1983). Ironies of Automation. *Automatica, 19, 6,* 775 779. http://dx.doi.org/10.1016/0005-1098(83)90046-8

Barland, J. (2015). Innovation for New Revenue Streams from Digital Readers: The Case of VG+. *The Journal of Media Innovations, 2, 1,* 123-130. http://dx.doi.org/10.5617/jmi.v2i1.952

Barrat, J. (2013). *Our Final Invention: Artificial Intelligence and the End of the Human Era*. New York: Thomas Dunne Books.

Battelle, J. (2006). *The Search: How Google and Its Rivals Rewrote the Rules of Business and Transformed Our Culture*. Boston: Nicholas Brealey Publishing.

Bender, W., Chesnais, P., Elo, S., Shaw, A., Shaw, M. (1996). Enriching communities: harbingers of news in the future. *IBM Systems Journal, 35, 3-4,* 369 – 380. https://doi.org/10.1147/sj.353.0369

Berman, S.J., Abraham, S., Battino, B., Shipnuck, L., Neus, A. (2007). New business models for the new media world. *Strategy & Leadership, 35, 4,* 23-30. https://doi.org/10.1108/10878570710761354

Bharat, K. (2006, January 23). And now, News. *Google official blog*. Retrieved from https://googleblog.blogspot.no/2006/01/and-now-news.html

Bilton, R. (2017, July 21). A new report examines the state of automated journalism in Europe — and what's holding it back. *NiemanLab.org*. Retrieved from http://www.niemanlab.org/2017/07/a-new-report-examines-the-state-of-automated-journalism-in-europe-and-whats-holding-it-back/

Boyles, J. L. (2016). The Isolation of Innovation. Restructuring the digital newsroom through intrapreneurship. *Digital Journalism, 4, 2,* 229-246, http://dx.doi.org/10.1080/21670811.2015.1022193

Bradshaw, P. (2014). Data Journalism. In Ethics for Digital Journalists: Emerging Best Practices, edited by Lawrie Zion, and David Craig, 202–219. New York: Routledge.

Bucher, T. (2016). 'Machines don't have instincts': Articulating the computational in journalism. *New Media & Society, 19, 6,* 918 – 933. https://doi.org/10.1177/1461444815624182

Calandrino, J. A., Kilzer, A., Narayanan, A. , Felten, E. W., Shmatikov, V. (2011). "You Might Also Like": Privacy Risks of Collaborative Filtering. *IEEE Symposium on Security and Privacy, Berkeley, CA, 2011,* 231-246. https://doi.org/10.1109/SP.2011.40

Cao, J. (2017, September 13). IBM's Rometty Sees AI Changing, Not Eliminating, Future Jobs. *Bloomberg*. Retrieved from https://www.bloomberg.com/news/articles/2017-09-13/ibm-s-rometty-sees-ai-changing-not-eliminating-future-jobs

Carlson, M. (2015). The Robotic Reporter. *Digital Journalism, 3, 3,* 416-431. http://dx.doi.org/10.1080/21670811.2014.976412

Chesnais, P.R., Mucklo, M.J., Sheena, J.A. (1995). The Fishwrap Personalized News System. *Proceedings of the Second International Workshop on Community Networking*. http://dx.doi.org/10.1109/CN.1995.509583

Christensen, C.M. (1997). *The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail*. Boston: Harvard Business School Press.

Clerwall, C. (2014). Enter the Robot Journalist. Journalism Practice, 8, 5, 519-531. http://dx.doi.org/10.1080/17512786.2014.883116

Coddington, M. (2015). Qualifying Journalism's Quantitative Turn. Digital Journalism, 3, 3, 331-348. http://dx.doi.org/10.1080/21670811.2014.976400

Cohen, N.S. (2015). From Pink Slips to Pink Slime: Transforming Media Labor in a Digital Age. *The Communication Review, 18, 2*, 98-122. https://doi.org/10.1080/10714421.2015.1031996

van Dalen, A. (2012). The Algorithms Behind the Headlines: How Machine-written News Redefines the Core Skills of Human Journalists. *Journalism Practice, 6*, 648–658. http://dx.doi.org/10.1080/17512786.2012.667268

Danziger S, Levav J, Avnaim-Pesso L. (2011). Extraneous factors in judicial decisions. *Proceedings of the National Academy of Sciences of the United States of America, 108, 17*, 6889-6892. doi:10.1073/pnas.1018033108.

Diakopoulos, N., Koliska, M. (2016): Algorithmic Transparency in the News Media, *Digital Journalism, 1-20*. http://dx.doi.org/10.1080/21670811.2016.1208053

Dogruel, L. (2013). Opening the Black Box. The Conceptualising of Media Innovation. In Storsul, T., Krumsvik, A.H. (Eds.), *Media Innovations. A Multidisciplinary Study of Change (pp. 9-12)*. Göteborg: Nordicom.

Dogruel, L. (2014). What is so Special about Media Innovations? A Characterization of the Field. *The Journal of Media Innovations, 1, 1*, 52-69. http://dx.doi.org/10.5617/jmi.v1i1.665

Dörr, K.N., Hollnbuchner, K. (2016). Ethical Challenges of Algorithmic Journalism. *Digital Journalism, 5, 4*, 404-419. http://dx.doi.org/10.1080/21670811.2016.1167612

The Economist. (2006, August 24). Who killed the newspaper? *The Economist*. Retrieved from http://www.economist.com/node/7830218

Ekdale, B., Singer, J. B., Tully, M., & Harmsen, S. (2015). Making Change: Diffusion of Technological, Relational, and Cultural Innovation in the Newsroom. *Journalism & Mass Communication Quarterly, 92, 4*, 938 –958. https://doi.org/10.1177/1077699015596337

Ertzaas, P., Bakken, J. (2016, January 24). Så mye gjeld har din kommune. *VG.no*. Retrieved from http://www.vg.no/nyheter/innenriks/saa-mye-gjeld-har-din-kommune/a/ 23600343/

Ess, C. (2009). *Digital Media Ethics*. London: Polity Press.

Fagerberg, J., Mowery, D., and Nelson, R (eds.) (2004). *The Oxford Handbook of Innovation*. Oxford: Oxford University Press.

Fink, J., Koenemann, J., Noller, S., Schwab, I. (2002). Putting Personalization into Practice. *Communications of the ACM, 45, 5*, 41-42. https://doi.org/10.1145/506218.506242

Fitzgerald, L., Ferlie, E., Wood, M., Hawkins, C. (2002). Interlocking Interactions, the Diffusion of Innovations in Health Care. *Human Relations, 55, 12*, 1429-1449. https://doi.org/10.1177/0018726021287822 13

Friedman, B, Kahn, P.H., Borning, A. (2008). Value Sensitive Design and Information Systems. In Himma, K.E., Tavani, H.T., *The Handbook of Information and Computer Ethics, 69–102*. New Jersey: Wiley.

Ganapati, P. (2010, August 16). Reverse-Engineering of Human Brain Likely by 2030, Expert Predicts. *Wired*. Retrieved from http://www.wired.com/2010/08/reverseengineering-brain-kurzweil/

Garrison, B. (2001). Diffusion of online information technologies in newspaper newsrooms. *Journalism, 2*, 221–239. https://doi.org/10.1177/146488490100200206

Gillespie, T. (2014). The Relevance of Algorithms. In Media Technologies. Paths Forward in Social Research, edited by Tarleton Gillespie, Pablo Boczkowski, and Kirsten Foot, 167–194. London: MIT Press.

Glascott, M.G. (2017, March 14). What is Natural Language Generation? Narrative Science. Retrieved from https://www.narrativescience.com/Resources/Resource-Library/Article-Detail-Page/what-is-natural-language-generation

Goertzel, B. (2007). Human-level artificial general intelligence and the possibility of a technological singularity: A reaction to Ray Kurzweil's The Singularity Is Near, and McDermott's critique of Kurzweil. *Artificial Intelligence, 171, 18*, 1161–1173. http://dx.doi.org/10.1016/j.artint.2007.10.011

Goertzel, B., Pennachin, C. (2007). *Artificial general intelligence*. Berlin: Springer.

Goertzel, B., Pennachin, C., Geisweiller, N. (2014). *Engineering General Intelligence, Part 2*. Paris: Atlantis Press.

Goodfellow, I, Bengio, A., Courville, A. (2016). *Deep Learning*. Cambridge: The MIT Press. Retrieved from http://www.deeplearningbook.org/

Graefe, A. (2016, January, 7). Guide to Automated Journalism. *Tow Center for Digital Journalism*. Retrieved from http://towcenter.org/research/guide-to-automated-journalism/

Graefe, A., Haim, M., Haarmann, B., Brosius, H.-B. (2016). Readers' perception of computer-generated news: Credibility, expertise, and readability. *Journalism*, 1-16. https://doi.org/10.1177/1464884916641269

Greenhalgh, T., Robert, G., Bate, P., Macfarlane, F., Kyriakidou, O. (2008). *Diffusion of Innovations in Health Service Organisations: A Systematic Literature Review*. Blackwell Publishing.

Greenslade, R. (2016, September 20). Why Facebook is public enemy number one for newspapers, and journalism. *The Guardian*. Retrieved from https://www.theguardian.com/media/greenslade/2016/sep/20/why-facebook-is-publicenemy-number-one-for-newspapers-and-journalism

Hagen, A. (2017, May 19). Polaris Media jobber aktivt for at alle deres mediehus skal ta i bruk personaliseringsteknologi. *Journalisten*. Retrieved from https://journalisten.no/2017/05/polaris-media-vil-tilby-personaliseringteknologi-til-alle-sine-aviser

Hamilton, J. T., Turner, F. (2009). Accountability through Algorithm: Developing the Field of Computational Journalism. *Report given at the Behavioral Sciences Summer Workshop, Stanford, CA, July 27–31*. Retrieved from: http://web.stanford.edu/~fturner/Hamilton%20Turner%20Acc%20by%20Alg%20Final.pdf

Harari, Y.N. (2015). Sapiens: A Brief History of Humankind. NY: Harper Collins.

Henn, S. (2012, October 31). The Night A Computer Predicted The Next President. *NPR.org*. Retrieved from http://www.npr.org/sections/alltechconsidered/2012/10/31/163951263/the-night-a-computer-predicted-the-next-president

Hindman, M. (2012). Personalization and the Future of News. *EUI Working Papers. RSCAS 2012/56. Robert Schuman Centre for Advanced Studies. Global Governance Programme-29*. Retrieved from http://hdl.handle.net/1814/24296

Hovy, D. Spruit, S.L. (2016). The social impact of natural language processing. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, volume 2, pages 591–598.

IEEE, editor. (2017). Ethically Aligned Design: A Vision For Prioritizing Wellbeing With Artificial Intelligence And Autonomous Systems. Overview, version II. *IEEE – advanced Technology for Humanity*. Retrieved from http://standards.ieee.org/develop/indconn/ec/ead_brochure_v2.pdf

IFJ (1954). IFJ Declaration of Principles on the Conduct of Journalists. IFJ.org. Retrieved from http://www.ifj.org/about-ifj/ifj-code-of-principles/

Jacquette, D. (2007). *Journalistic ethics: moral responsibility in the media*. Upper Saddle River: Pearson Education.

Jensen, M.H. (2016a, January 25). VG med robot: 428 artikler ser ut som én. *Journalisten*. Retrieved from https://journalisten.no/2016/01/vg-tar-robotene-i-bruk

Jensen, M.H. (2016b, June 17). Personaliserte nyheter bekymrer nordmenn. *Journalisten*. Retrieved from https://journalisten.no/2016/06/personaliserte-nyheter-bekymrer-nordmenn

Johansen, G. S. (2016, April 13). Avspark i mai: NTB dekker Eliteserien med robotjournalist. *Journalisten*. http://journalisten.no/2016/04/robot-inntar-norsk-sportsjournalistikk.

Jung, J., Song, H., Kim, Y., Im, H., Oh, S. (2017). Intrusion of software robots into journalism: The public's and journalists' perceptions of news written by algorithms and human journalists. *Computers in Human Behavior, 71*, 291-298. https://doi.org/10.1016/j.chb.2017.02.022

van der Kaa, H. A. J., & Krahmer, E. J. (2014). Journalist versus news consumer: The perceived credibility of machine written news. *In Proceedings of the Computation+Journalism conference New York*. Retrieved from https://pure.uvt.nl/portal/files/4314960/cj2014_session4_paper2.pdf

Kammer, A. (2013). The mediatization of journalism. MedieKultur, 54, 141-158. http://dx.doi.org/10.7146/mediekultur.v29i54.17385

Karlsen, J., Stavelin, E. (2014). Computational Journalism in Norwegian Newsrooms, *Journalism Practice, 8, 1*, 34-48. http://dx.doi.org/10.1080/17512786.2013.813190

Kent, T. (2015, March 4). An ethical checklist for robot journalism. *Mediashift*. Retrieved from http://mediashift.org/2015/03/an-ethical-checklist-for-robot-journalism/

Kline, S.J., Rosenberg, N. (1986). An Overview of Innovation. In Landau, R., Rosenberg, N. (Eds.), *The Positive Sum Strategy: Harnessing Technology for Economic Growth (pp. 275-304).* Washington, DC: National Academy Press.

Kraemer, F., van Overveld, K., Peterson. M. (2011). Is there an ethics of algorithms? *Ethics and Information Technology, 13, 3*, 251–260.

Kvale, S., Brinkmann, S. (2009). *InterViews: Learning the Craft of Qualitative Research Interviewing*. SAGE publications.

Küng, L. (2013). Innovation, Technology and Organisational Change. Legacy Media's Big Challenges. An Introduction. In Storsul, T., Krumsvik, A.H. (Eds.), *Media Innovations. A Multidisciplinary Study of Change (pp. 9-12).* Göteborg: Nordicom.

Latar, N.L. (2015). The robot journalist in the age of social physics: the end of human journalism? In Einav, G. (Ed.) The new world of transitioned media (65-80). New York: Springer.

Lavie, T., Sela, M., Oppenheim, I., Inbar, O., Meyer, J. (2010). User attitudes towards news content personalization. *International Journal of Human-Computer Studies, 68, 8*, 483-495. http://dx.doi.org/10.1016/j.ijhcs.2009.09.011

Levy, S. (2012, April 24). Can an algorithm write a better news story than a human reporter? Wired. Retrieved from https://www.wired.com/2012/04/can-an-algorithm-write-a-better-news-story-than-a-human-reporter/

Lewis-Kraus, G. (2016. December 14). The Great A.I. Awakening. *The New York Times Magazine*. Retrieved from https://www.nytimes.com/2016/12/14/magazine/the-great-ai-awakening.html

Linden, C.-G. (2017a). Decades of Automation in the Newsroom, *Digital Journalism, 5, 2,* 123-140. http://dx.doi.org/10.1080/21670811.2016.1160791

Linden, C.-G. (2017b). Algorithms for journalism: The future of news work. *The Journal of Media Innovations, 4, 1,* 60-76. http://dx.doi.org/10.5617/jmi.v4i1.2420

Manyika, J, Chui, M., Miremadi, M., Bughin, J., George, K., Willmott, P., Dewhurst, M. (2017, January). *A future that works: automation, employment, and productivity*. Retrieved from http://www.mckinsey.com/global-themes/digital-disruption/harnessing-automation-for-a-future-that-works

Marconi, F. (2016, April 27). Imagining a newsroom powered by artificial intelligence. *Tow Center for Digital Journalism*. Retrieved from https://towcenter.org/imagining-a-newsroom-powered-by-artificial-intelligence/

Marconi, F., Siegman, A., Machine Journalist (2017). The future of augmented journalism. A guide for newsrooms in the age of smart machines. Retrieved from https://insights.ap.org/uploads/images/the-future-of-augmented-journalism_ap-report.pdf

McBride, K., Rosenstiel. T. (2013). New Guiding Principles for a New Era of Journalism. In The New Ethics of Journalism, edited by McBride, K., Rosenstiel, T., 1–6. Thousand Oaks, CA: CQ Press.

McDonald, D., Bolc, L. (1988). *Natural Language Generation Systems*. NY: Springer.

Michalsen, G.L. (2016, May 18). Norges første robotjournalist sendte ut fotballreferater. *Aftenposten*. Retrieved from https://www.aftenposten.no/kultur/i/L06l9/Norges-forste-robotjournalist-sendte-ut-fotballreferater

Micó, J.L., Masip, P., Domingo, D. (2013). To wish impossible things: Convergence as a process of diffusion of innovations in an actor-network. International Communication Gazette, 75, 1, 118-137. https://doi.org/10.1177/1748048512461765

Miller, R. (2015, January 29). AP's 'robot journalists' are writing their own stories now. *The Verge*. Retrieved from https://www.theverge.com/2015/1/29/7939067/ap-journalism-automation-robots-financial-reporting

Montal, T., Reich, Z. (2017). I, Robot. You, Journalist. Who is the Author? *Digital Journalism, 5, 7,* 829-849. http://dx.doi.org/10.1080/21670811.2016.1209083

Murray-Hundley. D. (2017, July 14). The grumpy guide to AI, machine learning and deep learning. *Techworld*. Retrieved from https://www.techworld.com/tech-innovation/grumpy-guide-ai-machine-learning-deep-learning-3661324/

Negroponte, N. (1995). *Being Digital*. NY: Alfred A. Knopf.

Nelson, R.R., Winter, S.G. (1982). *An Evolutionary Theory of Economic Change*. Cambridge: The Belknap Press Of Harvard University Press.

Newton, C. (2017, July 19). Google introduces the feed, a personalized stream of news on iOS and Android. *The Verge.* Retrieved from https://www.theverge.com/2017/7/19/15994156/google-feed-personalized-news-stream-android-ios-app

Ombelet, P.-J.; Kuczerawy, A., Valcke, P. (2016). Employing Robot Journalists: Legal Implications, Considerations and Recommendations. *Proceedings of the 25th International Conference Companion on World Wide Web, 731-736*. https://doi.org/ 10.1145/2872518.2890093.

Oremus, W. (2014, March 17). The First News Report on the L.A. Earthquake Was Written by a Robot. *Slate*. Retrieved from http://www.slate.com/blogs/future_tense/2014/03/17/quakebot_los_angeles_times_robot_journalist_writes_article_on_la_earthquake.html

Otterdal, M.S., Ruud, G.T. (2017). Autostory. Et essay om hvordan automatisering forandrer journalistikken. Oslo: Cappelen Damm.

Pariser, E. (2011). *The Filter Bubble: What The Internet Is Hiding From You*. Penguin UK.

Pavlik, J. V. (2013). Innovation and the future of journalism. Digital journalism, 1, 2, 181–193. http://dx.doi.org/10.1080/21670811.2012.756666

Pohl, J. (2015). Artificial Superintelligence: Extinction or Nirvana? Presented on InterSymp-2015 – 27th International Conference on Systems Research, Informatics and Cybernetics (IIAS), at Baden-Baden, Germany. Retrieved from https://www.researchgate.net/profile/Jens_Pohl/publication/281748315_Artificial_Superintelligence_Extinction_or_Nirvana/links/55f723c908aeafc8abfa21ca.pdf

Reiter, E. (2017). What is NLG, and how does it relate to NLP and other forms of AI? Arria.com. Retrieved from http://blog.arria.com/what-is-nlg-and-how-does-it-relate-to-nlp-and-other-forms-of-ai

Reuters Institute for the Study of Journalism. (2016). *Reuters Institute Digital News Report 2016*. Oxford: Newman, N., Fletcher, R., Levy, D.A.L., Nielsen, R.K.

Riedl. M.O. (2016). Computational Narrative Intelligence: A Human-Centered Goal for Artificial Intelligence. *Proceedings of the CHI 2016 Workshop on Human Centered Machine Learning, San Jose, California, 2016*. Retrieved from http://www.cc.gatech.edu/~riedl/pubs/chi-hcml16.pdf

Rogers, E. M. (2003). *Diffusion of Innovations* (5th ed.). New York: Free Press.

Rogers, E. M. (2004). A Prospective and Retrospective Look at the Diffusion Model. *Journal of Health Communication, 9, 1*, 13-19. http://dx.doi.org/10.1080/10810730490271449

Rotman, D. (2013, June 12). How Technology Is Destroying Jobs. *MIT Technology Review*. Retrieved from https://www.technologyreview.com/s/515926/how-technology-is-destroying-jobs/

Rubin, H.J., Rubin, I.S. (2005). *Qualitative interviewing: The art of hearing data*. Thousand Oaks, CA: Sage.

Ryan, B., Gross, N. C. (1943). The diffusion of hybrid seed corn in two Iowa communities. *Rural Sociology, 8, 1*, 15-24. Retrieved from http://chla.library.cornell.edu/cgi/t/text/pageviewer-idx?c=chla;rgn=full%20text;idno=5075626_4294_001;view=image;seq=17

Schumpeter, J.A. (1934). The Theory of Economic Development: An Inquiry into Profits, Capital, Credit, Interest, and the Business Cycle. *Harvard Economic Studies, 46*. Harvard University Press.

Schumpeter, J.A. (1942). *Capitalism, Socialism and Democracy*. NY: Harper Perennial.

Shalev-Shwartz, S., Ben-David, S. (2014). *Understanding Machine Learning: From Theory to Algorithms.* Cambridge University Press.

Smiley, C., Schilder, F., Plachouras, V., Leidner. J.L. (2017). Say the Right Thing Right: Ethics Issues in Natural Language Generation Systems. Proceedings of the First Workshop on Ethics in Natural Language Processing, pages 103–108, Valencia, Spain, April 4th, 2017. Retrieved from http://www.ethicsinnlp.org/workshop/EthNLP-2017.pdf#page=115

Smith, B., & Linden, G. (2017). Two Decades of Recommender Systems at Amazon. com. *IEEE Internet Computing, 21, 3*, 12-18. Retrieved from http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=7927889

Spielkamp, M. (2017, June 12). Inspecting Algorithms for Bias. MIT Technology Review. Retrieved from https://www.technologyreview.com/s/607955/inspecting-algorithms-for-bias/

Storsul, T., Krumsvik, A.H. (2013). What is Media Innovation? In Storsul, T., Krumsvik, A.H. (Eds.), *Media Innovations. A Multidisciplinary Study of Change (pp. 9-12)*. Göteborg: Nordicom.

Sundar, S.S., Marathe, S.S. (2010). Personalization versus Customization: The Importance of Agency, Privacy, and Power Usage. *Human Communication Research, 36, 3*, 298-322. http://dx.doi.org/10.1111/j.1468-2958.2010.01377.x

Thurman, N., Schifferes, S. (2012). The future of personalization at news websites. *Journalism Studies, 13, 5-6*, 775-790. http://dx.doi.org/10.1080/1461670X.2012.664341

Tidd, J., Bessant, J. and Pavitt, K. (2005). *Managing Innovation. Integrating technological, market and organizational change (Vol. 3)*. West Sussex: John Wiley & Sons.

Tran, M. (2006, August 18). Robots write the news. *The Guardian*. Retrieved from https://www.theguardian.com/news/blog/2006/aug/18/robotswriteth

Tsvelik, A. (2012). *Life in The Impossible World*. St.-Petersburg: Ivan Limbakh Publishing.

Turing, A.M. (1950). Computing Machinery and Intelligence. *Mind, 59, 236,* 433-460. Retrieved from http://www.jstor.org/stable/2251299

Tørresen, J. (2013). *Hva er kunstig intelligens*. Oslo: Universitetsforlaget.

Vaagan, R.W., Barland, J. (2015). *Entreprenørskap og ledelse i media*. Oslo: Cappelen Damm.

Vincent, J. (2016, March 24). Twitter taught Microsoft's AI chatbot to be a racist asshole in less than a day. *The Verge*. Retrieved from https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist

Vinton, K. (2017, July 27). Jeff Bezos Overtakes Bill Gates To Become World's Richest Man. *Forbes*. Retrieved from https://www.forbes.com/sites/katevinton/2017/07/27/jeff-bezos-overtakes-bill-gates-to-become-worlds-richest-man/#5e962c13791c

Weeks, L. (2014). Media Law and Copyright Implications of Automated Journalism. *Journal of intellectual property and entertainment law, 4, 1*, 67-94. Retrieved from http://jipel.law.nyu.edu/vol-4-no-1-3-weeks/

Wright, A. (2015). Algorithmic Authors. Communications of the ACM, 58, 11, 12-14. http://dx.doi.org/10.1145/2820421
Young, M. L., Hermida, A. (2014). From Mr. and Mrs. Outlier To Central Tendencies. *Digital Journalism* (online first), http://dx.doi.org/10.1080/21670811.2014.976409.

Zion, L., Craig, D. (2014). *Ethics for Digital Journalists. Emerging Best Practices*. New York: Routledge.

Interviews:

Bakken, J., Newsroom developer, *Verdens Gang AS*, Oslo, Norway. Interview on 10 August 2016 at *VG* headquarters in Oslo.

Barland, J., Associate professor, *Norwegian University of Science and Technology*, Gjøvik, Norway. Interview on August 16, 2016 at the University of Oslo.

Bense, H., CEO, *Textomatic*, Dortmund, Germany. Skype interview on 26 September 2017.

Feulner, F., Chief visionary officer, *AX Semantics*, Stuttgart, Germany. Interview on February 23, 2017 at Hochschule der Medien, Stuttgart, Germany.

Grimnes, G.A., Chief technical officer at *Orbit* (Bonn, Germany) and *Bakken & Bæck*, Oslo 2016. Skype interview on 22 November 2016.

Johannesson, H., Chief sports editor, *MittMedia*, Gävle, Sweden. Skype interview on 15 September 2017.

Karlsson, S., CEO, *United Robots*, Malmö, Sweden. Skype interview on 22 September 2017.

Ruud, G.T., Head of the development, *NTB*, the co-author of "Autostory". Interview on 12 October 2017 at *NTB* headquarters in Oslo, Norway.

Śliwowska, E., Director of commercial operations, *Retresco*, Berlin, Germany. Skype interview on 21, 22 September 2017.

Volan, I., Readership development editor, *Dagens Næringsliv* (at the time of the interview Director of innovation and newsroom development, *Adresseavisen*, Trondheim, Norway). Skype interview on 16 August 2016.

Aabech, M., News developer, *NTB*, Oslo, Norway. Interview on 29 August 2016 at NTB headquarters in Oslo, Norway. A second follow-up interview when also Geir Terje Ruud above was present, was conducted on 12 October 2017 at NTB headquarters in Oslo.

Appendix

Appendix 1 – NSD receipt

Anders Fagerjord
Institutt for medier og kommunikasjon Universitetet i Oslo
Postboks 1093 Blindern
0317 OSLO

Vår dato: 22.06.2016 Vår ref: 48807 / 3 / AMS Deres dato: Deres ref:

TILBAKEMELDING PÅ MELDING OM BEHANDLING AV PERSONOPPLYSNINGER

Vi viser til melding om behandling av personopplysninger, mottatt 30.05.2016. Meldingen gjelder prosjektet:

48807	*Robotic journalism. The coexistence of algorithmic "writers" and human journalists in modern newsrooms*
Behandlingsansvarlig	*Universitetet i Oslo, ved institusjonens øverste leder*
Daglig ansvarlig	*Anders Fagerjord*
Student	*Ekaterina Pashevich*

Personvernombudet har vurdert prosjektet og finner at behandlingen av personopplysninger er meldepliktig i henhold til personopplysningsloven § 31. Behandlingen tilfredsstiller kravene i personopplysningsloven.

Personvernombudets vurdering forutsetter at prosjektet gjennomføres i tråd med opplysningene gitt i meldeskjemaet, korrespondanse med ombudet, ombudets kommentarer samt personopplysningsloven og helseregisterloven med forskrifter. Behandlingen av personopplysninger kan settes i gang.

Det gjøres oppmerksom på at det skal gis ny melding dersom behandlingen endres i forhold til de opplysninger som ligger til grunn for personvernombudets vurdering. Endringsmeldinger gis via et eget skjema, http://www.nsd.uib.no/personvern/meldeplikt/skjema.html. Det skal også gis melding etter tre år dersom prosjektet fortsatt pågår. Meldinger skal skje skriftlig til ombudet.

Personvernombudet har lagt ut opplysninger om prosjektet i en offentlig database, http://pvo.nsd.no/prosjekt.

Personvernombudet vil ved prosjektets avslutning, 01.07.2017, rette en henvendelse angående status for behandlingen av personopplysninger.

Vennlig hilsen

Kjersti Haugstvedt

Anne-Mette Somby

Kontaktperson: Anne-Mette Somby tlf: 55 58 24 10

Dokumentet er elektronisk produsert og godkjent ved NSDs rutiner for elektronisk godkjenning.

Personvernombudet for forskning

Prosjektvurdering - Kommentar

Prosjektnr: 48807

Ifølge prosjektmeldingen skal utvalget informeres muntlig om prosjektet og samtykke til deltakelse. For å tilfredsstille kravet om et informert samtykke etter loven, må utvalget informeres om følgende:

- hvilken institusjon som er ansvarlig
- prosjektets formål/problemstilling
- hvilke metoder som skal benyttes for datainnsamling
- hvilke typer opplysninger som samles inn
- at opplysningene behandles konfidensielt og hvem som vil ha tilgang
- at det er frivillig å delta og at man kan trekke seg når som helst uten begrunnelse
- dato for forventet prosjektslutt
- at data anonymiseres ved prosjektslutt
- hvorvidt enkeltpersoner vil kunne gjenkjennes i den ferdige oppgaven
- kontaktopplysninger til student/veileder.

Personvernombudet legger til grunn at forsker etterfølger Universitetet i Oslo sine interne rutiner for datasikkerhet. Dersom personopplysninger skal lagres på privat pc/mobile enheter, bør opplysningene krypteres tilstrekkelig.

Det oppgis at personopplysninger skal publiseres. Personvernombudet legger til grunn at det foreligger eksplisitt samtykke fra den enkelte til dette. Vi anbefaler at deltakerne gis anledning til å lese igjennom egne opplysninger og godkjenne disse før publisering.

Forventet prosjektslutt er 01.07.2017. Ifølge prosjektmeldingen skal innsamlede opplysninger da anonymiseres. Anonymisering innebærer å bearbeide datamaterialet slik at ingen enkeltpersoner kan gjenkjennes. Det gjøres ved å:
- slette direkte personopplysninger (som navn/koblingsnøkkel)
- slette/omskrive indirekte personopplysninger (identifiserende sammenstilling av bakgrunnsopplysninger som f.eks. bosted/arbeidssted, alder og kjønn)