D2.2

Initial white paper on the social, economic, and political impact of media AI technologies

Project Title	AI4Media - A European Excellence Centre for Media, Society and Democracy
Contract No.	951911
Instrument	Research and Innovation Action
Thematic Priority	H2020-EU.2.1.1. - INDUSTRIAL LEADERSHIP - Leadership in enabling and industrial technologies - Information and Communication Technologies (ICT) / ICT-48-2020 - Towards a vibrant European network of AI excellence centres
Start of Project	1 September 2020
Duration	48 months

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

info@ai4media.eu www.ai4media.eu

Deliverable title	Initial white paper on the social, economic, and political impact of media AI technologies
Deliverable number	D2.2
Deliverable version	1.0
Previous version(s)	-
Contractual date of delivery	28 February 2022
Actual date of delivery	28 February 2022
Deliverable filename	AI4Media_D2.2_final.docx
Nature of deliverable	Report
Dissemination level	Public
Number of pages	124
Work Package	WP2
Task(s)	T2.2
Partner responsible	UvA
Author(s)	Anna Schjøtt Hansen (UvA) Natali Helberger (UvA), Tobias Blanke (UvA), Rasa Bočytė (NISV)
Editor	Anna Schjøtt Hansen (UvA)
EC Project Officer	Evangelia Markidou

Abstract	Deliverable D2.2 *'Initial white paper on the social, economic, and political impact of media AI technologies'* provides an overview of the some of the core discussions of AI for media from a media studies/social science perspective, identifying the main potentials and challenges connected with AI applications across the media cycle. These concrete challenges are then discussed more widely in terms of how they might impact society (socially, economically, or politically) and what mitigative measures will be important to ensure the use of AI in the media sector remain responsible and that it positively affects society. The whitepaper is based on a thorough literature review of academic journals published by scholars within the field of humanities, social science, media and legal studies as well as reports developed either with a specific focus on AI in the media sector or with a broader outlook on AI in society. Furthermore, a range of examples of concrete AI applications are described to provide context for the reader and some of the mediated responses to the applications.
Keywords	Artificial Intelligence, Media, Impacts (economic, social, political), Responsible AI, Algorithmic accountability, Audience measurement systems, Diversity, Content production, Content distribution, Content Moderation, Content search, Fact-checking, Personalisation, Subscription models, Advertisement, Audiovisual archives, Bias, Discrimination, Media

	independence, Labour displacement, Privacy, Transparency, Accountability, Liability, Content manipulation, Mis/Dis-information.

Copyright

Contributors

NAME	ORGANISATION
Anna Schjøtt Hansen	UvA
Natali Helberger	UvA
Tobias Blanke	UvA
Rasa Bočytė	NISV

Peer Reviews

NAME	ORGANISATION
Rasa Bočytė / Johan Oomen	NISV
Lidia Dutkiewicz	KUL

Revision History

VERSION	DATE	REVIEWER	MODIFICATIONS
0.1	26/11/2021	Natali Helberger, Tobias Blanke	Initial feedback on structure and content in comment form.
0.2	18/01/2022	Natali Helberger, Tobias Blanke, Henning Müller, Rasa Bočytė	Direct editing, comment feedback and direct contributions to assigned sections.
0.3	14/02/2022	Rasa Bočytė / Johan Oomen, Lidia Dutkiewicz	Direct editing, comment feedback and throughout the entire document.
1.0	28/02/2022	Anna Schjøtt Hansen	Final version ready for submission

Index of Contents

Index of Figures

1 Executive Summary

Deliverable D2.2 *'Initial white paper on the social, economic, and political impact of media AI technologies'* provides an overview of the core discussions around the subject of AI for media from a media studies / social science perspective, identifying the main potentials and challenges connected with AI applications across the media cycle. These concrete challenges are then discussed more widely in terms of how they might impact society (socially, economically, or politically) and what mitigative measures will be important to ensure that the use of AI in the media sector remains responsible and that it positively affects society.

The white paper is intended as a 'reader's guide' for media professionals, AI developers working in the media sector and researchers interested in AI and media. It is based on a thorough literature review of academic journals published by scholars within the field of humanities, social science, media studies and to a degree legal studies, as well as reports developed either with a specific focus on AI in the media sector or with a broader outlook on AI in society. Furthermore, a range of examples of concrete AI applications is described to provide context for the reader and some of the mediated responses to the applications.

The following key points of consideration for the industry, policy makers and researchers regarding the future of AI for media were identified:

- **The need for more domain-specific, social and/or cultural expertise in the development process of AI systems for media.** All AI projects in the media sector should strive for diversity in the team (e.g., in terms of backgrounds, ethnicities or gender) to ensure that the decisions made regarding datasets, classification or metrics are made on a well-founded and reflective basis. Critically, domain knowledge should be prioritised together with social and cultural knowledge in qualifying these decisions.

- **The need to foster support, tools, and resources for responsible AI practices in the media sector.** Over the last years more awareness has been gained about the need for work with the biases of AI systems, now there is a need to develop concrete tools to support the media organisations in their work as well as foster support and resources for responsible AI practices – something that is challenged with the constant call for optimisation and efficiency within media organisations.

- **The need for new best practices on how to produce just AI systems in the media sector.** Currently, the examples of AI projects promoting data justice are scarce. If the sector is to begin a conversation on ways to achieve this, examples of best practices will be needed. This could be in the form of industry research collaborations.

- **The need for regulation that supports and fosters responsible AI practices in the media sector, rather than attempt to constrain the use.** Often regulatory measures are focused on banning dangerous uses of technologies, there will be a need for policies that rather than constraining provides incentives to adopt responsible AI practices in organisations, because as seen this is difficult with the current conditions in the sector.

- **The need for domain-specific, open-source and non-commercial datasets for training AI systems.** As many AI projects today rely on open-source and 'golden standard' datasets created without consideration for cultural and societal sensitivities and that have proven to induce certain unwanted biases. For the media sector to mitigative the negative effects of such biases and instead induce 'good' or more just biases, domain specific open-source datasets are needed, where there has been time and resources for thorough considerations of what biases to induce by a diverse team.

- **The need for responsible, domain-specific infrastructures to support responsible AI practices.** Due to the high reliance on commercialised and platform infrastructures in the development of AI in the media sector, it will be important to develop alternative data and content infrastructures that perhaps better accommodate the European values and are specific to the media sector.

- **The need for more engagement with media asset management (MAMs) vendors in the audiovisual sector.** This will be important to ensure that they offer more flexible, agile, and modular solutions that respond to the needs of the sector and the recent technological advancements in AI will be needed in the future.

- **The need for best practices and policies of 'diversity by design'.** Currently, limited knowledge and best practice exists on how to make the evaluation of whether, for example, a recommender system is successful – not only in a commercial sense. New best practices on how to make such decisions without benchmarking with, for example, purely commercial actors and how to include domain-specific measures of diversity in the projects (e.g., filling the gaps of user knowledge etc.), are needed (e.g., through concrete policies on diversity by design). Furthermore, there is a need for big media companies to be first movers and set the example for the rest of the sector and push this responsible development.

- **The need for a critical awareness of economic 'patrons' of the media sector and how they affect the development in the media sector.** Currently, limited research exists on the role of 'media patrons' and how they affect the future of the media sector. It will be important that more research is conducted, but also that researchers in fact can get access to these processes, as that is currently highly difficult.

- **The need for funding schemes oriented in EU values.** To counteract the growing role of (US-based) platforms in stimulating development, it will be important to develop similar funding schemes that better encompass EU values and the societal function of media.

- **The need for funding schemes and initiatives focusing on media diversity.** It will be important to counteract the trend in private funding identified by Fanta (2020) where established media organisations remain the main beneficiaries of funding for innovation. To not further the increasing competitive divides in the media sector, funding should be specifically oriented towards furthering media diversity.

- **The need for an increased focus on global AI divides and their consequences.** In general, more knowledge is needed on the severity of the AI divide between the global north and south. It will be important to explore the extent of the issue and its implications further.

- **The development of AI models for diverse languages or adaptive models.** To improve the overall access to AI benefits, AI models for large foreign and minority languages should be developed together with adaptive models that can be more efficiently reused for other languages. This could also produce new insights and highlight cultural biases/differences, which in turn could be used to make AI models for the more common languages more accurate.

- **The need for more research and policies addressing potential displacement patterns resulting from AI.** As the increased reliance of AI might result in certain jobs disappearing (e.g., routine tasks) in the media sector as well as across other sectors, providing a societal problem of unemployment. It will, therefore, be important that societal mechanisms and policies are developed to handle the citizens who will be left jobless and in need of specific upskilling.

- **The need for an increased focus on data and AI in media education.** The changes in the media professions also require action from the educational sector who must support students in developing the right skills for the labour market, including increased skills in data and in understanding how AI systems work as well as awareness of the problems connected to these technologies, as misconceptions of 'algorithmic objectivity' still flourish.

- **The need for more research on AI is changing labour conditions and growing power asymmetries in the media sector.** It will be important to understand how the introduction of AI is enhancing already increasing workplace asymmetries, for example, through the use of performances measurements and with what impacts on the individual and society and how it is producing shifts of power within these organisations, valorising technical staff and their approaches.

- **The need for meaningful oversight for media professionals.** There continues to be a strong emphasis on keeping a 'human-in-the-loop, both in practice and in policies, for most AI applications, to ensure control and oversight. However, this ambition is challenged by the fact that many of these systems remain difficult to have oversight over due to their opacity and scale. To solve this problem and fulfil this ambition it will be important to support the development of 'explainable AI' and human interface design.

- **The need for more best practices of responsible data practices in the media sector.** As the extensive use of data continues to grow in the media sector, it will be vital that new best practices are developed to support responsible data strategies that protect the rights of the individual.

- **The need for best practices and policies regarding disclosure of AI systems for the media sector.** As the question of who produced or curated an article is no longer limited to, for example, journalists, editors, and producers, it will be vital to introduce new guidelines on how to disclose the utilisation of AI in these processes to protect the individual's right to transparency.

- **The need for explainable and transparent AI solutions that can help users understand how AI systems work and makes decision.** As users increasingly are partly serviced by AI systems in their media experience, it is important that they have access to understandable explanations of what the system does and on the basis on what data to uphold their right to, for example, object to the way the decision was made (i.e., agency to act).

- **The need for clearer regulation and guidelines on the liability question regarding AI.** There is a need to help media organisations navigate the liability question that arises from the use of AI systems.

- **The need for mitigative and adaptive AI systems to counteract misinformation.** To protect the legitimacy of media organisations and the integrity of the online deliberative spaces, it will be important to develop AI systems to assist in content moderation and fact checking efforts. These must be highly adaptive to be effective and counter act adversarial tactics by groups who spread misinformation.

- **The need for more transparency in moderation systems and AI fact checking systems.** Currently the AI systems used to identify misinformation on social media platforms remain untransparent in their workings and the people who experience consequences do not always have access to a satisfying explanation of why, for example, their profile was deleted or to a complaint mechanism. As many fact checkers are today part of strategic partnerships with Facebook, the need to be transparent will become even more important to sustain legitimacy in these institutions that now serve and important societal function.

- **The need for more knowledge on fact checking as a social practice and its effects in the deliberate landscape.** As fact checking becomes an important societal function, it will be important to gain more in-depth knowledge in how they construct 'factual' accounts as well as what the consequences of potentially countering epistemologies of the truth might mean for the deliberative space and societal polarisation.

2 Introduction and scope

Artificial intelligence (AI) is today increasingly part of decision-making processes on everything from what treatment patients at a hospital should receive to what movies will be recommended on the front page of Netflix. While AI has a more than 70-year long history with its origin in the 1940s as part of the second world war code-breaking efforts, it is only within the last decade that AI has gained political and public attention beyond fleeting famous moments like when IBM's Deep Blue beat grandmaster Garry Kasparov in chess (IBM, 2012). This newly gained attention is a result of AI moving out of the R&D labs and into the 'real world'. A transition that has been made possible by the increasing amount of accessible digital data, to the point where the amount of data pertaining to each individual has never been more extensive, nor so individuated and relationally structured (Athique, 2018).

This is also the case in the media sector, where adaptation of AI in media production and distribution has been rapidly growing over the last five years. The 2019 report by LSE director, Charlie Beckett, for example, show that four in ten news organisations surveyed already had deployed AI in their day-to-day practices (based on a total of 71 organisations) (Beckett, 2019). While the report concludes that AI currently is 'additional, supplementary and catalytic, not yet transformational' (p. 12) to the journalistic practices, the future impacts of AI on the media profession and society remain uncertain. This uncertainty has resulted in a host of questions and concerns regarding what the social, economic, and political implications of the turn to AI in the media sector might be, which is the starting point for this whitepaper. We recognise that many of these implications are not solely due to AI, but have often become more intensified or visible with the development of AI.

The goal of the whitepaper is to map these ongoing discussions as they unfold in academic literature and industry reports, identify knowledge gaps and state-of-the-art. Concretely, the whitepaper will highlight current trends in AI applications in the media sector and their potentials – for the industry or society – but also discuss the critical questions and concerns that have been raised regarding these applications. The latter will also help to shed light on what mitigative measures will be needed to ensure that the use of AI for media is developed and applied in a responsible manner.

The aim of the whitepaper is to function as a 'reader's guide' for media professionals, AI developers working in the media sector and researchers interested in AI and media, which introduces the readers to some of the core discussions of AI for media from a media studies perspective and guide them towards relevant in-depth explorations of the different AI applications discussed. Thereby, providing a broader overview of the state of the discussion and best practices, while recognising that there will be applications we do not discuss and that we in this format cannot account for all the complexities related to each of the discussions raised. Instead, in Appendix A, readers can find a suggested further reading list that points to more sector and topic-specific resources. Equally, it is also important to note that the whitepaper has a predominantly ‘western’ focus, as most of the literature deals with cases studies from the US and European countries at the forefront of digital innovation, and that the discussions would most likely be different if we investigated other regions. Looking towards, for example, Russia where governmental rather than commercial pressures have been connected with transformations in the media sector. It would, therefore, be relevant for further work to explore the global differences in the discussions regarding AI. However, as AI4Media is a European Union project, this more euro-centric perspective has been chosen for this specific whitepaper.

As this field is also a rapidly developing one, where new applications of AI are constantly appearing along with new research on what the societal implications of these applications might be, this whitepaper can only ever be a snapshot of these discussions. However, we will use this whitepaper, together with other activities or reports produced in the AI4Media project, as the foundation for a ‘living document’ as part of the AI4Media Observatory. The whitepaper will also be updated again in 2023, where we will include a new section that will focus on specific challenges identified in part one, which then are discussed during industry workshops. This will be described more in the methodology section below.

2.1 Methodology and structure

This whitepaper is divided into three parts. The first (chapter 3) consists of a **thorough literature review**, including both industry and think tank reports as well as scholarly literature. The aim of this part is, as outlined above, to provide an overview of **the state of discussion of AI for media**, including the specific concerns raised in these discussions (technical, political, social, or economic).

The second part is a discussion (chapter 4), which identifies the **general impacts** that are highlighted throughout the review, with the aim of understanding the unique challenges that the media sector phases in the future. This also allows us to distil a range of specific recommendations for the media sector and point to the areas where more knowledge is needed.

The third part (chapter 5) will be added in the second version of this whitepaper to be published in December 2023 (Deliverable 2.5), which will be based on **several industry workshops to be carried out in the following years.** The aim of these workshops is to inform and validate the second version of this whitepaper with both specific case studies (best practices) and through discussions of the concrete challenges (technical, economic, social, or legal) the invited participants have encountered when working with AI within their respective organisations. These insights will, therefore, complement the literature review with concrete industry insights and ensure that emerging discussions are included.

2.1.1 Literature review: methodological considerations

As the literature review is the core component in this version of the whitepaper, the methodological consideration behind it will be briefly outlined here. As already described above, the literature review builds on insights from both industry reports and papers from peer-reviewed academic journals. This whitepaper does not claim to provide a comprehensive review of all literature on AI or AI for media, rather it is a selective review that focuses specifically on reports and papers that explore the impacts, potentials, and challenges of AI in media organisations.

Both the reports and the academic journal papers were identified through a snowballing method where the involved humanities and social science experts from the University of Amsterdam, with the assistance of other experts from the AI4Media Consortium, identified core papers and reports, which they considered good starting points for the scholarly literature. These were predominately from the field of media studies (encompassing both journalism and audience studies) as well as other relevant papers from the broader fields of humanities, social science and to a degree legal studies. From these selected core papers and reports, more papers were identified based on their bibliographies and when certain cited texts continued to appear, these were equally used as core texts to identify further readings. The list of papers and reports was also iteratively reviewed during the process to ensure that important papers or reports were not left out. Furthermore, topic-specific literature searches were conducted when topics seemed underrepresented in the identified literature to make sure this was not a bias in the collected material and their infield sourcing practices, but in fact representative of how much knowledge is available on the topic. For all papers and reports three criteria of relevance were determined: 1) they should explicitly deal with AI (either specific technologies or more widely), 2) they specifically related AI to media practices (e.g., archiving, production, or distribution), and 3) they contributed with insights regarding the societal impacts (economic, political, or social) of AI on the media industry (profession) or the societal role of media institutions.

It is important to note that this whitepaper does not emerge in a vacuum, rather it is part of a growing landscape of concerns with AI represented in a range of reports that explore the impact of AI in specific sectors including the media sector but also more in general in relation to overarching concerns such as inequality, bias, rights, and ethics. We have, therefore, also selected a range of more 'general' AI report and papers, that address the impacts of AI either more widely or with respect to specific technologies, societal changes, or other sectors (an overview of these can be found in Appendix B). These are used in some cases in the first part to qualify discussions, but specifically in the second part to introduce broader perspectives on the challenges faced in the media sector and as a comparison to highlight how the challenges might be in some cases easier for the media sector and in other perhaps more dire. The 'general' AI reports were selected based on an identification of core institutions that are highly influential in the field of ensuring responsible AI, which here includes AI for Now Institute, the Ada Lovelace

Institute, Alan Turing Institute, and the European Union (through the commissioning of reports). An overview can be found in Appendix B.

2.2 Core concepts

In the following, we briefly define the two core concepts of this report, namely media and artificial intelligence (AI), and how we operationalise them in the context of this whitepaper. This clarification is important as both the concepts of media and AI are used and defined in many ways. In this section, we do not provide an in-depth discussion of the many definitions, their similarities, differences, and the implications of this definitional unclarity, but focus on clarifying our use.

2.2.1 Media organisations

The term media is a widely used term in public discourse, encompassing all forms of mass communication across print news, broadcasters, and the internet (e.g., social media) and sometimes even entertainment (e.g., Netflix) or TV more in general. In this whitepaper a narrower understanding of media is adopted, focusing on what can be characterised as traditional mass media outlets, which includes local, regional, national news organisations and Public Service Media (PSM) as well as media archives (often part of the responsibilities of PSMs) as these represent what is traditionally viewed as the core democratic institutions – `the lifeblood of democracy` (Fenton, 2010). In disseminating their applications of AI and the societal impacts, the whitepaper will look across the practices and types of content produced by such organisations, including their role in imagining, producing, distributing, and archiving media content. Due to the growing influence of social media platforms as distributors of information in society, such platforms are also considered in the following, but limited to their role in distributing information (news) and the following dissemination of news (see e.g., Nielsen and Schrøder, 2014; Westlund, 2014), their other functions in society are not considered here (i.e., social, commercial, or democratic through deliberation). In Appendix A we have included sources that have reviewed the impact in the wider media landscape, including for example film or the creative industry (incl. e.g., games, publishing, and music), as this will not be covered here, but is also represented in AI4Media use cases.

2.2.2 Artificial Intelligence (AI)

Artificial Intelligence (AI) as media is also a rather elusive and encompassing term, which often becomes an umbrella term covering an array of different techniques ranging from rule-based algorithms to machine learning (ML) and deep learning (DL) or Natural Language Generation

(NLG). The AI deployed in today`s society and in media is what is often referred to as ‘narrow AI’, meaning it is trained to solve a specific task (e.g., personalise front page). This is in opposition to ‘general AI’, which describe systems that can adapt and ‘learn’ new tasks as they develop. In this white paper, we adopt a function-oriented rather than technical definition of AI, defining it as the automation of tasks or decisions (either fully or partly) that would previously have required the intelligence of a human being (i.e., a media professional), to paraphrase the original definition of John McCarthy, who was the first to coin the term AI (McCarthy *et al.,* 1955). Thereby, including all the various kinds of AI applications that might be utilised by media organisations. In the sections of the review, we highlight specific definitions of AI applications when relevant, but in some cases, the discussions do not engage that concretely with the technologies and they are predominately defined through their function in these cases.

3 Potentials and challenges of AI for media

Part one of this white paper explores the potentials and challenges of AI for media across the various stages of the media cycles: ideation and content gathering, production, distribution, deliberation and finally archiving. Each of these sub-chapters presents examples of popular uses of AI applications in that stage of the media cycle and relates these to the potential impacts (positive and negative) discussed in the reviewed literature. Through this an overview of the key discussions can be attained, which will set the stage for the second part of the whitepaper.

3.1 AI in ideation and content gathering

Within media organisations, AI is becoming more used in what we here refer to as the cycle of ideation and content gathering, which we see as the processes leading up to the actual production of media content, be it written, tv or audio content. In these processes, AI is increasingly leveraged to identify trends or spot stories, provide insights into the audiences' preferences, serve to become aware of existing media biases and counteract such biases as well as provide new tools for investigative journalists that can detect data patterns that were previously unidentifiable or would require a large amount of manual work. However, there are also challenges to the role of media organisations as these tools can for example reinforce existing biases or produce new ethical questions. AI systems have also not only become supportive tools for investigative journalists, but rather the very object of enquiry through new forms of 'algorithmic accountability reporting' (Diakopoulos, 2015), which aims at holding large platforms like Facebook responsible for the workings of their systems. In these projects, AI is often also used in the process of finding the evidence for the story and we, therefore, include it in this section. In the following we discuss:

- How AI is utilised to identify news stories and in investigative reporting
- How AI has become the topic of algorithmic accountability reporting
- How AI plays a role in ideation through AI driven audience measurement systems
- How AI is used to mitigate existing media biases and support content and source diversity

3.1.1 AI in story discovery

AI is seen as having great potential for supporting the ideation process by being able to **sift through large amounts of data** and **identify potential news stories** or **track social media for trending topics** (Beckett, 2019). While notions of computer-assisted reporting (CAR) going back to the 1950s and now more increasingly data journalism has been used to describe the emerging practices of journalists using computational tools both in ideation and production of content (Coddington, 2015), AI promises to further advance such methods in new ways.

One way is through the spotting of trending or breaking stories – particularly on social media, which today has become a core part of sourcing practices amongst media professionals (Schifferes *et al.*, 2014; Thurman *et al.*, 2016; Thurman, 2018). Social media have been proven to be the site where news stories in many cases break. The story of the Hudson River plane crash in 2009 was, for example, posted on Twitter 15 minutes before the first news stories started emerging (Thurman *et al.*, 2016). Making social media an increasingly important site for story discovery, but also one that can be difficult to monitor due to the vast amounts of content circulating on these platforms. As a result, new tools – some AI-driven – are emerging to support media professionals with this exact task. Examples of such tools are the Newswhip's Spike, which is a real-time media monitoring tool aimed at spotting stories while they are still 'small' and Geofeedia a location-based monitoring tool that helps discover location-specific stories (for in-depth analysis of these tools, see Thurman, 2018). Another example was the European Union's 'Social Sensor project', which aimed at developing a tool equipped to spot and verify stories originating from social media (for an in-depth description of the project, see Schifferes *et al.*, 2014; Thurman *et al.*, 2016). Due to the increasing amounts of mis- and disinformation[1] circulating on social media platforms (see e.g., Burgess, Vis and Bruns, 2012), the latter function has become increasingly important to ensure accurate reporting based on social media reporting. **The use of AI in verification of content authenticity in this part of the media cycle is, therefore, also becoming increasingly important and holds great potential for more accurate reporting in situations where events develop quickly.** We revisit the concrete applications of AI for this purpose in section 3.4.2 on AI in fact checking practices.

Media organisations have in some cases also developed their own tools, such as Reuters who developed the News Tracer system, which in real-time analyses all published tweets on Twitter and (i) filters (e.g., removes spam, checks the credibility of the tweet) and sorts the tweets into thematic clusters, (ii) assigns a topic to them (e.g., politics or sports), and (iii) using NLP, produces a summary of what the cluster contains, which is presented to a journalist (Stray, 2016a). This tool is used to quickly identify and report on breaking news as it proved to be at least half an hour quicker than journalists in discovering new stories through social media (Stray, 2019). The core features are the ways in which it can help discern whether a trending hashtag might be cause for a news story and immediately support the media professional in assessing the credibility of the story (Stray, 2016a). With the fast-moving news cycle today, these tools hold great economic potential for media organisations as being first with a story is a differentiating measure between news organisations and can also produce more traffic. Furthermore, it can also provide the opportunity to find different stories than a journalist would traditionally have found. This is also societally important as furthering diversity in coverage is at the core of the

[1] For definitions on the differences between mis- and disinformation see: https://guides.lib.uw.edu/c.php?g=345925&p=7772376

democratic ideals of media institutions of providing diverse and balanced coverage of societal events.

Another way AI is being leveraged to support story discovery is in relation to investigative reporting (Broussard, 2015; Stray, 2016a, 2016b, 2019). Investigative reporting is one of the most labour-intensive practices of media organisations where the ideation process is both difficult and, in many cases, does not actually turn into a story (Broussard, 2015). The key potential discussed here is **to increase the 'fourth estate' or 'watchdog' function of journalism** by employing AI tools to sift through the large amounts of government data and documents produced at both national and local levels in search of interesting data connections that can then be explored by a journalist (Stray, 2016a). AI tools could provide the **opportunity for discovering stories that would previously have been too resource-heavy** to pursue or **detect patterns in big data that would be impossible for humans to grasp**, producing potentially new forms of stories from what has been seen before (Hansen *et al.*, 2017; Stray, 2019). This potential of leveraging efficient AI systems to increase the accountability function of media is placed in the backdrop of the economic struggles of the media industry, where in-depth and cost-heavy stories are often disbanded due to the time and resources involved and because of the need for journalist to spend time constantly producing content rather than research (Broussard, 2015).

Jonathan Stray (2019) lists the most common AI tools utilised in investigative journalism. The first is *supervised document classification*, which was used by the Atlanta Journal Constitution in their 'License to Betray' story (Teegardin *et al.*, 2021) as they scraped more than 100.000 doctor disciplinary records in order to uncover how doctors who had sexually abused patients in many cases were allowed to continue to practice (Stray, 2016b, 2019). The second is *language analysis* (encompassing a range of NLP techniques including topic modelling, clustering, word embeddings and sentiment analysis), which was used by the international news agency Reuters to produce their story 'The Echo Chamber' (Biskupic, Roberets and Shiffman, 2014), which showed how almost all cases in the supreme courts in the US were argued by the same group of lawyers (Stray, 2016b, 2019). The third is *lead generation*, where the system points to potential avenues of enquiry based on data, rather than conclusive evidence, such as Buzzfeed's tool which identified law enforcement planes circling over cities and alerted journalists to produce concrete stories relating to the known fact that US government surveillance planes have been previously caught capturing surveillance footage in this manner (Stray, 2019).

Meredith Broussard (2015) has also in a more in-depth manner described her process of experimenting with a 'Story Discovery Engine' specifically, which could analyse education data in the Philadelphia school district in the US based on her initial idea that there might be an issue regarding the access to books at schools and the correlation with how the schools perform in national tests. The project was called 'Stacked Up' and included both the story identification tool and the stories that were produced based on it. The potential here is to provide investigative reporters with tools to represent the requested data to quickly identify ideas for stories and what sources to contact. As Broussard describes, based on the tool she was able to generate ten ideas in only half an hour, which could potentially lead to 30 or more articles that could illuminate the social problems connected to the access to schoolbooks. While she does point to how the tool is limited to an education data focus, it could be used to explore other fields as well and make the ideation process more efficient and more inclusive to more inexperienced staff who normally have difficulty entering the role of an investigative reporter.

Challenges of AI in story discovery

Several concerns have been discussed regarding the use of AI in story discovery on social media or the media environment more widely. First, studies have shown how **such tools are better equipped to monitor updates on stories that have already broken, rather than being able to discover original breaking stories,** ultimately removing the 'competitive edge' they promise (Thurman *et al.*, 2016; Thurman, 2018). An overreliance of such tools might also **risk a media coverage that overemphasises topics that are already present in the public debate**, rather than raising new and important or overlooked topics, which could be highly detrimental to ensuring both critical and diverse media coverage (Thurman, 2018; Stray, 2019). This potential risk of a 'popularity bias' is not the only form of bias that might risk being reinforced with the increasing use of AI story discovery tools. In their study of the 'SocialSensor project', Neil Thurman, Steve Schifferes, Richard Fletcher, Nic Newman, Stephen Hunt, and Aljosha Karim Schapals (2016) show how they use different techniques to give the system 'a nose for news', including training it to monitor known 'newshounds' (e.g., bloggers, journalists, media outlets etc.) or prioritise stories that spike short term and include people. In the article, they discuss how all such decisions could be criticised and that there are pros and cons to potential changes in how the system is developed. However, their main criticism is how currently, **these systems reinforce the existing biases in the media coverage and landscape**, which is both, for example, western focused, sensationalist and male-dominated (Thurman *et al.*, 2016; Thurman, 2018). Such systems, could, therefore, have a negative social impact on society by potentially amplifying rather than confronting existing biases.

A critical question deserving of raising with these systems is, therefore, whether they should and could be designed to counteract such existing biases – a question we return to further down as another trend is to use AI to improve the diversity of media coverage. It is also highlighted how the data coming from social media is biased towards certain groups in society. Twitter is, for example, disproportionally popular among professionals working in media and politics. This requires that **media professionals both understand the limits of these data and the AI tools that rely on them and remain critical towards what in fact they represent** (Hansen *et al.*, 2017; Thurman, 2018; Stray, 2019). As many of these monitoring tools are increasingly developed by the social media platforms themselves, it also becomes further pertinent to be critical towards the output they produce (Thurman, 2018). Neil Thurman (2018) also highlights, how many of these systems have in-built verification systems to help ensure that the content is 'real' and while this is an important supporting tool for media professionals that are grappling with ever-increasing amounts of misleading content, it must not become a crutch that makes media professionals be critical toward the output they are receiving from these systems. This remains pertinent, as these verification systems are not flawless and an **overreliance on them might risk a higher rate of misleading media content from otherwise credible media organisations**, putting at stake the general societal trust in media institutions (for an example of how such systems work, see Schifferes *et al.*, 2014). We return to the concrete challenges regarding such AI systems in section 3.4.2.

Moving on to potential challenges related to AI in investigative reporting, Jonathan Stray (2019) overall points to how during conferences, in reports and journals it continues to be the same handful of examples that are utilised to illustrate the promise of AI in this field. This illustrates both how this is still an emerging field, but also how there might be concrete challenges that impede such promises and the positive social and economic impacts in becoming reality. One of the **challenges relating to AI in story discovery is data availability**. This challenge is already beginning to be addressed, as more public data is becoming available and the topic of data availability has become part of the political agenda, with for example the Obama administration who in 2009 stated that they provide increased data access and availability (Obama, 2009 in Broussard, 2015). Equally, the EU has launched the initiative 'Media Data Space', which aims at making more media data available (see the AI4Media report *'Overview & Analysis of the AI Policy Initiatives in EU level'* for details on this initiative), but as will be discussed below the challenge has yet to be fully solved.

Jonathan Stray (2019), for example, describes how there are still national corporate registries that require payment for access to the data (most common in tax havens such as Hong Kong and Cyprus). Much public data (both current and historical) also remains in paper form, which minimises the usefulness of AI as journalist will still have to do much of the 'grunt' work, which was hoped to be avoided by employing AI. However, AI might also be the cure to this challenge by allowing easier digitalisation of hard copies. As already noted above, data unavailability affects this process in two ways both in feeding the model relevant data but also because of the lack of standardised training sets for this type of work, a problem that is enhanced due to the fact that within investigative journalism, the goal is to produce unique stories, so new training sets will be needed frequently (Stray, 2019). As the example of the 'Stacked Up' project showed, the tool was good at identifying stories relating to a very specific question of the correlation between schoolbooks and national tests, and while this of course could be expanded to other school districts and maybe also adapted to other wider enquires, further development would be needed, making reuse of AI tools more complicated within these media practices. This is not to say that more 'general' tools that for example summarise and cluster topics based on government documents will not be useful. They too will assist media ideation, but the fact that it might be difficult to develop general AI applications that can be used to support multiple investigative stories does challenge the potential of cost efficiency discussed above. Jonathan

Stray (2019) also notes that the salary differences between journalists and data scientists might increase the cost of producing the story if new tools must be built regularly, which could also undermine the positive economic impact.

A more general concern relating to privacy also arises out of both the use of AI in story discovery more in general and investigative reporting, as they both entail the increased monitoring of individuals, either through their social media accounts or through other supplementing technologies. The New York Times for example used facial recognition to gather the data of who was present in the audience during Donald Trump's inauguration together with financial campaign data (Hansen *et al.*, 2017). Equally, they have produced the 'Who The Hill' tool, where readers could upload photos to see if they contained the faces of US members of congress in images uploaded by readers. This tool, while not an active reporting tool, did lead to a story of congresswoman Claudia Tenney who was identified in a photograph taken at a fundraising party (Stray, 2019). Particularly in the case of images, a relating risk is misreporting and wrongly accusations based on photo manipulations, so-called 'deep fakes', where images or audio material is manipulated to show people in situations, they have not taken part in or saying something that in fact have not said. Similarly, as with the verification mechanisms, it remains pertinent that AI outputs are not solely relied on, but thoroughly investigated by media professionals as well to avoid such situations. Neil Thurman (2018) also critically highlights the potential misuse of social media content monitoring of individuals, because following his study of the different monitoring tools, it was revealed in 2016 that the, above mentioned, location monitoring tool 'Geofeedia was being used by over 500 "law enforcement and public safety agencies", including to monitor "activists and protesters"' (Thurman, 2018). While this resulted in Geofeedia being denied future access to both Twitter and Facebook as they had transgressed the terms of use, it illustrates how such monitoring can be used to infringe on privacy and to strategically survey certain groups.

EXAMPLES OF AI USES IN STORY DISCOVERY

The **'News Tracer'** developed in 2014 by Reuters, which filters and sort Tweets into topical clusters and assigns them both a topic e.g., politics, sports or business and further awards them a newsworthiness score.

The **'Leprosy of the Land'** project by developed by Texty, which utilised a machine learning model to identify traces of mining activity in satellite images covering an area of 70.000 square kilometres and discover sites of illegal amber mining in Ukraine.

The **'Stacked Up'** project developed by Meredith Broussard, which explores the question of whether school children in the school district of Philadelphia have enough books – a high stake question considering the importance put on standardised test in the US.

The **'Doctors & Sex Abuse project'** at the Atlanta Journal-Constitution included employing a bot to scrape regulators websites and a machine learning program to analyse the documents in order to detect patterns of physician sexual misconduct, which was identified and later reported on.

The Los Angeles Times story **'LAPD underreported serious assaults, skewing crime stats for 8 years'** was developed using a machine learning tool that could identify when a police report at the Los Angeles Police Department had been downgraded from aggravated to simple assault.

The **'Who the Hill'** tool developed by the New York Times uses computer vision and facial recognition to identify congress members in photos uploaded by users, which so far have only led to one concrete article

See more examples on AI uses in investigative journalism on: **https://investigate.ai/**

Figure 1: Examples of AI uses in story discovery

3.1.2 AI in algorithmic accountability reporting

A relating topic to AI in investigative journalism, is also how AI has become the object of journalistic reporting, what Nicholas Diakopolous (2015) calls 'algorithmic accountability reporting'. In the article, he specifically explored the method of 'reverse engineering' as a way of producing a data foundation for such stories. One example is journalist Michael Keller from the Daily Beast who noted how misspellings of 'rape' and 'abortion' was not autocorrected on his iPhone, which gave him the idea to explore the workings of the underlying algorithm both by getting access to the API but also by running a range of user simulations, where a script acted like human users in order to produce an overview of how the algorithm worked (Diakopoulos, 2015). Equally, the Wall Street Journal in 2012 exposed how many e-commerce platforms were discriminating when offering a price to different users. This story was also backed up by utilising computational methods to simulate different users visiting different sites to identify discriminatory patterns. This can be seen as a new way for media organisations to act like a 'watchdog' by utilising computational methods to explore the workings of AI systems, which will be important as these systems become increasingly used and it is important that they, as public institutions, and figures are held accountable for their actions. There are also non-profit organisations like 'Algorithm Watch' who have this as their mission, as they analyse automated decision systems and their impacts on society. There are also journalistic organisations like 'ProPublica' who have also begun to specialise more in doing these types of stories.

Challenges of algorithmic accountability reporting

One of the core challenges to this form of reporting is access to the workings of the algorithms underlying the major social media platforms, making it difficult for reporters and civil society to hold these actors accountable. Algorithm Watch, for example, in August 2021 described how they had been threatened by Facebook (now Meta) after they had launched a project that monitored the newsfeed on Instagram. It was based on user consent, as Instagram users could install a plugin in their browser, which then allowed Algorithm Watch to scrape their Instagram newsfeed (Kayser-Bril, 2021). This method was both to ensure ethical practice, but also because Meta's APIs on both Facebook and Instagram are notoriously difficult to interact with (for more in this see e.g., the public statement by the initiative Social Science One, 2019), compared to, for example, Twitter. Meta, however, argued that this was a breach of their 'terms of use' and that they would move towards formal (i.e., lawsuit) means of protest if the practice continued. Algorithm Watch also reported how the New York University's Ad Observatory had been shut down the week before they received the threat by Facebook, illustrating how there is a general issue of researchers and reporters being denied access to data and be able hold these actors accountable based on claims of 'terms of use' or trade secrecy. This is becoming a global societal challenge as neither scientist nor reporters can get proper access to these highly influential systems.

Similarly to the challenges discussed above regarding story discovery, 'reverse engineering' has also been criticised for only focusing on the outputs and the consequences of AI rather than on the bigger picture of their workings and when such consequences are made (for a critique of the method of reverse engineering, see Ziewitz, 2016). Illustrating not only the need for more access to source code but also studies of how decisions are made within these organisations on how to build these systems and on what grounds. Another critique echoes the one above of being aware of not over relying on AI tools when critically examining these practices as the tools used to reverse engineer might equally hold flaws that must be critically addressed.

EXAMPLES OF ALGORITHMIC ACCOUNTABILITY REPORTING

The Story **'Websites Vary Prices, Deals Based on Users' Information'** by the Wall Street Journal used computational methods to explore how AI systems would affect the price presented to users based on their user information

The Story **'The Apple 'Kill List': What Your iPhone Doesn't Want You to Type'** by the Daily Beast explored the word's that the autocorrect system on iPhones does not correct.

The Story **'Reverse-Engineering Obama's Message Machine'** by ProPublica explored how the Obama Campaign used an AI system to target voters with different emails.

Figure 2: Examples of algorithmic accountability reporting

3.1.3 AI driven audience measurement systems

AI is also increasingly used in media organisations in the form of AI-driven audience metric systems, which presents to the media professionals what content is performing well and which is not (see e.g., Anderson, 2011; Møller Hartley, 2013; Tandoc, 2014; Christin, 2018; Kristensen, 2021). These systems are now an important guiding factor in many media organisations in the ideation cycle as the performance numbers can be used to, for example, make the case for why an article should be written – It has become an important valorisation measure in the media field. A study of five Dutch online new sites, for example, showed how there was a direct correlation (four out of five times) between what stories performed well and what follow-up stories were pursued (Welbers *et al.*, 2016). Equally, the performance can also be decisive of how prominently media content is featured (Lee, Lewis and Powers, 2014), something we also return to in the following section on distribution. When implementing such systems, media organisations in some cases build their audience measurement systems in-house, but in many cases, they rely on commercial systems such as Chartbeat or Google Analytics. The idea of audience measurement is not new as media organisations, and particularly not in what was originally the advertising department or today the marketing department, have always been interested in knowing who their audience was and more importantly what their preferences were (Kristensen, 2021). However, the methods have changed as historically audience insights were based on surveys, media diaries etc., which gave overall impressions of the preferences allowing, for example, for the construction of 'reader profiles' (Willig, 2010), but now audiences can be represented with unseen granularity when leveraging big data and AI (Christin, 2020).

The potential of these systems is generally considered to be the possibility of being more responsive to users and what they find interesting, which is often seen as contrasting the more traditional 'paternalistic' role of media organisations, where media professionals ultimately determined what the audiences needed to see or read based on their 'gut feeling' and expertise, disregarding the preferences of the audiences completely (Anderson, 2011; Willig, 2011). As Herbert Gans originally described back in 1979 (Gans, 2004), the journalists he observed even intentionally disregarded any information about the audience, as considered their choice more valid (and democratic) than the audiences. The potential is, therefore, **to push media professionals out of their ivory towers where they unchallenged decide what is of societal importance**. The granular knowledge of audiences is also considered to be an economic potential as high traffic numbers have become an increasingly important measure in online advertisement, which have proven to be a challenging economic arena for media organisations (Christin, 2018).

Last, **audience measurements are often described as ways of empowering and giving more agency to the audience**, moving it out of its passive role and giving it a generative role in the media environment (Anderson, 2011). Here, for example, it has been discussed how the audiences become new 'gatekeepers' (see e.g., Tandoc, 2014; Vu, 2014) or how they are increasingly becoming 'produsers' of media content (a merge of producers and users), as the boundary between producers and consumers is becoming increasingly permeable (Bruns, 2009). However, the question of whether the users are empowered have been raised in critical voices regarding how they are often unaware of the degree of tracking and how their data is concretely utilised. Equally, concerns have been raised of how an 'obsession with traffic' (Anderson, 2011) might impoverish the online media content environment. We turn to the latter concern, while we save the discussion on tracking practices in section on personalised content curation and distribution.

Challenges of AI driven audience measurement systems

There are potential negative impacts of AI-driven audience measurement systems on both the availability of quality content online and on the digital media ecosystem (Christin, 2018). Angele Christin (2018, 2020) describes how quantification processes in general are highly transformative as they produce both standardised and quantifiable ideas of high quality, which in turn can change the organisational practices. As an example, she describes how university ranking systems induced a range of new practices that were aimed at sustaining the university ranking (Christin, 2018). In the UK, the negative impacts of such rankings have been clear as university grades, which were linked to the rankings, inflated over the last decade, leading to an integrity crisis of the universities as institutions (Lambert, 2019).

While Angele Christin (2018, 2020) in her study illustrates how the interpretation and reactive practices to audience measurements are different in two newsrooms, in the US and France, new practices emerge in both places that might also **undermine the quality and integrity of the online media landscape**. Here the main concern relates to how economic and managerial pressure to produce content that creates traffic and engagement, **might induce a shift towards certain types of content or ideas, namely the content that is deemed newsworthy due to its sensational characteristics or personal character** – what Jannie Hartley (2011) also collate under a joined 'audience criterion' for news. Several studies have shown how such specific online news criteria are emerging (while not replacing the old) such as 'shareability', which changing the selection of news content that is produced (Kristensen, 2021). **The prioritisation of more traffic driving beneficial articles is seen as a challenge to the democratic practice of media organisations of cultivating well-informed citizens** as it might lead to the deselection of more 'boring' content that for example engage in political events. The political impact of this could be less political participation in societies as some studies have pointed to the direct connection between of consumption of political media content and political engagement (see e.g., Lee and Wei, 2008).

The above relates more to the changing organisational practices, but quantification also affects the individual. Chris Anderson (2011), for example, describes how this tension between serving the audience and serving democratic ideals create personal tensions for media professionals, who now must find compromises. This is echoed in the studies by Jannie Hartley (2011, 2013) and Angele Christin (2018), who also highlight how previous studies have shown how the constant tracking of employee performance can create a quite threatening work environment, which could induce a behaviour of 'gaming' the system and aim of getting the right scores over everything else, which can have **detrimental effects for both the quality of the media content and personal consequences as negative or positive scores can affect the individual's mental state** (for a discussion of monitoring in the workplace, see Campolo *et al.*, 2017). Journalists in Chris Anderson's (2011), Angele Christin's (2018) and more recently Caitlin Petre's (2021) studies also express concerns of being hired or fired based on the numbers and state that these numbers as used as justifications for such decisions. This illustrates how professional value and success is also increasingly tied to these measurement systems, putting additional pressure on employees to score well. Something, that is partly realised in some media organisations where individual performance scores are used in evaluative meetings with managers and during layoff rounds, they become a major point of concern. Thereby, **impacting the social welfare in the workplace.**

These concerns can be argued to have little to do with the AI system itself, rather it relates to it existence and the management strategies surrounding it. However, as Lisa Merete Kristensen (Kristensen, 2021) expertly shows in her in-depth study of the development of an audience measurement system, the way these metrics are decided is a process of negotiation and their final output becomes highly determinative for what content scores high or not. As 80% of publishers (in 2018) in the US were using Chartbeat as their preferred audience measurement system, this shifts power to external providers in determining what the standards of a 'good' performance are (Christin, 2018, 2020). **Placing such infrastructures at the centre of valorising media content, requires thorough interrogation** of what values are chosen and why – and perhaps critical questioning of these values in efforts to sustain the quality of the media landscape and avoid potential negative impacts on society and democracy that can be produced by such systems. Particularly, as these systems are often built with commercial aims, rather than necessarily editorial – for example, in the case of Google Analytics, which is also used in marketing in general.

EXAMPLES OF AI DRIVEN AUDIENCE MEASUREMENT SYSTEMS

'Chartbeat' is one of the most well-known analytics tools for the publishing industry that provides real-time analytics on how the audience is engaging with the content. They partner with among other major new outlets land broadcasters like the New York Times, Le Monde and CNN.

'Google Analytics' is a product offered by Google that allows the users to track engagement analytics on and across their sites (including social media) as well as measure advertisement impact.

Figure 3: Examples of AI driven audience measurement systems

3.1.4 AI in supporting media coverage diversity

The last major trend of AI in the cycle of ideation and content gathering phase is **the use of AI to improve the diversity of media coverage by identifying and making media professionals aware of biases** – be it gender, race, or political orientation. An example of how AI can help mitigate such existing biases is the 'JanetBot', which was developed by the Financial Times (FT) in the US in 2017 as part of a collaboration between FT Labs, Data Analytics, Editorial Technology and the FT Newsroom. In real-time the AI system analyses the images presented on FT's online frontpage and classifies images with people as either 'man', 'woman' or 'undefined'. The aim is to make the editorial staff more aware of the gender ratio on the site, a goal that was motivated by audience feedback stating how there were too many 'men in suits' on the front page. This potential was also the key focus in the 2020 Journalism AI Collab, a global collaborative experiment that was launched in June 2020 by the initiative JournalismAI, which is run by the LSE Polis think tank and supported by Google's Digital News initiative. Here eight news organisations joined forces and experimented with how to understand biases in newsrooms and developed two concrete ways of mitigating them using computer vision and natural language processing (NLP) technologies (AIJO Project, 2021). The first experiment aimed at exploring gender differences in the images used on front pages and illustrated quite a significant gender gap of who is depicted in the news (AIJO Project - Experimenting with Computer Vision, 2021). The second experiment used the 'Gender Gap Tracker', which is an automated software system created by the Discourse Processing Lab at Simon Fraser University that measures the number of quotations of men and women, to explore the gender differences in news articles, again illustrating a significant gender gap both in the number of women quoted and the length of their quotes in comparison to male quotes (AIJO Project, Experimenting with Natural Language Processing, 2021).

The positive impact of AI could be immense as previous studies have shown, for example, how media coverage in both broadcast and written content highly favours male white sources – overlooking women, other races, and minority groups. Particularly in sports reporting, the difference in both the sheer number of references to female athletes and how they are referenced is remarkable. A study by Cornell Bowers CIS, where they too used NLP in their method of identification, showed how female tennis players were more often asked personal questions in comparison to their male colleagues at press conferences and follow up interviews (Steele, 2022). Equally, a study published by Cambridge University Press conducted during the 2016 Olympic games in Rio illustrated how women were less mentioned during the event and when mentioned, it was to a much higher degree than with their male colleagues related to their marital status or appearances (Cambridge University Press Research, 2016).

This problem with biases in the media coverage referred to as media biases have long been a research topic within the field of humanities and social science, where scholars have explored how the choice of events to cover, the sources used, the framing of the story, the word choice and labelling of sources and even the position in the paper or in a broadcast have contributed to biased reporting (for review, see Hamborg, Donnay and Gipp, 2019). The importance and problematic nature of such biases have also been highlighted multiple times, as for example, specific framing of societal topics can potentially sway the public opinion as the case was with Tobacco in the 1960s (Hamborg, Donnay and Gipp, 2019). However, the analyses of media biases have always been highly labour intense work, requiring that hundreds of texts were read, analysed and categorised, but now AI can be used to automate such processes making it not only easier to scale up such analyses, but also to do it in real-time as the example with the 'JanetBot' shows, where it can actually **help the media professionals rearrange existing content or provoke their thinking in what stories or segments to produce next -** and potentially also by giving in **real-time recommendations during the production process regarding choices of words or tone** (e.g., let a sports reporter know when they are using gendered language). This might also prove economically sensible as the coverage of currently overlooked minority groups or topics might **produce new reader and viewer bases**. Equally, increased transparency of such biases through disclosure of such biases by media organisations or even as a benchmarking feature amongst media organisations (for details, see Leppänen, Tuulonen and Sirén-Heikel, 2020). This would help audiences better navigate the media landscape and specifically search out varied opinions on a topic, which could have positive societal and democratic impacts, particularly as audiences still have an expectation that media represent an objective 'reality' (Reese and Shoemaker, 2016).

Challenges of AI in supporting media coverage diversity

While the potential of AI as discussed is immense, there are also certain challenges that arise. Particularly, **the challenge of not creating new or amplifying other biases in the attempt of mitigating, for example, gender biases.** In the examples used above, a cis-gendered reality is maintained in both the 'JanetBot' and the JournalismAI experiments, where you are either a man or woman – or potentially undefined, which ultimately can discriminate against or at a minimum overlook other gender orientations (Crawford and Paglen, 2021). However, there is no easy solution to this challenge as many of the biases attempted to be mitigated, such as notions of 'gender' of 'race' are not stable categories. Rather they are culturally and historically sensitive and the annotation and labelling process connected with AI will, therefore, always risks further enhancing or simply reproducing existing stereotypes, prejudices and racism (Leslie, 2020). Equally, such question raises many **new ethical questions that must be considered, such as which genders are appropriate to classify and what impacts could such classifications have.**

This will, therefore, likely remain a persistent challenge, which was also recognised in the JournalismAI experiment, where they did recognise the problem with their classification, but also had to work pragmatically and start somewhere (AIJO Project - Experimenting with Computer Vision, 2021). However, the recognition of this is an important first step, as awareness of how such classification is a highly political and social enterprise that must be interrogated will be essential to avoid negative societal impacts (Crawford and Paglen, 2021), something we return to throughout this report. In the review article by Felix Hamborg, Karsten Donnay, and Bela Gipp (2019), the authors suggest how a way to improve the use of AI in understanding and mitigating biases is by drawing inspiration from the quantitative and qualitative methods in which social scientists long have analysed these biases (including how they classify and analyse biases). They argue that much progress could be made by closing the knowledge gap between computer scientists, social science researchers and media practitioners in developing these new systems and tools.

In the JournalismAI experiment, they also raise another important challenge, namely that the error margin for wrongly labelling female faces as male was significantly higher than the other way around (AIJO Project - Experimenting with Computer Vision, 2021). This relates to a **general**

challenge of biased training sets – even those that are considered 'golden standard' (Leslie, 2020). The report 'Understanding bias in facial recognition technologies' by David Leslie (2020), published by The Alan Turing Institute, described how Microsoft's 'FaceDetect' model, which showed an overall error rate of 6,3% regarding genders, which seemingly was acceptable, but in fact, it had a 0% error rate for white males, while it was 20,8% for dark-skinned females, which highlights the need to interrogate not just the general error margins but how they affect subgroups differently. When developing such systems, it will be pertinent to interrogate the training sets used to train them to ensure that such inherent or historical biases are not sustained or amplified. We return to this discussion and the need for mitigative strategies and resources to foster responsible AI practices in chapter 4.

EXAMPLES OF AI SUPPORTING MEDIA COVERAGE DIVERSITY

The **'Gender Gap Tracker'** is an AI tool based on Natural Language Processing and was developed as part of a collaboration between the non-profit 'Informed Opinions' and Simon Fraser University. The tool quantifies the differences in quotes by men and women across seven of the larger news outlets in Canada.

The **'JanetBot'** was developed by the Financial Times (FT) in the US in 2017 as part of a collaboration between FT Labs, Data Analytics, Editorial Technology and the FT Newsroom. It is built using computer vision to identity images on the FT website as either 'man', 'woman' or 'undefined'. It is designed to support the editorial staff in providing a more gender diverse front page by making them aware of the ongoing ratio in gender represented in the online site.

The **'She Said, He Said Bot'** developed by the Financial Times followed the JanetBot to further explore support the focus on gender equality by editors on their online sites by identifying the gender of the sources cited in articles by classifying pronouns and first names. The next step is doing it before the articles are published within the CMS-system, so that the potential imbalances are made apparent to the editor prior to publishing.

The **'Stanford Cable TV News Analyzer'** enables researchers and the public to query the AI-driven tool to explore who appears and how much talking time they are afforded. It was developed by the Computer Graphics Lab at Stanford University in collaboration with the John S. Knight Fellowship Program.

Figure 4: Examples of AI supporting media coverage diversity

3.2 AI in media content production

While the boundaries between content gathering and content production in practice are overlapping, this section focuses on the opportunities and challenges of using AI specifically in the production phase, when media professionals are writing the news or producing the news broadcast or another TV segment. Here AI is leveraged in supportive functions by for example producing 'rough cuts' of audiovisual content, which are then finished by media professionals or by providing spell-check services or image suggestions to media professionals who are in the process of producing content. AI is also increasingly used to fully automate written content production, particularly in the news sector, where AI systems today are producing financial news or sports reporting. Full automation has yet to really become a core feature of audiovisual content production, which results in this section being slightly skewed towards the production of written content, as this is a major trend, which has received much scholarly attention. However, a few examples of fully automated audiovisual content production will be discussed and many of the challenges raised in the section on the automation of written content will equally apply to the audiovisual sector. In the following we discuss:

- How AI is used to support media content production
- How AI is used in written media content production
- How AI is used in audiovisual media content production

3.2.1 AI as support tools in media content production

In the report 'New Powers, New Responsibilies' by Charlie Beckett (2019), he describes how a range of AI-driven tools are already used to support media production, such as Grammarly for spellcheck and Deepl.com for translation of text, thereby, easing the work of the producers and editors. Another trend is the development of automated tagging systems that employ NLP technology to read and suggest relevant tags to be given to the article. A task that was previously carried out by editors (Beckett, 2019). The New York Times are already making this process even more intelligent as part of the 'Editor' project, which currently augments the work of journalists in annotating the produced content. Building on this, they are also experimenting with how such methods of fine-grained annotation can further enhance their ways of distributing the content, a potential we return to in the following section.

The **potential of this use of AI is to save time, so that media professionals do not have to spend time on doing these routine tasks**, which are highly necessary if the **content is to be searchable** in, for example, media archives after they are published. Equally, the production of good meta-data can be enormously important for content to be discoverable on search engines, such as Google News (see e.g., Kristensen, Forthcoming). We return to the potential of metadata producing tools in the section on archival uses of AI, where similar methods not only can make natively digital content more searchable, but also historic media content. Equally, such AI systems can be used to offer suggestions of what visual content should accompany the news article, such as the 'Panel' tool offered by GettyImages. This is again an efficiency measure but can also be inducive of more diverse imagery on online sites as journalists are prone to use the images, they already know are there rather than spend time on searching through large image databases.

AI is also increasingly used to support the production of audiovisual content by broadcasters to transcribe audio, which allows for the automated subtitling of content, to edit the sound and automated text summarisation to produce descriptions of the content (for a good overview of uses of AI in the audiovisual sector, see Rehm, 2020). The key potential expressed by the public broadcasters is like that of news outlets, namely that AI will **support and make processes of editing and producing the content more time efficient** (Rehm, 2020). This is particularly the case with subtitles, which are a time-intensive task to produce and if subtitles are to be produced in multiple languages, multilingual translators are necessary. Utilising AI technologies in this instance can, therefore, improve the accessibility to media content by potentially catering more nationalities and significantly reduce time spent on translation. Furthermore, such automated **translation tools can generate more (economic) value out of each piece of produced content**, as it enables other media organisations across national boundaries to easily reuse the content produced in other countries – also potentially **enhancing the diversity of the media content** available to the users.

Another important use is also to make 'rough cuts' based on the audiovisual data foundation (e.g., video or audio) or do automated image retouching, to again make the work of editing a TV or radio segment more efficient (Amato *et al.*, 2019; Rehm, 2020). This is also the focus in one of the AI4Media use cases 'AI in Vision - High Quality Video Production & Content Automation', where specifically the focus on timely coverage is considered a potential, as footage from, for example, social media platforms can quickly be edited. This can make the **distribution of audiovisual content timelier and more relevant for citizens due to faster publishing times**, by minimising editing time. This will be particularly important during breaking news situations, where live feeds can be quickly edited adding, for example, names of the people or locations presented in images based on image recognition AI. Equally, the production and selection of automated thumbnails is currently a widely discussed application, that can help find the most visually appealing thumbnail that will temp the viewers to click on the content, creating both engagement with the content and saving time, as finding the right thumbnail can be time consuming (for more on this see, Oury, 2020).

Challenges of supportive AI tools in content production

The AI tools used to support the production process in media organisations have, as Carl Gustav Linden (2017) points to, gone more 'under the radar' in terms of getting public attention or producing resistance or discussion amongst the professionals involved in their use (at least not publicly). Equally, they have not been the focal point of much academic literature. Taina Bucher (2017, 2018) in her study of five Nordic newsrooms notes **how the supportive systems are not really questioned but considered inherently helpful by taking over tasks and are not discussed critically in comparison to more directly transformative AI applications** such as recommender systems. The lack of interest in and unquestioned use of these tools could, therefore, be seen as one of the challenges connected to them as it means that we have very little knowledge in their impacts (negative and positive) and it also means that the tools remain uninterrogated users, which could reproduce certain biases inherent in those tools as discussed above.

Another challenge that is generally discussed in relation to the use of AI and language tools, which here relate to translation tools, grammar tools and tagging and indexing tools as well as transcription tools, is that they are **predominately trained on English datasets and generally perform significantly worse on other languages** (Cordell, 2020). This could produce a divide between media organisations that provide English spoken and written content compared to other outlets that produce content in the local mother tongue – this will be a significant challenge for smaller languages where the media landscape is smaller and developing such tools will be expensive. To succeed it will, therefore, require much more industry-wide collaboration and the inclusion of research institutes to develop such tools and make them available. One example of this problem could be solved, was a project organised by the Danish New agency Ritzau who together with 13 Danish media organisations and Swedish developer iMatrics developed a tagging system that would be trained on a Danish dataset provided by the news organisations (the supplied articles would be tagged by hired annotators according to a developed annotating structure) (Ritzau Bureau, 2020).

EXAMPLES OF AI SUPPORT TOOLS FOR PRODUCTION

'Grammarly' is an AI driven grammar and spell check tool that can be used, for example, through a plugin in a browser or in word and support spelling in English.

The **'Editor project'** run by The New York Times focuses on automating the annotation and tagging processes of content. Processes that can enable better recommendations of content, search engine optimisation and add targeting.

The tool **'Panels'** offered by GettyImages uses AI to recommend appropriate visual content for an online article.

Figure 5: Examples of AI support tools for production

3.2.2 AI in written media content production

The automation of written content is one of the AI applications that has the longest history in the media sector, dating back to the 1970s where computers began to produce weather forecast reports (Graefe, 2016). Over the last decade, the use of AI in the production of news in data-heavy domains has become increasingly popular, beginning with areas such as financial news and sports reporting where reliable data is readily available, but examples of uses in crime coverage, earthquake warnings, and recently also entertainment or 'gossip' news are also evident (Young and Hermida, 2015; Haim and Graefe, 2017). This utilisation of AI was in 2015 listed as a top newsroom trend by The Worlds Editors forum and optimistic forecasts by AI-service providers in this field have stated that half or up to 90 percent of all news content will be automated in the future (Graefe, 2016). However, this is still a field that is developing and with a limited number of providers (for an overview, of service providers for the media sector, see Dörr, 2016; Graefe, 2016).

This application of AI has already been discussed under a range of more popular names, such as 'robot' or 'automated' or 'algorithmic' journalism (see e.g., van Dalen, 2012; Carlson, 2015; Dörr, 2016), while also sometimes referred to through its more technological name, namely Natural Language Generation (NLG). NLG is commonly defined as 'the (semi)automated process of natural language generation by the selection of electronic data from private or public databases (input), the assignment of relevance of pre-selected or non-selected data characteristics, the processing and structuring of the relevant data sets to a semantic structure (throughput), and the publishing of the final text on an online or offline platform with a certain reach (output)' (Dörr, 2016: 3). In short, this means, that the AI system scrapes existing data sources (e.g., sports results) and produces an article based on an interpretation of that data, but beforehand it is trained on journalistic sample articles for it to both learn the right tone and to understand interpretative patterns (e.g., in the case of sports reporting to understand what different scores might indicate, such as a close game or big win) (Graefe, 2016). Several major potentials are connected to this application of AI in the media sector, such as **faster**, **more precise, and extensive news coverage**, the **personalisation of content production** and an overall **improvement of the news coverage by allowing journalists to spend their time on quality reporting rather than routine tasks.**

To begin with the latter, this **potential of journalists having more time to do in-depth reporting** and work on their stories is highlighted across the reviewed material. One of the major service providers of these AI solutions, Automated Insights claimed to be freeing up 20 percent of Associated Press' sports reporters time when they helped them implement an NLG system. The system automated the production of NCAA Division I men's basketball previews basketball, allowing the reporters to spend their time on in-depth stories relating to the games. Such AI applications could, therefore, have positive societal impacts as they could support an overall enrichment of the media environment as more in-depth reporting could be made available.

A second potential is **increasing the speed, scale, and accuracy of the news coverage.** The use of AI for content production allows media organisations to cover, for example, quarterly earnings the moment they are released. As an example, Associated Press was able to release an article on Apple's quarterly earnings, only minutes after they were released. This was possible using an NLG service provided by Automated Insights, which produced the stories based on data from Zacks Investment Research (Graefe *et al.*, 2018). The 'Quakebot' developed by The Los Angeles Times, which is based on data from the U.S. Geological Survey's Earthquake Notification Service, produces a brief article relating to the earthquake including the location, magnitude and time of the quake, also received media attention in 2013 when it was the first to break a story of an earthquake in Southern California (Graefe, 2016). Illustrating how this can both **allow media organisations to be first with a story, which can have positive economic effects**, but also **provide audiences with even more timely information**.

The automation of quarterly earnings at Associated Press also enabled them to produce 3700 quarterly earnings reports on both US and Canadian companies in a year (Thurman et al., 2017), which was a more than tenfold increase in reports (Graefe, 2016), illustrating the scaling potential related to automation of content production. While an ever-increasing amount of media content might not be a purely positive thing, as we discuss in the following part of this section, **scaling can allow for a fairer news coverage**. This was the specific aim of the Los Angeles Times 'Homicide Report' project, which was initially launched as a blog written by journalist Jill Leovy back in 2007 and later was leveraged by an AI system that scrapes reports from the coroner's offices to report on all murders in L.A. County (Young and Hermida, 2015). The project was launched to counter the general tendency in crime reporting to cover the sensational murders and the murders of affluent white citizens – the LA Times, for example, previously only covered ten percent of the annual murders – by covering all murders. The aim was to provide a fairer and more comprehensive coverage and to illustrate how the majority of homicides in fact affect other societal groups such as the black communities (Young and Hermida, 2015). However, doing this with manpower proved to be too labour intense a task, but could be achieved with AI and today all murders in L.A. County are reported on as soon as the data is available in the L.A. County coroner's database and are also added to an interactive map of the area (Young and Hermida, 2015). This use of AI, therefore, holds the potential to in a unique way mitigate media bias, by allowing for a more comprehensive coverage and ensuring that minority groups (racial, gender etc.) become better represented in the news.

Scaling also enables **serving wider audiences with news and potentially catering to underserved communities or localities**, which would normally have been economically unfeasible. An example of utilising AI for this specific purpose is the 'Heliograf', which was developed and implemented by the Washington Post to initially cover the 2016 Olympics in Rio by producing articles when the US won medals, but which is now also used to cover local high school sports and most recently also to update podcasts with result changes and deliver state-specific results based on the listener's location during the 2020 presidential election in the US.

The French newspaper Le Monde also utilised AI in this manner during their 2015 election. Here they collaborated with NLG service provider Syllabus to automate election reports in an effort to make the reporting more locally sensitive as they could now produce local pages for each of the 36000 cities on the electoral map (Beckett, 2019). This use of AI holds great democratic potential by providing citizens with valuable information on a local level, something that is especially important in a media landscape characterised by increasing consolidation and diminishing local coverage. The leveraging of AI could, therefore, have important social and political impacts in increasing the local participation in politics and ensuring citizens in 'outlying' areas to make them feel they are also part of the media agenda, which could counteract the growing societal polarisation between cities and the rural areas, which is a growing problem across the globe.

A relating potential is how AI in this context could also be leveraged to **provide not only more local content but also personalised and on-demand content**. Andreas Graefe (2016), for example, describes how Automated Insights already produces individual player reports for all players in Yahoo Fantasy Football (a highly popular online game) and how Narrative science can tailor financial reports to the individual costumers wishes. Another example is PersaLog, which is an AI system that personalises the produced media content to the user's location so that the temperature measure in the article is adapted to degrees for a reader in Europe and Fahrenheit for a reader in the US (Adar *et al.*, 2017). Graefe (2016), hypothesises how sports reporting could even include tonal differences for a won or lost game depending on the readers sport team affiliation and how users would be able to ask questions based on the content and receive more information on the topic that was generated by an AI, which could support the usefulness of the site and perhaps make news more attractive to a wider audience, something we discuss in the next section on distribution where chatbots are already providing similar functions based on pre-scripted content.

Last, another potential positive effect of AI in news production **is increased accuracy in the reporting**. Andreas Graefe (2016) describes how the former vice president and managing editor for entertainment, sports and interactive media at the Associated Press, Lou Ferrara, said to him in an interview how automation had decreased the error rate from about seven to about one percent and that 'the automated reports almost never have grammatical or misspelling errors' and 'the errors that do remain are due to mistakes in the source data.' (Graefe 2016: 19). Accuracy is important for media organisations to remain trustworthy to the audience, particularly in today's media environment characterised by increasing amounts of 'fake' or misleading content. Reducing errors is, therefore, something that is highly value. We return to the latter comment on how mistakes are related to the source data below in the discussions on challenges, as it can sometimes be difficult to achieve a balance between speed and accuracy.

Challenges of AI in written media content production

To begin where we ended, the accuracy of AI in content production is praised, but errors do occur – in fact, more often than one might think, as Graefe (2016) illustrates in his report by stating how thousands of automated articles have been corrected post-publication. These errors are most often a result of errors in or missing data in the source database from which the AI system generates its story or due to a misinterpretation of the data. Graefe (2016) uses the example of the second quarter earnings report for Netflix in July 2015 to illustrate how this can be problematic as an automated article wrongly reported that Netflix had not met their expectations and that the share price had fallen by 71 percent. In fact, Netflix had more than met expectations, but the technology had missed this as it had not correctly interpreted the information that Netflix's shares had undergone a seven-to-one split (Graefe 2016: 24).

Such **errors in reporting can have negative impacts both for those who are being reported on if not corrected quickly and for the overall credibility and reputation of the media organisation** who published the report - something that could be highly detrimental in the current media landscape, which is already characterised by declining trust in media organisations (Schiffrin, 2019). This use of AI in content production, therefore, requires the need for mitigative measures against such errors, which include oversight and verification of the generated reports - the continued need for a 'human-in-the-loop' (Milosavljević and Vobič, 2019; Pasquale, 2020). However, that too is challenging because as Graefe (2016) shows even when there is human oversight, errors can happen, such as in case of the 'Quakebot' that wrongly reported three earthquakes in 2015. Here the editor in charge of verifying the data before publishing the article placed too much trust in the AI system to be 'right' and did not in fact interrogate the data. It will, therefore, be important that the human oversight remains not only there but also critical towards the system and do not become over-reliant on the AI, adhering to the existing ideals for source verification in journalism (Thurman, Dörr and Kunert, 2017). However, this raises a new challenge as **it can be difficult for, for example, editors to be able to in a meaningful way evaluate the results produced by the AI system** and in fact challenge its results due to the opacity and complexity of these systems. A challenge that has led to an increased focus on the development of 'explainable AI' and human interface design (for a good overview on the discussion on explainable AI, see Wieringa, 2020).

The potential of scale also enhances this challenge, as **the sheer amount of automated content produced makes it difficult to sustain human oversight**. Associated Press, who uses Automated Insights Wordsmith service to automate text creation for quarterly earnings reports no longer monitors the produced content as it was too time-consuming for editors (Dörr and Hollnbuchner, 2017). The automated content, therefore, becomes exempt from the traditional oversight of editors to ensure quality and accuracy, which can further enhance the risk of mistakes going unchecked for longer time. To mitigate this challenge, Graefe (2016) suggest a detecting system that can identify outliers and have these go through editorial control to ensure higher accuracy while still retaining the efficiency benefits related to this AI application.

When **errors occur in automated content new legal questions are also raised, such as questions of responsibility and liability**. Historically, responsibility and liability have been connected to the individual journalist and editor, and in cases of anonymous articles extended to the media organisation (Dörr and Hollnbuchner, 2017). However, this 'simple' attribution of responsibility becomes more difficult with automated content, where new actors become part of the equation, such as data scientists involved in building an in-house system or external AI system providers, such as Automated Insights or Narrative Science (Dörr 2016; Dörr and Hollnbuchner, 2017). Monti (2019) describes how the 'easy' solution would be to make the automated content the responsibility of the editor or the media organisation as has been the case with anonymous articles, but also how that misses the important question of the programmers or external service liability in case of deformation. Here he highlights how editors can be held accountable for mistakes in the content output, but that if the fault is in the code (e.g., a persistent bias), then liability should be extended to such actors. Arguing for the need to develop standards to ensure such liability is clear as well as ethical guidelines for programmers working within the media sector. Dörr and Hollnbuchner (2017) also raise the question of how liability should be determined in cases where the collected data might infringe personal data and privacy rights as well as copyrighted material and thereby be in breach with legislation. Yet, another perspective is considered by Seth Lewis, Amy Kristin Sanders, and Casey Carmody (2019) who based on the context of the US First Amendment protection of speech, argue how it is with the current legislation almost impossible to win libel suits against a case of automated content. This produces a situation where libellous content produced by AI systems can go unchecked and where the rights of those who the libellous content harm are at risk.

The potential negative effects of errors in automated content production are also enhanced by the fact that experimental studies, both smaller and larger in scale and with different languages (German, Swedish and Dutch), have shown that readers find machine written content more credible than human written content, while the human-written ones are considered more readable (Clerwall, 2014; Van der Kaa and Krahmer, 2014; Haim and Graefe, 2017; Graefe *et al.*, 2018; Waddell, 2018). This is both the case if the source is intentionally declared wrong or is unknown to the reader and across all experiments, the differences in evaluations of the content were minimal, illustrating how it is becoming almost impossible to distinguish between humans and machines within this type of routine reporting. Haim and Graefe (2017) and Graefe et al. (2018) also consider how higher or lower expectations toward the human or machine might factor into the evaluation, which was questioned in the first studies as a possible reason behind the positive evaluations of credibility for the machine written articles, but the results remain the same. In fact, the opposite was the case as machine written texts that in the experiment were declared to be written by a human author were often evaluated worse (Graefe et al. 2018).

The fact, that it is becoming almost impossible to distinguish between human and machine produced content, Dörr and Hollnbuchner (2017) also argue **enhances the need to have full disclosure procedures and increased transparency of how the content was produced**. In the current proposal of the AI Act, it is also proposed that providers of systems will have to ensure that the systems they provide to users (media organizations, journalists, editors) are transparent as to whether they expose people to a machine produced content if it is not obvious (see ‘Article 52’). However, there is no mention of demands to the level of disclosure, meaning it can merely be a sentence stating, ‘this content was generated by an AI’, which is already the case in many instances of this use of AI. A related question is how to assign authorship to an automated article, a question that is raised by Montal and Reich (2017), which much like the liability question requires rethinking. Bylines and authorship have a long history in the media sector and are also connected with professional prestige, but now that articles no longer always have a human author, new policies for authorship are needed (Montal and Reich, 2017). In their study Montal and Reich (2017) show how there are discrepancies between how media organisations discuss this issue (based on interviews) where they underline readers right to transparency of the source, and their actual disclosure policies. Based on a quantitative content analysis they identify four levels of transparency ranging from full (byline by media organisation, algorithm (bot) or service vendor with a methodology description incl. e.g., data source, service vendor or programmers name), partial (byline by service vendor), low (byline by media organisations) or no transparency (no byline or by human reporter). The three latter have no further description of methodology or data source. Based on this rather varied approach taken by media organisations, the authors call for the need for policies to amend this and ensure increased disclosure transparency (how the story was selected and produced (see e.g., Karlsson, 2010) and algorithmic transparency (the specific methodology, construction, and limitations of the algorithm (see e.g., Diakopoulos, 2014). Monti (2019) highlights how it will also be important for the reader to be aware of if the data is sourced from a political actor, as this will be necessary for them to critically examine the content they are reading, compared to if the data is sourced from a governmental database, which is associated with higher degrees of objectivity. New forms of source ethics, therefore, also must emerge as this application of AI matures.

The above challenges have focused on the interrelated challenges emerging from having this new actor produce content and potentially false content from a more organisational view. Another string of scholarly work has focused on the challenges and limitations that media professionals see as prevalent and how they challenge for example the labour market or professional ideals. While the fear of potential displacement has been palpable in the media headlines, which is understandable with the steady decline in journalistic jobs over the last 30 years (Linden, 2017), media professionals in general do not see automated content as a major threat to their jobs as these systems remain limited to routine forms of reporting where structured data is available (see e.g., van Dalen, 2012; Carlson, 2015; Graefe, 2016). This type of reporting is specifically seen as not being able to deliver the important democratic function of news of delivering critical stories of societal matters (Thurman et al., 2017). Particularly, media professionals highlight that AI systems with their one data stream cannot provide nuanced multisided accounts of events, which is a key ideal for media organisations and that they do not interrogate the data, as the interesting story might lie in an anomaly in the data, which the AI system will not detect (Thurman, Dörr and Kunert, 2017). This relates back to the challenge discussed in the previous section of giving these systems a 'nose for news', because while a seasoned sports reporter might easily identify the story during a sports game, formalising this 'gut feeling' has proven difficult (Graefe, 2016). A sportswriter from the study by Neil Thurman, Konstantin Dörr and Jessica Kunert (2017) also highlights how when covering a football match, what happens on the pitch (i.e., the data stream the NLG technology uses) might not always be the most important information. It could be the riot that happened in front of the stadium minutes before or historical background of the match or future outcomes such as what the result might mean for the future of the club. Such limitations of AI in content production illustrate it is unlike, at least in the near future, that humans are fully replaced by AI systems, as some functions will remain necessarily 'human'.

However, studies have shown how automated content production induces the media professionals to reinterpret what skills matter for the profession as well as how genres and practices of writing will develop in the future (for the two latter, see e.g., Young and Hermida, 2015; R. Jones and Jones, 2019). Arjen van Dalen (2012), for example, found how journalists in the meeting with automation increasingly highlighted skills such as creativity and analysis over objectivity, preciseness, and speed, which have previously been core parts of the legitimising discourse and skills of journalism. As van Dalen (2012) argues, the journalists are in fact making journalism 'more human' by increasingly legitimising it through skills that are uniquely seen as human (e.g., humour or emotion) but in that they are also changing what skills will be important for future journalists. Here studies point to how some skills are beginning to be valorised more, namely technical and data skills amongst media professionals (see e.g., Lewis and Usher, 2013, 2014; Young and Hermida, 2015). Carl Gustav Linden (2017) who authored the article 'Decades of automation in the newsroom. Why are there still so many jobs in journalism' also argues that one of the reasons behind the limited threat of automation in the media sector is the strong ideology of this sector (i.e., the values used to legitimise the profession). However, Matt Carlson (2015, 2018) has also critically discussed how automation in many ways does challenge the authority that, for example, journalism has in today's societies and how this move towards more subjective qualities as the legitimating qualities of the field could further weaken journalist position as being capable of both choosing and writing societally relevant news. This could negatively affect the trust in media organisations and their role as public information providers, which as we also discuss further perhaps will become even more important today were mis- and disinformation flourishes.

EXAMPLES OF AI WRITTEN MEDIA CONTENT PRODUCTION

The **'Quakebot'** is an AI developed by the Los Angeles Times, which tracks notices from the U.S. Geological Survey and produces articles about the reported earthquake.

The **'Heliograph'** developed by the Washington Post was used to report during the 2016 Olympics and the US elections in 2016 and 2020.

The **'RADAR'** developed by the Press Association and Urbs media produces local news at a scale to reach smaller communities.

Figure 6: Examples of AI in written media content production

3.2.3 AI in audiovisual media content production

The use of AI in fully automated production of TV or other audiovisual types of content (e.g., radio, podcasts, or mixed media) has been much more experimentally deployed (Rehm, 2020) – and there are limited examples of this in scholarly publications and reports. While AI very often is employed as a supportive tool in the production process, as discussed above, to, for example, produce 'rough cuts' that can then be edited by a producer, by adding subtitles or creating automatic summaries of movies, the full automation of audiovisual content is still not a common use in the sector. One of the few examples of such use that have been publicised was when BBC experimented with letting an AI loose in their archives to produce a programme, which was shown on BBC4 and presented by Hannah Fry (BBC Four, 2018). The experiment was aimed at illustrating how the technology works and spark public discussion.

Equally, in 2020 Forbes announced the introduction of the prototype of 'a fully automated, yet presenter-led sports news summary system' (Chandler, 2021). The tool is developed with the Start-up Synthesia and works quite similarly to the so-called 'deep fakes' by combining pre-recorded videos of a news presenter with live data from English Premier League football matches, so that the site can deliver match specific reporting without having a reporter assigned to every match. Similarly, to the potential of automated news content this can allow for a wider and more inclusive audiovisual coverage of, for example, sports matches, where economic consideration currently limit what matches are covered.

Challenges of AI in audiovisual media content production

As of date, the authors have not been able to find specific studies that focus on automated content production in the audiovisual sector (e.g., public broadcasters, in radio or podcasts), which makes it difficult to discuss both the potentials and challenges of this utilisation of AI. This in itself illustrates a challenge with this development, namely a gap in knowledge that must be further explored if we are to understand the full impacts of this technology.

In one of the few sources relating to this topic, the 2020 report 'The use of Artificial Intelligence in the Audiovisual Sector' requested by the European Parliament's Committee on Culture and Education, Georg Rhem (2020) describes how the challenges of AI in this sector is not connected to their own use, but rather to how other actors who might utilise these technologies in the production of fake content and deep fakes. This use is seen as a threat as it could lead 'to an influx of uncontrolled, low-quality and untrustworthy content' (Rehm 2020: 7). Illustrating how continuing to remain trustworthy as a media institution is at the core of the concerns relating to AI. Rhem (2020) also highlights how public broadcasters are more hesitant to implement these tools in their workflows, keeping them at the experimental level and that they see more potential in bettering existing tools in comparison to applying new ones in their production. This lack of deployment could be one of the reasons for this application having received less attention, however, with major media companies like Forbes now breaking the barrier and directly implementing a prototype, it will become more pertinent to understand the impacts and their implication within this area of application.

Within legal scholarship **the potential challenges of the automation of creative processes and content, such as the making of a documentary, might pose for copyright law** has also been discussed. Such concerns are prompted by the fact that computers are no longer used to enable humans to produce art, music or other cultural artefacts – rather humans are enabling computer to produce such artefacts on their own (for overview of the historical discussion from a legal perspective, see Bridy, 2012).

EXAMPLES OF AI AUDIOVISUAL MEDIA CONTENT PRODUCTION

The **'Made by Machine: When AI Met the Archive'** was an experimental documentary produced by the BBC and made available for the audience as a scheduled programme on BBC4 and later online.

Reuters have used AI to produce the first prototype for automated video reports; **'a fully automated, yet presenter-led sports news summary system'**.

Figure 7: Examples of AI in audiovisual content production

3.3 AI in media content curation and distribution

Historically, media organisations have been characterised by a linear distribution, meaning that content was distributed ongoingly in a predetermined manner where each audience member receives the exact same content at the same time. In the case of broadcasters this meant the scheduling of programs on specific channels, while the printed paper represented 'today's news' - a format that was carried into the online paper originally, where content was published in a more ongoing manner but still on one shared front page (Carlson, 2015; Sørensen and Hutchinson, 2018). Today, the use of AI in distribution have made it possible to move away from such linear forms of distribution, towards forms of curated and personalised distribution. The AI technology used in this cycle is most commonly referred to as 'recommender systems', which is a general container term, covering both different AI systems (e.g., content or collaborative filtering systems) and forms of applications. However, across these differences, recommender systems have the shared purpose of being a filtering mechanism that provides an audience with recommendations of content based on a pre-determined logic. As Robin Burke, Alexander Felfernig, and Mehmet Göker (2011) write: 'Recommender systems are tools for interacting with large and complex information spaces' (p. 13). With the ever-increasing amounts of available online media content produced by the media organisations themselves as well as the abundance of content available online via search engines, curation apps and social media, such tools are becoming widely used in many media organisations. In the following, we discuss:

- How media organisations use AI to automate content curation and distribution either on their own sites or on external platforms (e.g., social media),
- How they use AI for the personalisation of content distribution and the use of 'chatbots' to both personalise and make distribution more interactive,
- How AI is used in intelligent advertisement and subscription walls, which is also part of the audiences' experience of the distribution of media content.

3.3.1 AI in content curation and distribution

A trend amongst media organisations has been the use of AI to increasingly automate parts of their content distribution on their own online sites. This automation can take many forms, but a classic example is a 'trending articles' feed on online media sites that based on user data present the most viewed or shared content to the user (Thurman, 2011). Such feeds can draw on live audience data from the media organisations own site or data regarding how their articles are performing on social media, presenting the audience with a list of what is popular on either channel. Equally, news feeds that offer the audience the most recently published content also appear on many sites. Such feeds, therefore, recommend or curate content for users based on specific qualities such as popularity or recency to make it **easier for the audience to navigate the content and find relevant content, continuing to orient people towards a shared public discussion.**

There has also been more radical automation of online sites. The large Swedish news site Svenska Dagbladet has, for example, completely automated their online news site, so that the placement of all articles is decided by an algorithm. The journalists simply assign a value of importance and lifetime to an article once it is finished and then the algorithm organises the website ongoingly based on the values. Editor in Chief Fredric Karén has stated that this has **significantly freed up the time of editors and journalists to focus more important tasks of writing and developing stories**, rather than constantly updating the website. Equally, he ascribes the algorithm a core role in reversing the economic development at the media economic, which in 2013 when he took over was highly challenged (TechSavvy, 2017). This illustrates how AI is also quite equipped to curate the front page in a manner that the audience find appealing. Equally, broadcasters are beginning to experiment with AI driven scheduling to maximise audiences. An example of this can be found at Spanish National TV (RTVE), where they experimented with an AI model that could predict what time slot would yield the highest viewing numbers for different programmes (Cibrián *et al.*, no date). The potential here is to **maximise viewership of the content, which can both have economic benefits in the case of commercial broadcasters**, but in fact also supports the **democratic role of broadcasters by reaching a wider audience than previously.** This is one interpretation of the ideal of universalism, which has characterised PSM's and which we discuss further in the following section on personalised content distribution.

AI is not only used to curate content on the media organisations own sites, but also to automate the distribution on social media platforms, which have become highly influential distribution channels as discussed in the introduction. Danish media BT in early 2021 fired their 'community managers' who had been responsible for distributing the company's media content on Facebook and Twitter. The community managers were being replaced by an AI tool, which based on how users are interacting with the content and how articles are performing decides when the optimal time is to publish the content on social media (Bruun-Hansen, 2021a). The potential of AI in automated distribution is, therefore, similarly to the automation of content production, **the hope to become more efficient, and to give journalists more time to focus on writing and thereby improve the quality of the media content available as well as gain better traffic and engagement on social media by making the distribution data-driven, which have economic benefits** for the media organisations.

Challenges of AI for content curation and distribution

One major concern with AI in distribution and curation could be, as seen above, the displacement of humans, as the case of BT illustrated how AI did result in the existing positions of 'community managers' becoming obsolete, which also could indicate that not only routine tasks of media professionals could disappear with AI, as such community manager roles are considered very important by other media organisations. Due to that importance, this move by BT received a lot of criticism from competing media organisations, who raised highlighted the task of 'community management' is much more than simply optimising distribution times, it also involves taking part in and moderating the debate that follows in the commentaries. Abandoning this part of the distribution responsibility might risk producing negative experiences for the sources quoted in the content as offensive comments directed towards them will not be dealt with (Bruun-Hansen, 2021b). In the next section on AI and deliberation, we also discuss how such moderator engagement can in fact be beneficial economically for the media organisation. The chief editor at BT responded to the critiques by highlighting how they would use software to moderate the commentary debate on social media, which removes offensive comments as a measure against this negative impact of the technology. We also further discuss the use of such software in the following section. At Svenska Dagbladet, the newsroom was also minimised following the introduction of AI in their distribution and curation practices. While this was ascribed to the fact that fewer printed papers were to be produced in the future, it could warrant exploration whether this was also connected with the efficiency gained through automation.

Another challenge relating to the use of AI to automate, for example 'most read' lists on media organisations websites is highlighted by Eun-Ju Lee and Edson Tandoc (2017) (Lee and Tandoc, 2017) as they describe how such aggregated lists do affect what the audience reads on the website, driving reader patterns down a shared route. **While this is not only a potential negative impact as it can promote choices of reading that lie outside what the reader normally would read (selective exposure), it can also create virality consumption patterns**, where, as discussed under the section of AI in audience measurement systems, what is read shifts towards content that is, for example, more sensationalist. This could be detrimental if that means that societally important information is overlooked by the audience – and ultimately also affects what content is produced through the algorithmic feedback-loop.

More in general, these developments induce the need to rethink the responsibility of the media vis-a-vis the audience, as AI and other digital technologies allow the media to steer and influence individual media diets, more than what has been the case with previous distribution technologies (for more on this new responsibility, see Helberger, 2011, 2016). Rethinking this responsibility in turn requires media organisations to critically revisit central professional values that for long time have been taken for granted, such as what it means to inform the audience, or provide the audience with a balanced and diverse media diet. Furthermore, it requires the media organisations to find a balance between harnessing the benefits of these new technologies to better inform citizens while not slipping into new forms of paternalism, audience surveillance and manipulation. We revisit these critical questions in the following section on personalised distribution of media content.

EXAMPLES OF AI IN CONTENT CURATION AND DISTRIBUTION

Svenska Dagbladet automated the entire frontpage to make content curation and distribution more efficient and audience oriented.

Spanish National TV (RTVE) broadcaster use AI for intelligent scheduling to ensure higher viewership.

Danish Tabloid BT uses AI to distribute and moderate content on social media.

Figure 8: Examples of AI in content curation and distribution

3.3.2 AI in personalised content curation and distribution

The idea of personalising the media content distribution has been discussed since Negroponte's vision of the 'Daily Me' in 1995 and even before that in science fiction with visions of having media distribution being adapted to the user's preferences and potentially also their station in life (Helberger, 2016). Today the use of recommender systems remains a key trend amongst media organisations to personalise the user's online feed (Newman, 2018; Beckett 2019), so that each user encounters news personally picked out for them or recommended TV or radio programs when entering a broadcaster's streaming app or website. Neil Thurman and Steve Schifferes (2012) in one of the initial studies of personalisation defined 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 and/ or implicitly determined preferences' (p. 776). As the quote reveals, recommender systems can be explicit or implicit (also known as self-selected personalisation or pre-selected personalization, see Zuiderveen Borgesius *et al.*, 2016). In the following, we explore how both applications have been utilised in the media sector and what opportunities that has provided, but we also discuss a third application, namely 'chatbots', as another form of personalised distribution, but which has the unique feature of 'interacting' with the users through informal conversation.

The usefulness and potential of personalisation is generally discussed in relation to the changing media landscape and the changing user expectations and habits (Helberger, 2016). Three core arguments are often prevalent. First, **the abundance of content in the internet era has induced an experience of 'information overload' amongst users, which personalisation can help alleviate by easing the user's navigation process through content** (Sørensen, 2020). This problem has also been intensified by the potential of automated content production as discussed above, producing this adverse effect that further enhances the potential of personalisation as a solution. Second, with the rise of personalised on-demand streaming services (e.g., Netflix) and social media platforms, **users have grown to expect a personalised**

content distribution service (Helberger, 2016), which induces the need for more traditional media organisations (i.e., public service broadcasters) to stay relevant in this changing trend in content distribution. Third, media organisations face a range of new competitors in the 'marketplace for attention' with the rise of new powerful intermediaries (e.g., Google and Facebook) who utilise personalisation when providing search results or suggestions on social media feeds (Helberger, 2016). As much traffic has already moved into these platforms, **personalisation presents an opportunity to retain traffic on the media organisation's own sites**. The latter illustrates how the potential of personalisation is highly connected to ensuring a higher relevance of content for the individual, which will ultimately have positive economic impacts as users who are engaging with the content are more likely to pay for a subscription and can also support ad sales. Something that was highlighted by MittMedia who saw how they both could retain paying readers longer as well as increasingly 'convert' visiting readers into paying customers, a potential that is highly valued due to the general economic challenges faced by the media sector.

Another potential often linked to personalisation is **the possibility to empower users, allowing them more autonomy in their news selection** – something that counters the traditional and critiqued 'paternalistic' role of media organisations, as also discussed previously (Anderson, 2011; Helberger 2016; Sørensen 2020). Personalisation could also **increase the user's engagement with media content**, something that is also highlighted by Natali Helberger (2011) who argues that there is a risk of users relying too heavily on intermediaries (who are only commercially oriented) if the media sector does not help the users in their new responsibility of selecting content. It will, therefore, be important that media organisations, similarly to the intermediaries, focus on increased and eased accessibility to news. As Natali Helberger (2011; 2016) argues, media organisations could in fact also **support democracy further with personalisation by 'filling the gaps' in the audiences' reading habits**, by serving them with content that is highly relevant right now, but which the reader might normally not be interested in. Thereby, use data analytics to not give the user more of the same, but more diversity, which could support the very foundational ideal of particularly PSM's of providing the audiences with diverse and universal access to media content. However, there is a trade-off with personalisation, where users either might purposely avoid types of content or continuously be served more of the same based on their reading habits, which might not be productive for democracy and for the public debate as we discuss in the following.

The already mentioned 'Editor' project by the New York Times was also further developed, so that tags could be automated not on the article level, but rather at the component level, so a recipe or an event described in the article. **This will allow for new ways of intelligently recommending other content to the user.** At British broadcaster BBC they are also exploring 'atomised journalism' where content is produced as small stand-alone pieces, rather than one cohesive piece to better cater the users – where some might want longer articles and others shorter, as well as allowing more reuse of content for future content that deals with the same topic (Jones and Jones, 2019). Both uses show how AI allows media organisations to rethink how they structure content and their distribution of it to **potentially reach new audiences and cater them in ways that they find more engaging.**

The **increased accessibility and engagement with media content by users are also the core potential connected with 'chatbots'**[2], which here uniquely become connected to their quality of being conversational, as chatbots in their engagement with users mimic everyday dialogue (Jones and Jones, 2019). Chatbots are applications that can either be a 'pop-up' on a site (as it is often seen with e.g., customer service chatbots), or which can be embedded into existing private messaging services (e.g., messenger or WhatsApp). The latter is currently what is mainly explored amongst media organisations (Ford and Hutchinson, 2019). Just like with the applications of personalisation more generally, there are also different types of chatbots, some are automated or semi-automated and rule-based, meaning they have pre-scripted and pre-organised inputs that a programmer has prepared, which they base their conversation on. Others are more 'intelligent' and use, for example, machine learning to act in a more autonomous manner (see e.g., Jones, 2018). The former is still the one that is predominately used within the media sector (Jones and Jones, 2019; Ford and Hutchinson).

Heather Ford and Jonathon Hutchinson (2019) explore how the Australian Broadcasting Corporation (ABC) has been experimenting with a chatbot, which offers daily summaries of three stories: one of public interest, a local story, and a feel-good story. The bot first greets the users when they engage with it and when presented with the selected stories, the reader can then ask questions for more details relating to the context of an incident or explanations of a word or event. The content selection is based on the preferences the user has given when beginning to 'converse' with the chatbot but can be adapted along the way by using the button 'more like this' (see Ford and Hutchinson, 2019 for in detail description). British Broadcaster BBC is also experimenting with different chatbots on both Twitter, Facebook, and Telegram. These included different elements of conversation, for example, summaries of news, push notifications and a Q&A format (Jones and Jones, 2019). The core potential of chatbots is how they can **remediate the relationship with the audience through this more informal tone and create engagement with the content amongst previously 'disconnected' users**. Thereby, **supporting the universalist mission of media organisations**, by bringing more people into the public sphere and discussion (Ford and Hutchinson, 2019; Jones and Jones, 2019). Heather Ford and Jonathon Hutchinson (2019) also highlights that in the ABC case, the chatbot was successful in getting previously 'disconnected' users to engage with their content (particularly the younger audience). The core reason why was that these audiences often felt 'talked down to' by the journalists and media discourse in general, but the informal conversational format produced

[2] Other forms of 'bots' have also been discussed in literature, such as more general 'news bots' that curate news on, for example, certain topics (see Lokot and Diakopoulos, 2016 for discussion on these applications).

new forms of trust in the media and a willingness to engage. Illustrating, again how the paternalist and elite role of the media organisations is challenged by users who now prefer different formats.

Hilde Van den Bulck and Halvard Moe (2018) in their explorations of PSM's and their uses and strategies regarding personalisation, also find that a 'majority of PSM are moving in the direction of digital and algorithmic personalisation, which they see as a tool to realise universality in new ways' (p. 890). Illustrating how many of these organisations do not see AI as a threat to the core values of universalism and diversity, which have traditionally been understood as provision of diverse content (understood as both ensuring different perspectives on public events as well as catering to multiple groups in society) to all citizens in a national context (Van den Bulck and Moe, 2018), but rather a different way to attain these ideals. However, as we will see in the following, this reinterpretation of the ideals with AI also brings forward several new challenges and is highly dependent on the design of these AI systems.

Challenges of AI in personalised content curation and distribution

To begin where we ended, personalisation of content distribution is in general considered to **produce new tensions relating to the core values of diversity and universalism**, which have historically been seen as central to media organisations and particularly Public Service Media (Helberger 2011; Sørensen and Hutchinson, 2018; Van den Bulck and Moe, 2018; Sørensen 2019). The very same values that were discussed above as potentially being reinvented or complemented with personalisation, but where **strong critiques have been made against inducing too individualistic a media consumption and producing 'filter bubbles' as well as how personalisation can infringe user privacy and the autonomy of users** (Sørensen, 2020). Natali Helberger (2015) nicely summed up these challenges when she argued that media organisations are 'at a crossroad where they must decide how personal, persuasive, and responsive their relationship to the audience should be, and what safeguards are needed to preserve autonomy, privacy, and the public sphere' (p. 1325). We in the following dive into these discussions.

The most persistent discussions regarding the threat of personalisation have concerned the production of so-called 'echo chambers' (Sunstein, 2004) and 'filter bubbles' (Pariser, 2011). The notion of 'echo chamber' specifically highlights the threat that users will only select and seek out information that aligned with their own view of the world now that information is readily available everywhere, locking them into reinforcing echo chambers of information (Sunstein, 2006) (also discussed as selective exposure in communication literature, see Zuiderveen Borgesius *et al.*, 2016). While the notion of 'filter bubble' points to the same threat, it focuses specifically on the role of the algorithmic filtering mechanisms underlying the content distribution of particularly the emerging intermediaries, such as Google and Meta, which have the power to isolate users in an information bubble with others that share their worldviews (Pariser, 2011). Both concepts, therefore, focus on the threat of losing diversity in the media consumption, something that historically and today is considered essential for democracy and social cohesion, and that has traditionally been the provisional responsibility of media organisations - particularly for PSMs.

The associated **negative impacts of both echo chambers and filter bubbles are generally considered to be an increased societal fragmentation and polarisation** as people due to this constant reinforcement of their opinions will develop more extreme viewpoints, inducing the distance between societal groupings (Zuiderveen Borgesius *et al.*, 2016). However, in the last few years, there have been several critical voices entering into and nuancing the debate regarding these concepts, emphasising how there is no empirical evidence supporting the algorithmic production of 'filter bubbles' (see e.g., Zuiderveen Borgesius *et al.*, 2016; Bruns, 2019) and that the 'inherently cross-media' consumption of users (Schrøder, 2011) also makes it unlikely that users will not be exposed to a range of media content (Zuiderveen Borgesius *et al.*, 2016; Möller *et al.*, 2018). Other studies have also shown how it is not media consumption, but other factors such as the national political system (e.g., two- or multi-party system) (see e.g., Zuiderveen Borgesius *et al.*, 2016; Dahlgren, 2020) or economic decline and rising inequality (see e.g., Stewart, McCarty and Bryson, 2020) that induce political and societal polarisation. Others have highlighted that recommender systems today are still rather crude and imprecise in their predictions and, therefore, do not pose an immediate threat of producing filter bubbles as a result, as users will likely still receive rather random suggestions (Zuiderveen Borgesius *et al.*, 2016; Sørensen, 2020). However, **scholars still caution that as recommender systems mature further and become more precise, it might be necessary to revisit these concerns** (Zuiderveen Borgesius *et al.*, 2016; Sørensen, 2020). Particularly a focus on how feedback-loops are created within the systems will be vital to understand whether the content path become inclosing and potentially harmful (for more on feedback-loops, see Bozdag, 2013; Knott *et al.*, 2021).

In terms of the wider discussion of how recommender systems might negatively affect the role of media organisations in providing a diverse coverage and living up to the ideal of universalism, scholars have in recent years also nuanced the debate. Judith Moller, Damian Trilling, Natali Helberger and Bram van Es (2018), for example, point to how the evaluation of diversity measurements varies depending on whether you ask a social scientist or computer scientist, and they also argue that a recommender system might in fact be less biased compared to a human editor. Jannick Sørensen (2020) makes a similar argument regarding universalism, stating it very much depends on what understanding of universalism is adopted and the potential negative impacts are highly dependent on what understanding of democracy underlies this understanding (e.g., procedural, deliberative, or participatory). The latter is also highlighted by Natali Helberger (2019) who in her analyses shows how different threats and opportunities can be identified regarding recommender systems depending on the theoretical understanding of democracy that underlies this evaluation. This discussion illustrates how the question of whether personalisation will have a positive or negative societal impact is not a simple one, rather what becomes evident is the need to find ways of designing these systems in ways, so they are supportive of democracy, but here too, a range of challenges and new questions arises.

Jannick Sørensen and Jonathan Hutchinson (2018) raise the key question of whether personalisation systems will be developed so that they embody the existing editorial values and ideals or mimic their commercial originators (personalisation systems originate from commercial organisations like Amazon, see Smith and Linden, 2017), or they will integrate media organisations even more into the commercial media ecology they are already reliant on (e.g., Google and Meta). Thereby, potentially pushing media organisations into a more commercial direction, where the potentials of AI to support democracy might be undermined in this context. Balazs Bodó (2019) in his study shows how media organisations do have a strategy oriented towards developing media specific systems when adapting these AI systems, but **the question of whether they succeed in translating the ideal and values of the media organisations into the AI system, remain unanswered** – something that scholars foresee will be a difficult task (see e.g., Dörr and Hollnbuchner, 2017; Sørensen and Hutchinson, 2018). In their ethnography of the development of a recommender system in a large regional news organisation in Denmark Anna Schjøtt Hansen and Jannie Møller Hartley (2021) point to how this does prove highly difficult during the process of development, where core values of societal importance, timeliness and localness must be reconfigured. While this is not necessarily a negative thing, as Fabian Muniesa (2011), with reference to algorithmic systems in the stock market, points to as this forces the involved to explicate and reflect on normally implicit values, it, however, in this instance produced new power asymmetries as the data scientist often ended up with the final say (Schjøtt Hansen and Hartley, 2021).

The **challenge of AI producing new power asymmetries** is also highlighted by Jannick Sørensen (2020) who has shown how there are innate tensions between the visions of diversity by editors and computer scientists. A growing power asymmetry between editors and computer scientist, could therefore skew the prioritisation of editorial values in the process of development. In the report 'Google, the media patron - How the digital giant ensnares journalism', Alexander Fanta and Ingo Dacwitz (2020) also shows how technological innovation projects in media organisations are predominately run by and decided upon within the technical or marketing departments in the media organisations, not necessarily involving journalists or editors. This can produce negative social and political impacts in society if the balance is skewed too much and the ethical and societal considerations of editors are not considered in the design of recommender systems, threatening how the balance of measures of diversity, for example, are decided upon. Also, because the computer scientists are often also connected to the commercial part of the media organisations, which could induce an increased focus on more commercial values at the expense of editorial ones, reraising and intensifying the historic conflict between the commercial and editorial departments (see e.g., Willig, 2011).

This potential for an **increasing power imbalance is often connected to the opacity of algorithms** (see e.g., Gillespie, 2014), which was also the case in the study by Schjøtt Hansen and Hartley (2021), where the involved editors found it difficult to really assess the results of the system and whether they were 'good' or 'bad' and what then should be changed in order to make it ‘better’, highlighting the problem raised above on the challenge of ensuring the realisation of a human-in-the-loop in a meaningful way. In this media organisation, the solution to maintain control became to keep editorial control over certain placements of the site and implement a mechanical filter that would ensure that the recommendation of the AI systems would not be too old or that a certain amount of it would be local content. This approach to upholding the democratic ideals of the media organisations and control is currently quite common amongst media organisations (Sørensen, 2020), as also the example with MittMedia illustrates. While this is a practical way of finding a balance, it does not help to answer the question of what makes a recommender system ‘good’ or democratic enough. A question that is cause to much negotiation and speculation in the media organisations but have yet to be systematically covered in research beyond the discussed notions above of how it could, for example, be utilised to enhance the exposure diversity. It would, therefore, be fruitful to gain more insights into how such measures are decided upon within the media organisations and attempt to develop some best practices to help guide future development processes.

The potential for an internal power shift in media organisations, with ‘technical’ staff gaining a more prominent voice is not the only power shift that is discussed. There is also a focus on the **external dependency on intermediaries in these development processes** (see e.g., Lindskow, 2016; Sørensen and Hutchinson, 2018; Van den Bulck and Moe, 2018; Schjøtt Hansen and Hartley, 2021). Here, for example, the continued reliance on external data collections systems or readymade solutions of recommender system from external parties, also raise critical questions, because **media organisations are becoming reliant on existing (often commercially oriented) measures and systems.**

Another highly discussed challenge posed by these different applications of AI in personalised distribution is how all processes depend on user data – either given explicitly or implicitly, raising serious **concerns over data protection and privacy** (for overview of this discussion, see Zuiderveen Borgesius, 2014). Explicit personalisation has often not worked as well as expected as users either do not want to spend energy on customising their site or they forget to maintain it as it changes (Sørensen, 2013; Kunert and Thurman, 2017), which is why implicit personalisation have become more dominant, but this way of personalising raises serious privacy and data protection questions as the data is often tracked, without the user's awareness. Here several concerns have arisen. First, **this produces increasing information asymmetries between the user and the media organisations who now have ever-increasing amounts of data pertaining to the individual**. Second, it is often **not transparent in what ways the user data is used both by the organisations and by third partie**s. Third, it is very **difficult to ‘track the trackers’ and ensure accountability** (Bodo *et al.*, 2017). This is also why regulatory measures such as ‘informed consent’ are being challenged as it is almost impossible for users to in fact give informed consent and it is very difficult to take part in society if you, for example, deny cookies or giving login information (Zuiderveen Borgesius, 2015). This issue was also raised by Natali Helberger (2013, 2015) regarding the ‘Dutch cookie wall’ incident, where the Dutch public broadcasters had implemented ‘hard’ cookie walls, which only left users with the choice of handing over their private data or not having access to media content, which was supposed to serve the public. The practice was objected by the Dutch Data Protection Authority who argued this limited the individual’s access to societally relevant media content and that the broadcasters had a monopoly in occupying this role, making it a breach in the individual's right to information (Helberger, 2015). This incident illustrated the growing asymmetry between users and information providers when it comes to data practices.

This is part of a much larger discussion as the trackers, algorithmic monitoring and collecting data have become a common practice across multiple sectors (Bodo *et al.*, 2017). What will be a **significant challenge for media organisations (as well as other publicly trusted institutions) is how to maintain trust in them as an organisation**, as users feel increasingly under surveillance and exploited. This could severely hurt the trust and relationship to the user, something that is uniquely important in today’s media landscape where trust in media and news is generally declining. Thereby, **undermining the media organisation's ability to ensure public information is not only available but also trusted and used as a source to be publicly informed as a citizen.** The focus on this challenge can also be detected in the media sector, where the BBC, for example, this year have announced that they are exploring new ways for users to be in better control of their data through ‘personal data stores’ (Sharp, 2021). Compared to other commercial media platforms such as Facebook and Google, who must ‘simply’ abide to regulation, media organisations as trusted institutions have a larger professional responsibility in finding solutions to the privacy challenge so that it is transparent to users how their data is used and where users can in fact have access to media content without necessarily being tracked. Something we return to in the following section where we explore commercial models in media organisations that also utilise personalisation.

One last, **emerging challenge relating to recommender systems are the user attitudes toward them.** When The New York Times first announced they were implementing the ‘Recommended for You’ box, it resulted in highly critical reader comments. Irene Costera-Meijer and Tim Groot Kormelink (2014) have equally found that users generally prefer articles by journalist and find particularly explicit personalisation to be too much effort. They also are too afraid of deselecting content as they fear missing out on important news. However, recommender systems and particularly implicit systems are connected to higher rates of conversion and engagement as the example of MittMedia shows, which could indicate that they in fact are more successful than the users might think. Neil Thurman, Judith Moeller, Natali Helberger and Damian Trilling (2019) also showed that users generally found article suggestions based on their past browsing history helpful, but there were significant concerns about privacy and trust in news. The latter also was clearly evident in a report by Arjen van Dalen from the University of Southern Denmark, which based on a survey in Denmark showed that Danes generally do not trust algorithms to select their news. Only, if the algorithms were working in collaboration with journalists or editors, did the majority of the users asked consider them beneficial (van Dalen, 2020).

The many user opinions of personalisation, also again illustrate **the need for disclosure and transparency practices in media organisations**, where it might not always be clear what is selected by an AI or an AI together with a media professional. The study by Arjen van Dalen (2020) illustrates how disclosure and transparency practices might in fact be beneficial for media organisations as the explanation of how journalist and editors are involved in the determining the recommendations (e.g., by giving timeliness values or by curating part of the page), might in fact reinduce more trust into media organisations.

EXAMPLES OF AI IN PERSONALISED CONTENT CURATION AND DISTRIBUTION

The **'Recommended for You'** box on The New York Times site, which offers unique recommendations to each user.

Full personalisation on Swedish regional news organisations MittMedia's sites, except for the top three placements.

The **ABC chatbot**, which converses with users over messenger and offers three pieces of news catered to them.

Figure 9: Examples of AI in personalised content curation and distribution

3.3.3 AI in intelligent subscription models and advertisement

Another way that media organisations are leveraging AI is through intelligent subscription models. While not directly distribution of content, the appearance of, for example, a paywall hinders access to certain forms of content and is part of the media organisations distributional strategies. The 2019 report by Charlie Beckett uses the example of how the Wall Street Journal's (WSJ) used machine learning to introduce a dynamic paywall, which allows different visitors different levels of access to the site depending on their likelihood of subscribing – a judgement that is made based on 60 variables including among other their preferred content type, length of visits and frequency of access and favourite device. With this use of AI, the WSJ moved away from the 'hard' paywall, where only subscribers could enter the site, to a now more dynamic model, that only present the paywall to the user at the moment the algorithm predicts the user will be most likely to prove willing to enter into a subscription (Wang, 2018). Some might, therefore, immediately, encounter a paywall, while others might browse several articles before encountering the paywall. New users can also be offered a 'guest pass' allowing them more access to the site in exchange for an email address, allowing the WSJ to better collect data on the user for the future.

A similar approach is also taken by Scandinavian publishing house Schibsted, which based on the activities of users who are logged in on one of their many online sites (based on 15 variables based on data from existing users who became subscribers), predicts the likelihood that they will become a paying subscriber (Corcoran, 2018). In both cases it has proven successful in **ensuring higher numbers of digital subscribers, which indicates an important economic potential for media organisations**, who are currently navigating the changing media landscape in search of new and more sustainable business models. Beyond personalising the access to the site, Schibsted has also used the prediction model to target groups of users who are predicted to be more likely to sign up for a subscription with ads on Facebook – a strategy that also has proven fruitful (Corcoran, 2018).

The example of targeted ads by Schibsted illustrates another very common use of AI in media organisations business strategies (Beckett, 2019). Also in the broadcasting world, where, for example, Korean broadcasters are increasingly inspired by the 'Smart AI Programming System' developed by Korean retailer Lotte. A system that automatically creates a schedule for its segments on Lotte Home Shopping (a TV shopping channel), based on a forecast of the timing and sales volumes of the different products (Lee, 2020). **Making it possible to schedule adverts in the TV-program more successfully**. In these systems media organisations are utilising the same data foundation from the subscription models or personalisation discussed above, the marketing departments of media organisations to not only target users with ads regarding subscription but also with product ads from external businesses. Historically, media organisations have been dependent on advertisement revenue and the use of AI to behaviourally target users with ads is a response to how this has become the general approach to advertising (Boerman, Kruikemeier and Zuiderveen Borgesius, 2017). An approach that was initiated by their now biggest competitors on the advertisement market, Google, and Meta and which have led to personalised ads becoming a condition of the ad market. As revenue from ads remains an integral part of the business model of many media organisations, they have also followed the trend in the market and are also increasingly relying on targeted advertisement using AI.

Challenges of AI in intelligent subscription models and advertisement

These AI applications have similarly to personalisation, **raised concerns about personal data and privacy**, as the personalisation of paywalls and advertisement is equally as data dependent as the personalisation of content distribution. However, in this commercial context they are intensified because, as Joseph Turow (2005) has shown, the commercial departments in media organisations play a key role **in pressuring the tracking practices to their limits to tailor commercial messages.** The potential negative impacts on public trust in media organisations, discussed above, might be further enhanced based on the countering goals of the commercial departments of optimising sales and the democratic values that guide media organisations, which, as also discussed above, might become more skewed with the implementation of AI. Interestingly, Dutch national broadcaster NPO in 2020 in fact stopped using targeted advertisement and saw an increase in revenue, rather than a decrease. As this is a single example and other factors might have been in play, it is impossible to generalise that this would be the case for all media organisations, but it does show that targeted advertisement might not be the only viable advertisement strategy and that others less privacy infringing might also work (Anderson, 2020).

The severity of this challenge can also be seen in the light of diminishing transparency in the media sector – a sector that historically was open about its owner relations, economic models and potential income, its employees and its media products, but now much of this information is no longer as publicly available, inducing increasing information asymmetries (Bodo *et al.*, 2017). **The lack of transparency of how user data is used for both editorial and commercial purposes, therefore, pose a serious threat to the credibility and trustworthiness of media organisations** and here the commercial pressure can play a unique role, which must be explored further and also addressed. This increased blurring between the editorial and commercial is also highlighted by Jessica Kunert and Neil Thurman (2019), who also describe how content personalisation features (e.g., a recommendation box) now also sometimes include ads, which are also presented in the style of editorial content and not clearly as an add. Again, illustrating how there is a growing transparency challenge in the media sector, where it is becoming increasingly difficult for users to both know what data is used to profile and target them, but also in differencing editorial content from commercial.

However, the intelligent targeting of users with paywalls and ads also **raise highly important ethical questions of discrimination and manipulation** (Helberger, 2016). While the WSJ states that users visiting their sites are not offered different prices, they simply encounter the paywall at different points in time (Corcoran, 2018) the fact that some users based on their behavioural patterns and geographical location can experience a 'free' media experience while others are 'blocked' in their access can be discussed as a form of 'behavioural discrimination', which can be connected to other form of discrimination (e.g., sex or race), which is part of the algorithmic calculation (Wachter, 2020).

These differential practices are highly disliked by the users, who in general find the practice personalised pricing unfair and manipulative (Zuiderveen Borgesius and Poort, 2017). The opaqueness surrounding these technologies, where neither the exact workings of the algorithm or the description of why the user is confronted with a paywall is available, again raises serious issues regarding transparency. Frederik Zuiderveen Borgesius and Joost Poort (2017) in their explorations of whether personalised pricing is covered under the European data protection law, find that it is and that companies are required to be transparent about both their data usage practices and if prices have been personalised. However, the question is whether it is abided to and in this case how returning to the discussion above on how such information is disclosed. If the level of transparency is increased, it might also challenge the potential positive economic impacts that were discussed above, as users might be dissuaded from using the site at all, which could have highly negative impacts on media consumption and democracy, which, as will be discussed in the next section, is already under pressure by the rising power of intermediaries and alternative media. The lack of transparency, particularly regarding targeted advertisement has also produced concerns of manipulation as users are not aware of why they are presented with certain ads, because as discussed above targeted content can in fact change people's position on political topics and their emotions.

EXAMPLES OF AI IN INTELLIGENT SUBSCRIPTION MODELS AND ADVERTISEMENT

The Wall Street Journal's dynamic paywall, which is based on an AI system that predicts the likelihood of subscribing based on 60 variables.

Scandinavian Schibsted's use of AI based on 15 variables to predict the user's likelihood of subscribing.

The use of AI in intelligent scheduling of commercials by Korean broadcasters to increase sales numbers based on add showings.

Figure 10: Examples of AI in intelligent subscription models and advertisement

3.4 AI in deliberation over content

AI is increasingly leveraged to moderate and qualify the public debate relating to media content or mediated information. This is done both by supporting the identification and removal of harmful comments on media sites and social media as well as supporting the work of fact-checkers in finding potentially false claims. This section, therefore, relates to a much larger discussion of content moderation in the ever-expanding information environment and particularly of the role of large social media platforms in moderating their feeds, who in response to increased critiques of the harmful effects that hateful, false or propagandist content can produce have increasingly been implementing moderation systems – where AI plays a key role (see e.g., Gillespie, 2020). In keeping with the scope of this whitepaper, we focus on the instances where AI is used by media organisations either to moderate the debate on their own sites or the debate on social media platforms that relate to the content they have posted. Furthermore, we focus on the work of fact-checking organisations that in many cases developed out of existing media organisations or act as a media in their daily work in combatting false information and the use of AI in this work. We, however, do draw on some of the discussions relating to content moderation more widely to qualify the potential and challenges related to the use of AI in this context. In the following we discuss:

- How AI is used in comment moderation, also relating it to wider content moderation practices
- How AI is used in fact-checking practices by both platforms and be fact checkers.

3.4.1 AI in comment moderation

In 2011 over 90 percent of online newspapers in the US had adopted commenting systems (Zamith and Lewis, 2014). However, the hope for a thriving online public sphere was soon abandoned due to the enormous amounts of hateful or uncivil comments that were posted in the commentary sections – leaving the hopes of users themselves being able to deliver on rational and inclusive debate behind (Hughey and Daniels, 2013; Zamith and Lewis, 2014; Gardiner, 2018). In their study of the comments posted on a local newspaper's online site Kevin Coe, Kate Kenski and Stephen Rains (2014) showed how one in five comments included some form of incivility. Equally, Becky Gardiner (2018), in her study of the 70 million comments left on articles published on The Guardians website, showed how online articles written by women or racial minorities (e.g., people of colour) received a disproportionate number of abusive comments no matter what the subject of the article was.

As a result, many media organisations have abandoned the commenting sections completely or implemented strict comment moderations policies to counteract the hateful comments (Hughey and Daniels, 2013; Gardiner, 2018). The labour involved in the moderation of these comments has, however, proved to be immense, which have led many media organisations to explore the use of AI systems to support their moderation processes (Wang, 2021). The Guardian has, for example, during the last few years implemented several initiatives, including a machine learning tool that identifies potential abusive comments, which then are reviewed by their human moderators, and they have also decided to not allow comments on all articles to minimise the number of comments that need to be reviewed (Gardiner, 2018). The importance of having a well-functioning comment moderation practice have been pointed to in several studies that have shown how abusive comments can affect the willingness of journalists to engage with certain topics (self-censor) (Binns, 2017). It can be harmful to the well-being of the media staff and dissuade females to raise their voice in media and even leave their jobs (Binns, 2017; Gardiner, 2018) and it can negatively affect the interpretation of the media content (e.g., polarising the risk perception over a certain topic) and the credibility of the authors as well as the overall reputation of the media organisations (Anderson *et al.*, 2014, 2018; Yeo *et al.*, 2019; Searles, Spencer and Duru, 2020; Wang, 2021). Furthermore, it can be mentally exhausting work for the comment moderators who daily encounter hateful speech, which can also affect how credible they view the media whom they work for (Wang, 2021).

While many media organisations, as stated above, have decided to fully eliminate comments to mitigate these negative effects, this too has negative impacts for media organisations. A study of the German newspaper Die Welt has shown that the commentary section in fact garners 10 percent of all page visits and that the users who visit these sites are more likely to become frequent visitors on the online news site (Sterzing, Oberholzer-Gee and Melas, 2017). However, at the same time, the human moderators were struggling to sort through all the comments and still have time to engage in the debates with the users, which is important as **engaging**

moderators also positively affects the user's willingness to pay for a subscription and their judgement of the deliberative nature of the comment section (Sterzing, Oberholzer-Gee and Melas, 2017; Wang, 2021). The solution at Die Welt was an AI tool that roughly sorts the comments into three categories; either publishable, to be rejected or potentially unsafe, allowing the human moderators to only focus their attention on the last category (Sterzing, Oberholzer-Gee and Melas, 2017). This decreased the amount of manual moderation by 70 percent freeing up the moderators to better engage in the debates (Sterzing, Oberholzer-Gee and Melas, 2017). **Leveraging AI in this context, therefore, opens the possibility of retaining the comment section while largely mitigating the negative impacts on both the media organisations, its professionals and on the dissemination of and deliberation around media content**. These systems allow for, for example, the identification of potentially abusive content, which can then be assessed by human moderators, the direct removal of abusive content or to notify the user in real-time that the content might be abusive and lead to removal, potentially making the user reconsider their choice of words (Llansó *et al.*, 2020).

The reasoning for removing the comment function has, however, not exclusively been the presence of abusive comments. In an article by the NiemanLab, they describe how media mastodonts like Recode, Reuters, Popular Science, The Week, Mic, The Verge, and USA Today's FTW had all removed their comment function, but that this decision was only partly based on the presence of abusive language as they also saw how many users were increasingly engaging with their content on social media platforms (e.g., Twitter and Facebook), making the commentary option on the websites obsolete (Ellis, 2015). **The move to social media platforms, however, does not alleviate the problem of hateful speech in the comments, oppositely such speech also flourishes in this context** (Gillespie, 2020; Analyse & Tal, 2021). While the comments are here perhaps seen as slightly more external to the content and the media organisations, it must be expected that similar negative impacts to those above can be expected from the presence of abusive comments on the media organisation's social media accounts.

Studies of online deliberation have also shown that hateful language induces certain people to refrain from taking part in the public debate online - and that the lack of participation is skewed as predominantly women are becoming more hesitant to partake in online debates. This was, for example, illustrated in among other a Plan International study from 2020 including 22 countries, which showed how half the woman surveyed had experienced harassment or abuse online. In that group, one in five had changed their behaviour following such incidents, either cutting down or fully stopping to engage in an online debate (Plan International, 2020). Such studies illustrate the detrimental effects this language has on online deliberation and the freedom of expression by individuals who no longer feel they can express their opinion (Llansó *et al.*, 2020). While, for example, Facebook and Twitter have developed its own AI solutions and moderation policies to counteract hate speech (that have been highly debated, which we return to), many media organisations now have their own 'community managers', as also discussed above, that moderate the comment strings on their social media accounts based on their own moderation policies, but many are increasingly supplementing this with either in-house or out-sourced AI solutions to either fully replace or support the moderators.

To understand the scale of this problem, a recent study by Danish consultancy 'Analyse & Tal' (Analysis & Numbers) showed that every fifth comment posted on either a Danish media organisation site or a Danish politician's site on Facebook included hateful speech (Analyse & Tal, 2021). Illustrating the dire necessity for media organisations to continue and further develop their moderation practices, which is why many are turning towards AI as one important strategy in mitigating these negative social impacts. However, such tools can be economically unfeasible for media organisations, particularly in a context like Denmark where few tools are specifically developed for the Danish language. The need for more local and open-source solutions was the reasoning behind a Danish research project at the Danish IT University, which developed a 'misogynist algorithm' that could detect misogynist language, which was seen as one of the core reasons for women to leave the public debate. This was made open source to be leveraged by media organisations (Johnson, 2021). The study by 'Analyse & Tal' also included the development of a Danish hate speech detection algorithm, which can be implemented by media organisations. We return to how small countries with 'small' languages face specific challenges in leveraging AI in the following sections where we also discuss how such AI application induce their own challenges.

While most AI-driven systems focus on the removal of abusive comments, it is also worth mentioning that there are other examples where AI is also used to highlight comments of high quality, thereby, encouraging users to write in this manner (see e.g., Diakopoulos and Naaman, 2011; Park *et al.*, 2016). Gillespie (2020) also highlights how such AI moderating systems could also specifically be designed to sort out the most horrific content (e.g., child pornography or beheadings) to protect the mental welfare of the human moderators who otherwise must sift through these forms of content.

Challenges of AI in comment moderation

There are several challenges that are often discussed in relation to AI in comment moderation, which often fall under the headlines of accuracy, biases, and censorship (see also the AI4Media report 'First generation of Human- and Society-centred AI algorithms', which in section 2 proposes concrete policy recommendations for content moderation online). To begin with the first, a good place to start is the supposedly inflated accuracy by Facebook (now Meta), which was highly debated earlier in 2021. Here an internal assessment allegedly revealed that Facebook only eliminated between three and five percent of user views of content containing hate speech – previously having claimed this number was in fact 97 percent (Daws, 2021). The following debate centred on how the different numbers had been calculated, illustrating how **'accuracy' is a fluid metric and how much hateful content remains out of sight of the AI tools.** The need for transparency around how measures of accuracy have been devised will be necessary to really assess the workings of these applications and, as will be discussed below, how such measures can hide discriminatory judgements.

In a similar vein, **the way that notions of 'hateful', 'harmful' or 'misogynist' are operationalised has received much critique**. Regarding the Danish developed 'misogynist algorithm', critics voiced concerns over how these AI applications will only be helpful to a degree because the full complexity of what 'misogyny' is cannot be operationalised (e.g., considering sarcasm or subtle insults), stating how AI tools are not very accurate when dealing with highly ambiguous concepts (Henningsen, 2021). Equally Hughes and Daniels (2013) point to how **in the operationalisation of such terms, often dominant understandings are used, in their case regarding racism, which might also sustain existing and historical biases.** Furthermore, the lack of contextual understanding of these AI systems poses an issue, because some language can be hurtful in one context while not in others (Llansó *et al.*, 2020). Something that was also highlighted in one of the sub-projects in the 'Responsible AI working' in GPAI in their report 'Responsible AI for Social Media Governance'. Here they explored how to use citizen involvement methods to gain a better understanding of what 'harmful' was for the citizens in New Zealand, illustrating how it is important to not just use universal measures or notions of harmful as they might be regional and culturally dependent (Knott *et al.*, 2021). Missing such regional or local nuances as well as, for example, satirical uses of a word remain something that AI systems have difficulty handling and the deletion of content in such cases could impoverish the online deliberation (Llanso *et al*., 2020; Gillespie, 2020). This is also why Gillespie (2020) critically highlights the limitations of AI systems at a scale and the importance of having human moderators oversee the flagged content to ensure such instances are caught. He, furthermore, suggests how AI systems could perhaps rather be specialised towards offering more contextual data on the posts (e.g., whether it is a serial offender) to better equip them in their interpretation (Gillespie, 2020).

Equally, **the accuracy of models can pose a challenge if a model does not dynamically evolve with the language, making it outdated very quickly, or if a model is too simple, focusing, for example, only on banned keywords** (Llanso *et al*., 2020, Gillespie, 2020). The first risks losing its efficiency and can also become subject to 'gaming' by users who will adapt their word choices to avoid being caught by AI systems. This could be detected during the Covid-19 pandemic, where critical voices against the Covid-19 measurements began using 'code language' to avoid having comments removed. In the US, social media users started using the word Pizza when referring to Pfizer (Collins and Zadrozny, 2021) and in Denmark many emoticons and intentional misspellings were used in the comments and posts by social media users (Kristensen, 2021). The second, simple form of moderation based on 'banned words' lists can have highly unfortunate effects, as the example of a Christian website that developed a system that would automatically exchange the word 'gay' to 'homosexual', which when the athlete Tyson Gay won an Olympic medal switched the athletes last name Gay to Homosexual (Akers, 2008). Danish journalist Torben Sangild also got his Facebook profile shut down after sharing a post with a critical article regarding the conspiracy theory surrounding QAnon, because the algorithm simply reacted on the word QAnon without understanding the context (Joergensen, 2021).

The lack of accuracy either due to context insensitivity, crude, or non-adaptive models, can have serious implications for freedom of expression online, particularly when content is directly removed. This is both the case when the system produces false positives (something is wrongly flagged as e.g., the example of Torben Sangild) or false negatives (something for example hateful is not caught). The first can wrongly censor citizens and without proper complaint mechanisms and due process this can have very detrimental effects on the individual's right to express themselves. The second, on the other hand, fails to mitigate the issue of uncivil language, potentially affecting people’s desire to take part in the debate as discussed above (Llanso *et al*., 2020). Gillespie (2020) also illustrates how the error rate increased significantly during the first Covid-19 lockdown where the human moderators were sent home and, for example, Twitter’s AI moderating system was left to its own devices. Illustrating the need for proper oversight with such systems to limit the potential false negatives and positives. The risk of both false positives and negatives might also not be fairly distributed as such algorithmic systems have the potential to be biased against certain groups or certain languages based on the dataset used to train the model, as the model will be most accurate on content that is most similar to what it was trained on. **Underrepresentation of certain groups may result in them being unfairly silenced or experiencing more abusive language than other groups** (Llanso *et al.,* 2020). A discussion that was reinvigorated when it came out how Facebook’s hate speech algorithm was highly biased, both in terms of working significantly worse on minority languages, but also that it was significantly worse at detecting verbal attacks on minority groups, for example, the black population, who in fact were often victims of the worst attacks (Dwoskin, Tiku and Timberg, 2021).

In the report by ‘Tal & Analyse’ they equally showed that in the Danish context the same was the case, where racial minority groups were the most targeted together with women and disabled (Tal & Analyse, 2021). When training the algorithm such biases in the division of attacks should, therefore, be considered and counteracted when developing AI moderation models to ensure at least fairness, if not justice as we will discuss in the last section on archiving. This also **highlights the need to interrogate the accuracy accounts stated by the developers** of such systems as the accuracy might be severely skewed, while overall seeming high. As AI applications are often trained on data from existing moderating practices by editors and sometimes complemented with user flagging practices, such datasets could also sustain existing biases as discussed in section on media biases. **The lack of access to both training data sets and models that work well on minority languages is also a general bias across many AI applications**. Google Maps, for example, produces highly different results depending on the language used in the query. In the case of hate speech, this puts pressure on local solutions to be developed in those countries as discussed above since the platforms own solutions do not efficiently cover those languages. Illustrating a severe issue, as predominantly the western developed world is serviced by these systems, while the global south, where large countries like India with multiple small languages receive a very poor service. This illustrates highly asymmetric access to the AI infrastructure for media organisations in these parts of the world.

Overall, the question of content moderation raises the important question of what content can and should be censored. While certain forms of censorship such as in the cases of child pornography is widely accepted and unquestioned, it will be critical to discuss what forms of moderation will be accepted in the future (Llanso *et al.*, 2020). Matthew Hughey and Jessie Daniels (2013) also discuss **another negative side effect of moderation, namely that it can induce a state of ignorance towards**, in their case, the existence of racism as users are no longer confronted with it as they were previously. Highlighting the importance of transparency, so that the users remain aware that there continues to be a problem with racism or misogyny. Such transparency can even have positive effects, as full transparency of both human or machine moderators has a positive impact on how users view the credibility and trustworthiness of the information presented to them (Yeo *et al.*, 2019; Wang, 2021). Sai Wang (2021) even points to an automation bias amongst users, who find machine moderation more accurate than human moderations, which opens the question of the perceived 'algorithmic objectivity', which is discussed later in the whitepaper.

EXAMPLES OF AI IN COMMENT MODERATION

The **'Misogynist algorithm'** was developed by three researchers at the IT University in Copenhagen, Denmark to detect hate speech directed at women on social media in the Danish context.

Meta's policies and AI tools that moderate content on their social media platforms (e.g., Facebook and Instagram).

Figure 11: Examples of AI in comment moderation

3.4.2 AI in fact-checking practices

In the early 2000s, another actor emerged in the media landscape, namely the independent fact-checkers, which today have become important institutions in this landscape (Graves and Cherubini, 2016). While these organisations initially emerged in the US, fact-checking organisations can now be found in more than 50 countries across the globe – and many of them have been established only within the last decade (Graves and Cherubini, 2016). Many of them are affiliated with an existing media outlet or began that way, but the majority are now independent institutions (sometimes part of civil society projects) fulfilling a specific democratic role in the media landscape, namely holding politicians, media organisations or private persons accountable for publicly voiced false statements (e.g., on social media, in TV or in written content) (Graves and Cherubini, 2016). These new institutions emerged as a response to the increasing amounts of online mis- and disinformation or more popularly ‘fake’ news that was circulating (Graves, 2018; Chambers, 2021). ‘Fake’ news is connected to negative societal impacts of manipulating political opinions and elections, furthering conspiracy thinking and inducing a state of ‘epistemic instability’ where previously publicly agreed-upon truths and trusted institutions are now being contested and challenged in their authority, inducing increased societal polarisation (Chambers, 2021). Studies have been shown that social media platforms like Facebook do tend to disproportionally amplify the circulation of ‘fake’ content, compared to news from either media organisations or partisan sources (Graves, 2018) and as much people’s access to facts today is mediated, this does pose a significant issue (Chambers, 2021). As a result, many discussions on fact-checking focus on how to develop **tools to help combat dis- and misinformation on these platforms and here AI is considered a potential solution to support fact-checkers in identifying claims and content to be checked** (Graves, 2018).

Lucas Graves (2018) highlights how many funding organisations are currently investing heavily in projects that address this issue through technological solutions. Such automated fact-checking (AFC) tools can be used in many ways, but currently the most prominent one is as a support tool for the fact-checking organisations. The AFC platform ‘ClaimBuster’, developed by the University of Texas-Arlington and later adapted by the Duke Reporters’ Lab, is for example used by fact-checkers at PolitiFact, FactCheck.org, the Washington Post, and the Associated Press to support them in their selection of claims to check (Graves 2018) (a full description of the tool can be found in Hassan *et al.*, 2017). The tool monitors the public debate across mediated sites (e.g., scraping transcripts, online news sites, selected social media feeds or automatically transcribing political debates in TV and monitoring these) and identifies claims that warrant concern and then alert fact-checkers of possible statements that could deserve checking (Graves, 2018; Adair *et al.*, 2019). **The potential of this is to make the fact-checking process more efficient by removing the tedious work of trawling through online media content in search for claims to check – allowing the fact-checkers to produce more fact-checks that can help improve the public debate** (Graves 2018; Adair *et al*., 2019). As stated above, Duke Reporter’s lab adapted this tool, but what is interesting was that the changes did not relate to the AI system itself, but the presentation of the results. Initially, the ClaimBuster produced automatic emails, which were sent to fact-checkers, but these were only turned into actual fact-checks in the first couple of months, then interest dwindled. The Duke Reporter’s Lab change was to produce a more curated list where a journalist would go through the list produced by the ClaimBuster and pick out the claims that were considered more relevant from a news perspective, leaving out the ‘irrelevant claims, often artifacts of ClaimBuster’s imperfect aim at finding newsy political claims’. This resulted in the tool being more useful for fact-checkers, illustrating the remaining need for a ‘human touch’ for the potential of such tools to be realised (Adair *et al*., 2019). One of the AI4Media run by Deutsche Welle and ATC uses cases also specifically focuses on supporting media professionals with verifying claims (see 'AI for Social Media and Against Disinformation').

There are also other **potentials linked to these systems, namely of authoritatively verifying claims automatically and potentially doing this in real-time** (Graves, 2018). This could both help to increase the amount of content that is fact-checked and support the dissemination of for example political debates, where the audiences would have access to verifications of what is said as it unfolds (Graves, 2018; Adair *et al*., 2019). However, as we return to it in the following section on challenges, there are several reasons why this potential currently is limited. One way that has already proven fruitful to further utilise these systems is checking claims against the existing database of fact-checks to see if the claim has already been checked and a previous article can be reused and also to identify 'reoffenders' who can then be flagged (Graves, 2018). This allows fact-checkers to save time in digging through the archives. This function has already been implemented in ClaimBuster, which checks claims against the libraries of known fact checking organisations (Graves, 2018).

The social media platform Facebook has also developed its own AI systems that are trained to identify misinformation across the content posted on Facebook. This is used to both delete content that violates Facebook's policies but also to identify claims that are then checked by a local, independent, and certified (in the International Fact-Checking Network (IFCN)) fact-checking organisation based in the country where the post originates from (Meta, 2020). Currently, Facebook has such partnerships in 14 countries and are expanding them on an ongoing basis. They claim that employing these organisations have helped to induce the distribution of fake content with about 80 percent on average (Meta, 2020) (claims that, however, have been widely criticised for exaggerating the success of the efforts). In this case, it is Facebook's algorithm that flags potentially false content, which the local fact-checking organisation is responsible for deciding what and how much of the flagged content they will fact-check. The content that is fact-checked will be clearly labelled as such on Facebook and link to the fact-check, thereby, again making it a collaboration between humans and the AI.

Challenges of AI in fact checking practices

As already stated above, there are several challenges that emerge when discussing the potential of leveraging AI to fully automate the verification of claims. First, these systems are limited in what types of claims they can identify. If a claim is too complex, implicitly implied, or for example makes a reference to a previous claim, these systems have difficulty identifying them (Graves, 2018). **If we were to fully rely on automated systems at this point without human selection, many potentially false claims might go undetected**. When it comes to automating the verification of claims, more challenges arise because debunking a claim is not always straight forward and it requires consulting multiple sources, which can sometimes only be included through interviews with for example researchers. This can be the case if no authoritative sources exist in data form and even if data is available to check the claim against it, it might not be in a format that is compatible with the system (Graves, 2018).

The ‘Chequeabot’ developed in Argentina, which specifically focuses on expanding the use of AFC tools to the Spanish language has a module that does check claims against public sources of information (e.g., national statistics on employment). However, **the system's understanding of claims remains rather simplistic and unable to verify all claims and the data sources are only available for certain areas of enquiry.** Both the technical solutions and the data quality, therefore, currently remain core barriers for realising this potential of AI in fact-checking (Graves, 2018). This challenge is further underlined by Lucas Graves and Federica Cherubini (2016) who note how data sources are even more scarce and difficult to get access to in countries that are under authoritarian rule, where fact-checking practices would be highly needed in supporting a more accurate debate. This and the example of the Chequeabot again illustrates how there are highly differentiated access to the potentials of AI and language specific infrastructures. The question of what sources to ‘trust’ also poses an issue in itself, as Lucas Graves (2018) points to how many of the developers behind these systems are weary about giving some institutions higher ‘truth’ ratings, as this would exclude others and all institutions can unintentionally provide false information. This would also limit the diversity in the public discourse if some sources were always heard while others are not.

In the factsheet Lucas Graves (2018) points to how AFC systems are often seen as a technical solution to a technical problem of minimising the spread of false information. However, this presents another challenge, because as pointed to by Mette Bengtsson and Anna Schjøtt Hansen (2021) who in their study of Danish Covid-19 protesters found that this ‘black and white’ image of fact-checking as a cure to fake news has overlooked that what is really at stake is different ways of viewing the truth – different ontologies amongst protesters and the Danish fact-checking organisation. Similarly, to early studies of how ‘facts’ was constructed (see e.g., Knorr-Cetina, 1984; Latour and Woolgar, 2013), to fully grasp the growing societal conflicts, simple technical solutions cannot stand alone, more in-depth studies of why such alternative truth constructions are emerging must be understood. Relating to this, Lucas Graves (2018) also does point to the controversy often connected to the practice of fact-checking, **which is a subjective practice, where ideals are set up for how the truth can be obtained – but such ideals can and should be challenged** (see also Graves and Cherubini 2016 for overview on the discussions of fact-checking practices).

The study of Danish Covid-19 protesters also showed how **fact-checks in fact had the adverse effect of spurring on the protesters to find new channels to circulate their ‘truths’ and experiment with adversarial strategies**, as discussed above with the use of intentional misspellings and emoticons to not get caught by the algorithm. This produces new technical challenges for these AI systems, which must constantly adapt to such language changes. This experience of heavy censorship amongst the protesters, both from fact-checkers and from not being heard in the mass media, was also highly conducive to further polarisation as they felt 'hunted' by an elite in society (Bengtsson and Schjøtt Hansen, 2021). This illustrates how **fighting media polarisation through fact-checking and content moderation practices, might have dangerous side-effects that counteract that very goal**. The scepticism of these protesters against the established system and the fact-checking practices was also further enhanced by the lack of transparency of how the Danish fact-check organisation 'TjekDet', who is Facebook's Danish local partner, used the algorithm provided by Facebook. As well as the lack of transparency in how penalties (e.g., losing account, having content removed or losing functions) related to being fact-checked. This points to larger issues discussed in relation to content moderation by platforms, which were discussed above.

EXAMPLES OF AI IN FACT-CHECKING

The **'Chequeabot'** developed in Argentina with a foundation in the Spanish language, checks claims against public sources of information (e.g., national statistics on employment).

THE **'ClaimBuster'** platform was developed by the University of Texas-Arlington, and later adapted by the Duke Reporters' Lab, is an external service used by among other fact-checkers at PolitiFact, FactCheck.org, the Washington Post, and the Associated Press to support them in their selection of claims to check.

Figure 12: Examples of AI in fact-checking

3.5 AI in audiovisual archives

We now reach the conclusion of the media cycle, which is the archiving on media content[3]. Many public broadcasters also function as media archives, such as the BBC in the UK or Denmark's Radio (DR) in Denmark, while other countries have specific organisations that serve this purpose, such as Netherlands Institute for Sound and Vision and the French Institut national de l'audiovisuel INA. The different institutional setups of the archives (e.g., closer ties to educational, culture or news sectors) may influence both the ideals guiding the archival practice, but the key function of storing media content is shared across these media organisations (Cechine, 2021c). Research into media archives and their use of AI have been growing in the later years (see appendix A for an overview of relevant resources). In this whitepaper we draw predominantly by the overview provided by Randi Cecchine (Cechine, 2021c, 2021b, 2021a) who through a range of blogpost has described the research she did as part of the master thesis at

[3] In the context of this white paper, we are primarily concerned with organisations that archive broadcaster content.

the University of Amsterdam's Preservation and Presentation of the Moving Image Master's program, documented a range of AI-practices by media archives. Together with the report in 'AI in the Audiovisual Sector' which also mentions examples of AI use in archiving practices (Rehm, 2020). Beyond we also draw on related work from other sectors, such as the recent review article, by Giovanni Colavizza, Tobias Blanke, Charles Jeurgens and Julia Noordegraaf (2021), which offers an overview of the potentials of AI within archival practices more in general and the report concerning the state of the art of AI usage within libraries by Ryan Cordell (2020). In the following we explore the potentials and challenges related to three applications of AI in the archival processes, namely in search, in discovery, in reuse of content and in research.

We should note the specific technological context in which audiovisual archives operate and which in many ways define the opportunities for employing AI on archival collections. To manage their vast amounts of data, these organisations rely on media asset management (MAMs) systems. In most cases, these are proprietary solutions bought from commercial vendors (to name a few, Vizrt, Tedial, AVID, Dalet and TransMedia Dynamics). They tend to be monolithic, closed-off systems that offer limited compatibility or integration with external tools. At the same time, archival organisations are increasingly interested to integrate AI-based tools (developed in house or from third parties) into their workflows. While such enterprise MAMs perform well in an archival context where reliability (contractually controlled through service level agreements with vendors) are key priorities, they often fail to offer such integration flexibility. Even in cases where integration is possible, this requires significant in-house human resources and new type of skills (e.g., data engineers) to oversee implementation and deployment, which is not achievable for many publicly funded archives. Hence, the adoption of AI solutions in audiovisual archives is determined by (i) the openness of vendor-operated MAMs to integrate external AI solutions while ensuring security, and (ii) the capacity of archives to implement AI solutions and tailor them to their specific needs. This issue is not limited to audiovisual archives and to some extent affects many media organisations who use similar MAM systems the previously described stages of the media cycle. We return to this issue in the next chapter of the whitepaper, while we in the following focus on:

- How AI is used to make archival content more searchable
- How AI is used to also make it easier to discover content – in more serendipitous ways
- How AI is used in repurposing content, allowing new creative ways of using existing media content and the personalisation of content
- How AI is used to conduct new forms of research based on archival resources.

3.5.1 AI in search

One of the key missions of media archives is to make the content accessible, which is echoed in an interview with Jake Berger from the BBC Archive: 'The BBC Archive's editorial mission is to open up the archive to as many people as possible in as many ways as possible and try to optimize how people can find what they're looking for, or how they can be introduced to things they didn't know [they were] looking for but would enjoy' (Cechine, 2021c). Questions of accessibility in archives are tightly linked to the availability of appropriate metadata. Currently, metadata is gathered through the following practices: (i) metadata coming directly from production (titles, broadcast dates, technical data imbedded in files), (ii) metadata added manually by archivists, (3) metadata generated (semi)automatically through leveraging AI solutions, (4) metadata added by end users (for instance, through crowdsourcing) (Oomen *et al.,* 2013). However, many archival collections currently have insufficient metadata to support complex search queries in the first place, for example, if they have been digitised but not annotated or indexed.

Even if basic metadata exist, the challenge with audiovisual collections, in comparison to text or images, is their linearity (e.g., existing metadata might describe an entire TV programme but not the content of its individual segments). **The ability to leverage AI to annotate archival material at a large scale and on a granular level is, therefore, immense in making media archives much more accessible.** The application of AI, for example, supports the production of tags on written content, the annotation of audiovisual content (e.g., with geographic locations, people, or objects), the production of automated summaries or automated transcription of speech, which all make content much more searchable (Rhem 2020; Cecchine, 2021c; Cordell 2021).

These AI-driven annotation methods can open audiovisual collection to search queries from a broad range of users, including professional content producers, artists, educators, researchers, and general audiences. Cordell (2021), for example, mentions the case of a German Broadcasting Archive who used AI (specifically a neural network) to tag concepts and people across more than 2.500 hours of video archive from the German Democratic Republic (GDR). While there was still a margin of error, the neural network managed to receive an average precision score over 50% on 66% of the concepts that were tagged, which does illustrate how such immense amounts of data can become more discoverable even if it is not correct every time. In the interview with Randi Cecchine (2021c), Jake Beger also shared an example of how he using AI believed to have found the first televised appearance of the band the Grateful Dead. A video clip that previously had not been annotated or indexed, making it impossible to find it. **AI can, therefore, allow long forgotten pieces of content to come to the surface.** The potential of these methods is also transferable into production processes, where AI can be used to, for example, auto tag media content or automatically add subtitles, as discussed above.

Increased searchability also allows for the identification of 'gaps' in the collections or to highlight historically more silenced voices. One of example of that in the more classic archival context is the project 'Unsilencing Dutch Colonial Archives', where AI is utilised to specifically search out individuals who were not heard much in the vast Dutch colonial archives, e.g., woman or the indigenous population in colonial realm (CREATE, 2021). It can, therefore, be used with the specific aim of highlighting the voices of minorities and reclaim forgotten moments in history. The potential for rediscovery and research also becomes the overarching possibility that allows for the following potential uses, which are discussed in the following sections after we discuss the challenges related to content discovery.

Challenges of AI in search

Several challenges emerge as a result of using AI to enrich the metadata of archival material in order to make it more searchable. One challenge relates to sensitive information, where AI can both be the potential venom and cure, because as Colavizza et al. (2021) point to **archival institutions must ensure that sensitive or personally identifiable information is not disclosed prematurely to users of the archival content** – a need that in EU is intensified with the General Data Protection Regulation (GDPR). However, the fact that such information is not easily identifiable across collections, often makes it impossible to grant users access to the material through a Freedom of Information (FOI) request or as a researcher (Colavizza et al. 2021). While this is less of an issue with audiovisual archives (when thinking in terms of broadcasters and not, for example, the archiving of social media data), there are still instances where material is sensitive and should not be searchable or reused (e.g., content from inside prisons or hospitals, or content that has been filmed during a trial) as this might be harmful for the people portrayed in the content. It will, therefore, be important to develop ways to ensure that such sensitive content is identified and handled in an appropriate manner when content is made increasingly searchable, much like sensitive content must be moderated in new productions as discussed above. Colavizza et al. (2021) drawing on Hutchinson (2018) propose how AI in the form of supervised machine learning could be utilised to identify privacy sensitive records, something that perhaps could equally be useful in media archival practices.

Another key **challenge relates to inherent biases in the archival collections or biases produced by the choices of what material to digitise**. Cordell (2021), drawing on the work of Ben Fagan (2016), argues that when making archival content more available you might risk emphasising existing biases, because as Fagan shows in his work, library collections in the US have been prone to prioritise the digitisation of newspapers that were predominantly white and oriented towards the middle-class. This happened at the expense of digitalising newspapers run by blacks or other minorities. While he also points to how this was not an intended exclusion, the fact was that the state-wide digitalisation based on geographic representations, meaning that the selection focused on geographic spread, allowing only the most popular papers in a state to be digitalised, thereby, ultimately excluding minorities. Similar issues relating to the lack of focus on racial and cultural representation could also apply to media archives who also must make prioritisations when digitalising their collections, thereby potentially making some media realities more present and accessible at the expense of others. As Cordell (2021) stresses **the problem of representation is nothing new, but AI-technologies could potentially amplify these past biases further or introduce new ones, unless measures are taken to act against it.**

A specific challenge here is how many **AI technologies used in the archival practices are developed in commercial contexts, which are currently rarely domain specific and predominantly caters to the 'dominant voices'.** As Cordell (2021) describes within heritage and library institutions there generally is a lack of domain specific data sets available to train AI-systems, which will work on e.g., different accents (regional or minority) and on language from different periods. Using the example of the Library of Congress's work on the 'Speech-to-text Viewer', which was to transcribe the audio from the Smithsonian Folkway Recordings, Cordell (2021) describes how an involved participant in the project said the accuracy in general was quite low, but when it came to regional dialects it was even worse, due to a very limited presence of such dialects in the dataset. This describes a larger challenge of most datasets being based on a 'standard English' most associated with the white middle class and data sets are also often trained on more contemporary material, which means that changes in language and its meaning will be lost as well as. Cecchine (Cechine, 2021b) equally describes how the lack of available training datasets for these kinds of projects often results in archives training AI technologies on their existing collections, but as shown above such data sets also have inherent biases. What is emphasised by both the authors is that it is important that the professionals involved in AI projects are aware and actively work to counter-act such biases.

Another issue arising with commercial AI technologies is that it induces concerns over intellectual property loss. Unclear terms of service adopted by commercial AI providers do not make it clear how the archival data provided by the archive to train algorithms will be stored, used and handled. Cecchine (2021a, 2021b) explains that multiple of the media archives she interviewed expressed concerns over using commercial tools for this exact reason even though they are often well developed and more robust. A response to this has been that e.g., Amazon Web Services (AWS) are now allowing companies to opt-out of sharing their data and the trained AI technology with Amazon, making commercial tools more alluring again for media archives as building individual solutions (either from scratch or by adapting existing open-source repositories) remain expensive, labour intensive and require expertise and infrastructures (Cecchine 2021a, 2021b). This is also why both Cordell (2021) and Cecchine (2021b) highlight the importance of both research institution collaboration but also the need for developing AI technologies specifically for archives and good datasets for training that can be shared across the organisations within this field to ensure it is built in ways that support the archival practices and archive's role in society.

EXAMPLES OF AI IN SEARCH

The **'Civil War Photo Sleuth project'** conducted at Virginia Tech University used Machine learning to collate data from the Library Congress, National Archives, National Portrait Gallery and other sources in an attempt to rediscover lost identities of soldiers in Civil War card-portraits.

The 'Unsilencing Dutch Colonial Archives' project aimed at finding and enhancing the minority voices in the archival collection. Currently it has been tested as a pilot under the name **'Unsilencing the VOC testaments'**.

Figure 13: Examples of AI in search

3.5.2 AI for content discovery

Once, metadata is increasingly available **AI can also be leveraged to make these collections more discoverable in other ways, by making connections in and across collections** (Cordell, 2021). This can be done both based on supervised training methods, where the AI finds connections between, for example, tagged themes or genres or based on unsupervised training methods, where the AI itself finds similarities across the media content. The latter can potentially allow for more serendipitous discovery than previously (Cordell, 2021) and potentially help alleviate 'confirmation bias' in archivist searches, where the searcher simply finds what they already knew (Collavizza 2021). Cordell (2021) uses the example of the 'National Neighbours' project, which leveraged AI to find similarities in the collection of the National Gallery of Art across 1.048 dimensions to illustrate the value of this along with the 'Neural Neighbours' project, which equally found connections between 27.000 historical photographs in the Meserve-Kunert Collection. The latter had a user interface, where if they clicked one photo, connected images were suggested to the user. Illustrating the potential of AI in increasing access to archival material and thereby upholding the democratic and societal role of such institutions as supporting historic and cultural memory, by allowing more people easier access to their societal past. Increased searchability and connectivity to archival collections could also ease the life of media professionals, who in the production process could search for the topic they are working on and be presented with previous content on the topic, making it easy to find the relevant background information to produce the story.

Increased data discoverability also enables media archives to develop new products and services for content dissemination. For instance, in a Horizon 2020 project ReTV, the Netherlands Institute for Sound and Vision designed a messenger service that delivered archival content to users in a fully automated manner. Here an AI-driven personalisation engine used fine-grained analysis of audiovisual content to suggest matchings between a user's profile and video content (Bocyte and Oomen, 2022). Based on their previous interaction, each user would receive a different selection of videos. Services such as this have the potential to increase the attractiveness of archival media collections for contemporary audiences, increasing the public's exposure to historical perspectives. From the perspective of an organisation preserving media collections, automatically annotated collections create opportunities to not only optimise content dissemination workflows, but also leverage the breadth of audiovisual archival content, especially the long tail of large-scale collections that might not be discovered and shared via manual methods. Overall, this can significantly strengthen the societal and economic impact that media archives can deliver through the increased exposure to and dissemination of their collections.

Challenges of AI in content discovery

A challenge pointed to by Cecchine (2021b) relates to the type of discovery that can be enabled by off-the-shelf solutions. **Many of the commercial AI technologies (e.g., built and trained by Amazon, Google or IBM) are not specifically designed for archives, but rather for business and profit purposes.** She references an interview with Jim Duran, Director of the Vanderbilt Television News Archive and Curator of Born-Digital Collections, who describes this challenge: 'We have access to pre-built tools, intended for a different type of use than archives and libraries. We are not trying to sell objects; we are trying to describe them to make them accessible. A lot of times the tools don't really match what we need; they give you results that don't quite fit. To get the most out of pre-built tools, we must transform our data to fit their algorithm' (Cechine, 2021b). **This could potentially impoverish the collections as only commercially relevant understandings of the collection will come to the foreground with commercial solutions, thereby limiting collection enrichment by excluding potential 'non-commercial' but relevant understandings.** Similarly, the previously mentioned ReTV project concluded that commercial recommendation algorithms that are tailored to increase the time users spend consuming content ("binging" behaviour) do not adequately cater to the needs of media archives (Bocyte, 2021). Especially publicly funded media archives prioritise building long-lasting relationships with their audiences and exposing them to a diversity of topics to increase their awareness about a variety of cultural, political, and social angles possible on a given story.

Hence, it is important that media archives assume a societal role in shaping AI technologies that are used across all economic sectors. In this direction, the position paper by the Netherlands AI Coalition workgroup on Culture and Media in their report 'The Art of AI for All' from 2022 argues for "Culture for AI" where historic and cultural meanings, experiences and values can be leveraged to shape the development and deployment of AI. The position paper argues that the media sector as a whole can act as a testbed for social consequences of AI technologies before they are rolled out in other domains (e.g., testing algorithms on diverse archival collections to monitor and correct biases).

EXAMPLES OF AI IN DISCOVERY

The **'Neural Neighbours'** project have utilised AI to allow users to explore over 27.000 photographs from the Meserve-Kunhardt Collection at Yale's Beinecke Rare Book and Manuscript Library.

'The Sensory Moving Image Archive' (SEMIA) project visualises the similarity of 103,273 shots from 6,969 archival videos. Each shot in terms of colour, movement, shape, and clutter, allowing users to discover and explore content based solely on visual qualities.

Figure 14: Examples of AI in discovery

3.5.3 AI in content reuse and repurposing

As collections become more accessible, **another potential to utilise AI arises, namely the potential to reuse or repurpose content in new ways**. Cecchine (2021c) uses the example of the 'Citizen DJ platform', which was developed as part of a collaboration between the (United States) Library of Congress and Brian Foo as part of the Library of Congress's Innovator In Residence Program. On this platform, users can create their own Hip Hop music out of audio and video collections, which had previously been annotated by AI, allowing them to explore the music and film archive in new interactive ways as well as recontextualising it into todays' societies. A trend towards 'demixing' of music, where AI has enabled the source separation of music tracks, allowing particular sounds to be reused (Amato et al., 2019). The value here is that the past music and film or TV segments can be brought to life and contribute to the production of new creative products. **Granular AI annotations enable media archives to serve their collections as raw material that can be used for sampling, inspiration and creative experimentation, giving a boost to young and emerging creators**. Something that can be seen as important to counteract what has been discussed as the homogenisation of music due to central platforms like Spotify, where certain forms of music thrive the 'Spotify core' (Morris, 2020), here bringing in 'old' sounds could potentially reinvoke ideals of music of the past and bring more diversity to the music stage.

In their study based at British broadcaster BBC Rhianne Jones and Bronwyn Jones (2019) also discuss the potential of repurposing content in the production and distribution by having an AI dynamically put together existing content based on user preferences. This will ultimately mean that the notion of 'an article' will be atomised as articles will be made up of different atomic parts either prepared by journalists or readily available in the archive. Much like with chatbots and personalisation technologies, **the value of this is to better cater to the individual, by serving them, for example, a shorter or longer version of a story.** While the BBC in this experiment was still working predominately with text bits specifically prepared for the AI, over time an archive of article atoms would be available and ready for reuse when certain topics re-emerge on the public agenda, making the archival practices central to these new repurposing practices.

Challenges of AI in content reuse and repurposing

With the potential for reuse of archival content **new challenges emerge regarding copyright, ethical use and manipulation of the content**. Cecchine (2021c) highlights the potential new challenges that arises with increased access and specifically when content can be repurposed by users, such as the example of the 'Citizen DJ Platform' mentioned above, namely, how to make sure that users do not breach copy-right. The way that specific project has solved it is by providing ethics and copyright guidelines that describe both how the content can be repurposed and why, as well as how to sample it in an ethical manner by, for example, considering the historical and cultural context of the audio. Archives, therefore, when they allow repurposing, must consider how they want to ensure that the content is not misused by users and how such potential misuse should be handled. In a similar vein, websites that allow users to produce 'deep fakes' (e.g., sites such as Deepfakes.com) are facing similar challenges, though even more intensified due to the potential manipulative or malicious use of that type of content. Here practices are not stabilised either and, in some cases, no ethical practice exists, in other situations you see ethical guidelines of use, like that of the 'Citizen DJ Platform', while there are also more 'extreme' practices, where all the content produced is given a watermark to ensure that it can be identified as fake content. Equally, AI systems have been used to verify images and provide watermarking post publishing, thereby, also being a solution to the problem. **As audiovisual archives have a long running tradition of protecting and upholding authenticity in their collections, it will be vital to find solutions to this challenge and have standards in place if more content is made available for reuse** – the latter is a question many archives are currently contemplating, exactly due to this potential of misuse (for discussions on, for example, colourisation of black and white films definitely fits under this topic, see Op Den Kamp, 2016).

In the BBC case discussed above by Jones and Jones (2019) another challenge emerged, namely how the efficient repurposing and production of content was dependent on a large archive of text developed for this purpose. The journalists tasked with writing the bits of content to be used when testing the AI system stated how it was much more time-consuming to produce small stand-alone pieces that could be mixed and matched, compared to writing a single article, as they constantly had to make sure that the text bit could in fact be read entirely on its own. **This could challenge the potential for gained efficiency by just creating new forms of routine tasks.** However, as these archives grow larger, the challenge might resolve itself as a large pool of ready-to-go material would be available, and equally, over time content creation processes might change to better support such repurposing approaches.

EXAMPLES OF AI IN CONTENT REUSE AND REPURPOSING

The **'Citizen DJ platform'** developed in collaboration between the United States Library of Congress and Brian Foo as part of the Library Congress's Innovator In Residence Program, allows users to create Hip Hop tracks with a base in the archive. Allowing for the recontextualising of music, film, and dialect collections.

The **'Jan Bot'** is an automated website that is placed in the Hallway of the Eye Collection Centre in Amsterdam and produces short films using clips from the collections of the Eye Film Museum in Amsterdam based on what topics are trending that day.

Figure 15: Examples of AI in content reuse and repurposing

3.5.4 AI in media archival research

The last key potential of AI in archives relates to how the accessibility and the development of **new tools to analyse archival data are opening new and important avenues for research where archival material become data rather than individual records.** Cecchine (2021c) emphasises important collaborations between media archives and universities or other research institutions in developing research infrastructures and tools to explore the vast amounts of data, such as the 'CLARIAH Media Suite project' in the Netherlands (which is the at the centre of one of the AI4Media use cases) or 'The BoB FOR AI project' from Learning on Screen, the British audiovisual resource for educators, where researchers can explore the types of questions the data might help to illuminate. **One of the core values of these new collaborations is that whole new research questions become answerable by opening up and making the data analysable by researchers.** Tobias Blanke and Jon Wilson (2017) show how through the use of AI tools (topic modelling and language models) they can discern and study specific epochs (a period of coherent language use) from textual archival material and the 'INA Segmenter', an open-source audio segmentation toolkit by the French Institut national de l'audiovisuel (INA), which allows for the segmentation of male and female voices across audiovisual archival content enabling researchers to explore gender differences in TV in quantitative ways. In a political context, researchers used the previously mentioned CLARIAH Media Suite and its tools for speech, face, and voice recognition to analyse the representation of politicians in the media during the Dutch general election in 2021 (MediaSuite, 2021).

Challenges of AI in media archival research

One of the key challenges in this AI use is the importance of transparency of the workings of the systems so that researchers who use them know how they work and how much these tools contribute to the meaning making processes (Aasman *et al.*, 2018). This is highly important as it becomes fundamental for researchers to in fact make conclusions based on their findings – otherwise they risk making wrong conclusions, much like the case was when media professionals used, for example, monitoring tools. This, therefore, requires that researchers like media professionals gain the necessary skills for understanding such results, placing new fields like digital humanities and social sciences at the forefront of critically examining and helping to produce such tools.

EXAMPLES OF AI IN MEDIA ARCHIVAL RESEARCH

The **'CLARIAH Media Suite'** is a part of the CLARIAH research infrastructure aimed at the humanities and social sciences, allowing researchers access to audiovisual data and tools.

The **'BoB for AI'** developed by Learning Screen is a tool that allows for the rediscovery and research into over 2.7 million television and radio podcasts, where for example the gender representation in UK media has been explored based on the data.

The **'Ina Segmenter'** is an audio segmentation toolkit developed by the French Institut national de l'audiovisuel (INA), which allows for the segmentation of male and female voices across audiovisual archival content to research gender representations.

Figure 16: Examples of AI in media archival research

4 Key societal concerns of AI for media

In this second part of the whitepaper, we move from the focus on the unique challenges and potentials that AI applications produce across the stages of the media cycle, to discussing the wider societal concerns that these uses of AI has induced. In this part we, therefore, discuss specific concerns, rather than concrete applications of AI. The concerns discussed have been distilled partly based on the above review, where certain concerns were continuously raised, but also partly from the more general literature on AI (see Appendix B). This part explores how general questions of, for example, labour and biases, are discussed both more widely and in relation to the media sector (here we draw on further literature) to flesh out the unique impacts that the sector faces. The aim of the section is to discuss how these more general concerns specifically affect the media sector and produce some core points of considerations for the industry, policy makers and researchers who engage with the sector. We in the following sections, discuss the following six wider societal concerns:

- Biases and discrimination
- Media (in)dependence and commercialisation
- Inequalities in access to AI
- Labour displacement, monitoring, and professional control
- Privacy, transparency, accountability, and liability
- Manipulation and mis-and disinformation as an institutional threat

4.1 Biases and discrimination

Throughout the media cycle, one of the reoccurring concerns is biases and how such biases might lead to discriminatory practices. AI is often discussed as a double-edged sword. One side, it is seen as tools to mitigate both conscious and unconscious biases in human judgement and decision-making, such as mitigating existing media biases as discussed above. Thereby, offering positive societal impacts relating to, for example, more diversity in coverage, which could improve the public debate and political awareness of previously overlooked societal issues. On the other hand, AI is also built by humans who make decisions on what data to include in the training dataset (which reflect existing societal biases) and how to design the AI system (e.g., by using standard algorithmic models or deciding on including certain metrics), which can replicate or even enhance existing biases by reinforcing certain ways of 'knowing' and 'seeing' in these systems (Campolo *et al.*, 2017; Littman *et al.*, 2021).

This is a risk of AI that has been widely discussed in general (see Appendix B), where applications of AI in, for example, predictive policing, the economic and public sector have been at the centre of the discussions as the implications of biases in these contexts are severe with, for example, the risk of wrongful incarceration or discriminatory actions against loan seekers, sustaining certain groups in society in poverty cycles (Bird *et al.*, 2020). While it is easy (and trendy) and blame AI for being the source of bias and discrimination, the above point must be considered, namely that humans are also biased. Studies in courtrooms have, for example, shown how judges generally are more lenient just after lunch compared to later in the day (Bryant, 2011). However, the increased use of AI systems has helped to reamplify the importance of this problem, which have led action and focus on mitigating the issue of the AI bias problem, through in some cases quite extensive measures such as bans of predictive policing systems in certain countries and calls for governmental caution when implementing automated decision-making systems (ADS), which are often implemented without proper testing and with poor designs (Whittaker *et al.,* 2018). However, as Florian Jaton (2021) has highlighted, it is also important not to 'wash out' the AI bias problem as a technical problem with AI systems, because there can be no AI without biases, when understood as contingent external referent (either data or ground truth databases), but this fact is often forgotten in the current debate. What will be important in the future is to better understand the morality of these systems and allow for room to discuss how these biases should be determined, which we also return to further down.

In the media sector, the discussion on AI produced biases focuses on how such biases could induce highly severe long term negative social and political impacts for societal and cultural groups, and for the special function of journalism to inform and act as a watchdog in our democracy. Such serious implications could, for example, arise because of certain gender representations and racially discriminatory patterns being maintained in the coverage, which could lead certain societal groups to feel underrepresented in the media landscape and disconnect from the public debate as well as produce a skewed portrayal of certain societal topics. Or in the case of price discrimination in the access to news, which could highly affect what news is consumed by citizens. Equally, problematic feedback loops from recommender systems could produce the risk of very individualised and closed off media consumption patterns, which could in the future negatively affect political fragmentation and polarisation or support the circulation of radicalising online content. Last, the proved biases and discriminatory effects of content moderation systems, for example, on Facebook, as well as the limited availability of such tools in certain languages could prove to have highly negative impacts on the public debate online and contribute to highlighting already dominant voices and further undermine minority voices online (e.g., racial or gender). Long-term effects that are critical since media accounts, while increasingly contested, remain considered as representations of 'the reality' by many of its audiences (Reese and Shoemaker, 2016). The problem that AI induced biases and discrimination poses for the representations of that reality for such institutions and the societal, political, and cultural groups who are dependent on such institutions, poses a serious long-term issue that cannot be left for the future. Therefore, it will be important to develop more knowledge on how biases emerge and strategies to monitor and understand what the positive and negative impacts might be, as the implications of biases are not always immediate but arise over time. This will be necessary if meaningful mitigative measures are to be developed to counteract the negative effects of biases in AI applications for media.

Some concrete points of entry to where biases could be mitigated have already been identified. One is, for example, the use of otherwise considered 'golden standard' datasets for training (see e.g., Leslie, 2020; Cordell, 2021; Crawford and Paglen, 2021). A study by IBM, for example recently showed six out of eight of the most used open-source datasets contained more male faces than females and has eighty percent light-skinned faces (Leslie, 2020). Among these, the ImageNet dataset, which is widely used to classify visual data was shown to only have two gender categories 'male' and 'female' and would use 'racists slurs and misogynistic terms' when describing the objects in the images (Crawford and Paglen 2021: 36). As David Leslie (2020) eloquently illustrates, many of such biases have historic roots (for instance, cameras have historically been ill-equipped to capture 'darker' objects, which reflected the cultural norms at the time, where whiteness prevailed as the ideal skin colour). However, such historic racists biases are now carried into facial recognition tools, which then can amplify the issues.

The most important mitigative strategy highlighted across the publications is an awareness of these biases and the focus on interrogating and changing the datasets for the better (see e.g., Cordell 2021; Crawford and Paglen, 2021). As Cordell (2021) highlights in the context of libraries: 'the current state of datasets cannot be dismissed as a regrettable but unavoidable reality' (p.15). Leslie (2020) echoes this plea stating a need to leave behind this apathy towards flaws in data sets and the unwillingness to be transparent about biases and benchmark datasets. This is also why Kate Crawford and Trevor Paglen (2021) in the project on 'Excavating AI' argue that building datasets and in the case of facial recognition annotating images, is an inherently social and political enterprise that demands interrogation and reflection on how choices are made. In these interrogation processes the need for the involvement of domain, social and cultural experts are also highlighted as an important mitigative strategy (see e.g., Cordell, 2021; Campolo *et al.*, 2017). Florian Jaton (2021) notes that also 'ground truthing' practices and ground truth datasets should receive more attention these ground truths are the basis for many working algorithms. Following the point raised by him that biases are unavoidable, many recent discussions have also moved away from the mitigation or attempts to eliminate biases, to rather produce 'socially good' biases that support social justice. This is often referred to as the move from data ethics to data justice (see Cordell 2021 for overview of this discussion). An ambition that could and should apply for media organisations as well.

This would both require interdisciplinarity in the teams working on AI in the media organisations and that resources are available and prioritised in the organisations. Currently, studies have shown the opposite, namely how the media professionals are limitedly involved in the development processes (Fanta and Dachwitz, 2020) and that when involved they often become overruled by technical arguments (Schjøtt Hansen and Hartley, 2021). This can be highly critical, because it can lead to a prioritisation of technical visions of, for example, core values of diversity and universalism (see Sørensen, 2020), but also to the acceptance of purely 'technical fixes' of biases, by for example developing fairer mathematical models, which is highly criticised in the AI Now Institute report from 2018 (Whittaker *et al.*, 2018). However, simply calling for the expertise of social scientists and humanities in tech projects is not enough, as Elena Marris (2022) writes in her critical article on WIRED. Here she highlights how many employees have these backgrounds, but their 'soft data' remains undervalued, meaning they do not in fact change much, as their input is disregarded in comparison with that of technical staff members. We, therefore, see how the 'power shift' identified the previous chapter, where technical expertise is prioritised can have very negative impacts on developing responsible AI in the media sector. What will be essential in the future of media will be to foster responsible AI practices across the design, implementation, monitoring and usage of AI, so that the consequences of, for example, certain biases are both discussed and monitored and later mitigated if necessary. However, this is a difficult task in most media organisations, where the expertise to actually assess and understand the consequences of AI might not be available and where there is a constant pressure to optimise workflows and be efficient, which makes easily available AI solutions the easy choice. Producing tools, strategies, and incentives to help media organisations to introduce such responsible AI practices will, therefore, be imperative.

It will also be important that both knowledge gained from 'excavation experiments' are shared and that datasets that have been excavated are shared in the industry, to provide better alternatives to the more generic open-source datasets available. This is also important because in lack of datasets, many media organisations use their own data foundation (e.g., articles, comment moderation examples etc.), which then will sustain the existing biases of the humans who made those judgements. Making shared datasets that have been interrogated thoroughly and only need small adaptations by the individual media organisations, could make it much more cost-efficient for media organisations to prioritise an increased focus on data biases as well as data justice, as discussed above. Furthermore, an increased awareness of the biased nature of AI systems amongst media professionals and audiences will be vital as currently many media professionals continue to believe in the superior objectivity of AI systems compared to humans (see e.g., Thurman, Dörr and Kunert, 2017). This continued belief in the 'objectivity of algorithms' and the 'technologically inflected promise of mechanical neutrality' (Gillespie, 2014: 181) could induce a limited critical stance towards these projects, as well as the AI-produced content as discussed above. Minimising the critical reflection needed when developing these AI systems and the content they have produced.

CORE POINTS OF CONSIDERATION FOR THE FUTURE

The need for more domain-specific, social and/or cultural expertise in the development process of AI systems for media. All AI projects in the media sector should strive for diversity in the team (e.g., in terms of backgrounds, ethnicities or gender) to ensure that the decisions made regarding datasets, classification or metrics are made on a well-founded and reflective basis. Critically, domain knowledge should be prioritised together with social and cultural knowledge in qualifying these decisions.

The need to foster support, tools, and resources for responsible AI practices in the media sector. Over the last years more awareness has been gained about the need for work with the biases of AI systems, now there is a need to develop concrete tools to support the media organisations in their work as well as foster support and resources for responsible AI practices – something that is challenged with the constant call for optimisation and efficiency within media organisations.

The need for new best practices on how to produce just AI systems in the media sector. Currently, the examples of AI projects promoting data justice are scarce. If the sector is to begin a conversation on ways to achieve this, examples of best practices will be needed. This could be in the form of industry research collaborations.

The need for regulation that supports and fosters responsible AI practices in the media sector, rather than attempt to constrain the use. Often regulatory measures are focused on banning dangerous uses of technologies, there will be a need for policies that rather than constraining provides incentives to adopt responsible AI practices in organisations, because as seen this is difficult with the current conditions in the sector.

The need for domain-specific, open-source and non-commercial datasets for training AI systems. As many AI projects today rely on open-source and 'golden standard' datasets created without consideration for cultural and societal sensitivities and that have proven to induce certain unwanted biases. For the media sector to mitigative the negative effects of such biases and instead induce 'good' or more just biases, domain specific open-source datasets are needed, where there has been time and resources for thorough considerations of what biases to induce by a diverse team.

Figure 17: Core points of considerations for the future - biases and discrimination

4.2 Media (in)dependence and commercialisation

Another concern that runs across the different media cycles is how the use of AI might induce an increased commercialisation of media organisations at the expense of societal responsibility as they become more deeply embedded into the platform economy (see e.g., Lindskow 2016; Van den Bulck and Sørensen 2020). This is not to state that media organisations have not always had a commercial side, private media organisations are a business and PSM's still need to provide legitimisation for their funding by, for example, illustrating their viewership. However, historically these two parts of media organisations have been separated, but over the last 50 years that separation has crumbled (see e.g., Willig, 2011, 2021). AI has proved to further intensify this classic conflict between the editorial and commercial side of media organisations by, for example, pushing the limits of data tracking practices pursued in media organisations (Turow, 2016) or by shifting power to commercial departments who more unquestioned than previously can affect how decisions are made through their knowledge of the infrastructures (e.g., AI or data) (Schjøtt Hansen and Hartley, 2021).

These concerns can be connected to larger discussions of the increased dependence of platform infrastructures in media practices (ranging from audience measurement data to distributional strategies), often discussed as the 'platformisation of news' (van Dijck, Poell and Waal, 2018), where the rise of new powerful intermediaries (Google, Meta etc.) have changed the relationship between media organisations and their audiences by placing themselves in the middle as central organs for the flow of media content (Newman, 2016). As well as the increased datafication of the media sector (and society as a whole), as data is increasingly valorised as part of multiple practices, becoming 'the new oil' in society (van Dijck, 2014; The Economist, 2017). AI is in a sense not producing new concerns but radicalising existing concerns regarding the societal impacts of platformisation and datafication on society, and specifically the role of media organisations.

Some of the potential impacts discussed regarding media organisations concerns how the increased valorisation of data and particularly audience data might impoverish the overall media landscape by affecting what forms of media content is produced by for example valorising certain genres (e.g., more sensationalist content) and de-valorising content that is not 'clickable' but of societal importance – also due to the importance of content circulating well on social media or ranking high in Google News. An increased prioritisation of such externally produced values could result in less production of political news, which is linked to less political participation (Lee and Wei, 2008) and through that have significant negative impact on the political environment in different countries and areas, or the further decline of local news offers. However, again it is never a one-sided argument, as there might also be positive impacts from a more responsive media sector who better understands what is important to the audience, which could lead to more people being able to relate to the media content, potentially improving the general engagement with and opinion of media organisations, which increasingly are considered elitist. Delivering, for example, better recommendations for users is, therefore, not in opposition to the ideal of universalist access to media content. Doing so, however, requires critical awareness of the goals and objectives recommendation and content moderation systems are optimised for, as well as solving difficult questions about the conceptualisation and formalisation of editorial values.

Another way to view the potential negative impacts of particularly platformisation is discussed in relation to how it places immense power in the hands of very few companies (Bird *et al.*, 2020)

– in a European context these are the five ‘tech giants’ Google, Facebook (now Meta), Microsoft, Apple, and Amazon who predominately provide both data and other technical infrastructures for, among other, the media sector[4]. The potential negative impacts of this power imbalance are discussed widely as it places enormous amounts of societal influence and political power in the hands of a few commercial actors, who do not bear a societal responsibility beyond complying with existing regulation, which for the internet, until now, imposes only minimum standards of respect for fundamental rights and democratic values.

The negative social impact from the economic, societal and political power of platforms were discussed lately with the leaking of the so-called ‘Facebook papers’, where it was revealed how Facebook (now Meta) did the absolute minimum to minimise harmful effects related to Instagram and to address the discriminatory effects of their content moderation system (Wells, Hortwitz and Seetharaman, 2021). The fact that very few actors hold much of today’s societal power might also have more long-term effects as discussed by Tarleton Gillespie (2014) who highlights how AI engines like Google Search, Facebook’s feed is impacting the way people participate in social and political discourse and get a sense of what is important. These tech giants and their deployed AI systems become governing forces of what information is available and through that also in legitimising certain forms of knowledge over others (Gillespie, 2014). They are, therefore, not just important intermediaries that induces shifts in ideation and production patterns of media organisations, they also affect the shape of the public discussion through their function of distributing content. An important question moving forward is if, and to what extent this situation will change, once the European Digital Services Act (DSA) has been adopted, which formulates broad obligations for at least some of the largest platforms to monitor their content moderation and recommendation algorithms, their community guidelines, and terms of use for any systemic risks for fundamental rights and society they may create.

Beyond, providing concrete digital communication infrastructure and having important intermediary functions in content distribution, these tech giants are also becoming increasingly vital economic patrons in providing support for digital innovation in the media sector. Both Google and Facebook have developed funding schemes in support of innovation, namely the Facebook Journalism Initiative (FJP) and Google’s Digital News Innovation Fund (DNI), followed by Google’s News Initiative (GNI). All promising to further the digital innovation at media organisations and ensure a sustainable future for the sector. These innovation programs further strengthen the economic and technological dependence between media organisations and the platforms or digital intermediaries. Alexander Fanta and Ingo Dachwitz (2020), for example, show how the DNI Fund over the years became not only a supplement for innovation in media

[4] The reason for mentioning those five and not the ‘big nine’ including the Asian tech giants Tencent, Beidu and Alibaba (van Dijck, Poell and Waal, 2018; Webb, 2019) is that they remain the most influential in the European market, while Tencent owned social media platform TikTok is beginning to shift that balance.

organisation but a vital cashflow driving the innovation[5]. The GNI program has significantly reduced the amount of money to be invested into independent innovation projects based in the media organisations, replacing it with other opportunities that place their own products and infrastructures more at the centre of the funding programmes (e.g., Subscribe with Google or GNI YouTube Innovation funding). Thereby, not only increasing the economic dependence, but also furthering the infrastructural independence between the media organisations and their wider platform ecology and economy (Fanta and Dachwitz, 2020).

Some of the negative impacts discussed based on this dependence is how it might threaten the editorial independence of media organisations, which is central to their societal accountability function. While Alexander Fanta and Ingo Dachwitz (2020) find that Google did not directly influence the media organisations as part of their DNI programme, the funding reportedly did lead to self-censorship in some of the media organisations who had received funding. Google also used the funding scheme as a bargaining chip during, for example, EU negotiations on future and stricter regulations (Fanta and Dachwitz, 2020). Furthermore, the economic dependence induced by the DNI programme might risk either stalling the innovation in the media sector or force media organisations to use the new types of funding, making media organisations more reliant on Google infrastructure and the affordances that they offer – potentially further enhancing some of the datafied or platformised dynamics discussed above.

Last, the more long-term effects of how such funding programmes might sway the overall direction of the development must be discussed. About a decade ago Seth Lewis (Lewis, 2011, 2012) discussed the role of the Knights Foundation in stimulating the innovation in the media sector in the US. Highlighting both how they discursively reconstructed the media landscape by changing their focus from ‘news’ to ‘information’. Equally, the foundation also increasingly began to fund more projects that had a more participatory orientation, compared to more classic media approaches. Based on his analysis, Lewis (2011; 2012) highlights the enormous power of these institutions in driving what the future of media looks like by both rhetorically framing it, but then also supporting that rhetorical frame through actual funding. Similar effects could be expected to happen based on the funding programmes offered by Facebook (now Meta) and Google and it must be critical assessed, which unfortunately is quite difficult as these programmes (the funding amounts and terms) are quite untransparent (see Fanta and Dachwitz 2020). It is worth noting that the European Commission’s ‘European Democracy Action Plan: making EU democracies stronger’ has a string focus on the media sector and also offers a specific funding programme to support the continued democratic role of media (European Commission, 2020), which illustrate that governmental funding schemes are emerging alongside the commercial funding. The AI4Media open calls equally represents ways to fund projects that aims to strengthen the democratic function of media with a foundation in European values.

[5] For transparency reasons it is here also important to note that several AI4Media partners have too received funding via these funds in the past.

Within the audiovisual sector, a very specific dependency was also identified above, which does not relate to the platforms, but their dependence media asset management (MAMs) systems, which are currently limiting the ways in which media archives can currently act in the AI landscape. For this sector, more engagement with vendors to ensure that they offer more flexible, agile, and modular solutions that respond to the needs of the sector and the recent technological advancements in AI will be needed in the future.

CORE POINTS OF CONSIDERATION FOR THE FUTURE

The need for responsible, domain-specific infrastructures to support responsible AI practices. Due to the high reliance on commercialised and platform infrastructures in the development of AI in the media sector, it will be important to develop alternative data and content infrastructures that perhaps better accommodate the European values and are specific to the media sector.

The need for more engagement with media asset management (MAMs) vendors in the audiovisual sector. This will be important to ensure that they offer more flexible, agile, and modular solutions that respond to the needs of the sector and the recent technological advancements in AI will be needed in the future.

The need for best practices and policies of 'diversity by design'. Currently, limited knowledge and best practice exists on how to make the evaluation of whether, for example, a recommender system is successful – not only in a commercial sense. New best practices on how to make such decisions without benchmarking with, for example, purely commercial actors and how to include domain-specific measures of diversity in the projects (e.g., filling the gaps of user knowledge etc.), are needed (e.g., through concrete policies on diversity by design). Furthermore, there is a need for big media companies to be first movers and set the example for the rest of the sector and push this responsible development.

The need for a critical awareness of economic 'patrons' of the media sector and how they affect the development in the media sector. Currently, limited research exists on the role of 'media patrons' and how they affect the future of the media sector. It will be important that more research is conducted, but also that researchers in fact can get access to these processes, as that is currently highly difficult.

The need for funding schemes oriented in EU values. To counteract the growing role of (US-based) platforms in stimulating development, it will be important to develop similar funding schemes that better encompass EU values and the societal function of media.

Figure 18: Core considerations for the future - media (in)dependence and commercialisation

4.3 Inequalities in access to AI

Another related concern regarding AI in the media sector is the inequalities in access to AI solutions by users and AI infrastructure by media organisations. To start with the former as it more directly relates to the discussion above, Beckett (2019) highlights how AI is unevenly distributed in the media sector, where particularly local and regional media with smaller budgets are lacking behind, which can reinforce the existing inequalities in the media sector. Particularly the local and regional media organisations have struggled to find their economic footing in this changing media landscape and consolidation has been a major trend over the many years, where smaller local media organisations are becoming part of large regional media groups (see e.g., Schultz, 2007). The inequalities relating to AI could further amplify this trend by further increasing the divide between local and regional as well as niche media organisations and the large economically more secure media organisations. As discussed above, AI in fact holds promises to reinvigorate the 'local journalism' through the potential scaling of automated content to cover small events and sports. However, this requires that regional and local actors break the barrier of gaining access to such tools and the skills to use them in a responsible way.

Alexander Fanta and Ingo Dachwitz (2020) also show how the funding by the Google DNI Fund, at least in the German context, is oriented towards already large (commercial) media organisations, while smaller start-ups, niche or non-profit organisations are less funded, illustrating how they perhaps selectively stimulate the innovation, making the rich richer, rather than diversifying the access to AI. Again, illustrating how this powerful intermediary elite can stimulate the access to AI into a certain direction, amplifying existing inequalities in the sector (see also Bird, 2020). This increasingly uneven access to AI could have serious social and economic impacts in society by diminishing diversity in the media offering available as certain media organisations unable to leverage the power of AI. An observation that is also highlighted by the interviewees in the survey by Beckett (2019), stating how it becomes a competitive catch up for the smaller media organisations. The same was emphasised by Reginald Chua, executive editor for editorial operations, data, and innovation at Thomson Reuters, who told Andreas Graefe (2016) 'You can't compete if you don't automate' (p.15).

To now return to the first point mentioned, namely how access to AI services by both professional and regular users is also highly unequal across language areas. This has been highlighted multiple times during the review, where the problem of, for example, training data only existing in English or tools predominately being developed in English was mentioned. Following the first discussion on biases, the tools available in other languages also often perform significantly worse. The benefits of AI are, therefore, not shared across the globe, and particular the divide between the global North and South is growing with the increased use of AI across all sectors and therefore also when it comes to media (Bird et al., 2020). The social and political implications of this are vast regarding several AI applications for media, for example, content moderation in large diverse countries like India suffer as they generally have less efficient content moderation and many minority languages experience worse performance, potentially keeping them from partaking in the public debate. There is a need to place increased focus on this inequality of access to ensure that media diversity is sustained and that the benefits of AI become shared, not only by those who are already in privileged positions.

CORE POINTS OF CONSIDERATION FOR THE FUTURE

The need for funding schemes and initiatives focusing on media diversity. It will be important to counteract the trend in private funding identified by Fanta (2020) where established media organisations remain the main beneficiaries of funding for innovation. To not further the increasing competitive divides in the media sector, funding should be specifically oriented towards furthering media diversity.

The need for an increased focus on global AI divides and their consequences. In general, more knowledge is needed on the severity of the AI divide between the global north and south. It will be important to explore the extent of the issue and its implications further.

The development of AI models for diverse languages or adaptive models. To improve the overall access to AI benefits, AI models for large foreign and minority languages should be developed together with adaptive models that can be more efficiently reused for other languages. This could also produce new insights and highlight cultural biases/differences, which in turn could be used to make AI models for the more common languages more accurate.

Figure 19: Core points of considerations for the future - inequalities in access to AI

4.4 Labour displacements, monitoring and professional control

One of the most discussed impacts of AI in the more general literature has been regarding the labour market and the prospect of mass job loss when tasks become increasingly automated (Campolo *et al.*, 2017), with sometimes very high estimates of the job losses to be expected, such as 38 million people in the US being in high risk of losing their job because of automation (Bird *et al*., 2020). While the fear of displacement has become more nuanced both since full automation of many jobs still lies far in the future (Campolo *et al.,* 2017) and as studies have shown that while some jobs will disappear, others will emerge with the growing AI industry (Bird *et al*., 2020). In the review, it was also possible to identify a palpable fear of displacement and a few examples of how AI had in fact led to layoffs of 'human' staff. However, in the media sector the fear of displacement in the discussion has also been nuanced to focus more on the changes AI might impose on the profession and how to maintain the legitimacy of the profession.

One of the impacts of AI has been an emphasised focus on 'technical' or 'data oriented' media professionals (see e.g., Lewis and Usher 2012; 2013), so an upskilling, rather than displacement, but also how the technology and data focus is increasingly legitimised through managerial shifts (see Young and Hermida, 2015) and result in the media professionals become more 'disposable' compared to employees with technical skills who are often much harder to recruit (Schjøtt Hansen and Hartley, 2021; Lewis and Usher, 2013). Such discussions illustrate how AI while focused on replacing routine tasks, might have wider implications for the profession as it changes the organisation more widely and what is valorised, meaning that the potential of better media content might not always be fulfilled. These concerns have yet to be fully explored in research, but indications of such shifts can be identified. This could, therefore, producing new asymmetries in the labour market of the media sector, where certain types of jobs and skills are being devaluated. This relates to a wider impact of AI in the labour market, where routine-based jobs are disappearing (e.g., classic blue-collar jobs) and are being replaced with jobs that need a different skillset. This might require new policies (economic and social) to address this potential societal gap created due to AI, where some groups will be left without jobs (Campolo *et al.*, 2017), such as potentially the more routine jobs in media. More research is needed to address to what extent the media sector is impacted by job losses because of AI

Beyond, impacting what skills are considered important, AI has also been a contributing factor, as part of the overall datafication of the media sector, to the increased importance of data in performance evaluations, recruitment processes and retention decisions, as highlighted by Angele Christin (2018, 2020) and Caitlin Pretre (2021). In the AI Now Institute report from 2017 this is highlighted as a general negative social impact of AI, as the increased reliance on such, often untransparent, tools have impoverished the working conditions in many places by amplifying the power asymmetry between employer and employee or by impacting recruitment processes (Campolo *et al.*, 2017). These professional changes can negatively affect mental health through for example dynamic data visualisations and reminders of goals, which might place stress on the individual to perform or even overperform due to its importance for keeping one's job (Campolo *et al.*, 2017). These systems are also generally developed with the employer and not employee in mind, placing the impacts mainly on the individual (Crawford *et al.*, 2019). Currently, little research is available on the degree of this problem in the media sector, where the focus has more been on how AI affected production and distribution patterns of media content and not how AI enhanced datafied work practices affected media professionals.

As discussed, even with these changes to the profession there remain a strong legitimacy of the role of the journalist, attributed to a very strong professional ideology (Linden, 2017; Deuze and Witschge, 2018). While it will be important to understand how, for example, notions of algorithmic objectivity or authority might challenge the legitimacy of the media sector (see Carlson, 2018 for this discussion), another important aspect will be to see how media professionals continue to exert control over these systems, either in the form of oversight as with automated content or retain elements of curational control on the online sites. Currently, the approach is still by a 'human-in-the-loop' approach (Milosavljević and Vobič, 2019), but as seen with, for example, automated content production oversight is in some cases already proving difficult and the potential negative implications of this must be discussed and solved in the future. Furthermore, the problem of how to ensure that this oversight is meaningful for the media professionals so that they in fact do not only have oversight, but can actually act on this information, will be a highly important focus area in order to foster responsible AI practices (see more on the importance of this in Green and Kak, 2021).

CORE POINTS OF CONSIDERATION FOR THE FUTURE

The need for more research and policies addressing potential displacement patterns resulting from AI. As the increased reliance of AI might result in certain jobs disappearing (e.g., routine tasks) in the media sector as well as across other sectors, providing a societal problem of unemployment. It will, therefore, be important that societal mechanisms and policies are developed to handle the citizens who will be left jobless and in need of specific upskilling.

The need for an increased focus on data and AI in media education. The changes in the media professions also require action from the educational sector who must support students in developing the right skills for the labour market, including increased skills in data and in understanding how AI systems work as well as awareness of the problems connected to these technologies, as misconceptions of 'algorithmic objectivity' still flourish.

The need for meaningful oversight for media professionals. There continues to be a strong emphasis on keeping a 'human-in-the-loop, both in practice and in policies, for most AI applications, to ensure control and oversight. However, this ambition is challenged by the fact that many of these systems remain difficult to have oversight over due to their opacity and scale. To solve this problem and fulfil this ambition it will be important to support the development of 'explainable AI' and human interface design.

The need for more research on AI is changing labour conditions and growing power asymmetries in the media sector. It will be important to understand how the introduction of AI is enhancing already increasing workplace asymmetries, for example, through the use of performances measurements and with what impacts on the individual and society and how it is producing shifts of power within these organisations, valorising technical staff and their approaches.

Figure 20: Core points of considerations for the future - labour displacement, monitoring and professional control

4.5 Privacy, transparency, accountability and liability

A plethora of new concerns regarding AI relate to the users' right to both privacy and transparency in 'who' they are interacting with, but also to how the introduction of AI produces new questions of accountability and liability was brought forward in the above review. This is a discussion that is also echoed in the wider discussions of AI (Campolo *et al.*, 2017; Whittaker *et al.*, 2018; Crawford *et al.*, 2019; Bird *et al.*, 2020; Ada Lovelace Institute, 2021). The potential social or economic impacts on individuals have been highly discussed, for example, in relation to how facial recognition technologies, can allow the identification of individuals across contexts – and even their moods and sexual orientation (Whittaker *et al*., 2018; Bird *et al*., 2020), raising questions of individuals right to privacy in public spaces. Or when the first cases of people being fired based on, for example, GPS data or when data from a pacemaker was used to geographically locate a citizen, leading to his conviction of arson, as a court case in recently set precedence for (Bird *et al.,* 2020). All these uses of the increasing amounts of trackable personal data have raised serious questions of how to protect the data rights of individuals. Something, that perhaps is less present in the debate regarding media, but that can raise new critical questions, for example, if recommender systems would start to use facial expression data to predict and in the case of the 'Who the Hill'. However, here the focus will be on the privacy concerns regarding the generally applied tracking practices in the media sector.

Much like the above discussion on biases, the individual impacts of the data tracking practices related to the use of AI in the media sector, might be much less severe than the examples given above. However, untransparent and potentially excessive tracking practices by media organisations could have fatal consequences for what Neil Richards (2008) has called intellectual privacy', as a critical precondition for not only the fundamental right to privacy but also the exercise of freedom of expression, and the trust in these organisations, impeding them from fulfilling the societal task of providing public information. Such harmful practices could include tracking without consent of the users or using the data for commercial purposes with third parties without consent. While there is already EU regulation that aims at limiting misuse of personal data (e.g., cookie consent and the GDPR), there is an ongoing discussion if such initiatives are adequate and sufficient to deal with the potential privacy implications of the wider proliferation of AI-driven applications in general, (Bird *et al*, 2020; Crawford *et al*., 2019), and in the media specifically (Irion and Helberger, 2017; Eskens, 2019). Under the GDPR, Member States can enact exemptions for 'journalistic purposes', which opens new questions of to what extent these exemptions do apply to the use of AI, for example, in the production or distribution of journalistic content, and if not, whether there is a need for comparable safeguards of the right to privacy, freedom of expression and data protection (Erdos, 2016). With the proposed AI Act and the Digital Services Act, it is likely that the already existing transparency obligations and safeguards against automated decision making, and profiling will be complemented with new transparency obligations.

Users, too, however, highlight the importance of transparency and effective means of exercising agency (Monzer *et al.*, 2020), It will, therefore, be important to increase transparency in data use by media organisations. The BBC's 'personal data stores', mentioned above, represent one example, where an overview of the use of data is made easily available and, more importantly, the user can react based on this data and retract their consent for certain uses. Another example that is often brought up in the wider discussion is the use of 'Data Trusts', which are independent external institutions that help steward the individual's data rights (see interim report '<u>Enabling data sharing for social benefit through data trusts'</u> developed as part of the Global Partnership of AI (GPAI) for an in-depth discussion of this).

In the above review, Article 52 of the draft AI Act was also referenced multiple times, which point to the right of individuals to know whether they are interacting with an AI system. This is again a wide-reaching discussion that both emphasise the right to know whether a user interacts with an AI, but also, for example, in the context of healthcare, whether one can request not to have an AI involved in the process of diagnosis[6] (see e.g., Plough and Holm, 2020). For the media sector, as discussed, the disclosure practices are widely differentiated amongst media organisations (Bastian *et al.*, 2020), illustrating the need for more harmonisation of how media organisations should approach this new challenge. This includes transparency in disclosing when AI has been involved in the process of producing or curating content, but also in how the system came to its decision. The latter is something that is highly discussed, as the concrete working of AI systems are often very complex and opaque. One of the core issues is that the workings of AI are often 'hidden' behind arguments of trade secrecy and protection against misuse of the systems (e.g., the possibility of gaming the systems) (Gillespie, 2014; Bird *et al*., 2020). However as highlighted by Eleanor Bird, Jasmin Fox-Skelly, Nicola Jenner, Ruth Larbey, Emma Weitkamp, and Alan Winfield (2020) this can affect both civil society, researchers, and media organisations, as also discussed above, from being able to hold the organisations accountable. This situation could change with the upcoming DSA and the Digital Markets Act, both stipulating additional transparency obligations on the functioning of ranking and recommendation algorithms, both vis-a-vis professional users as well as end users. Having said so, to be meaningful, transparency interventions need to be accompanied by real choice for users to exercise choice and agency.

Another related question that is highly discussed is the question of liability regarding AI system, because one thing is disclosing that AI, for example, produced a piece of content, but it is another to determine liability, because many new actors are now involved in this question (e.g., external service providers, in-house developers etc.). Currently there are still no clear policies or guidelines on this question, which could have negative impacts on media organisations or individual media professionals. Equally, as Seth Lewis, Amy Kristin Sanders, and Casey Carmody (2017) point to, there is also a risk that current regulation will allow a loophole for AI produced content in the case of for example personal deformation suits. It will, therefore, become highly important to develop more clarity for media organisations on how to act on this question, and how to translate any legal accountability obligations into organisational practices and internal divisions of responsibility between editors, journalists, data, and economic departments.

[6] This is currently not covered in the AI Act, but in the the GDPR, Art. 22, albeit in limited cases.

CORE POINTS OF CONSIDERATION FOR THE FUTURE

The need for more best practices of responsible data practices in the media sector. As the extensive use of data continues to grow in the media sector, it will be vital that new best practices are developed to support responsible data strategies that protect the rights of the individual.

The need for best practices and policies regarding disclosure of AI systems for the media sector. As the question of who produced or curated an article is no longer limited to, for example, journalists, editors, and producers, it will be vital to introduce new guidelines on how to disclose the utilisation of AI in these processes to protect the individual's right to transparency.

The need for explainable and transparent AI solutions that can help users understand how AI systems work and makes decision. As users increasingly are partly serviced by AI systems in their media experience, it is important that they have access to understandable explanations of what the system does and on the basis on what data to uphold their right to, for example, object to the way the decision was made (i.e., agency to act).

The need for clearer regulation and guidelines on the liability question regarding AI. There is a need to help media organisations navigate the liability question that arises from the use of AI systems.

Figure 21: Core considerations for the future - privacy, transparency, accountability, and liability

4.6 Manipulation and mis- and disinformation as an institutional threat

Across the review, there was also a growing concern amongst the media organisations regarding manipulation of content and misinformation. While this was not related specifically to their own work, the negative impacts of the growing amounts of misinformation were seen as highly detrimental to the trust in the media sector, as evident in the survey by Georg Rehm (2020). This is an institutional threat to the existing media landscape, whose legitimacy is increasingly contested as part of this development. It is also a threat to individual users' freedoms, such as freedom of expression, the freedom to hold opinions and what the Council of Europe has coined the 'cognitive autonomy' of individuals (Council of Europe, 2019). Equally, the fact checking genre and independent fact checking institutions are a result of this growing problem and become a new part of the media landscape. This discussion is, therefore, also slightly different as many of the AI systems that are utilised to mitigate such misinformation are developed by social media platforms or to assist fact checking organisations.

The impact of manipulative uses of AI became more publicly discussed following the Cambridge Analytica Scandal, where it was proven that through algorithmic ad targeting, they knowingly attempted to manipulate the national elections in both the US and in the UK. Today other related topics of, for example, the manipulation of markets through artificial training agents or the manipulation of content itself, such as 'deepfakes' where images and audio is manipulated to show someone saying something they have never said. A famous example is the deepfake of

Barack Obama that circulated in 2018 produced by Buzzfeed (Vincent, 2018). Equally, the use of AI (often in the form of bots) to spread 'fake' and propaganda content to manipulate the public opinion on different societal topics by inflating the size of smaller political groups or ideas, is now seen as one of the core societal impacts of AI (Bird *et al*., 2020). The latter became extremely apparent during the Covid-19 pandemic, the war of 'truths' in the online environment became intensified. However, what also became clear here is that simply removing content might also not always be the most appropriate strategy, as discussed by Bengtsson and Schjøtt Hansen (2021). Here the censorship proved to in fact radicalise the anti-systemic thinking amongst Covid-19 sceptics to the point of becoming believers of conspiracy theorist. Equally, the many adversarial tactics utilised by the sceptics challenge the very function of many of the AI tools implemented to mitigate misinformation on social media.

The increasing focus on removing mis- and disinformation with the assistance of AI systems also raise important discussions regarding freedom of expression, as new guidelines for appropriate forms of censorship must be discussed as well as the potential risks of false positive and negatives in these processes and the lack of complaint mechanisms or satisfactory explanations of why content was deleted (see Llanso *et al*., 2020; Gillespie, 2020 for more on this discussion), but also a surge of national regulations, some of which walking a difficult balance between the protection of public interest and interference with fundamental freedoms (Ó Fathaigh, Helberger and Appelman, 2021). Equally, as the practice of fact checking, and particularly AI assisted fact checking grows these practices must also be more explored, as this remain a highly subjective practice, but which is gaining societal importance. Here both the need for more transparency in the workings of the AI systems used to identify misinformation will become important, particularly as they become intertwined with fact checking organisations through strategic partnerships, such as the ones initiated by Facebook (now Meta).

Across the review, there was also a growing concern amongst the media organisations regarding manipulation of content and misinformation. While this was not related specifically to their own work, the negative impacts of the growing amounts of misinformation were seen as highly detrimental to the trust in the media sector, as evident in the survey by Georg Rehm (2020). This is an institutional threat to the existing media landscape, whose legitimacy is increasingly contested as part of this development. It is also a threat to individual users' freedoms, such as freedom of expression, the freedom to hold opinions and what the Council of Europe has coined the 'cognitive autonomy' of individuals (Council of Europe, 2019). Equally, the fact checking genre and independent fact checking institutions are a result of this growing problem and become a new part of the media landscape. This discussion is, therefore, also slightly different as many of the AI systems that are utilised to mitigate such misinformation are developed by social media platforms or to assist fact checking organisations.

The impact of manipulative uses of AI became more publicly discussed following the Cambridge Analytica Scandal, where it was proven that through algorithmic ad targeting, they knowingly attempted to manipulate the national elections in both the US and in the UK. Today other related topics of, for example, the manipulation of markets through artificial training agents or the manipulation of content itself, such as 'deepfakes' where images and audio is manipulated to show someone saying something they have never said. A famous example is the deepfake of Barack Obama that circulated in 2018 produced by Buzzfeed (Vincent, 2018). Equally, the use of AI (often in the form of bots) to spread 'fake' and propaganda content to manipulate the public opinion on different societal topics by inflating the size of smaller political groups or ideas, is now seen as one of the core societal impacts of AI (Bird *et al.*, 2020). The latter became extremely apparent during the Covid-19 pandemic, the war of 'truths' in the online environment became intensified. However, what also became clear here is that simply removing content might also not always be the most appropriate strategy, as discussed by Bengtsson and Schjøtt Hansen (2021). Here the censorship proved to in fact radicalise the anti-systemic thinking amongst Covid-19 sceptics to the point of becoming believers of conspiracy theorist. Equally, the many adversarial tactics utilised by the sceptics challenge the very function of many of the AI tools implemented to mitigate misinformation on social media.

The increasing focus on removing mis- and disinformation with the assistance of AI systems also raise important discussions regarding freedom of expression, as new guidelines for appropriate forms of censorship must be discussed as well as the potential risks of false positive and negatives in these processes and the lack of complaint mechanisms or satisfactory explanations of why content was deleted (see Llanso *et al*., 2020; Gillespie, 2020 for more on this discussion), but also a surge of national regulations, some of which walking a difficult balance between the protection of public interest and interference with fundamental freedoms (Ó Fathaigh, Helberger and Appelman, 2021). Equally, as the practice of fact checking, and particularly AI assisted fact checking grows these practices must also be more explored, as this remain a highly subjective practice, but which is gaining societal importance. Here both the need for more transparency in the workings of the AI systems used to identify misinformation will become important, particularly as they become intertwined with fact checking organisations through strategic partnerships, such as the ones initiated by Facebook.

CORE POINTS OF CONSIDERATION FOR THE FUTURE

The need for mitigative and adaptive AI systems to counteract misinformation. To protect the legitimacy of media organisations and the integrity of the online deliberative spaces, it will be important to develop AI systems to assist in content moderation and fact checking efforts. These must be highly adaptive to be effective and counter act adversarial tactics by groups who spread misinformation.

The need for more transparency in moderation systems and AI fact checking systems. Currently the AI systems used to identify misinformation on social media platforms remain untransparent in their workings and the people who experience consequences do not always have access to a satisfying explanation of why, for example, their profile was deleted or to a complaint mechanism. As many fact checkers are today part of strategic partnerships with Facebook, the need to be transparent will become even more important to sustain legitimacy in these institutions that now serve and important societal function.

The need for more knowledge on fact checking as a social practice and its effects in the deliberate landscape. As fact checking becomes an important societal function, it will be important to gain more in-depth knowledge in how they construct ‘factual’ accounts as well as what the consequences of potentially countering epistemologies of the truth might mean for the deliberative space and societal polarisation.

Figure 22: Core points of consideration for the future - manipulation, mis- and disinformation as an institutional threat

4.7 Summing up

Here we provide an overview to sum-up how the core topics of discussions are characterised, respectively in the more general debate on AI and in the media specific debate.

	General	**Media**
Biases and discrimination	Focuses on the potential impacts of biased AI systems stemming from both the training data and the design of the system, which can lead to discrimination against certain groups in society and sustain economic divides in society	Biases are discussed both as a potential solution to mitigating existing biases but also in terms of how it could have severe long-term effects on the role of media in society and the democratic practices it cultivates.
(In)Dependence	Focuses on the increasing role of big tech platforms in society through, for example, their access to user data and the dependence on access to their infrastructure, which AI is dependent on in many cases. This is often termed the platformisation of society.	The platformisation of society also applies to the media sector, who are dependent on e.g., social media in their distribution and are entangled in commercial data infrastructure needed for AI. Specifically, the effects of such dependencies on the media independence, is a major concern.
Access	Focuses on the skewed access to AI infrastructure both in terms of global benefits of AI (locations and language) and in financial and computational power to develop AI solutions, which negatively affects the market and the sharing of benefits.	Access is also discussed in terms of who benefits from AI for media, namely dominantly the global north and English-speaking context, but also in terms of how it is widening creating a competitional divide between smaller and larger media organisations, minimising media diversity.
Labour	Focuses on the impacts of AI on the labour market, ranging from the impacts on job losses and shifts in demands of the labour market to the potential impacts discriminatory or controlling technologies might have on the welfare and opportunities for employees or jobseekers.	The labour discussion relates to profession transformations (e.g., changes in beats and value of technical skills) and how this is produces new power asymmetries, together with increasing monitoring practices of performance.
Control	Focuses on the importance of control and human oversight with AI to ensure that negative impacts of AI are discovered and how this is also key in ensuring trust in these technologies e.g., in the health sector.	In media certain control strategies have been employed e.g., by only personalising part of the site, but the oversight and control are challenged both by the scale and opaqueness of AI.
Privacy	Focuses on the increased surveillance and infringements on the privacy rights induced by AI as well as the potential conflicts of data rights and AI systems, which might impact the user's ability to be both capable of and have the opportunity to decide over how their personal data is used.	The privacy discussion in media relate mostly to data privacy, where the conflict between commercial and democratic ideals intersects. Media organisations must consider their responsibility regarding data privacy models and new best practices of responsible data practices are needed.
Transparency	Focuses on the importance of ensuring transparency of the workings of AI system for the impacted (both individuals or organisations) have the possibility of interrogating the systems and explore the consequences to provide grounds for accountability.	Transparency is mainly discussed regarding the level of disclosure that media organisations should have when employing AI and the need for more access to the workings of other AI systems as the foundation for 'algorithmic accountability reporting'.
Accountability and liability	Focuses on the need for accountability mechanisms to ensure that people or organisations that have been negatively impacted can hold someone liable.	Accountability is mainly discussed in relation to how to place responsibility now that new actors enter the media landscape (e.g., service providers of AI).
Manipulation	Focuses on the potential impacts of manipulated content (e.g., deep fakes) or the manipulation of political elections through targeted advertisement or by artificially enhancing the prominence of certain voices and topics in the public debate. This can lead to negative effects on the workings of democracy but also negative effects on individuals who might be victims of fraud.	Manipulation is highly present in the media discussion of AI, but mainly regarding how other actors through the manipulation of content or public discourse can affect the trustworthiness of the media organisations. As a core actor in the fight against disinformation, the development of tools to support the work of media professionals is important.

Figure 23: Table that sums up core societal concerns - more in general and for the media sector

5 Industry workshops

This part will be added in the extended version to be published in December 2023.

From the spring of 2022, a series of industry workshops will be carried out by partners in the AI4Media consortium, to among other provide input to the updated version of this whitepaper.

Each of these workshops will focus on a specific topic or problem identified in the above review and invite relevant industry actors to join a workshop together with researchers to discuss the issue they are facing and new best practices for that topic. The topics will be chosen on an ongoing basis and relevant media organisations from the consortium as well as external organisations will be invited to participate by the organisers.

These workshops will be useful in producing concrete examples of responsible AI practices through case studies and concrete recommendations for other media organisations facing similar issues. The outputs of the workshops will also feed into delivering concrete policy recommendation for the media sector.

6 Conclusion and forward gaze

In this initial whitepaper on the social, economic, and political impact of media AI technologies we have explored the concrete potentials and challenges of different applications across the stages of the media cycle. This helps to produce an overview of the state of discussion regarding AI for media, and points to focal points for future work aimed at dealing with some of the challenges posed, such as access to public documents or the operationalising of values in AI systems. However, based on this section of the review it is also possible to conclude that more knowledge is needed, as in some cases there were indications of some of the potential, such as efficiency and increased quality of media content (by freeing up the time of media professionals). It will be important in the future to better understand and document what the concrete benefits of AI are as well as the negative consequences they might have, as in current literature remain to large degrees speculative and the long-term implications have yet to materialise.

The second part of the whitepaper further helped to shed light on what the wider impacts of AI in the media sector might be and where action must be taken on a wider level, rather than regarding the individual application. Here several concerns were discussed, including: (i) Biases and discrimination, where the importance of not simply 'blaming' biased AI, but foster responsible AI practices was highlighted. (ii) Media (in)dependence and commercialisation, where the importance of understanding and mitigating the negative effects of platform dependency (in terms of both infrastructure and economy) was highlighted. (iii) Inequalities in access to AI, where the language and global north and south divide in AI was highlighted as well as the risk of leaving behind local media outlets who do not have the capacity to develop AI themselves. (iv) Labour displacement, monitoring and professional control, where it was highlighted that while AI promises to only replace routine tasks, there are indications of a devaluation of creative work while technical skills are increasingly valued. Equally, the increasing monitoring and focus on performance induced by data and AI systems in the workplace (such as audience metrics), and its possible negative impacts, was highlighted. (v) Privacy, transparency, accountability, and liability, where it was highlighted how AI raises many new questions regarding how to legally ensure the rights of individuals, transparency as well as accountability practices regarding AI for media. (vi) Manipulation and mis- and disinformation as institutional threat, where both the concrete impacts this might have in the trust of media organisations were highlighted as well as the consequences for a well-functioning democracy.

Based on these discussions, we distilled several core points of considerations for industry professionals, policy makers and researchers, which can be found both in that section and in the executive summary. These will be used as a starting point in the work in the consortium going forward, both in terms of framing the coming industry workshops, but also as a foundation for the work of developing concrete policy recommendations that can support a responsible development and uptake of AI in the media sector, which is the mission of the AI4media project.

7 Appendices

7.1 Appendix A: In-depth readings on specific sector and applications of AI for the media and related industries

7.1.1 Readings focusing on the journalism sector

- AI Journalism Starter Pack. A practical guide designed to help news organisations learn about the opportunities offered by AI to support their journalism.

- New powers, new responsibilities. A global survey of journalism and artificial intelligence. Looks at the use of AI across the media cycles as well as more general questions of responsibility and strategy.

- Guide to Automated Journalism. Looks specifically at the use of NLP/NLG applications in the journalism sectors, exploring the potentials and implications for journalistic organisations, journalists, and the public.

- How Artificial Intelligence Can Help Us Crack More Panama Papers Stories. Article by the International Consortium of Investigative Journalism (ICIJ) that explores the uses of AI in investigative journalism.

7.1.2 Readings focusing on the audiovisual sector

- The use of Artificial Intelligence in the Audiovisual Sector. Looks at the use of AI across the media cycles in the audiovisual sector and explores more general questions and concerns for the sector.

- Artificial intelligence in the audiovisual sector. That explores the concrete challenges raised regarding AI in the audiovisual sector, providing a valuable overview of questions of freedom of speech, cultural diversity, copyright, targeted advertisement, and personality rights.

7.1.3 Readings focusing on audiovisual archives

- AudioVisual Data in DH. Digital Humanities Quarterly issue dedicated to AV data with several examples of AI-driven projects.

- AI in relation to GLAMs. A EuropeanaTech task force have explored the role and impact of artificial intelligence in the digital cultural heritage domain, especially with regards to collections analysis and improvement

- Audiovisual Data in Digital Humanities. VIEW journal issue dedicated to AV and digital humanities.

- Automatic Annotations and Enrichments for Audiovisual Archives. Conference contribution presented at Special Session on Artificial Intelligence and Digital Heritage.

7.1.4 Readings focusing on AI use in libraries

- Machine Learning + Libraries: A Report on the State of the Field. Explores the potentials and challenges related to the use of AI in the library sector.

7.1.5 Readings focusing on the creative industries

- AI in the media and creative industries. Report that explores the potentials and challenges of the use of AI in the media and creative industries.
- Artificial Intelligence in the Creative Industries: A Review. This paper reviews the current state of the art in artificial intelligence (AI) technologies and applications in the context of the creative industries.

- Ghost in the (Hollywood) machine: Emergent applications of artificial intelligence in the film industry. This article examines the nascence of artificial intelligence (AI) applications in the film industry at the greenlighting stage, where decisions are made as to the feasibility and earning potential of film projects.

7.1.6 Readings focusing on the economic aspects of AI in the media sector

- Google, the media patron How the digital giant ensnares journalism. Report that explores the role of Google in funding innovation in the German media sector.

- How do emerging technologies affect the creative economy? Report that explores emerging technologies among other AI's effects on the creative economy—art, journalism, music, and more.

7.2 Appendix B: Further readings on AI in a wider societal context

- The AI Now 2017 report, has a general focus on societal impacts of AI, but with emphasis on automation and labour, bias and inclusion, rights and liberties and ethics and governance.

- The AI Now 2018 report, has a general focus on societal impacts of AI, but with emphasis on AI and accountability, surveillance, protection of rights, fairness, biases and discrimination as well as ethics.

- The AI Now 2019 report, has a general focus on societal impacts of AI, but with emphasis on policies for ensuring responsible use of AI, regulation of and adoption of AI by governmental actors, race and gender disparities and the negative impacts on climate change and lack of access to AI in the global South.

- Algorithmic accountability for the public sector by the Ada Lovelace Institute explores the initial wave of algorithmic accountability policy for the public sector.

- The ethics of artificial intelligence, commissioned by the European Parliament deals with the ethical implications and moral questions that arise from the development and implementation of artificial intelligence (AI) technologies.

- Understanding bias in facial recognition technologies, published by the Turing Institute explores the broader ethical questions around the potential proliferation of pervasive face-based surveillance infrastructures and makes some recommendations for cultivating more responsible approaches to the development and governance of these technologies.

- Gathering Strength, Gathering Storms: The One Hundred Year Study on Artificial Intelligence (AI100) 2021 Study Panel Report, published by Stanford University, sets forward 14 questions regarding AI, which has been discussed and answered by relevant panels.

8 References

Aasman, S. *et al.* (2018) 'Tales of a Tool Encounter: Exploring Video Annotation for Doing Media History', *VIEW Journal of European Television History and Culture*, 7(14), pp. 73–87. doi:10.18146/2213-0969.2018.jethc154.

Ada Lovelace Institute (2021) *Algorithmic accountability for the public sector: Learning from the first wave of policy implementation*. London, UK: The Ada Lovelace Institute. Available at: https://www.adalovelaceinstitute.org/report/algorithmic-accountability-public-sector/ (Accessed: 23 November 2021).

Adair, B. *et al.* (2019) 'The human touch in automated fact-checking', in *Proceedings of the computation+ journalism symposium*.

Adar, E. *et al.* (2017) 'PersaLog: Personalization of news article content', in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pp. 3188–3200.

AIJO Project (2021) *Home, The AIJO Project*. Available at: https://www.aijoproject.com (Accessed: 18 January 2022).

AIJO Project - Experimenting with Computer Vision (2021) *Experimenting with Computer Vision, The AIJO Project*. Available at: https://www.aijoproject.com/post/experiment-documentation (Accessed: 18 January 2022).

AIJO Project, Experimenting with Natural Language Processing (2021) *Experimenting with Natural Language Processing, The AIJO Project*. Available at: https://www.aijoproject.com/post/experimenting-with-nlp (Accessed: 13 February 2022).

Akers, M.A. (2008) 'Christian Site's Ban on "G" Word Sends Homosexual to Olympics'. Available at: http://voices.washingtonpost.com/sleuth/2008/07/christian_sites_ban_on_g_word.html (Accessed: 18 January 2022).

Amato, G. *et al.* (2019) 'AI in the media and creative industries', *arXiv:1905.04175 [cs]* [Preprint]. Available at: http://arxiv.org/abs/1905.04175 (Accessed: 21 September 2021).

Analyse & Tal (2021) *Angreb i den offentlige debat på Facebook*. Copenhagen, Denmark: TrygFonden. Available at: https://strapi.ogtal.dk/uploads/966f1ebcfa9942d3aef338e9920611f4.pdf.

Anderson, A.A. *et al.* (2014) 'The "Nasty Effect:" Online Incivility and Risk Perceptions of Emerging Technologies', *Journal of Computer-Mediated Communication*, 19(3), pp. 373–387. doi:10.1111/jcc4.12009.

Anderson, A.A. *et al.* (2018) 'Toxic talk: How online incivility can undermine perceptions of media', *International Journal of Public Opinion Research*, 30(1), pp. 156–168.

Anderson, C. (2011) 'Between creative and quantified audiences: Web metrics and changing patterns of newswork in local US newsrooms', *Journalism*, 12(5), pp. 550–566. doi:10.1177/1464884911402451.

Anderson, T. (2020) 'Dutch national broadcaster saw ad revenue rise when it stopped tracking users. It's meant to work like that, right?' Available at: https://www.theregister.com/2020/07/03/stop_tracking_increase_revenue_effectiveness/ (Accessed: 24 February 2022).

Athique, A. (2018) 'The dynamics and potentials of big data for audience research', *Media, Culture & Society*, 40(1), pp. 59–74. doi:10.1177/0163443717693681.

Bastian, M. *et al.* (2020) 'Explanations of news personalisation across countries and media types', *Internet Policy Review*, 9(4). Available at: https://policyreview.info/articles/analysis/explanations-news-personalisation-across-countries-and-media-types (Accessed: 14 February 2022).

BBC Four (2018) *BBC Four - Made by Machine: When AI Met the Archive*, *BBC*. Available at: https://www.bbc.co.uk/programmes/b0bhwk3p (Accessed: 18 January 2022).

Beckett, C. (2019) *New powers, new responsibilities: A global survey of journalism and artificial intelligence.* Polis, London School of Economics and Political Science. Available at: https://blogs. lse. ac. uk/polis/2019/11/18/new-powers-new-responsibilities.

Bengtsson, M., Schjøtt Hansen, A. (2021, august 18th). *Battling for deliberative spaces through hybrid quantification logics during the COVID-19 pandemic* [Conference presentation as part of panel titled 'Datafied publics? Exploring datafication of journalism']. NordMedia2021 18-20th of August, Iceland (virtual).

Binns, A. (2017) 'Fair game? Journalists' experiences of online abuse', *Journal of Applied Journalism & Media Studies*, 6(2), pp. 183–206. doi:10.1386/ajms.6.2.183_1.

Bird, E. *et al.* (2020) *The ethics of artificial intelligence: Issues and initiatives*. EPRS | European Parliamentary Research Service. Available at: https://www.europarl.europa.eu/RegData/etudes/STUD/2020/634452/EPRS_STU(2020)634452_EN.pdf.

Biskupic, J., Roberets, J. and Shiffman, J. (2014) 'The Echo Chamber', *Reuters*. Available at: http://www.reuters.com/investigates/special-report/scotus/ (Accessed: 24 November 2021).

Blanke, T. and Wilson, J. (2017) 'Identifying epochs in text archives', in *2017 IEEE International Conference on Big Data (Big Data)*. IEEE, pp. 2219–2224.

Bocyte, R. (2021) '4u2 Messenger: Personalising the Distribution of Television Archives', *ReTV: Re-Inventing TV for the Interactive Age*, 24 February. Available at: https://retv-project.eu/2021/02/24/4u2-messenger-personalising-the-distribution-of-television-archives/ (Accessed: 13 February 2022).

Bocyte, R. and Oomen, J. (2022) 'Content Adaptation, Personalisation and Fine-grained Retrieval: Applying AI to Support Engagement with and Reuse of Archival Content at Scale', in. *Special Session on Artificial Intelligence and Digital Heritage: Challenges and Opportunities*, pp. 506–511. Available at: https://www.scitepress.org/Link.aspx?doi=10.5220/0009188505060511 (Accessed: 13 February 2022).

Bodo, B. *et al.* (2017) 'Tackling the Algorithmic Control Crisis - The Technical, Legal, and Ethical Challenges of Research into Algorithmic Agents', *Yale Journal of Law and Technology*, 19, pp. 133–181.

Bodó, B. (2019) 'Selling News to Audiences – A Qualitative Inquiry into the Emerging Logics of Algorithmic News Personalization in European Quality News Media', *Digital Journalism*, 7(8), pp. 1054–1075. doi:10.1080/21670811.2019.1624185.

Boerman, S.C., Kruikemeier, S. and Zuiderveen Borgesius, F.J. (2017) 'Online behavioral advertising: A literature review and research agenda', *Journal of advertising*, 46(3), pp. 363–376.

Bozdag, E. (2013) 'Bias in algorithmic filtering and personalization', *Ethics and Information Technology*, 15(3), pp. 209–227. doi:10.1007/s10676-013-9321-6.

Bridy, A. (2012) 'Coding creativity: copyright and the artificially intelligent author', *Stan. Tech. L. Rev.*, p. 5.

Broussard, M. (2015) 'Artificial Intelligence for Investigative Reporting: Using an expert system to enhance journalists' ability to discover original public affairs stories', *Digital Journalism*, 3(6), pp. 814–831. doi:10.1080/21670811.2014.985497.

Bruns, A. (2009) 'From Prosumer to Produser: Understanding User-Led Content Creation', in *Transforming Audiences 2009*. *Transforming Audiences 2009*. Available at: https://eprints.qut.edu.au/27370/ (Accessed: 24 November 2021).

Bruns, A. (2019) *Are Filter Bubbles Real?* Wiley.

Bruun-Hansen, K. (2021a) 'B.T. fyrer: Erstatter community managers med robotter', *Journalisten*, 25 February. Available at: https://journalisten.dk/b-t-fyrer-erstatter-community-managers-med-robotter/ (Accessed: 18 January 2022).

Bruun-Hansen, K. (2021b) 'Hvem skal tøjle B.T.'s debat, når robotterne overtager SoMe?', *Journalisten*, 26 February. Available at: https://journalisten.dk/hvem-skal-toejle-b-t-s-debat-naar-robotterne-overtager-some/ (Accessed: 18 January 2022).

Bryant, B. (2011) 'Judges are more lenient after taking a break, study finds', *The Guardian*, 11 April. Available at: https://www.theguardian.com/law/2011/apr/11/judges-lenient-break (Accessed: 23 November 2021).

Bucher, T. (2017) '"Machines don't have instincts": Articulating the computational in journalism', *New Media & Society*, 19(6), pp. 918–933. doi:10.1177/1461444815624182.

Bucher, T. (2018) *If... then: Algorithmic power and politics*. Oxford University Press.

Burgess, J., Vis, F. and Bruns, A. (2012) 'How many fake Sandy pictures were really shared on social media?', *The Guardian*, 6 November. Available at: http://www.theguardian.com/news/datablog/2012/nov/06/fake-sandy-pictures-social-media (Accessed: 13 February 2022).

Burke, R., Felfernig, A. and Göker, M.H. (2011) 'Recommender systems: An overview', *Ai Magazine*, 32(3), pp. 13–18.

Cambridge University Press Research (2016) 'Aesthetics, athletics and the olympics', *Cambridge University Press Research*. Available at: https://www.cambridge.org/about-us/news/aest (Accessed: 18 January 2022).

Campolo, A. *et al.* (2017) *AI Now Report 2017*. AI Now Institute at New York University. Available at: https://ainowinstitute.org/AI_Now_2017_Report.pdf.

Carlson, M. (2015) 'The Robotic Reporter: Automated journalism and the redefinition of labor, compositional forms, and journalistic authority', *Digital Journalism*, 3(3), pp. 416–431. doi:10.1080/21670811.2014.976412.

Carlson, M. (2018) 'Automating judgment? Algorithmic judgment, news knowledge, and journalistic professionalism', *New Media & Society*, 20(5), pp. 1755–1772. doi:10.1177/1461444817706684.

Cechine, R. (2021a) 'AI in Audiovisual Archives: What Kinds of Analysis Are Possible? | Sound and Vision'. Available at: https://beeldengeluid.nl/en/knowledge/blog/ai-audiovisual-archives-what-kinds-analysis-are-possible (Accessed: 26 November 2021).

Cechine, R. (2021b) 'Bringing AI into the Audiovisual Archive | Sound and Vision'. Available at: https://beeldengeluid.nl/en/knowledge/blog/bringing-ai-audiovisual-archive (Accessed: 26 November 2021).

Cechine, R. (2021c) 'The Potential for AI in Audiovisual Archives | Sound and Vision'. Available at: https://beeldengeluid.nl/en/knowledge/blog/potential-ai-audiovisual-archives (Accessed: 26 November 2021).

Chambers, S. (2021) 'Truth, Deliberative Democracy, and the Virtues of Accuracy: Is Fake News Destroying the Public Sphere?', *Political Studies*, 69(1), pp. 147–163. doi:10.1177/0032321719890811.

Chandler, S. (2021) 'Reuters Uses AI To Prototype First Ever Automated Video Reports', *Forbes.com*. Available at: https://www.forbes.com/sites/simonchandler/2020/02/07/reuters-uses-ai-to-prototype-first-ever-automated-video-reports/?sh=3f1aeb607a2a (Accessed: 18 January 2022).

Christin, A. (2018) 'Counting clicks: Quantification and variation in web journalism in the United States and France', *American Journal of Sociology*, 123(5), pp. 1382–1415.

Christin, A. (2020) *Metrics at Work*. Princeton University Press. Available at: https://press.princeton.edu/books/ebook/9780691200002/metrics-at-work (Accessed: 13 February 2022).

Cibrián, E. *et al.* (no date) 'Artificial Intelligence and Machine Learning for commercial analysis in the audiovisual sector: a case study of designing TV schedules'. Available at: http://www.kr.inf.uc3m.es/artificial-intelligence-and-machine-learning-for-commercial-analysis-in-the-audiovisual-sector-a-case-study-of-designing-tv-schedules/ (Accessed: 18 January 2022).

Clerwall, C. (2014) 'Enter the Robot Journalist', *Journalism Practice*, 8(5), pp. 519–531. doi:10.1080/17512786.2014.883116.

Coddington, M. (2015) 'Clarifying Journalism's Quantitative Turn: A typology for evaluating data journalism, computational journalism, and computer-assisted reporting', *Digital Journalism*, 3(3), pp. 331–348. doi:10.1080/21670811.2014.976400.

Coe, K., Kenski, K. and Rains, S.A. (2014) 'Online and Uncivil? Patterns and Determinants of Incivility in Newspaper Website Comments', *Journal of Communication*, 64(4), pp. 658–679. doi:10.1111/jcom.12104.

Colavizza, G. *et al.* (2021) 'Archives and AI: An Overview of Current Debates and Future Perspectives', *arXiv preprint arXiv:2105.01117* [Preprint].

Collins, B. and Zadrozny, B. (2021) 'Who is the "pizza king"? The secret language being used by anti-vaccine groups to skirt detection', *NBC News*. Available at: https://www.nbcnews.com/tech/tech-news/anti-vaccine-groups-changing-dance-parties-facebook-avoid-detection-rcna1480 (Accessed: 18 January 2022).

Corcoran, L. (2018) 'Not all news site visitors are created equal. Schibsted is trying to predict the ones who will pay up', *Nieman Lab*. Available at: https://www.niemanlab.org/2018/02/not-all-news-site-visitors-are-created-equal-schibsted-is-trying-to-predict-the-ones-who-will-pay-up/ (Accessed: 18 January 2022).

Cordell, R. (2020) *Machine Learning + Libraries: A Report on the State of the Field*. LC Labs Library of Congress.

Council of Europe (2019) *Declaration by the Committee of Ministers on the manipulative capabilities of algorithmic processes*, *Data Protection*. Available at: https://www.coe.int/en/web/data-protection/newsroom/-/asset_publisher/7oll6Oj8pbV8/content/declaration-by-the-committee-of-ministers-on-the-manipulative-capabilities-of-algorithmic-processes (Accessed: 14 February 2022).

Crawford, K. *et al.* (2019) *The AI Now Report 2019*. AI Now Institute at New York University. Available at: https://ainowinstitute.org/AI_Now_2016_Report.pdf.

Crawford, K. and Paglen, T. (2021) 'Excavating AI: The politics of images in machine learning training sets', *AI & SOCIETY*, pp. 1–12.

CREATE (2021) 'Education - unsilencing the VOC testaments'. Available at: https://www.create.humanities.uva.nl/education/unsilencing-the-archive/ (Accessed: 13 February 2022).

Dahlgren, P.M. (2020) 'Media echo chambers: Selective exposure and confirmation bias in media use, and its consequences for political polarization'.

van Dalen, A. (2012) 'THE ALGORITHMS BEHIND THE HEADLINES: How machine-written news redefines the core skills of human journalists', *Journalism Practice*, 6(5–6), pp. 648–658. doi:10.1080/17512786.2012.667268.

van Dalen, A. (2020) *MFI undersøger Synet på nyhedsalgoritmer*. Mediernes forsknings- og innovationscenter, MFI. Available at: file:///Users/gast/Downloads/MFI%20SYNET%20PAA%20NYHEDSALGORITMER%20(1).pdf.

Daws, R. (2021) 'Facebook claims its AI reduced hate by 50% despite internal documents highlighting failures', *AI News*, 18 October. Available at: https://artificialintelligence-news.com/2021/10/18/facebook-claims-ai-reduced-hate-50-internal-documents-highlighting-failures/ (Accessed: 16 December 2021).

Deuze, M. and Witschge, T. (2018) 'Beyond journalism: Theorizing the transformation of journalism', *Journalism*, 19(2), pp. 165–181. doi:10.1177/1464884916688550.

Diakopoulos, N. (2014) 'Algorithmic accountability reporting: On the investigation of black boxes'.

Diakopoulos, N. (2015) 'Algorithmic Accountability: Journalistic investigation of computational power structures', *Digital Journalism*, 3(3), pp. 398–415. doi:10.1080/21670811.2014.976411.

Diakopoulos, N. and Naaman, M. (2011) 'Towards quality discourse in online news comments', in *Proceedings of the ACM 2011 conference on Computer supported cooperative work*, pp. 133–142.

van Dijck, J. (2014) 'Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology', *Surveillance & Society*, 12(2), pp. 197–208. doi:10.24908/ss.v12i2.4776.

van Dijck, J., Poell, T. and Waal, M. de (2018) *The Platform Society: Public Values in a Connective World*. Oxford University Press.

Dörr, K.N. (2016) 'Mapping the field of Algorithmic Journalism', *Digital Journalism*, 4(6), pp. 700–722. doi:10.1080/21670811.2015.1096748.

Dörr, K.N. and Hollnbuchner, K. (2017) 'Ethical Challenges of Algorithmic Journalism', *Digital Journalism*, 5(4), pp. 404–419. doi:10.1080/21670811.2016.1167612.

Dwoskin, E., Tiku, N. and Timberg, C. (2021) 'Facebook's race-blind practices around hate speech came at the expense of Black users, new documents show', *Washington Post*, 21 November. Available at: https://www.washingtonpost.com/technology/2021/11/21/facebook-algorithm-biased-race/ (Accessed: 18 January 2022).

Ellemark, K. (2018) 'Därför förenklar Mittmedia startsidorna och skapar personaliserade flöden', *mittmedia*, 25 September. Available at: https://medium.com/mittmedia/d%C3%A4rf%C3%B6r-f%C3%B6renklar-mittmedia-startsidorna-och-skapar-personaliserat-fl%C3%B6d-677d98e546a1 (Accessed: 18 January 2022).

Ellis, J. (2015) 'What happened after 7 news sites got rid of reader comments', *Nieman Lab*. Available at: https://www.niemanlab.org/2015/09/what-happened-after-7-news-sites-got-rid-of-reader-comments/ (Accessed: 16 December 2021).

EPSTEIN, E.J. (2000) *News from Nowhere. Television and the News. Ivan R*. Dee Publisher, Chicago.(Publicación original en Random House, Nueva York, 1973).

Erdos, D. (2016) *Statutory Regulation of Professional Journalism Under European Data Protection: Down But Not Out?* SSRN Scholarly Paper ID 2817823. Rochester, NY: Social Science Research Network. doi:10.2139/ssrn.2817823.

Eskens, S. (2019) 'A right to reset your user profile and more: GDPR-rights for personalized news consumers', *International Data Privacy Law*, 9(3), pp. 153–172. doi:10.1093/idpl/ipz007.

European Commission (2020) *European Democracy Action Plan*, *European Commission - European Commission*. Available at: https://ec.europa.eu/commission/presscorner/detail/en/ip_20_2250 (Accessed: 13 February 2022).

Fanta, A. and Dachwitz, I. (2020) *Google, the media patron. How the digital giant ensnares journalism.* preprint. SocArXiv. doi:10.31235/osf.io/3qbp9.

Fenton, N. (2010) 'Drowning or Waving? New Media, Journalism and Democracy', in Fenton, N., *New Media, Old News: Journalism & Democracy in the Digital Age*. 1 Oliver's Yard, 55 City Road, London EC1Y 1SP United Kingdom: SAGE Publications Ltd, pp. 3–16. doi:10.4135/9781446280010.n1.

Ford, H. and Hutchinson, J. (2019) 'Newsbots that mediate journalist and audience relationships', *Digital Journalism*, 7(8), pp. 1013–1031.

Gardiner, B. (2018) '"It'sa terrible way to go to work:" what 70 million readers' comments on the Guardian revealed about hostility to women and minorities online', *Feminist Media Studies*, 18(4), pp. 592–608.

Gillespie, T. (2014) 'The relevance of algorithms', *Media technologies: Essays on communication, materiality, and society*, 167(2014), p. 167.

Gillespie, T. (2020) 'Content moderation, AI, and the question of scale', *Big Data & Society*, 7(2), p. 2053951720943234. doi:10.1177/2053951720943234.

Graefe, A. (2016) *Guide to automated journalism*. US: The Tow Center for Digital Journalism.

Graefe, A. *et al.* (2018) 'Readers' perception of computer-generated news: Credibility, expertise, and readability', *Journalism*, 19(5), pp. 595–610. doi:10.1177/1464884916641269.

Graves, L. (2018) *Understanding the promise and limits of automated fact-checking*. Oxford: Reuters Institute for the Study of Journalism. Available at: https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2018-02/graves_factsheet_180226%20FINAL.pdf.

Graves, L. and Cherubini, F. (2016) *The rise of fact-checking sites in Europe*. Oxford: Reuters Institute for the Study of Journalism.

Green, B. and Kak, A. (2021) 'The False Comfort of Human Oversight as an Antidote to A.I. Harm', *Slate*, 15 June. Available at: https://slate.com/technology/2021/06/human-oversight-artificial-intelligence-laws.html (Accessed: 24 February 2022).

Groot Kormelink, T. and Costera Meijer, I. (2014) 'Tailor-made news: Meeting the demands of news users on mobile and social media', *Journalism Studies*, 15(5), pp. 632–641.

Haim, M. and Graefe, A. (2017) 'Automated News', *Digital Journalism*, 5(8), pp. 1044–1059. doi:10.1080/21670811.2017.1345643.

Hamborg, F., Donnay, K. and Gipp, B. (2019) 'Automated identification of media bias in news articles: an interdisciplinary literature review', *International Journal on Digital Libraries*, 20(4), pp. 391–415. doi:10.1007/s00799-018-0261-y.

Hansen, M. *et al.* (2017) *Artificial Intelligence: Practice and Implications for Journalism*. Tow Center for Digital Journalism, Columbia University. Available at: https://doi.org/10.7916/D8X92PRD (Accessed: 24 November 2021).

Hartley, J.M. (2011) *Radikalisering af kampzonen: en analyse af netjournalistisk praksis og selvforståelse i spændingsfeltet mellem idealer og publikum*. PhD Thesis. Roskilde Universitet. Available at: https://forskning.ruc.dk/files/33188606/Afhandling_Hartley.pdf.

Hassan, N. *et al.* (2017) 'Toward Automated Fact-Checking: Detecting Check-worthy Factual Claims by ClaimBuster', in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY, USA: Association for Computing Machinery (KDD '17), pp. 1803–1812. doi:10.1145/3097983.3098131.

Helberger, N. (2011) 'Diversity by design', *Journal of Information Policy*, 1(1), pp. 441–469.

Helberger, N. (2013) 'Freedom of expression and the Dutch cookie-wall', *Available at SSRN 2197251* [Preprint].

Helberger, N. (2015) 'Public service media| merely facilitating or actively stimulating diverse media choices? public service media at the crossroad', *International Journal of Communication*, 9, p. 17.

Helberger, N. (2016) 'Policy implications from algorithmic profiling and the changing relationship between newsreaders and the media', *Javnost-The Public*, 23(2), pp. 188–203.

Helberger, N. (2019) 'On the Democratic Role of News Recommenders', *Digital Journalism*, 7(8), pp. 993–1012. doi:10.1080/21670811.2019.1623700.

Henningsen, L. (2021) 'Forskere fra København vil bekæmpe kvindehad, men risikerer at ramme uskyldige', *Politiken*, 4 August. Available at: https://politiken.dk/kultur/art8311629/Forskere-fra-K%C3%B8benhavn-vil-bek%C3%A6mpe-kvindehad-men-risikerer-at-ramme-uskyldige (Accessed: 16 December 2021).

Hughey, M.W. and Daniels, J. (2013) 'Racist comments at online news sites: a methodological dilemma for discourse analysis', *Media, Culture & Society*, 35(3), pp. 332–347. doi:10.1177/0163443712472089.

IBM (2012) *Deep Blue*, *Deep Blue*. IBM Corporation. Available at: http://www-03.ibm.com/ibm/history/ibm100/us/en/icons/deepblue/ (Accessed: 23 November 2021).

Irion, K. and Helberger, N. (2017) 'Smart TV and the online media sector: User privacy in view of changing market realities', *Telecommunications Policy*, 41(3), pp. 170–184. doi:10.1016/j.telpol.2016.12.013.

Jaton, F. (2021) 'Assessing biases, relaxing moralism: On ground-truthing practices in machine learning design and application', *Big Data & Society*, 8(1), p. 20539517211013570. doi:10.1177/20539517211013569.

Joergensen, A. (2021) 'Torben Sangild er blevet smidt af Facebook uden nogen forklaring', *Journalisten*, 25 January. Available at: https://journalisten.dk/torben-sangild-er-blevet-smidt-af-facebook-uden-nogen-forklaring/ (Accessed: 16 December 2021).

Johnson, K. (2021) 'This New Way to Train AI Could Curb Online Harassment | WIRED', *WIRED*. Available at: https://www.wired.com/story/new-way-train-ai-curb-online-harassment/?fbclid=IwAR1d8Yi-nVVAbxlhmbymJLY5kIexVS0rZLrXu9EGWWPSE4OC9Zn_X3mXnxg (Accessed: 16 December 2021).

Jones, B. and Jones, R. (2019) 'Public Service Chatbots: Automating Conversation with BBC News', *Digital Journalism*, 7(8), pp. 1032–1053. doi:10.1080/21670811.2019.1609371.

Jones, R. (2018) 'NatWest Bank tests Cora, an AI bot that will answer customer questions', *The Guardian*, 21 February. Available at: https://www.theguardian.com/money/2018/feb/21/natwest-bank-tests-cora-an-ai-bot-that-will-answer-customer-questions (Accessed: 13 February 2022).

Jones, R. and Jones, B. (2019) ‘Atomising the News: The (In)Flexibility of Structured Journalism’, *Digital Journalism*, 7(8), pp. 1157–1179. doi:10.1080/21670811.2019.1609372.

Karlsson, M. (2010) ‘Rituals of Transparency’, *Journalism Studies*, 11(4), pp. 535–545. doi:10.1080/14616701003638400.

Kayser-Bril, N. (2021) ‘AlgorithmWatch forced to shut down Instagram monitoring project after threats from Facebook’, *AlgorithmWatch*. Available at: https://algorithmwatch.org/en/instagram-research-shut-down-by-facebook/ (Accessed: 24 November 2021).

Knorr-Cetina, K.D. (1984) ‘The fabrication of facts: Toward a microsociology of scientific knowledge’.

Knott, A. *et al.* (2021) *Responsible AI for Social Media Governance: A proposed collaborative method for studying the effects of social media recommender systems on users*. GPAI. Available at: https://gpai.ai/projects/responsible-ai/social-media-governance/responsible-ai-for-social-media-governance.pdf.

Kristensen, L.M. (2021) ‘Audience Metrics: Operationalizing News Value for the Digital Newsroom’, *Journalism Practice*, pp. 1–18. doi:10.1080/17512786.2021.1954058.

Kunert, J. and Thurman, N. (2019) ‘The form of content personalisation at mainstream, transatlantic news outlets: 2010–2016’, *Journalism Practice*, 13(7), pp. 759–780.

Lambert, H. (2019) ‘The great university con: how the British degree lost its value’, *New Statesman*, 21 August. Available at: https://www.newstatesman.com/long-reads/2019/08/the-great-university-con-how-the-british-degree-lost-its-value (Accessed: 18 January 2022).

Latour, B. and Woolgar, S. (2013) *Laboratory life*. Princeton University Press.

Lee, A.M., Lewis, S.C. and Powers, M. (2014) ‘Audience Clicks and News Placement: A Study of Time-Lagged Influence in Online Journalism’, *Communication Research*, 41(4), pp. 505–530. doi:10.1177/0093650212467031.

Lee, E.-J. and Tandoc Jr, E.C. (2017) ‘When news meets the audience: How audience feedback online affects news production and consumption’, *Human communication research*, 43(4), pp. 436–449.

Lee, J. (2020) ‘AI Is Transforming Broadcasting’, *WIRED Korea*, 22 May. Available at: https://www.wired.kr/news/articleView.html?idxno=1238 (Accessed: 18 January 2022).

Lee, T.-T. and Wei, L. (2008) ‘How Newspaper Readership Affects Political Participation’, *Newspaper Research Journal*, 29. doi:10.1177/073953290802900302.

Leppänen, L., Tuulonen, H. and Sirén-Heikel, S. (2020) ‘Automated Journalism as a Source of and a Diagnostic Device for Bias in Reporting’. doi:10.17645/mac.v8i3.3022.

Leslie, D. (2020) *Understanding Bias in Facial Recognition Technologies*. UK: Alan Turing Institute. Available at: https://www.ssrn.com/abstract=3705658 (Accessed: 24 November 2021).

Lewis, S.C. (2011) 'The Knight News Challenge: How it works, what succeeds, and why that matters for the shaping of journalism innovation', in *International Symposium on Online Journalism, Austin, TX*.

Lewis, S.C. (2012) 'From Journalism to Information: The Transformation of the Knight Foundation and News Innovation', *Mass Communication and Society*, 15(3), pp. 309–334. doi:10.1080/15205436.2011.611607.

Lewis, S.C., Sanders, A.K. and Carmody, C. (2019) 'Libel by Algorithm? Automated Journalism and the Threat of Legal Liability', *Journalism & Mass Communication Quarterly*, 96(1), pp. 60–81. doi:10.1177/1077699018755983.

Lewis, S.C. and Usher, N. (2013) 'Open source and journalism: toward new frameworks for imagining news innovation', *Media, Culture & Society*, 35(5), pp. 602–619. doi:10.1177/0163443713485494.

Lewis, S.C. and Usher, N. (2014) 'Code, Collaboration, And The Future Of Journalism', *Digital Journalism*, 2(3), pp. 383–393. doi:10.1080/21670811.2014.895504.

Linden, C.-G. (2017) 'Decades of Automation in the Newsroom: Why are there still so many jobs in journalism?', *Digital Journalism*, 5(2), pp. 123–140. doi:10.1080/21670811.2016.1160791.

Lindskow, K. (2016) *Exploring digital news publishing business models: A production network approach*. Copenhagen Business School [Phd].

Littman, M.L. *et al.* (2021) *Gathering Strength, Gathering Storms: The One Hundred Year Study on Artificial Intelligence (AI100) 2021 Study Panel Report | One Hundred Year Study on Artificial Intelligence (AI100)*. Available at: https://ai100.stanford.edu/2021-report/gathering-strength-gathering-storms-one-hundred-year-study-artificial-intelligence (Accessed: 23 November 2021).

Llansó, E. *et al.* (2020) *Artificial intelligence, content moderation, and freedom of expression*. Annenberg Public Policy Center of the University of Pennsylvania.

Lokot, T. and Diakopoulos, N. (2016) 'News Bots: Automating news and information dissemination on Twitter', *Digital Journalism*, 4(6), pp. 682–699.

McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (2006). A proposal for the dartmouth summer research project on artificial intelligence, august 31, 1955. *AI magazine*, *27*(4), 12-12.

MediaSuite (2021) '"Who's speaking?"- Politicians and parties in the media during the Dutch election campaign 2021'. Available at: https://mediasuitedatastories.clariah.nl/elections-2021-first-results/ (Accessed: 13 February 2022).

Meta (2020) 'Partnering with Third-Party Fact-Checkers', *Facebook's Approach to Misinformation: Partnering with Third-Party Fact-Checkers*. Available at: https://www.facebook.com/journalismproject/programs/third-party-fact-checking/selecting-partners (Accessed: 13 February 2022).

Milosavljević, M. and Vobič, I. (2019) 'Human still in the loop: Editors reconsider the ideals of professional journalism through automation', *Digital journalism*, 7(8), pp. 1098–1116.

Møller Hartley, J. (2013) 'The online journalist between ideals and audiences: towards a (more) audience-driven and source-detached journalism?', *Journalism Practice*, 7(5), pp. 572–587.

Möller, J. *et al.* (2018) 'Do not blame it on the algorithm: an empirical assessment of multiple recommender systems and their impact on content diversity', *Information, Communication & Society*, 21(7), pp. 959–977. doi:10.1080/1369118X.2018.1444076.

Montal, T. and Reich, Z. (2017) 'I, Robot. You, Journalist. Who is the Author?', *Digital Journalism*, 5(7), pp. 829–849. doi:10.1080/21670811.2016.1209083.

Monti, M. (2019) *Automated Journalism and Freedom of Information: Ethical and Juridical Problems Related to AI in the Press Field*. SSRN Scholarly Paper ID 3318460. Rochester, NY: Social Science Research Network. Available at: https://papers.ssrn.com/abstract=3318460 (Accessed: 26 November 2021).

Monzer, C. *et al.* (2020) 'User Perspectives on the News Personalisation Process: Agency, Trust and Utility as Building Blocks', *Digital Journalism*, 8(9), pp. 1142–1162. doi:10.1080/21670811.2020.1773291.

Muniesa, F. (2011) 'Is a stock exchange a computer solution?: explicitness, algorithms and the Arizona stock exchange', *International Journal of Actor-Network Theory and Technological Innovation (IJANTTI)*, 3(1), pp. 1–15.

Newman, N. (2016) *Journalism, media and technology predictions 2016*. Reuters Institute for the Study of Journalism. Available at: https://ora.ox.ac.uk/objects/uuid:f15fac34-bafb-4883-898c-a53ade027e32.

Nielsen, R.K. and Schrøder, K.C. (2014) 'The relative importance of social media for accessing, finding, and engaging with news: An eight-country cross-media comparison', *Digital journalism*, 2(4), pp. 472–489.

Ó Fathaigh, R., Helberger, N. and Appelman, N. (2021) 'The perils of legally defining disinformation', *Internet Policy Review*, 10(4). doi:10.14763/2021.4.1584.

Oomen, J. *et al.* (2013) 'Symbiosis between the TRECVid benchmark and video libraries at the Netherlands Institute for Sound and Vision', *International Journal on Digital Libraries*, 13(2), pp. 91–104. doi:10.1007/s00799-012-0102-3.

Op Den Kamp, C. (2016) 'Too Good to be Forgotten. The Copyright Dichotomy and the Public-Sector Audiovisual Archive.', *International Association of Sound and Audiovisual Archives (IASA) Journal*, 46, pp. 33–41.

Oury, S. (2020) 'Automating the process of selecting thumbnails', *Medium*, 19 October. Available at: https://medium.com/@simonoury/automating-the-process-of-selecting-thumbnails-15e99702d468 (Accessed: 24 February 2022).

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

Park, D. *et al.* (2016) 'Supporting comment moderators in identifying high quality online news comments', in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pp. 1114–1125.

Pasquale, F. (2020) *New laws of robotics*. Harvard University Press.

Petre, C. (2021) *All the News That's Fit to Click*. US: Princeton University Press. Available at: https://press.princeton.edu/books/hardcover/9780691177649/all-the-news-thats-fit-to-click.

Plan International (2020) *Abuse and harassment driving girls off Facebook, Instagram and Twitter*, *Plan International*. Available at: https://plan-international.org/news/2020-10-05-abuse-and-harassment-driving-girls-facebook-instagram-and-twitter (Accessed: 16 December 2021).

Reese, S.D. and Shoemaker, P.J. (2016) 'A Media Sociology for the Networked Public Sphere: The Hierarchy of Influences Model', *Mass Communication and Society*, 19(4), pp. 389–410. doi:10.1080/15205436.2016.1174268.

Rehm, G. (2020) *The use of artificial intelligence in the audiovisual sector: concomitant expertise for INI report : research for CULT Committee*. LU: EU Publications. Available at: https://data.europa.eu/doi/10.2861/294829 (Accessed: 24 November 2021).

Richards, N.M. (2008) 'Intellectual Privacy', *Texas Law Review*, 87. Available at: https://papers.ssrn.com/abstract=1108268 (Accessed: 14 February 2022).

Ritzau Bureau (2020) '13 danske medier skaber metadata med kunstig intelligens | Ritzaus Bureau A/S'. Available at: https://via.ritzau.dk/pressemeddelelse/13-danske-medier-skaber-metadata-med-kunstig-intelligens?publisherId=10&releaseId=13586969 (Accessed: 18 January 2022).

Schifferes, S. *et al.* (2014) 'Identifying and verifying news through social media: Developing a user-centred tool for professional journalists', *Digital journalism*, 2(3), pp. 406–418.

Schiffrin, A. (2019) *Credibility and Trust in Journalism*, *Oxford Research Encyclopedia of Communication*. doi:10.1093/acrefore/9780190228613.013.794.

Schjøtt Hansen, A. and Hartley, J.M. (2021) ‘Designing What’s News: An Ethnography of a Personalization Algorithm and the Data-Driven (Re)Assembling of the News’, *Digital Journalism*, 0(0), pp. 1–19. doi:10.1080/21670811.2021.1988861.

Schrøder, K.C. (2011) ‘Audiences are Inherently Cross-Media: Audience Studies and the Cross-Media Challenge’, *CM Komunikacija i mediji*, 6(18), pp. 5–27.

Schultz, I. (2007) ‘Fra Partipresse over Omnibuspresse til Segmentpresse’, *Journalistica* [Preprint], (5). doi:10.7146/journalistica.v2i5.1807.

Searles, K., Spencer, S. and Duru, A. (2020) ‘Don’t read the comments: the effects of abusive comments on perceptions of women authors’ credibility’, *Information, Communication & Society*, 23(7), pp. 947–962. doi:10.1080/1369118X.2018.1534985.

Sharp, E. (2021) ‘Personal data stores: building and trialling trusted data services’, *BBC R&D*. Available at: https://www.bbc.co.uk/rd/blog/2021-09-personal-data-store-research (Accessed: 18 January 2022).

Smith, B. and Linden, G. (2017) ‘Two decades of recommender systems at Amazon. com’, *Ieee internet computing*, 21(3), pp. 12–18.

Social Science One (2019) ‘Public statement from the Co-Chairs and European Advisory Committee of Social Science One’. Available at: https://socialscience.one/blog/public-statement-european-advisory-committee-social-science-one (Accessed: 24 February 2022).

Sørensen, J.K. (2013) ‘PSB goes personal: The failure of personalised PSB web pages’, *MedieKultur: Journal of media and communication research*, 29(55), pp. 28-p.

Sørensen, J.K. (2020) ‘Personalised Universalism in the Age of Algorithms’, in Lowe, G., Savage, P., and Medina, M. (eds) *Universalism in Public Service Media*. Göteborg: Nordicom (Re-Visionary-Interpretations of the Public Enterprise (RIPE)), pp. 191–205.

Sørensen, J.K. and Hutchinson, J. (2018) ‘Algorithms and public service media’, in *Public Service Media in the Networked Society: RIPE@ 2017*. Nordicom, pp. 91–106.

Steele, B. (2022) ‘CIS Study Finds Gender Bias in Sports Journalism | Cornell Computing and Information Science’, *Cornell Bowers CIS*. Available at: https://cis.cornell.edu/cis-study-finds-gender-bias-sports-journalism (Accessed: 18 January 2022).

Sterzing, A., Oberholzer-Gee, F. and Melas, H. (2017) ‘Comment Moderation by Algorithm: The Management of Online Comments at the German Newspaper “Die Welt”’, in *Perspectives on Harmful Speech Online*, p. 42. Available at: http://nrs.harvard.edu/urn-3:HUL.InstRepos:33746096.

Stewart, A.J., McCarty, N. and Bryson, J.J. (2020) ‘Polarization under rising inequality and economic decline’, *Science advances*, 6(50), p. eabd4201.

Stray, J. (2016a) 'The age of the cyborg', *Columbia Journalism Review*. Available at: https://www.cjr.org/analysis/cyborg_virtual_reality_reuters_tracer.php (Accessed: 24 November 2021).

Stray, J. (2016b) 'What Do Journalists Do with Documents? Field Notes for Natural Language Processing Researchers', in *Computation þJournalism Symposium. Palo Alto, CA: Stanford University. https://journalism. stanford. edu/cj2016/files/What do journalists do with documents. pdf*.

Stray, J. (2019) 'Making Artificial Intelligence Work for Investigative Journalism', *Digital Journalism*, 7(8), pp. 1076–1097. doi:10.1080/21670811.2019.1630289.

Sunstein, C.R. (2004) 'Democracy and filtering', *Communications of the ACM*, 47(12), pp. 57–59.

Tandoc, E.C. (2014) 'Journalism is twerking? How web analytics is changing the process of gatekeeping', *New Media & Society*, 16(4), pp. 559–575. doi:10.1177/1461444814530541.

TechSavvy (2017) 'Algoritme styrer nyhederne hos svensk storavis • TechSavvy', *TechSavvy*, 28 March. Available at: https://techsavvy.media/algoritme-styrer-nyhederne-hos-svensk-storavis/ (Accessed: 18 January 2022).

Teegardin, C. *et al.* (2021) 'License to Betry', *The Atlanta Journal-Constitution*. Available at: http://doctors.ajc.com/part_1_license_to_betray/ (Accessed: 24 November 2021).

The Economist (2017) 'The world's most valuable resource is no longer oil, but data', *The Economist*, 6 May. Available at: https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data (Accessed: 13 February 2022).

Thurman, N. (2011) 'Making "The Daily Me": Technology, economics and habit in the mainstream assimilation of personalized news', *Journalism*, 12(4), pp. 395–415. doi:10.1177/1464884910388228.

Thurman, N. *et al.* (2016) 'Giving Computers a Nose for News: Exploring the limits of story detection and verification', *Digital Journalism*, 4(7), pp. 838–848. doi:10.1080/21670811.2016.1149436.

Thurman, N. (2018) 'Social media, surveillance, and news work: On the apps promising journalists a "crystal ball"', *Digital Journalism*, 6(1), pp. 76–97.

Thurman, N. *et al.* (2019) 'My Friends, Editors, Algorithms, and I: Examining audience attitudes to news selection', *Digital Journalism*, 7(4), pp. 447–469. doi:10.1080/21670811.2018.1493936.

Thurman, N., Dörr, K. and Kunert, J. (2017) 'When Reporters Get Hands-on with Robo-Writing', *Digital Journalism*, 5(10), pp. 1240–1259. doi:10.1080/21670811.2017.1289819.

Thurman, N. and Schifferes, S. (2012) 'THE FUTURE OF PERSONALIZATION AT NEWS WEBSITES: Lessons from a longitudinal study', *Journalism Studies*, 13(5–6), pp. 775–790. doi:10.1080/1461670X.2012.664341.

Tuchman, G. (1973) ‘Making News by Doing Work: Routinizing the Unexpected’, *American Journal of Sociology*, 79(1), pp. 110–131.

Turow, J. (2005) ‘Audience construction and culture production: Marketing surveillance in the digital age’, *The ANNALS of the American Academy of Political and Social Science*, 597(1), pp. 103–121.

Turow, J. (2016) *Media Today: Mass Communication in a Converging World*. 6th edn. New York: Routledge. doi:10.4324/9781315681726.

Van den Bulck, H. and Moe, H. (2018) ‘Public service media, universality and personalisation through algorithms: mapping strategies and exploring dilemmas’, *Media, Culture & Society*, 40(6), pp. 875–892. doi:10.1177/0163443717734407.

Van der Kaa, H. and Krahmer, E. (2014) ‘Journalist versus news consumer: The perceived credibility of machine written news’, in *Proceedings of the computation+ journalism conference, Columbia university, New York*, p. 25.

Vincent, J. (2018) ‘Watch Jordan Peele use AI to make Barack Obama deliver a PSA about fake news’, *The Verge*, 17 April. Available at: https://www.theverge.com/tldr/2018/4/17/17247334/ai-fake-news-video-barack-obama-jordan-peele-buzzfeed (Accessed: 23 November 2021).

Vu, H.T. (2014) ‘The online audience as gatekeeper: The influence of reader metrics on news editorial selection’, *Journalism*, 15(8), pp. 1094–1110. doi:10.1177/1464884913504259.

Wachter, S. (2020) ‘Affinity Profiling and Discrimination by Association in Online Behavioral Advertising’, *Berkeley Technology Law Journal*, 35, p. 367.

Waddell, T.F. (2018) ‘A Robot Wrote This?’, *Digital Journalism*, 6(2), pp. 236–255. doi:10.1080/21670811.2017.1384319.

Wang, S. (2018) ‘After years of testing, The Wall Street Journal has built a paywall that bends to the individual reader’, *Nieman Lab*. Available at: https://www.niemanlab.org/2018/02/after-years-of-testing-the-wall-street-journal-has-built-a-paywall-that-bends-to-the-individual-reader/ (Accessed: 18 January 2022).

Wang, S. (2021) ‘Moderating uncivil user comments by humans or machines? The effects of moderation agent on perceptions of bias and credibility in news content’, *Digital journalism*, 9(1), pp. 64–83.

Webb, A. (2019) *The Big Nine: How the Tech Titans and Their Thinking Machines Could Warp Humanity*. PublicAffairs.

Welbers, K. *et al.* (2016) ‘News selection criteria in the digital age: Professional norms versus online audience metrics’, *Journalism*, 17(8), pp. 1037–1053. doi:10.1177/1464884915595474.

Wells, G., Hortwitz, J. and Seetharaman, D. (2021) ‘Facebook Knows Instagram Is Toxic for Teen Girls, Company Documents Show’, *Wall Street Journal*, 14 September. Available at: https://www.wsj.com/articles/facebook-knows-instagram-is-toxic-for-teen-girls-company-documents-show-11631620739 (Accessed: 23 November 2021).

Westlund, O. (2014) ‘The production and consumption of news in an age of mobile media’, *The Routledge companion to mobile media*, 135145.

Whittaker, M. *et al.* (2018) *AI now report 2018*. AI Now Institute at New York University New York. Available at: https://ainowinstitute.org/AI_Now_2018_Report.pdf.

Wieringa, M. (2020) ‘What to account for when accounting for algorithms: a systematic literature review on algorithmic accountability’, in *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. New York, NY, USA: Association for Computing Machinery (FAT* ’20), pp. 1–18. doi:10.1145/3351095.3372833.

Willig, I. (2010) ‘Constructing the audience: a study of segmentation in the Danish press’, *Northern Lights: Film & Media Studies Yearbook*, 8(1), pp. 93–114. doi:10.1386/nl.8.93_1.

Willig, I. (2011) ‘The Journalistic Gut Feeling: Journalistic Doxa, News Habitus and Orthodox News Values’, in Berkowitz, D.A. (ed.) *Cultural meanings of news*. Thousand Oaks: SAGE Publications, pp. 83–98.

Willig, I. (2021) ‘From audiences to data points: The role of media agencies in the platformization of the news media industry’, *Media, Culture & Society*, p. 01634437211029861. doi:10.1177/01634437211029861.

Yeo, S.K. *et al.* (2019) ‘The effect of comment moderation on perceived bias in science news’, *Information, Communication & Society*, 22(1), pp. 129–146.

Young, M.L. and Hermida, A. (2015) ‘From Mr. and Mrs. Outlier To Central Tendencies’, *Digital Journalism*, 3(3), pp. 381–397. doi:10.1080/21670811.2014.976409.

Zamith, R. and Lewis, S.C. (2014) ‘From Public Spaces to Public Sphere: Rethinking systems for reader comments on online news sites’, *Digital Journalism*, 2(4), pp. 558–574. doi:10.1080/21670811.2014.882066.

Ziewitz, M. (2016) ‘Governing Algorithms: Myth, Mess, and Methods’, *Science, Technology, & Human Values*, 41(1), pp. 3–16. doi:10.1177/0162243915608948.

Zuiderveen Borgesius, F. (2015) ‘Informed Consent: We Can Do Better to Defend Privacy’, *IEEE Security Privacy*, 13(2), pp. 103–107. doi:10.1109/MSP.2015.34.

Zuiderveen Borgesius, F. and Poort, J. (2017) ‘Online Price Discrimination and EU Data Privacy Law’, *Journal of Consumer Policy*, 40(3), pp. 347–366. doi:10.1007/s10603-017-9354-z.

Zuiderveen Borgesius, F.J. (2014) *Improving privacy protection in the area of behavioural targeting*. PhD Thesis. University of Amsterdam. Available at:

https://dare.uva.nl/search?identifier=c74bdba6-616c-4cd9-925e-33a5858935e5 (Accessed: 13 February 2022).

Zuiderveen Borgesius, F.J. *et al.* (2016) 'Should we worry about filter bubbles?', *Internet Policy Review*, 5(1). doi:10.14763/2016.1.401.

AI4media

ARTIFICIAL INTELLIGENCE FOR THE MEDIA AND SOCIETY

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

info@ai4media.eu www.ai4media.eu