Ref. Ares(2023)8875168 - 26/12/2023

AI4media
ARTIFICIAL INTELLIGENCE FOR THE MEDIA AND SOCIETY

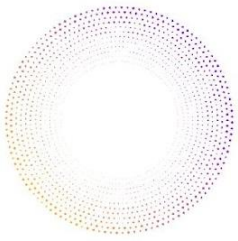
D2.5

Final 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	Final white paper on the social, economic, and political impact of media AI technologies
Deliverable number	D2.5
Deliverable version	1.0
Previous version(s)	-
Contractual date of delivery	31 December 2023
Actual date of delivery	26 December 2023
Deliverable filename	AI4Media_D2.5_Final white paper on the social, economic, and political impact of media AI technologies_final
Nature of deliverable	Report
Dissemination level	Public
Number of pages	62
Work Package	WP2
Task(s)	T2.4
Partner responsible	UvA
Author(s)	Anna Schjøtt Hansen (UvA), Noémie Krack (KUL), Lidia Dutkiewicz (KUL), Rasa Bočytė (NISV), Johan Oomen (NISV), Natali Helberger (UvA), Georg Thallinger (JR), Aleksandra Kuczerawy (KUL), Emine Ozge Yildirim (KUL)
Editor	Anna Schjøtt Hansen (UvA)
EC Project Officer	Evangelia Markidou

Abstract	The '*Final white paper on the social, economic, and political impact of media AI technologies*' builds on the literature based insights outlined in the 'Initial white paper on the social, economic, and political impact of media AI technologies' (D2.2) published in February 2022. It extends the initial insights by (1) outlining the aims and functions of the newly launched AI Media Observatory, (2) describing the findings from three industry workshops that each explored one of the challenges that was identified in the initial version of the whitepaper, and (3) discussing how policy and legislation could better support the responsible use and development of AI in media.

Keywords	Artificial Intelligence, Media, Societal impact, Observatory, Content moderation, Recommender systems, Audiovisual archives, Policy recommendations

www.ai4media.eu

info@ai4media.eu

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

Copyright

© Copyright 2023 AI4Media Consortium

This document may not be copied, reproduced, or modified in whole or in part for any purpose without written permission from the AI4Media Consortium. In addition to such written permission to copy, reproduce, or modify this document in whole or part, an acknowledgement of the authors of the document and all applicable portions of the copyright notice must be clearly referenced.

All rights reserved.

Contributors

NAME	ORGANISATION
Anna Schjøtt Hansen	UvA
Noémie Krack	KUL
Lidia Dutkiewicz	KUL
Rasa Bočytė	NISV
Johan Oomen	NISV
Natali Helberger	UvA
Georg Thallinger	JR
Aleksandra Kuczerawy	KUL
Emine Ozge Yildirim	KUL

Peer Reviews

NAME	ORGANISATION
Danae Tsabouraki	ATC
Patrick Aichroth	FhG
Filareti Tsalakanidou	CERTH

Revision History

VERSION	DATE	REVIEWER	MODIFICATIONS
0.1	2022/2023	Anna Schjøtt Hansen, Noémie Krack, Lidia Dutkiewicz, Rasa Bočytė, Johan Oomen, Natali Helberger, Georg Thallinger, Aleksandra Kuczerawy, Emine Ozge Yildirim	Production of short reports and dissemination materials based on workshops
0.2	06/07/2023	Anna Schjøtt Hansen, Tobias Blanke and Natali Helberger	Producing outline for the final whitepaper
0.3	25/10/2023	Anna Schjøtt Hansen	Production of draft sections regarding the workshops that had not been produced earlier

0.4	11/11/2023	Natali Helberger, Lidia Dutkiewicz, Noemie Krack	Feedback and revisions on short reports draft sections regarding workshops
0.5	13/11/2023	WP2 members	Reading of full draft by WP2 members. Edits, including adding more visual representations (e.g., screenshots of the short reports) and clarifying the text and adding to parts of the sections. Furthermore, the whitepaper was made more consistent in terminology
0.6	01/12/2023	Danae Tsabouraki Patrick Aichroth Filareti Tsalakanidou	Reading of the final draft. Edits include improving language, adding additional sections regarding the Observatory to provide transparency about its operation and adding an additional section on the policy workshops
0.7	22/12/2023	Anna Schjøtt Hansen	Final version after implementing internal review comments
1.0	26/12/2023	Filareti Tsalakanidou	Final version ready for submission

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

Table of Abbreviations and Acronyms

Abbreviation	Meaning
AI	Artificial Intelligence
API	Application Programming Interface
DPPA	United Nations Department of Political and Peacebuilding Affairs
EC	European Commission
EU	European Union
FIAT/FTA	Fédération Internationale des Archives de Télévision / The International Federation of Television Archives
GDPR	General Data Protection Regulation
LLM	Large Language Model
MAM	Media Asset Management
ML	Machine Learning

Index of Contents

1 Executive Summary 11

2 Introduction 16

2.1 Methodology 16

2.1.1 Industry workshops: Qualifying challenges and research gaps 17

2.1.2 Policy workshops: Qualifying the pilot policies for the use of AI in the media sector 18

2.2 Workshop dissemination 20

3 AI Media Observatory – a knowledge platform 22

3.1 Introducing the AI Media Observatory 23

3.1.1 Background on the implementation 25

3.1.2 Day-to-day work of the Observatory 26

3.1.3 Content submission 27

3.1.4 Expert directory 29

3.1.5 Dissemination and Reach of the Observatory 29

3.1.6 Long-term sustainability of the Observatory 31

4 Qualifying concrete challenges in media AI 32

4.1 Workshop 1: Measuring the success of recommender systems 32

4.1.1 Key insights: Three core questions 32

4.1.2 Good practices 35

4.1.3 Towards adaptable and process-oriented assessment frameworks 36

4.1.4 Policy recommendations 37

4.1.5 Results in brief 38

4.2 Workshop 2: Identifying common challenges regarding the use of AI in content moderation 38

4.2.1 Key insights: Six core themes of AI-assisted moderation 39

4.2.2 Core challenges for using AI in content moderation 40

4.2.3 Potential ways forward 42

4.2.4 Results in brief 43

4.3 Workshop 3: Current challenges and future paths for AI in audiovisual archives 44

4.3.1 Key Insights: AI readiness and strategies 45

4.3.2 Potentials and barriers 46

4.3.3 Five recommendation areas 47

4.3.4 Journey map: Building AI in-house 50
4.3.5 Journey map: Buying AI from a third party vendor 51
4.3.6 Results in brief 52
5 Policy workshops 53
5.1.1 Exploring the tensions of AI regulation 53
5.1.2 Exploring the Wishes for AI Regulation 55
5.1.3 Results in brief 55
6 Policy discussion 57
6.1 Six cross-cutting policy needs 57
6.1.1 Impose and support good practices for transparency 57
6.1.2 Support research in and of AI solutions 58
6.1.3 Stimulating responsible development of AI 58
6.1.4 Mitigating AI divides 59
6.1.5 Mitigating power imbalances 59
6.1.6 Global and societally focused policies 60
6.2 Future impact of recommendations 60
7 Conclusions 61

Index of Tables

Table 1: Overview of users and pageviews in the Observatory. 30
Table 2: Case study description - Recommenders for news. 35
Table 3: Potentials and challenges of AI in audiovisual archives. 47

Index of Figures

Figure 1: Overview of methodology. 17
Figure 2: Screenshots of the short report from the first policy workshop. 20
Figure 3: Screenshots of the flyer disseminating the work of the consortium. 21
Figure 4: Visualisation of AI Media Observatory used for dissemination. 23
Figure 5: Screenshot of Your AI Media Feed. 23
Figure 6: Screenshot of Let's Talk AI and Media. 24
Figure 7: Screenshot of Find an AI Media Expert. 24
Figure 8: Screenshot of the brochure. 26
Figure 9: Screenshot of content submission form. 27
Figure 10: Visualisation of the content submission workflow. 28
Figure 11: Screenshot of expert sign-up form. 29
Figure 12: Key statistics for the AI Media Observatory (October 2023). 30
Figure 13: Screenshot of the short report outlining the insights from the workshop on recommender systems. 38
Figure 14: Screenshot of the short report outlining the insights from the workshop on AI-assisted content moderation. 44
Figure 15: Photos from the workshop on AI in audiovisual archives. 45
Figure 16: Screenshot of the short report outlining the insights from the workshop on AI for audiovisual archives. 52
Figure 17: Photos from the second policy workshop in Pisa. 53
Figure 18: Screenshots of the short reports outlining the insights from the policy workshops. 56

1 Executive Summary

The 'Final white paper on the social, economic, and political impact of media AI technologies' builds on the literature-based insights outlined in the 'Initial white paper on the social, economic, and political impact of media AI technologies' (D2.2) published in February 2022 and disseminated in a short Factsheet. It extends the initial insights by (1) outlining the aims and functions of the newly launched AI Media Observatory as a 'living document' and knowledge platform that maps the social economic and political impact of media AI technologies, (2) describing the findings from three industry workshops that each explored one of the challenges that was identified in the initial version of the whitepaper, and (3) discussing how policy and legislation could better support the responsible use and development of AI in media based on the industry workshops and a series of policy workshops that were aimed at qualifying the 'Pilot Policy Recommendations for the use of AI in the Media Sector' (D2.4), published in August 2022.

(1) In short, the AI Media Observatory aims to support the ongoing efforts of the multidisciplinary community of professionals who are working towards ensuring the responsible use of AI in the media sector. As reports will always present a 'snapshot' of the current discussions and research, the Observatory enables ongoing monitoring, curation and interpretation of the social, economic and political impacts of AI in media, which is highly important due to the rapid pace of AI development and implementation across societies.

The Observatory aims to fulfil this goal by curating relevant content that provides expert perspectives on the potentials and challenges that AI poses for the media sector through its sections 'Your AI Media Feed' for written content and 'Let's Talk AI and Media' for audiovisual content. It also provides an easy overview of relevant experts in the field through the 'Find your AI Media Expert' directory, as an important outreach component.

(2) Three industry workshops were conducted in 2022 and 2023, which contributed with unique insights into (i) how to measure the success of recommender systems for media, (ii) what the common challenges relating to the use of AI in content moderation are, and (iii) the current challenges and future paths of using AI in audiovisual archives. The findings from these workshops have both been disseminated as short individual reports and are presented in the report highlighting both potential challenges, ways forward and concrete policy recommendations.

The first industry workshop provided insights into the need for the following conditions to ensure more responsible measurements of success in recommender systems:

- **Improving knowledge and negotiating power when working with third-party providers:** This could include the ability to ask for optimizations beyond click-rate through accuracy to alleviate the current market gap in which only certain media organisations have the agency to adjust their recommenders.
- **Better benchmarking practices:** This could include the ability to benchmark against other media systems and with metrics beyond accuracy, which could help better illuminate the public value of recommender systems.

- **Better industry-academia collaborations:** This could include collaborations on technical solutions but also to help develop value frameworks, which could help reduce the costs of the projects and support more responsible development.

The second industry workshop provided insights into three potential ways forward that could support the use of AI in content moderation:

- **Local and open-source solutions:** One way forward focused on providing better conditions for making local and open-source AI solutions that could complement the AI-assisted moderation systems on, for example, large platforms to minimise the moderation gap that currently exists between English and non-English speaking contexts. These conditions could be provided via either targeted funding or policy but also would require new forms of collaboration between third-party providers and large platforms, who are both important sources of data and where such solutions could have a high impact. Such local solutions could also better consider the specificities of the cultural context in which the moderation would take place, thereby, also alleviating the dislocation of moderation.
- **Transfer learning across regional languages:** To enable a way forward with such solutions, the participants looked towards transfer learning and the ability to provide regional models rather than purely local models, as this would require less training data to retrain models in similar languages. Such experimentation had already been carried out by two participants with high accuracy rates, where only 10-20% labelled data was needed to retrain the models, which would place much less stress on small or minority languages to be fully responsible for building their own systems and training them on complete datasets.
- **Publicly available datasets:** The last way forward, which also would be incremental for any of the solutions to be attainable, was to create publicly available datasets, as this remained the core barrier for most of the projects described during the workshop. While a few examples of datasets relating to sexism have been produced, they remain English-focused, and the quality of available data varies significantly across different geographic contexts. High-quality datasets would better enable the foundational training of regional models, which could then be retrained on smaller amounts of locally labelled and contextual data.

The third industry workshop provided insights into five recommendation areas for the responsible use of AI in audiovisual archives:

- **Leapfrogging via digitisation:** Many organisations are still in the process of digitising their vast audiovisual collections. This process was often viewed as a challenge and an economic burden, but in the context of AI, it could also be seen as a unique opportunity for organisations to leapfrog into AI. The digitisation efforts across archives in past years were predominantly focused on digitisation for preservation purposes; the future digitisation processes can learn from these past efforts but also have a unique opportunity to focus on digitisation for access by combining the digitisation processes with AI projects. This would entail including the needs of end-users in the design of

digitisation projects. This can help to define AI techniques that will make digitised data accessible and searchable in meaningful ways for the target audiences of the organisations.

- **The strength of local & regional collaboration:** Many of the consulted organisations described experiencing the same challenges with implementing AI, relating to language, costs and human resources. This could provide unique opportunities in the AI context as many of the countries share linguistic and cultural similarities and collaborative approaches would make AI more accessible and scalable across multiple archives. Current examples of the benefits of such collaborations can be seen in the newly established network for Nordic AI Journalism or the Associated Press' efforts to help facilitate scalable projects amongst local newsrooms. What became apparent during the workshop was the willingness to enter such regional collaborations. Next to regional collaborations, there are also unexplored opportunities to build partnerships with local actors active in the field of AI - such as universities and start-ups - who can provide resources and skills that archives might not have in-house.
- **The unique challenge and opportunity of MAMs:** The centrality of Media Asset Management (MAM) systems in the archival workflow provides both a unique challenge and an opportunity for using AI. On one hand, it can constrain the use of open-source and third-party solutions as these need to integrate with the MAM systems which are often not flexible enough to support this. On the other hand, MAM system providers could play a valuable role in making AI applications more easily accessible for archives via easy plug-in solutions. Closer interaction with MAM systems providers (for instance, via participation in networks such as FIAT/FTAs Media Management Commission) can help in bridging the gap between the specific needs of archives in the region and the available solutions.
- **AI & organisational strategy:** It has become clear that organisational strategies often do not provide guidance and specific considerations for the use of AI. This could relate to procurement guidelines, accountability mechanisms and quality control procedures. Given the societal harms that the neglectful application of AI can produce (for instance, perpetuating the marginalisation of societal groups), a clear organisational strategy on AI is key. AI policies in audiovisual archives should be guided and aligned with their public mission but should not restrict experimentation with AI.
- **AI & capacity building:** Implementation of AI puts additional pressure on resources and expertise that not all organisations currently have. As a starting point, archives should ensure that staff in different departments are equipped to constructively participate in conversations about the use of AI, especially in relation to potential risks, procurements and quality assurance processes, and adherence to organisational policy and legal requirements. Next to this, new positions - such as AI engineers and data scientists - could be introduced to allow archives to experiment with existing AI tools or independently develop their own solutions without reliance on vendors. Since not all organisations have the resources to do this, regional and sectoral partnerships can be used to pool resources and benefit from expertise in local universities, start-ups or other archives. Such collaborations can be particularly important for developing localised and context-specific AI solutions.

(3) Three policy workshops were conducted in 2023, which invited relevant stakeholders to 'provotype policies'[1] as a way to qualify the 'Pilot Policy Recommendations for the use of AI in the Media Sector'. These workshops provided cross-stakeholder insights into which legal conditions are needed for the responsible use and development of AI for media. The findings from these workshops will be disseminated in short reports. The findings from the first and second workshops have already been published. The last will be published in early 2024.

The findings from the policy workshops together with the insights from the industry workshops are also used to provide a final policy discussion, which highlights six cross-cutting policy needs, which were deemed to be the most urgent and important based on their impact on the media industry and how strongly and often they were brought up by the participating stakeholders. These, in short, included:

- The need for **policies that impose and support good transparency practices**; including the need to support both **(1)** *internal transparency* to address the intelligibility gap[2] in organisations where the end-users of the systems (e.g., journalists) do not have enough knowledge and confidence to challenge the workings of the systems, **(2)** require *external transparency* for user-facing AI applications and **(3)** demand increased *third party transparency* to enable buyers of AI solutions to make an informed decision.
- The need for **policies that support research in and of AI solutions**; including the need to *give access to data sets and APIs* for research and investigative purposes, support the production of *open and shareable datasets* and enable *long-term and sustainable funding schemes* that support the development of AI solutions.
- The need for **policies that stimulate the responsible development of AI**; including the need to support the *upskilling of employees and providing guidelines on responsible **practices***, the need to enable *sustainable funding schemes* that support societal interests and have a long-term focus and the need to stimulate *public-private collaborations* to enable solutions being developed with the support of researchers.
- The need for **policies that address and mitigate the growing AI divide**; including the need for supporting *organisational capacity building* particularly for small organisations and those situated in geographical regions with historical disadvantages, the need for supporting increased *local and regional collaborations* to limit the costs of developing AI and the need to support the *development of open datasets and AI solutions for minority languages* to counteract the fact that most AI solutions are developed for and work best in English
- The need for **policies that aim to address the current power imbalances in the AI landscape**; including the need for supporting *alternative funding schemes* that allow organisations developing AI to be independent of funding from big tech companies, the need for supporting the *development of public and open-source critical AI infrastructures and models* (e.g., LLMs) to ensure a diverse landscape of AI

[1] Boer, L., & Donovan, J. (2012). Provotypes for participatory innovation. In *Proceedings of the designing interactive systems conference* (pp. 388-397).

[2] https://www.tandfonline.com/doi/full/10.1080/21670811.2022.2145328

developments and the need to enable organisations to have more *negotiation and bargaining power* with third-party providers.

- The need for **policies that are more globally and societally focused**; including the need to support *a global and collective approach to AI governance* and the need to focus on *societal risks*, as opposed to individual risks and rights to ensure that long-term harms to society are always considered and accounted for when building AI solutions.

2 Introduction

This whitepaper builds on the initial insights outlined in the 'Initial white paper on the social, economic, and political impact of media AI technologies' (D2.2) published in February 2022, which has also been disseminated in a short fact sheet format focusing on the 'Key societal concerns of AI applications in Media'.

Whereas the goal of the initial whitepaper was to map these ongoing discussions as they unfold in academic literature and industry reports and identify knowledge gaps and state-of-the-art, this extended whitepaper presents the newly launched AI Media Observatory as a portal to provide ongoing insights into the societal impacts of AI. Beyond this, the whitepaper extends and qualifies the previous version by providing unique insights into specific challenges that were identified in the first version.

The 'Final white paper on the social, economic, and political impact of media AI technologies' first introduces the newly launched **AI Media Observatory**, which will serve as a 'living document' that maps the social economic and political impact of media AI technologies.

Then the whitepaper describes the findings from three **industry workshops** that were conducted in 2022 and 2023. The topics for the workshops were chosen based on the many challenges identified and described in the previous whitepaper, and here the emphasis was on choosing three challenges where very limited literature existed to guide good practices. The workshops explored the following three topics: (1) Measuring the success of recommender systems for media, (2) Identifying common challenges relating to the use of AI in content moderation, and (3) The current challenges and future paths of AI in audiovisual archives.

The whitepaper then proceeds to discuss how policy and legislation could better support the responsible use and development of AI in media based on findings from the three **industry workshops** together with the findings from a series of three **policy workshops** that invited different relevant stakeholder groups to 'provotype policies' as a way to provide new insights into and qualify the 'Pilot Policy Recommendations for the use of AI in the Media Sector'(D2.4), published in 2022 by the consortium, which will be updated in 2024 (D2.6).

2.1 Methodology

The main components of this deliverable consist of (1) the AI Media Observatory, (2) three industry workshops, and (3) three policy workshops, where the latter two were the main research methodologies supporting the findings of the whitepaper. In section 3.1.1, we introduce the background and process of developing the Observatory as well as its final functionalities. Below, we outline the methodologies behind the two series of workshops, respectively the industry and policy workshops.

As the initial whitepaper was based on academic literature to provide an overview of the state-of-the-art regarding AI in the media industry, it was important in the extended version to qualify those findings with researchers and industry experts to further qualify where the concrete

challenges and risks for using AI lie, but also to enable discussions of how to move forward in responsible ways. This was why the overarching methodology of workshops was chosen, as this provided us with a way to convene relevant stakeholders physically or online to discuss in-depth the specific challenge identified in the initial version of the whitepaper. Different workshop exercises were designed to help facilitate the discussion and to help provide insights into where the main 'pain points' were for them and what forms of support were needed to enable them to work responsibly with AI.

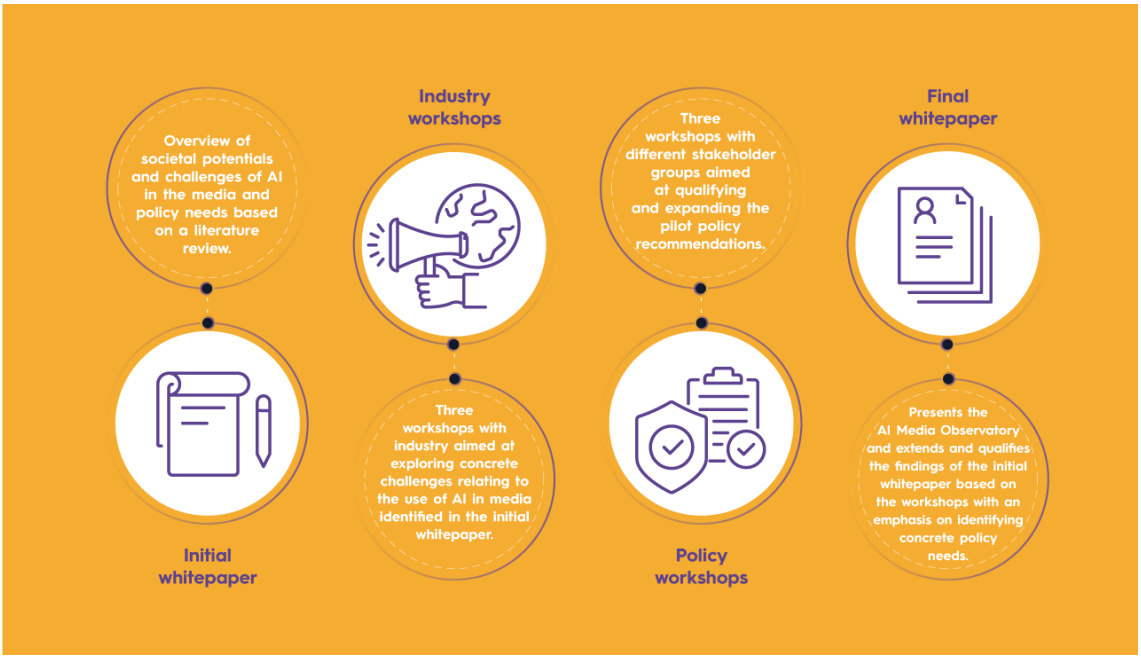


Figure 1: Overview of methodology.

2.1.1 Industry workshops: Qualifying challenges and research gaps

Between 2022 and 2023 three different industry workshops were conducted, under the topics of (1) Measuring the success of recommender systems for media, (2) Identifying common challenges relating to the use of AI in content moderation, and (3) The current challenges and future paths of AI in audiovisual archives.

Two of the workshops were conducted as online workshops to ensure participants from all over Europe could participate and one was conducted in person as it had a regional focus, which enabled us to convene the relevant stakeholders in one location.

The methodology for each workshop was developed in close collaboration with partners of the consortium with expertise within the concrete challenge to ensure a relevant setup for each workshop. This included drawing particularly on the expertise of KUL and NISV, as well as expertise within WP2. Under the findings from the workshops in the following sections, the concrete methodology of each is described more in detail. This section provides a short overview.

The two online workshops focusing on recommender systems and content moderation were both two hours long and were invite-only, as the focus was on getting the right people in the room, which included media organisations, platforms, and AI developers. They also shared a similar setup, where the workshop started with an inspirational talk by a key expert on the topic, followed by a round table with the participants before proceeding with a discussion of potential good practices.

The physical workshop focusing on AI in audiovisual archives took place over two days at the headquarters of HRT (Croatian Radiotelevision) in Zagreb, Croatia and included participants from seven organisations from the Balkan and Mediterranean regions. The decision to conduct an extended, in-person workshop in this region stemmed from a notable gap highlighted in the initial whitepaper. This gap revealed two key issues: firstly, there is a general lack of knowledge about AI in audiovisual archives, and secondly, there exists a substantial geographical disparity in the current literature on AI in the media sector. The majority of this literature has predominantly focused on the United States and Northern Europe, overlooking other regions. Conducting the workshop with participants from the Balkan and Mediterranean regions was a way to surface knowledge on the concrete challenges faced in currently understudied areas – both in terms of topic and geography.

2.1.2 Policy workshops: Qualifying the pilot policies for the use of AI in the media sector

In 2023, three policy workshops were conducted with three different stakeholder groups. The aim of these workshops was to disseminate and qualify the 'Pilot Policy Recommendations for the use of AI in the Media Sector' (D2.4), which were initially published in August 2022.

The pilot policy recommendations represented the attempt to provide policy guidance on how to better regulate AI in the context of the media sector and are planned to be updated and delivered in their final version in August 2024 (D2.6). The pilot policy recommendations were based on previous deliverables in the consortium, including the initial whitepaper, which also identified specific policy recommendations, as well as surveys and public stakeholders' consultation. As a result, the insights from the workshops are used here to re-engage with the policy recommendations raised in the initial whitepaper and provide an updated list of core policy needs for the media sector.

With this aim in mind, the three policy workshops were designed to engage the participants in active discussion over the legal challenges that were highlighted in the initial pilot recommendations and to offer them the chance to provide input on what policies they found to be missing from their vantage point in the sector.

Oppositely to the industry workshops above, these policy workshops all followed the same format and instead focused on conducting the workshop with different stakeholders. The format was designed based on the structure of the pilot policy recommendations, which highlighted both challenges and recommendations for three areas, namely (1) Media companies, (2) Academia and researchers and (3) Legal and Societal.

The first workshop was conducted during the Joint Computation + Journalism European Data & Computational Journalism Conference 2023, in Zurich, Switzerland in June 2023, where the participants were representatives from media organisations or media researchers.

The second workshop was conducted during the 9th General Assembly of the AI4Media consortium, in Pisa, Italy, in October 2023, where the participants consisted of consortium partners and, as a result, included many technical experts as well as use case partners.

The third workshop was conducted with the AI, Media, and Democracy Lab at the University of Amsterdam, Netherlands in November 2023, where the participants included media and legal scholars working in the intersection of AI and media.

The workshops were divided into three parts that are outlined below:

1. First, the participants were given a presentation of the 'Pilot Policy Recommendations for the use of AI in the Media Sector' (D2.4).
2. Second, they engaged in the first exercise titled 'Provotyping Policies'. This exercise was inspired by the use of provotypes within participatory design, which are understood as: "'types' that embody tensions surrounding an area of interest, to support collaborative analysis of that area and to collaboratively explore design possibilities"[3]. The participants were asked to move to three different corners of the room, which represented the media landscape, the research landscape and the regulatory and policy landscape. In each corner, they were presented with three provocative statements that were aimed at highlighting specific tensions in the current AI and media landscape. The participants would spend ten minutes in each corner and were asked to read, react and discuss the statements and provide post-its, where they would highlight their agreement or disagreement with the statements.
3. Third, they were introduced to a last exercise 'The Wishing Well', where they were asked to first write down three wishes for policies that they felt were highly needed. These could be based on previous discussions or personal experience within their field. Once everyone had written three wishes, they were divided into smaller groups and asked to group the wishes into themes. As the last part of the exercise, they were asked as a group to rank each of the overarching themes from most important to least important and finally present this to the other groups.

After each of the workshops, a short report was put together summarising the inputs from the specific workshop to highlight the unique insights from that stakeholder group (Figure 2). These were published under the Results in Brief section on the AI4Media site and disseminated via the AI Media Observatory. These short reports contain a ranked regulation wishlist, which illustrates which policies were considered most vital by the participants as well as an overview of the current tensions regarding AI's role in the media sector and the role of regulation. The short report from the first and second workshops has already been published, while the last will be published in the coming months. In this whitepaper, the results from the three policy workshops

[3] Boer, L., & Donovan, J. (2012). Provotypes for participatory innovation. In *Proceedings of the designing interactive systems conference* (pp. 388-397).

are brought in conversation with each other to distil cross-workshop findings and point to the core policy needs of the sector.

Figure 2: Screenshots of the short report from the first policy workshop.

2.2 Workshop dissemination

The workshops were seen as important avenues to disseminate the work already conducted by the consortium in WP2. To this end, a flyer was developed (see Figure 3), which could be shared with participants in the workshops. Beyond the dissemination of the work in WP2 and AI4Media more broadly, the workshops also facilitated further invitations to speak about the work within AI4Media.

After the second policy workshop (organised at the 9th General Assembly), the Dutch-based research project 'Designing Responsible AI for Media Applications' (DRAMA) reached out and asked if a smaller version of the workshop could be conducted at their closing event on November 21st, 2023. Here, Anna Schjøtt Hansen (UvA) and Noémie Krack (KUL) participated and conducted a half-hour presentation and short exercise that was based on the initial workshop setup. The participants were, furthermore, encouraged to provide input to future policy recommendations via a collaborative document. The participation was also disseminated on social media, like X and LinkedIn.

The European Broadcasting Union also reached out to Noémie Krack (KUL) to learn more about AI media applications challenges and the AI4Media project. On November 20th, she participated in the European Social Dialogue in the audio-visual sector, a meeting jointly coordinated by EBU (European Broadcasting Union) and EURO-MEI (UNI-Europa Media, Entertainment & Arts). The social dialogue is the opportunity for workers and employer representatives to meet and discuss the main issues in the media sector. Noémie addressed the legal and ethical challenges associated with the use of generative AI in the media sector. The event was also the opportunity to share awareness about AI4Media's realm of action and gather input for the final policy

recommendations from the exchanges with the participants. They were, furthermore, also encouraged to provide input to future policy recommendations via a collaborative document. The employees' representatives were particularly concerned about the impact of technology on employment and workers' conditions in the media sector. They are also concerned about the impact of AI on culture, art and human creativity. Copyright considerations were also a key discussion item with the numerous uncertainties with the generative AI developments in the sector such as the body scan of actors, generated voice dubbing, and new music generated based on artist's sounds or performances. The idea of power dynamics and humans at the centre of technology development were also discussed.

Figure 3: Screenshots of the flyer disseminating the work of the consortium.

3 AI Media Observatory – a knowledge platform

As noted in the initial whitepaper, a report will forever remain a snapshot of the ongoing discussions of the social, economic, and political implications of AI in the media sector. This issue was highly apparent with the release of ChatGPT by OpenAI in November 2022, which illustrated the pace at which changes and new implications can emerge in the intersection of AI and media.

The high quality and easy accessibility of the application led to many new uses emerging rapidly, which in some cases have become highly problematic such as uses in school or college exams. In the media sector, this also led to many new policies being drafted, to keep up with the development[4].

The newly launched AI Media Observatory represents the solution to this issue as it provides a 'living document' that can ongoingly capture and provide insights into the unfolding discussions, implications, and ways forward in the AI media field.

The Observatory is the culmination of three years of work in the AI4Media consortium and was initially launched in a 'beta-version' earlier this year in April and officially launched in October 2023 together with several outlined workflows that will enable the continued management of the Observatory. These will be described more in detail below.

The overarching goal of the Observatory is to support the ongoing efforts of the multidisciplinary community of professionals who are working towards ensuring the responsible use of AI in the media sector. It fulfils this goal in different ways, namely (1) by curating and providing access to relevant written and audiovisual content that explores the impact of AI in the media sector, and (2) by providing an overview of relevant experts in the field, which can be found and contacted through the expert directory.

[4] See e.g., https://generative-ai-newsroom.com/towards-guidelines-for-guidelines-on-the-use-of-generative-ai-in-newsrooms-55b0c2c1d960

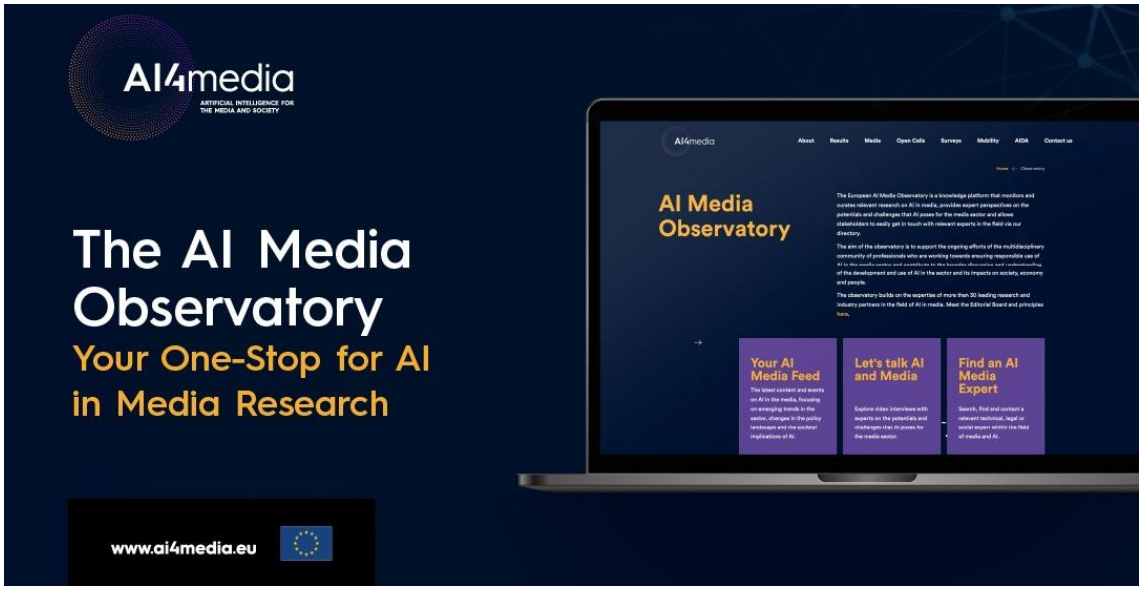


Figure 4: Visualisation of AI Media Observatory used for dissemination.

3.1 Introducing the AI Media Observatory

The AI Media Observatory consists of three main components that together serve to make it a one-stop-shop for AI Media research, where both media organisations, civil society, policy makers and the research community can find relevant information, experts, and events.

- 'Your AI Media Feed' features content that engages with technical trends within the field of AI, ethical approaches to developing AI, legislative changes and their impacts, or the societal and environmental implications of AI (Figure 5).

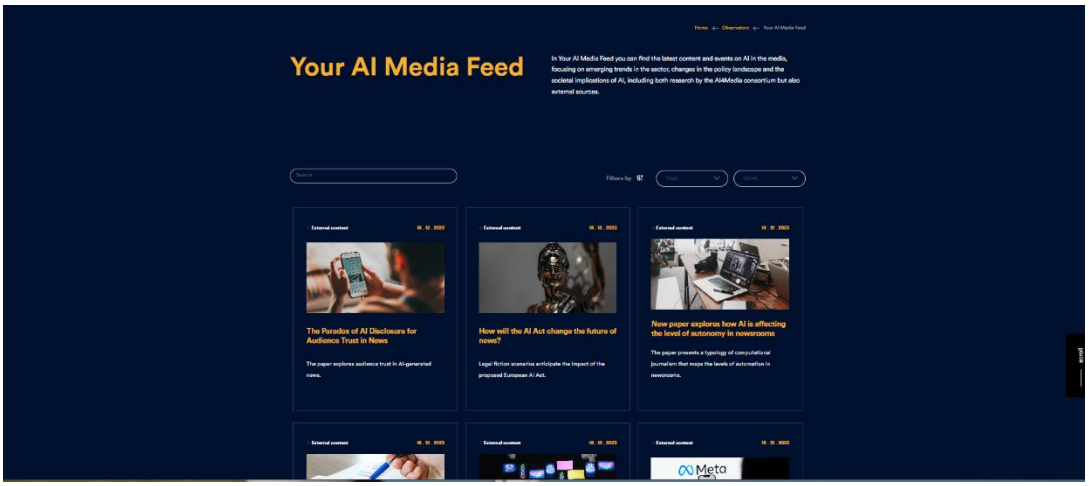


Figure 5: Screenshot of Your AI Media Feed.

- 'Let's Talk AI and Media' features audiovisual content, such as roundtables or keynotes that engage with technical trends within the field of AI, ethical approaches to developing AI, legislative changes and their impacts or the societal and environmental implications of AI (Figure 6).

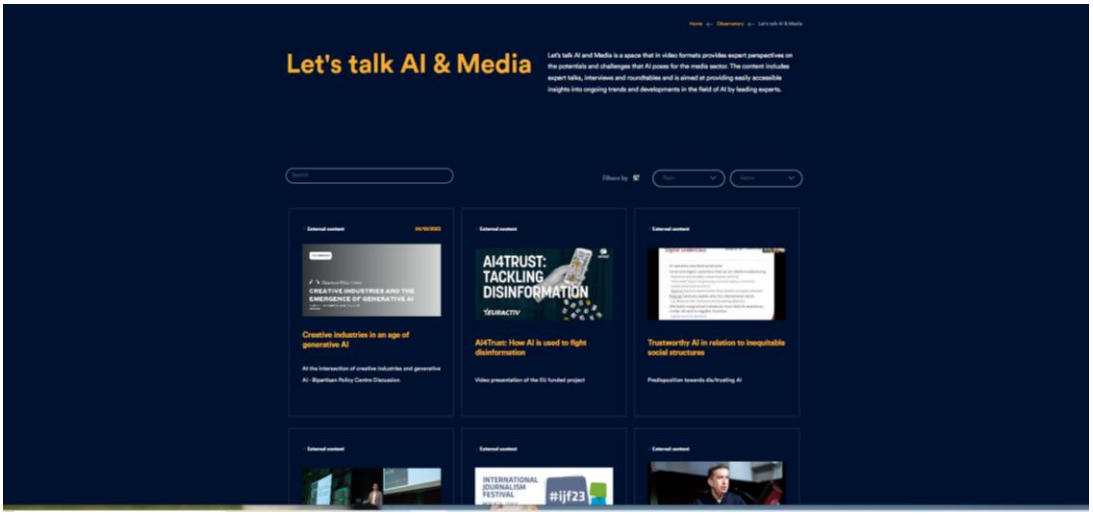


Figure 6: Screenshot of Let's Talk AI and Media.

- 'Find your AI Media Expert' features an overview of relevant social, technical, and legal experts that work in the intersection of AI and media. They are featured with a short description and contact information to enable different stakeholders to reach out to relevant experts (Figure 7).

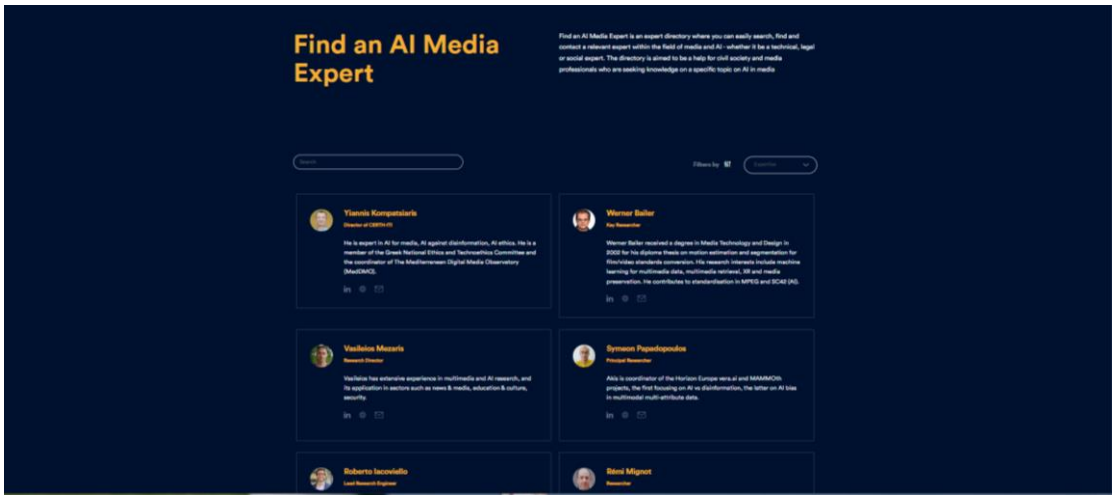


Figure 7: Screenshot of Find an AI Media Expert.

Across these different parts of the Observatory, it is possible to find information that engages with the various facets of the media sector and the impact of AI, such as the implications for the gaming industry, the news industry, and the societal discourse as well as insights into how researchers can utilise AI in their research or good practices on developing responsible AI. Furthermore, the Observatory also features upcoming relevant events on AI and media.

3.1.1 Background on the implementation

The implementation of the AI Media Observatory was a highly collaborative process, where UvA, KUL and NISV focused more on defining the functionalities and processes of the Observatory while LOBA focused on design and development. The function of the Observatory was ongoingly discussed in the WP2 meetings and overtime crystalised into an idea of an online site that would monitor, aggregate, study, and interpret relevant information with the purpose of supporting a better understanding of media AI developments and their impact on society, economy, and people. Beyond this, it was decided that it would be important that there was a dialogue and outreach component, which led to the establishment of the expert directory.

The concrete design work was initiated in the spring of 2022 and several iterations were made over the following year. The initial discussion was carried out within WP2. Once the idea had matured, it was presented to the wider consortium in September 2022 during the 6th plenary, which provided the first occasion for feedback that was used to develop the first prototype of the Observatory. The website prototype was then presented again to the consortium at the following plenary meeting for further feedback.

In April 2023, the 'beta-version' of the Observatory went live on the AI4Media website together with the editorial policies and workflows (will be described below). This allowed for testing the procedures for content submission and upload and provided time to identify bugs on the site and implement further features. In October 2023, the final version of the Observatory was launched and was widely disseminated via a brochure (see Figure 8), social media, emails, events and newsletters (see Figure 4).

Figure 8: Screenshot of the brochure.

3.1.2 Day-to-day work of the Observatory

The content and experts featured on the Observatory are curated by an **editorial board** that convenes on a monthly basis to evaluate the content that is suggested and assess the applications of experts. The editorial board is composed of partners from the consortium and currently consists of:

- **Editor-in-chief:** Anna Schjøtt Hansen, PhD Candidate at the University of Amsterdam
- **Editorial oversight and support:** Candela Bravo, European Project Consultant
- **Editor:** Frédéric Precioso, Professor at Universite Cote d'Azur
- **Editor:** Rasa Bocyte, Senior Advisor for Research Collaborations at The Netherlands Institute for Sound and Vision
- **Editor:** Noémie Krack, Researcher at the Centre for IT & IP Law at KU Leuven
- **Editor:** Lidia Dutkiewicz, Researcher and PhD Candidate at the Centre for IT & IP Law at KU Leuven

The possibility of joining the editorial board was open to the consortium, aiming to have both technical, social, legal and use case partners join.

3.1.3 Content submission

As the Observatory is a curated knowledge platform, it depends on suggestions of quality content. These suggestions currently mainly come from partners in the AI4Media consortium who are asked on a monthly basis to submit and author content to be featured on the Observatory. However, to ensure inclusivity and openness the public is also welcome to submit suggestions via an online form on the website (Figure 9).

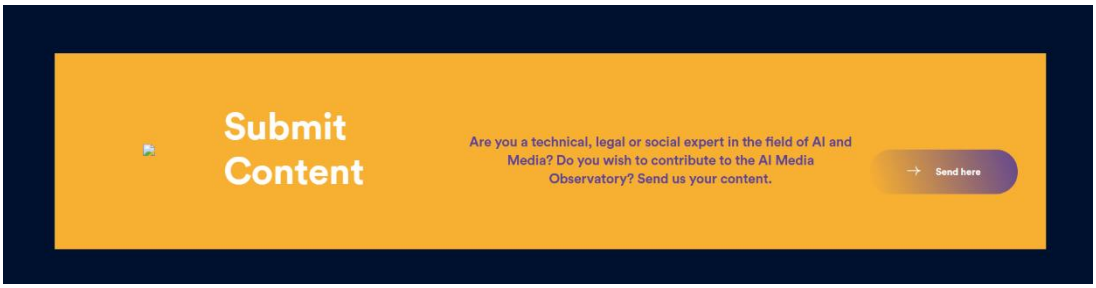


Figure 9: Screenshot of content submission form.

The content submission can entail simply submitting a link to a relevant piece of content together with a short header and subheader that provides a clear description of the content and its relevance. For consortium partners, there is also the possibility of submitting authored content, such as blog posts or a piece on a certain trend. To support the production of authored content the editorial board developed additional resources that guide the consortium partners through this process and an editorial calendar has been established where each month two partners are asked to submit both external content and author content for the Observatory.

To ensure a transparent process, the suggested content is evaluated based on the editorial principles that were developed in WP2, which are now outlined, but also available online here. Overarchingly, the AI Media Observatory publishes content that focuses on the impact of AI in the media sector. This could include but is not limited to, impacts of legislation changes, economic impacts in the sector (e.g., layoffs or investments), societal impacts (e.g., potentials of AI in battling misinformation) or social and ethical approaches to AI.

This means that academic papers with a purely technical focus are not considered relevant, but technical perspectives presented in more digestible formats and with a focus on their impact in the sector are welcomed.

All submitted content should relate to one of the following five overarching topics:

- **Policy & Legislation:** Content that provides insights into new policies or legislation of AI that affect the media sector.
- **Social & Ethical AI**: Content that addresses different approaches or principles for how to achieve socially and ethically responsible AI solutions in the media sector. This could include questions of transparency, discrimination or explainability.

- **Environmental impact**: Content that addresses questions of the environmental impact of AI and possible solutions to improve the carbon footprint of AI in the media sector.
- **Trends & technical innovation**: Content that provides insights into new AI trends or technical innovations that might both positively or negatively impact the media sector.
- **Societal impacts:** Content that provides insights into both the positive and negative impacts of AI in society in the context of the media sector. This could, for example, include the impact of generative AI in relation to trust in news or disinformation.

Submitters are welcome to propose content in different genres (e.g., reports, blog posts, news articles, commentaries etc.) and relating to different topics that abide by the general principle outlined above.

The choice of images accompanying the content should follow the guidelines stipulated by Better Images of AI, which have provided guidance on how to avoid stereotypical and problematic representations of AI.

Once content is submitted by either the consortium or the public, it is reviewed by one of the editors of the editorial board to ensure it abides by the editorial principles. The reviewer decides whether to publish the content and then LOBA takes over the process of uploading the content. If a reviewer is in doubt regarding a piece of content they can ask a fellow member of the editorial board to provide a second review. The process is outlined below in Figure 10.

Figure 10: Visualisation of the content submission workflow.

3.1.4 Expert directory

The expert directory is open to everyone, and relevant experts can easily apply via a form on the website (Figure 11). Being featured as an expert enables visibility towards an expert field and industry actors and recognition of being a key expert within the area of AI and media.

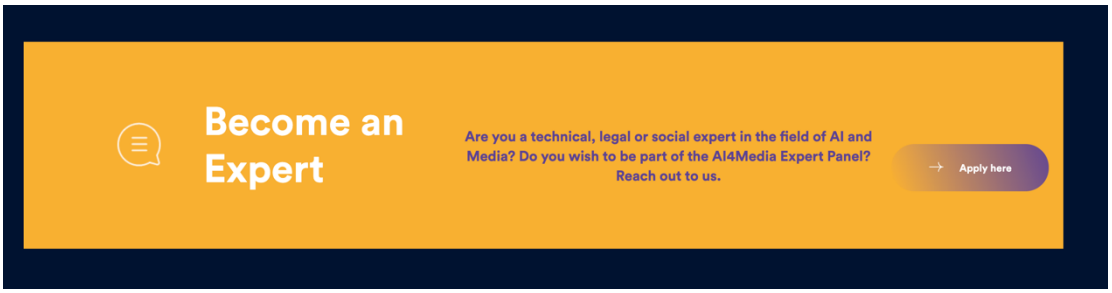


Figure 11: Screenshot of expert sign-up form.

Similar to the content submission process, the editorial board oversees the submissions of expert profiles and conducts an eligibility check. Here an eligibility criterion was established to guide this assessment, which includes that the expert must hold an expertise that is directly relevant for AI in the media sector, not either 'purely' AI or Media expertise.

As part of the implementation of the Observatory and particularly the expert directory, a specific 'Legal Notice & Privacy Policy' was developed by KUL to ensure good legal practice both in the content curation but also in the handling of personal information of the experts.

3.1.5 Dissemination and Reach of the Observatory

The Observatory currently features 74 pieces of content and 23 experts. The Observatory has been disseminated via different channels. This has included bringing the flyer to different events, a dissemination email for consortium partners to use, social media posts regarding the Observatory and a soon-to-be-published blog post on the AI Alliance Blog. Furthermore, LOBA is currently running several targeted social media campaigns planned to end in February 2024. A full report on the effects of the campaigns will be produced then, but the initial numbers already illustrate a significant growth in reach as shown in Table 1 below.

Table 1: Overview of users and pageviews in the Observatory.

	Users	Page Views
2023 July	50	96
2023 August	33	58
2023 September	46	117
2023 October	61	196
2023 November	2,050	2,907
TOTAL	**2,468**	**3,781**

During the 9th project meeting in October 2023, LOBA presented some initial statistics for the period where the Observatory was in its initial beta-version (Figure 12). However, already here there were a good number of visitors and particularly two pieces of content were gaining a lot of views:

- Initial white paper on the social, economic, and political impact of media AI technologies (365 page views).
- The AI4Media Strategic Research Agenda on AI for the Media Industry (301 page views).

Figure 12: Key statistics for the AI Media Observatory (October 2023).

There were some limitations regarding the numbers, as LOBA could at this time only provide numbers for AI4Media-hosted content, not external content.

3.1.6 Long-term sustainability of the Observatory

The long-term sustainability of the Observatory beyond the AI4Media project's finalisation is currently being discussed within WP2, where several potential routes are being discussed. These include:

- Ensuring the commitment of several consortium partners to sustain the editorial board for a minimum of five years and commit to author content.
- Finding funding to cover the continued maintenance of the Observatory by LOBA (the domain is currently owned for five years).
- Connecting the Observatory to other EU initiatives, such as the AI-on-demand platform.
- Applying for additional funding to expand the Observatory to have further functionalities and more focus on outreach (e.g., events and partnerships).

The final sustainability plan is aimed to be finalised in the spring of 2024, so that there is a clear direction once the AI4Media project ends in August 2024. Currently, this work is dependent on understanding better what the costs of maintenance will be and what options for funding there are. These options are already discussed but discussions are planned to become more concrete in early 2024.

4 Qualifying concrete challenges in media AI

In the following sections, the insights from the three industry workshops focused on qualifying identified media AI challenges will be outlined. These are all being disseminated in easily accessible formats like factsheets.

4.1 Workshop 1: Measuring the success of recommender systems

This section provides insights into the current challenges, potentials and good practices for measuring the success of recommender systems for news media. It is based on an online workshop organised in April 2022 with six commercial and public service media organisations from Europe. The results have also been disseminated in this short report.

The workshop was designed based on insights from the 'Initial white paper on the social, economic, and political impact of media AI technologies' (D2.2), where evaluating recommender systems was identified as a core challenge for media. As a result, the central question in the workshop was: 'When can we consider a recommender system successful and how can it be measured?'. The workshop included:

- A short case study presentation by the project leader of a recommender project at a Danish media organisation who presented insights from their recent tests and evaluations of their recommender systems.
- A roundtable by all participants where they each reflected on how they had previously engaged with the question of what quality criteria are important when evaluating recommender systems in the media context.
- A shared discussion focusing on sharing knowledge and developing some initial core principles or questions that should be considered by the media and policymakers.

4.1.1 Key insights: Three core questions

During the workshop, three core questions emerged that characterised the discussion on how to measure the success of recommender systems in media, namely (1) what to optimise recommenders for, (2) how to balance the 'costs' recommenders and (3) how to include the voice of the audience. These three questions are now discussed based on the insights from the workshop.

Optimising toward what goal?

Participants raised the question of what should drive the optimisation of the recommender systems deployed by media, which raised a related question of what metrics to use in evaluating the systems' success.

Suggestions for drivers of optimisation:

- Measuring the success of a recommender system not only by what it shows to us but also by what it shouldn't show to us to avoid societally harmful effects of recommenders (e.g., inducing rabbit holes or harmful feedback loops).
- Measuring the success not only on an individual level but also on how the system manages to capture community interest (e.g., addressing minority groups).
- Measuring the success by the diversity in content discovery, so that users are exposed to a wide range of content they might not typically come across.
- Measuring the success according to editorial values, so that the system recommends in a way that aligns with the local editorial mission of the media organisation while remaining aware of problematic feedback loops this might produce.

Challenges for evaluation:

- Moving beyond click-based accuracy metrics (e.g., diversity novelty, coverage, serendipity etc.), as this remains the common standard to test and evaluate the performance of the systems against, particularly when deploying systems by third-party providers.
- Operationalizing alternative metrics proves a challenge as editorial values can be hard to translate into metrics and require a close editorial-tech collaboration, which is often even more difficult when deploying systems by third-party providers. Being able to do so also requires investments on the side of the organisation, in skills, expertise, as well as time and room for experimentation.
- Lack of benchmarks, to test how the system performs in comparison to other systems that optimise towards media values, as existing benchmarks also remain focused on click-through rates or have simplistic understandings of, for example, diversity.

How to balance the 'costs' of recommenders?

During the workshop the participants also raised the question of how to balance the cost of recommenders - both in terms of the economic costs of building and scaling recommender systems, but also in terms of the potential 'costs' of privacy these systems might produce for the audience, who will have to surrender their data for these systems to run effectively.

Suggestions for ways of balancing implications:

- Making infrastructures as efficient as possible to reduce costs of scaling recommender systems and being conscious of what is minimally required to technically deliver on the proposed mission.
- Approaching data management as a service in which users have a choice of how their data is used and managed (e.g., what type of data or opt-out options), to also maintain trust in media, which can be threatened by misuse or untransparent use of personal data.

- Being conscious of data decisions and considering using the minimal amount of data needed and avoiding using personal data, when possible, as opposed to engaging with data practices through a lens of " the more data the better".
- Being transparent to the audience regarding what data is collected and how it is used within the systems.

Challenges in balancing implications:

- Deploying ML models at scale requires high costs for data storage and training, which can be challenging even for large well-resourced media organisations.
- Cloud computing is often viewed as one solution to reduce costs as it provides options for scaling up and down but also induces new dependencies on commercial infrastructures and risks adding greater costs to, for example, audience privacy.
- Balancing the cost of transparency for the user experience, as large technical descriptions might produce barriers to audience engagement without delivering the goal of transparency.

Where is the voice of the audience?

The last core question that was raised by the participants was how to better include the voices of the audience in the development of recommender systems, but it was also questioned to what extent this should be the case. The latter relates to how audiences might also skew the distribution of content and produce problematic feedback loops as a result of audiences tending to prefer reading what they know and like (confirmation bias).

Suggestions for including the audience's voice:

- Ensuring strong editorial and technical collaboration via cross-disciplinary teams where the editors can speak on behalf of the audiences' interests.
- Including the audience in defining the optimization goals for the recommender system via, for example, focus groups with an emphasis on getting insights from marginalised groups whose perspectives are often threatened to be left outside of these discussions of how to assess recommenders.
- User control, not only over data collection and usage, but also over the recommendations themselves via, for example, different ways to tweak the optimising goals of the recommenders towards more topic or genre diversity, and being able to provide feedback.

Challenges in balancing the role of the audience's voice:

- Balancing the audience's voice against editorial concerns with the aim to ensure that the mission of the publication remains present, as the aim of media is also to push perspectives and democratically inform about multiple perspectives (at a minimum by exposing the audience to such diverse content).

- How to meaningfully include audience perspectives, provides a challenge as often the groups that are most at risk of being marginalised are harder to reach and the discussion can easily become too abstract.

4.1.2 Good practices

During the workshop, several good practices were discussed together with some potential recommendations for how policy could support these best practices. Below the report presents a case study that was presented at the workshop that outlines an evaluation framework for assessing and testing recommender systems (Table 2). Then it provides an overview of good practices for pursuing an adaptable and process-oriented assessment frameworks.

Table 2: Case study description - Recommenders for news.

Case study: A value-driven framework for the evaluation and design of AI systems for news

The case study is based on a presentation by a large Danish commercial media organisation who are developing an in-house recommender system with the aim to (1) deliver a more engaging and informing news experience, (2) deliver AI systems that are aligned with the outlet's editorial mission and minimise the dependence on the tech giants by building in-house, and (3) push for healthy norm setting for the use of AI in news, which is currently still in development.

As part of the project, the media organisation has developed a value-driven framework that outlines four different value domains, which need to be considered and balanced during the design and evaluation of AI systems. These include:

- Moral responsibility via ethical AI values (e.g., transparency/explainability, fairness, justice, privacy, and avoiding harm).
- Public Service via journalistic values (e.g., truthfulness, objectivity, credibility, pluralism, relevance, identification, and sensation).
- Economic value creation via business values (cost-effectiveness, superior value to attractive customer segments and strategic independence).
- Technical Excellence via technical values (best performance against optimization goal, efficiency in resource use, such as data and computing)

These four value domains can help to inform the compromises needed to be made throughout the process of developing a recommender system and to balance the known challenges of AI, such as the lack of ethical and journalistic considerations in commercial recommender systems.

4.1.3 Towards adaptable and process-oriented assessment frameworks

One insight that stood out during the workshop was the need for more adaptable and process-oriented assessment frameworks of recommender systems. It was highlighted how each media organisation will have their own values and goals that need to be foregrounded in the assessment framework, but also how it would have to be more of an assessment pipeline or procedure because assessment must be made throughout the developing process, not only in the final stages. Below there is an outline of seven evaluative questions that should be asked across the pipeline and what compromises are raised.

Mission definition: Initially it should be discussed what values are essential to the media and how these can be optimised toward in the future recommender systems. This process should be inclusive in the sense of including representatives from the different parts of the organisation (management, legal, marketing, technical, journalists, etc.). This could be represented in a framework, such as the one presented in Table 2 above.

- **Compromise(s):** The framework can help to assess what values can be negotiated and which are non-negotiable, but it will also be important to ensure how questions of audience voice will be addressed.

Buy from a third-party provider or build in-house: This decision should be based on the above goals and values, in-house expertise and resources, business case and whether the system will be mission critical (see also Council of Europe Procurement Checklist in the Guidelines on the responsible implementation of artificial intelligence (AI) systems in journalism).

- **Compromise(s):** Such decisions might be impacted by cost and resource efficiency concerns, and it will be important to assess when it will not be acceptable to use a third-party provider.

Data collection and cleaning: Consider the dataset to ensure the data is not biassed in ways that counteract the goals set out, and that training datasets or the pool of information to draw on is sufficiently diverse (e.g., due to the composition of the dataset) and assess how the data choices will impact the performance of the model.

- **Compromise(s):** Data collection and cleaning might be impacted by the quality of the data available, which might not be of high enough quality and could lead to compromises in expected performance. In the consideration of what data is the 'best' to use there might also emerge a trade-off, where 'good' data will affect the values of data privacy of the audience.

Model selection, training, and testing: Consider different recommender models (e.g., collaborative or content filtering) and how they can contribute differently to the goals. Testing multiple models allows for an assessment of how they produce different effects.

- **Compromise(s):** resource constraints might impede testing a large variety, so it will be important to first understand what is at a minimum required to gain the needed insights. Equally, training the models is expensive and there might be compromises in how much training is possible economically and what this might mean for the performance of the model.

Testing across locations: Testing the different models across locations on the online news sites, such as how the recommender performs on the front page versus on sub-sections on the site. Such testing might reveal how they might perform better or worse in some locations and where they might induce unwanted harm, allowing for a discussion of how to compose the recommender infrastructure to best reach the goals set out.

- **Compromise(s):** There might be conflicting goals between increasing, for example, engagement and diversity, so evaluations at this stage will need to compromise on what values matter at what parts of the news site.

Testing beyond accuracy: Testing the models for different metrics also enables nuance regarding how some models might enable other values beyond engagement, such as diversity in exposure and consumption.

- **Compromise(s):** It might be difficult to operationalize measures beyond accuracy and some compromises might arise in how and what can be measured.

Scaling: When scaling, it will be necessary to assess what computational power will be needed and what efficiency measures are attainable.

- **Compromise(s):** Here it will be necessary to decide where technical compromises can be made in terms of running the systems live or online a few times a day or assess whether a lower performance might be acceptable if it significantly reduces costs.

4.1.4 Policy recommendations

During the workshop, concrete suggestions for how to create the conditions for measuring the success of recommender systems more efficiently were also discussed and three main conditions were highlighted to be better supported by policy.

1. **Improving knowledge and negotiating power when working with third-party providers**: This could include the ability to ask for optimizations beyond click-rate through accuracy to alleviate the current market gap, in which only certain media organisations have the agency to adjust their recommenders.
2. **Better benchmarking practices**: This could include the ability to benchmark against other media systems and with metrics beyond accuracy, which could help better illuminate the public value of recommender systems.
3. **Better industry-academia collaborations:** This could include collaborations on technical solutions but also to help develop value frameworks, which could help reduce the costs of the projects and support more responsible AI development.

4.1.5 Results in brief

As a way to make the results of the workshop easily disseminable, they were summarised in a short report that was published in December 2023 (see Figure 13). The short report will be used for dissemination at upcoming events and via social media.

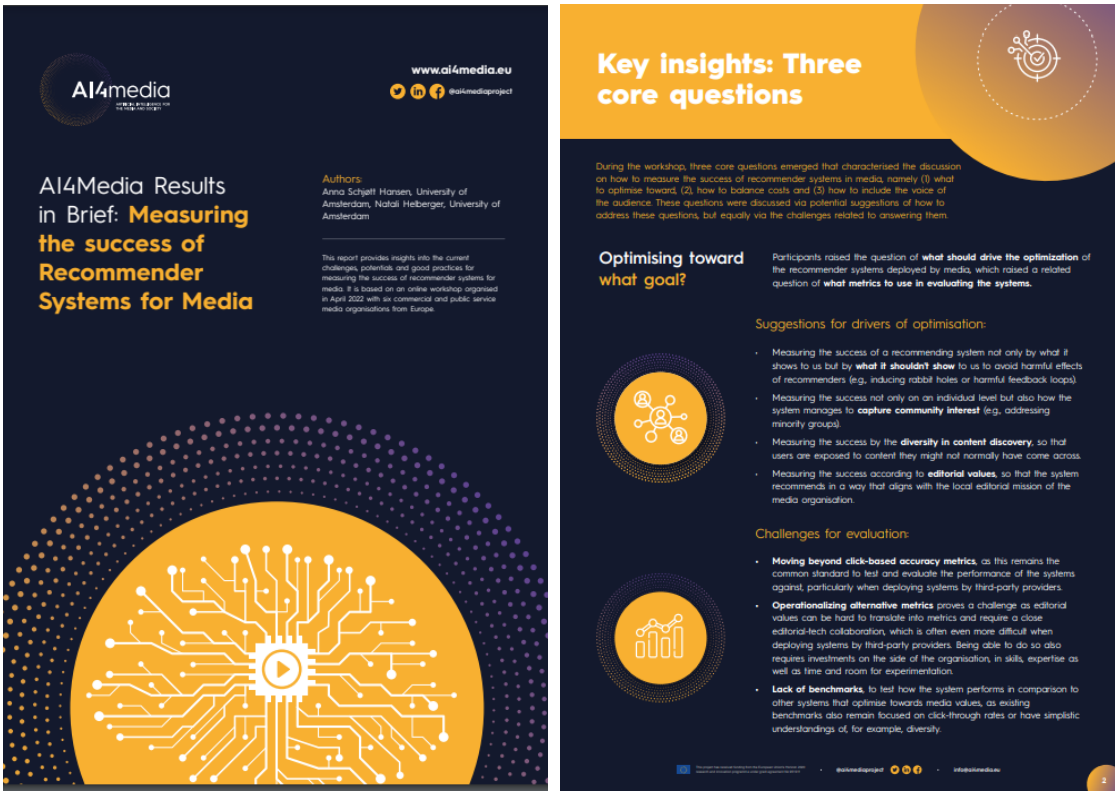


Figure 13: Screenshot of the short report outlining the insights from the workshop on recommender systems.

4.2 Workshop 2: Identifying common challenges regarding the use of AI in content moderation

This section provides insights into the common challenges faced by industry actors using AI in content moderation. It is based on an online workshop organised in February 2023 with nine industry actors representing both small and large organisations based in Europe, who were invited via the network of the organisers and the wider consortium. The findings from the workshop have also previously informed the 'Report on Policy for Content Moderation' (i.e. deliverable D6.2) and have also been disseminated in a short report.

The workshop aimed to identify the common challenges faced by industry actors who engage with AI in the context of content (comment) moderation and learn from their respective experiences on the use of AI systems assisting their content moderation efforts.

The workshop was organised by KUL and UVA on February 6th, 2023 and was conducted under the Chatham House Rules. A participant list is provided below that provides some contextual information regarding the participants. The workshop unfolded in the following way:

- **Introductory talk:** The workshop started with a short introductory talk by Distinguished University Professor of Law and Digital Technology, with a special focus on AI, Natali Helberger from the University of Amsterdam (UvA).
- **Round table:** Each participant shared what they consider their main challenge when working with AI-enabled content moderation (e.g., technical, or ethical challenges).
- **Discussion of good practices:** Based on the round table, the last part of the workshop focused on identifying potential good practices. Two discussants were also present to help and guide the discussion: Bernhard Rieder, Associate professor in New Media and Digital Culture at the UvA and Aleksandra Kuczerawy, postdoctoral researcher at KU Leuven focusing on online Content Moderation and the Rule of Law.

The participants included:

- A European company producing image recognition solutions for developers and businesses.
- A European consultancy doing content moderation analysis.
- An AI4Media-funded open call project working on adaptable comment filtering.
- A prominent newspaper from Austria.
- A UK company developing socially Responsible AI for Online Safety. They develop AI-powered tools to find and stop toxic content.
- A German local broadcast media production and distribution company doing responsible journalism and professional entertainment.
- An American technology company that owns a very large online platform.
- A European company developing trustworthy, transparent and explainable human-centred AI solutions that read and understand large amounts of text.
- A researcher from a well-known university in the Netherlands and consultant for the United Nations Department of Political and Peacebuilding Affairs (DPPA) Innovation Cell.

4.2.1 Key insights: Six core themes of AI-assisted moderation

During the workshop, six overarching themes emerged that described both what AI-assisted moderation can facilitate, but also pointed to issues of human oversight, transparency and inequalities in the quality of AI-assisted moderation.

Moderation at scale: AI systems remain key to ensuring moderation at scale and proactively classifying content or comments that violate the policies of either news outlets or social media. This can enable prioritisation of the worst content, which enables human moderators to quickly remove the most heinous content first but also protects the human moderators from seeing violent and degrading content, which can be filtered immediately. Thereby, producing better working conditions for moderators by minimising their workload and protecting their mental health.

Hybrid solutions remain key: AI-facilitated moderation solutions cannot stand alone, they require human oversight to ensure the quality of the moderation and assess boundary cases where factors might lead to false positives (e.g., humour or satire). Human oversight also remains key as an accountability mechanism in case moderation decisions are contested. However, AI supported moderation can also help to counteract individual moderator bias.

Supporting constructive debate: AI systems also offer different strategies of moderation by enabling the identification of 'good' content that provides constructive input to the debate. Beyond identifying and potentially removing or down-prioritising problematic content, AI can help to prioritize, for example, constructive comments as a way of positively reinforcing the practices in the communities and supporting constructive debate.

Putting numbers behind the abuse: AI systems can also contribute to putting numbers behind the abuse by illustrating, for example, the number of verbal attacks experienced by politicians or other public figures. Currently, limited evidence exists to illustrate the scale of these issues.

Correcting the moderation gap: An important aim will also be to bridge the current moderation gap, where English-speaking contexts and spaces are currently much better moderated via AI systems, as opposed to non-English-speaking contexts. This gap was seen as also connected to the lack of willingness by large platforms to collaborate with third-party providers who develop local solutions. This gap leads to a dislocation of moderation where non-English speaking countries both experience over and under-moderation. The under-moderation emerges because of language issues, as current AI systems are predominately trained on English data, whereas the over-moderation occurs as a result of Western – often American – understandings and policies of moderation (e.g., of hate speech) imposed on these contexts.

Moderation transparency: There is a growing need to be more transparent about the ways in which AI is used in moderation practices, also induced by emerging laws in, for example, the EU. This could include publishing guidelines but also being more transparent about the ways in which the system classifies and the following process of human oversight, as well as better possibilities of contesting the decisions. However, this also requires clear transparency and explainability of the systems to the moderators for them to correct false decisions.

4.2.2 Core challenges for using AI in content moderation

Several challenges were raised in relation to the use of AI in content moderation where the most predominant related to data, both in terms of access and diversity in data and noise in the data, but also data protection concerns. Equally, the lack of contextual understanding and clear definitions produced major challenges for the industry actors as well as unique issues related to evaluation and live moderation.

Data challenges

Data was considered the most important challenge, particularly because the field is now drawing from the same language models (e.g., Bert), which makes the data even more important to innovate the systems. These challenges included the following:

- **Access to datasets:** The participants described how accessing quality datasets was an issue, as there was a lack of publicly available datasets to train the AI systems on and often the data was not from the exact context that the AI system was to be used in. For example, one participant described how they were building a moderation tool for Facebook, they could harvest data from Facebook's API, but that did not include the content that had been moderated already, which would serve as high-quality training data for their model.
- **Diversity in data:** The participants also referred to issues of diversity in data relating to a lack of data on particularly small or minority languages and particular geographic regions. Particularly trustworthy and high-quality data was seen as highly difficult to attain when moving beyond high-resource languages.
- **Dealing with noisy data:** Another issue raised was problems with noisy data which resulted from inconsistent practices of moderation by human moderators. This could both be seen as a subjectivity bias but also related to the lack of clear and shared definitions of, for example, hate speech as well as other factors such as fatigue, complexity of content or difference in training. Equally, many of the datasets risked perpetuating existing data by drawing on 'real' moderation data, which should also be mitigated when using such data.
- **Data protection and sharing data:** Data was also connected with challenges of ensuring the data subject's data protection rights, for example, via anonymization. The GDPR was seen as a challenge for sharing datasets between multiple organisations.

Definitional challenges

Another group of challenges related to the problems of defining what should be moderated and accounting for local contexts as well as finding a balance to uphold free speech. These challenges included:

- **Defining content to be moderated:** A main challenge related to the difficulties in defining the different forms of content to be moderated, such as hate speech, because the definition would often vary across different geographic contexts and also not be a stable definition, but rather evolving with societal and cultural changes, which is also highly situational.
- **Including the context:** The emphasis on contextual understandings of what requires moderation challenges of the use of AI, as it requires local retraining and ongoing adjustment, as societal events, such as war, change what can and cannot be posted.
- **Balancing moderation and freedom of expression:** As many policies on moderation extend beyond what is legally defined as illegal speech and include problematic or hateful speech, a new critical question emerges on how to ensure safety, while also protecting free speech. This balancing is also related to the accuracy of the AI systems, as high numbers of false positives might threaten free speech, which might particularly occur in the moderation of small or minority languages with poorer training data.

Challenges in evaluating AI systems for moderation

There are also specific challenges related to the lack of shared evaluation frameworks and difficulties in gaining insights into how these systems work. These challenges include:

- **Lack of shared evaluation frameworks:** Some emerging evaluation frameworks exist for AI systems used for moderation, such as the 'hate check', which tests for what functions and attributes the system has. However, these frameworks are not applicable to all systems. Equally, there is a lack of shared benchmarks to help assess whether the systems work and how they work for different user groups (e.g., minority groups).
- **Lack of researcher access:** Another challenge relates to the minimal access to information on how these systems work, particularly on large platforms where limited researcher access is given. As these systems moderate public speech, it will be important to be able to understand how norms of speech are evolving as well as have an accountability function that assesses the function of these systems. This would also better enable public deliberation over what should and should not be considered problematic speech, which currently is left to the organisations building and deploying these tools.

Challenges for live moderation

Specific challenges also emerge or are exacerbated when AI systems are used for live moderation, such as challenging the infrastructure and computational power needed and also the human oversight.

- **AI infrastructure:** When AI systems are used to moderate speech live in, for example, comments sections, it requires a more complex infrastructure and more computational power, which might not be accessible to all organisations, inducing a gap in the access to real-time moderation in certain contexts.
- **Lack of human oversight:** Using AI systems in live content moderation also challenges the aim to always have human oversight and the potential ramifications of mistakes made by AI systems.

4.2.3 Potential ways forward

The workshop also provided insights into potential ways forward that could enable the responsible use of AI systems for content moderation and alleviate some of the identified challenges. Particularly three recommendations were made relating to producing more local and open-source solutions, using transfer learning to alleviate the current moderation gap, and producing better conditions for sharing datasets.

- **Local and open-source solutions:** One way forward focused on providing better conditions for making local and open-source AI solutions that could complement the AI-assisted moderation systems on, for example, large platforms to minimise the moderation gap that currently exists between English and non-English contexts. These conditions could be provided via either targeted funding or policy but also would require

new forms of collaboration between third-party providers and large platforms, who are both important sources of data and where such solutions could have a high impact. Such local solutions could also better consider the specificities of the cultural context in which the moderation would take place, thereby, also alleviating the dislocation of moderation.

- **Transfer learning across regional languages:** To enable a way forward with such solutions, the participants looked towards transfer learning and the ability to provide regional models rather than purely local models, as this would require less training data to retrain models in similar languages. Such experimentation had already been carried out by two participants with high accuracy rates, where only 10-20% of labelled data was needed to retrain the models, which would place much less stress on small or minority languages to be fully responsible for building their own system and training them on complete datasets.

- **Publicly available datasets:** The last way forward, which also would be incremental for any of the solutions to be attainable, was to create publicly available datasets, as this remained the core barrier for most of the projects described during the workshop. While a few examples, such as datasets relating to sexism, have been produced, they remain English-focused, and the quality of available data varies significantly across different geographic contexts. High-quality datasets would better enable the foundational training of regional models, which could then be retrained on smaller amounts of locally labelled and contextual data.

4.2.4 Results in brief

As a way to make the results of the workshop easily disseminable, they were summarised in a short report that was published in December 2023 (see Figure 14). The short report will be used for dissemination at upcoming events and via social media.

Figure 14: *Screenshot of the short report outlining the insights from the workshop on AI-assisted content moderation.*

4.3 Workshop 3: Current challenges and future paths for AI in audiovisual archives

This section provides insights into the current challenges, potentials and good practices for implementing and using AI in audiovisual archives in the Balkan and Mediterranean region, which previously has been less explored in AI4Media research. The insights in this section are based on discussions from a one-and-a-half-day workshop on 23-24 March 2023, hosted by the Croatian Public Broadcaster (HRT) in Zagreb and co-organised with COPEAM, a non-profit association devoted to the promotion of dialogue and cultural integration in the Mediterranean Region among other through the involvement of the major players of the audiovisual sector, among which the public service radio and TV broadcasters from 26 countries in the area. 12 participants from seven different countries in Eastern Europe and the Mediterranean participated in the workshop, including Romania, Croatia, North Macedonia, Turkey, Serbia, and Kosovo. The results have been disseminated on social media, in a short report and at the 54th IASA CONFERENCE & 4th ICTMD FORUM.

The workshop involved different sessions:

Day one: AI in the audiovisual archives

- Introduction round with participant presentations, focusing on the question of where they are in their AI journey.
- Introduction to good practices for AI in audiovisual archives by Johan Oomen, Head of Research at the Netherlands Institute for Sound & Vision, Georg Thallinger and Hannes Fassold, Joanneum Research (partners in AI4Media).
- Exercise #1: Mapping the potentials and challenges of using AI in audiovisual archives.
- The AI Pipeline - legal, social & technical perspectives. Roundtable discussion presentations from legal, social and technical experts from AI4Media on the critical questions relating to AI in audiovisual archives
- Exercise #2: Ranking critical questions for your organisation - what conditions are needed for you to really engage with AI.
- Inspirational session with a presentation by the Institut national de l'audiovisuel (INA).

Day two: Build or buy?

- Talk on Procurement guidelines by Prof. Natali Helberger and Stan Piasecki, University of Amsterdam.
- Exercise #3: Where can we compromise?
- Group discussion: three compromises and three critical values

Figure 15: Photos from the workshop on AI in audiovisual archives.

4.3.1 Key Insights: AI readiness and strategies

Previous reports have found that audiovisual archives remain focused on experimenting with rather than implementing AI directly into archival workflows[5]. This sentiment was confirmed by seven organisations from the Balkan and Mediterranean region who took part in the workshop "AI for Audiovisual Archives: Current Challenges and Future Paths". It is clear that there is a growing awareness of AI amongst these organisations. Many have started developing AI strategies as well as initiated various experiments and proof-of-concept implementations. These include the use of face recognition to identify public personalities, generating subtitles and sign

[5] See https://op.europa.eu/en/publication-detail/-/publication/20b9f607-f7c4-11ea-991b-01aa75ed71a1

language avatars, and using sentiment analysis to analyse user interaction with content, shots and face detection. We identified **four themes** that define the way participating audiovisual archives engage with AI:

- **Always Busy Digitising:** This sentiment was very strong amongst all participants - organisations feel pressure and urgency to digitise their audiovisual collections while also dealing with limited resources. The organisational focus on digitisation limits the possibility of engaging in research projects around AI. At the same time, there is a clear understanding that digitisation is not the end solution, but a prerequisite to the most uses of AI for archival content distribution and reuse. Beyond this, it is clear that new challenges emerge as digital collections grow, such as ensuring sufficient and protected storage solutions and the need for new infrastructure to better manage digital collections.

- **Public Mission of Archives:** The role that AI can play is shaped by the public mission that media archives serve. Organisations are eager to use AI with the goal of increasing their societal contributions, for instance, by developing services specifically for minority groups or education purposes.

- **Archivists as Data Scientists:** The fundamental shift from managing physical collections to managing data asks archivists to acquire new skills that are close to those of data scientists - including data analysis and visualisation, programming, and a basic understanding of how AI systems are designed.

- **Human vs Artificial Intelligence:** A recurring theme throughout the workshop was the interaction between humans and AI systems. What kind of skills do humans need to meaningfully interact with AI? What information must an AI system provide to its users in order to deliver easily interpretable outcomes? This is crucial to address questions about the trustworthiness and transparency of AI systems.

4.3.2 Potentials and barriers

The workshop highlighted several core potentials and challenges of AI that both align with previous findings but also add new regional and archive-specific nuances (see Table 3).

Table 3: Potentials and challenges of AI in audiovisual archives.

POTENTIALS	CHALLENGES
Improving the searchability of the archive Enriching the archival collections (Speech-to-text, Facial recognition, object recognition).	**Lack of AI models** in local languages.
Improving preservation and restoration of collections Increasing accessibility (e.g., sign language, subtitles) and providing new services directed at minority groups.	**Pressure to spend available resources** are focused on digitisation, not on AI.
Increasing the valorisation of the archive and providing new revenue streams.	**Lack of existing AI solutions** that can integrate into currently used tools and workflows.
Enhancing the impact of public service media through new and better-tailored services.	**High costs of developing or purchasing** AI systems from third-party providers
Improving the accessibility and visibility of exponentially growing collections.	**Expectation to deliver "perfect" solutions**, little room for error and experimentation.
Enabling an increased focus on diversity issues (e.g., by providing information on speaking times of politicians and the representation of different gender or minority groups).	**Lack of legal clarity** about AI implementation in a legal landscape in flux (particularly blurry boundaries for European but non-EU countries).
	Limited access to human resources and (local) expertise relating to AI Lack of sufficient and interoperable metadata to efficiently implement AI systems.

4.3.3 Five recommendation areas

The workshop identified five areas of recommendations that can support the future responsible use of AI in audiovisual archives and guide policy. They are directed both at practitioners in audiovisual archives as well as policymakers.

Leapfrogging via digitisation: Many organisations are still in the process of digitising their vast audiovisual collections. This process was often viewed as a challenge and an economic burden, but in the context of AI, it could also be seen as a unique opportunity for organisations to leapfrog into AI. The digitisation efforts across archives in past years were predominantly focused on digitisation for preservation purposes, the future digitisation processes can learn from these past efforts but also have a unique opportunity to focus on digitisation for access by combining the digitisation processes with AI projects. Taking the perspective of accessibility

means including the needs of end-users in the design of digitisation projects. This can help to define AI techniques that will make digitised data accessible and searchable in meaningful ways for the target audiences of the institutions.

- **Good Practice Recommendation**: In applying for funding for digitisation processes, include use cases for AI tools directly in the project descriptions. This might also open up new streams of funding oriented towards AI usage.
- **Policy Recommendation:** Increased focus on supporting combined AI and digitisation projects. Ensure that end-users (content producers, researchers, educators, etc.) are involved in AI projects by design - either as collaborators or as stakeholders consulted throughout the process.

The strength of local & regional collaboration: Many of the consulted organisations described experiencing the same challenges with implementing AI, relating to language, costs and human resources. This could provide unique opportunities in the AI context as many of the countries share linguistic and cultural similarities and collaborative approaches would make AI more accessible and scalable across multiple archives. Current examples of the benefits of such collaborations can be seen in the newly established network for Nordic AI Journalism or the Associated Press' efforts to help facilitate scalable projects amongst local newsrooms. What became apparent during the workshop was the willingness to enter such regional collaborations. Next to regional collaborations, there are also unexplored opportunities to build partnerships with local actors active in the field of AI - such as universities and start-ups - who can provide resources and skills that archives might not have in-house.

- **Good Practice Recommendation:** Be open towards and work actively with setting up knowledge exchanges or smaller collaborative projects with other local or regional public broadcasters / audiovisual archives as well as research organisations or universities.
- **Policy Recommendation:** Support the establishment of funding schemes that can help enable regional and local collaborations on AI that can allow for the development of scalable and open-source AI solutions to be used in multiple organisations.

The unique challenge and opportunity of MAMs: The centrality of Media Asset Management (MAM) systems in the archival workflow provides both a unique challenge and an opportunity for using AI. On one hand, it can constrain the use of open-source and third-party solutions as these need to integrate with the MAM systems, which are often not flexible enough to support this. On the other hand, MAM system providers could play a valuable role in making AI applications more easily accessible for archives via easy plug-in solutions. Closer interaction with MAM systems providers (for instance, via participation in networks such as FIAT/FTAs Media Management Commission) can help in bridging the gap between the specific needs of archives in the region and the available solutions.

- **Good Practice Recommendation:** When entering into the procurement process of a MAM system, aim to be clear about future AI needs and discuss these with potential

service providers to ensure the viability of the specifications that are part of the procurement process.

- **Policy Recommendation:** The need for open mechanisms to facilitate close collaborations between MAM providers and archives to ensure relevant AI applications can be integrated into and are made available in MAM systems (e.g., via standardised interfaces).

AI & organisational strategy: It has become clear that organisational strategies often do not provide guidance and specific considerations for the use of AI. This could relate to procurement guidelines, accountability mechanisms and quality control procedures. Given the societal harms that the neglectful application of AI can produce (for instance, perpetuating the marginalisation of societal groups), a clear organisational strategy on AI is key. AI policies in audiovisual archives should be guided and aligned with their public mission but should not restrict experimentation with AI.

- **Good Practice Recommendation:** Draft an organisation-wide strategy on responsible use of AI and/or add this to existing policies. Inspiration can be found in existing guidelines such as the EU's 'Ethics guidelines for trustworthy AI'.
- **Policy Recommendations:** Develop sector-wide responsible AI guidelines that take into account the unique role and requirements of audiovisual archives.

AI & capacity building: Implementation of AI puts additional pressure on resources and expertise that not all organisations currently have. As a starting point, archives should ensure that staff in different departments are equipped to constructively participate in conversations about the use of AI, especially in relation to potential risks, procurements and quality assurance processes, adherence to organisational policy and legal requirements. Next to this, new positions - such as AI engineers and data scientists - could be introduced to allow archives to experiment with existing AI tools or independently develop their own solutions without reliance on vendors. Since not all organisations have the resources to do this, regional and sectoral partnerships can be used to pool resources and benefit from expertise in local universities, start-ups or other archives. Such collaborations can be particularly important for developing localised and context-specific AI solutions.

- **Good Practice Recommendation:** Invest in upskilling of existing employees in AI (e.g., develop local training courses or utilise existing available sources, such as Elements of AI) and consider whether there is a need for new skill sets internally. Identify potential regional/sectoral collaborators who can share their knowledge or provide complementary resources and expertise to work on localised AI solutions.
- **Policy Recommendations:** Facilitate opportunities for collaboration and regional and sectoral partnerships in regions that currently have limited access to local AI providers.

4.3.4 Journey map: Building AI in-house

During the workshop, AI4Media partners presented key insights related to core moments in the process of building AI, which we in this context divide into five interrelated and iterative steps. The journey map below summarises and highlights the core considerations /questions that audiovisual archives should consider at each of the five steps when deciding to build an AI system in-house. The list of questions raised is not exhaustive but can help to ensure that key questions are raised.

Problem definition:

- What problem are we trying to solve with AI and does this particular problem require AI?
- If we decide to use AI, what legal frameworks apply - are we using AI in a legal sense?
- How do we ensure that the chosen solution reflects the public values of the organisation?

Model selection:

- Are there open-source solutions on the market that are suitable for our use case (incl. having an accessible licensing agreement e.g., terms of re-licencing)?
- Does training the model include the use of "personal data" (e.g., according to GDPR) or copyrighted information and do we have legal grounds to use them?
- How can we implement transparency for end-users (e.g., producing model cards)?

Data collection, cleaning, and training:

- Do we have a suitable training dataset and if not, should we make or buy data annotation (hereunder consider needed expertise and ethics of data labelling companies)?
- How do we ensure that we critically reflect on the biases of our dataset (e.g., legal and linguistic nuances when annotating the data).
- How can we disclose the data practices transparently to the end-users? (e.g., datasheets for datasets).

Evaluation and optimisation:

- What measures can we use to assess the model's performance for a specific task and are there ways to benchmark the results?
- Have we evaluated the model from different perspectives beyond assessing its accuracy (e.g., qualitative evaluations by domain experts)?

- How do we comply with ethical principles for Trustworthy AI (e.g., do we explicitly acknowledge and evaluate the trade-offs between them?)?

Deployment:

- How do we sustainably scale the solution (incl. considerations of cloud versus local computing)?
- How do we ensure the system can be maintained, both technically but also in terms of ensuring continued quality of the output?
- How do we ensure accountability measures (e.g., human-in-the-loop or end-user disclosure policies)?

4.3.5 Journey map: Buying AI from a third party vendor

During the workshop the participants in groups discussed a concrete scenario of buying an AI system and determined where compromises could be made and where they had strong demands. The following journey is a compilation of the insights produced in the three groups that can help guide other audiovisual archives that are considering AI applications.

Team: In the process of selecting the provider, an interdisciplinary team from different departments is needed with the presence of at minimum the following expertise: (1) end-user of the intended AI solution (e.g., cataloguer or editorial staff) who can define the problem and the needed solution, (2) technological expert who can assess and evaluate the technical solution (e.g., engineer), (3) financial expert who can assess budget possibilities, (4) public procurement expert to help with the tendering process and legal expertise to ensure legal compliance.

Deciding the specs and aligning with values: In the process of selecting the provider, and before moving ahead with the tendering process, it will be important to decide on the non-negotiable specifications of the solution. This could include the level of accuracy or local language proficiency, but also considering negotiable preferences, such as local providers. In this process, the team should consult local procurement guidelines/ethical principles for the use of AI and ensure ethical considerations and public values are included in the final decision specifications. If there are no such guidelines available, a separate process to discuss the ethical considerations should be carried out and new guidelines should be developed.

Providers: A good collaboration with future providers was highlighted as a key requirement. It is advised to start engaging in conversations with multiple providers in the early stages of the process to co-develop the final specifications of the project and develop long-term collaborations rather than one-off projects to ensure continued learning. Concrete

requirements of providers included considerations for: (1) locality (both in terms of local knowledge and potential legal requirements for working with international providers), (2) ethicality (does the provider live up to the same ethical standards as public broadcasters), (3) maturity (does the provider have previous experience with projects like this and proven solutions).

Terms of service: In negotiating terms of service, several non-negotiable terms were highlighted, including (1) Data ownership and sovereignty (data is stored and processed locally but the buyer can gain temporary access to data for training purposes), (2) GDPR compliance, (3) Modularity of the solution to ensure sustainability and the ability to integrate with future innovations, (4) Warranties and liabilities for non-performance, (5) Possibilities of service level agreement, (6) Proof of ethical conduct of the provider as well as potential transparency measures needed to ensure responsible use and disclosure to users.

4.3.6 Results in brief

As a way to make the results of the workshop easily to disseminate, they were summarised in a short report that was published in May 2023 (see Figure 16). The results have been disseminated at the policy workshops and other relevant events, such as the 54th IASA CONFERENCE & 4th ICTMD FORUM. The co-organiser COPEAM also shared this short report within their network and NISV participated in an event organised by COPEAM to also disseminate the results to their members.

Figure 16: Screenshot of the short report outlining the insights from the workshop on AI for audiovisual archives.

5 Policy workshops

The policy workshops were designed to mainly feed into the final policy recommendations (D2.6) and the provotyping exercise (see methodology in section 2.1.2) was also designed around the three challenge areas identified in the pilot policy recommendations. These challenges included: challenges for media companies, challenges for academia and researchers and legal and societal challenges. In this section, there will be a brief outline of the outputs produced as part of these workshops and also the differences that emerged between them due to the different stakeholder groups involved. This leads to a discussion that brings together insights from the industry and policy workshops.

Figure 17: Photos from the second policy workshop in Pisa.

5.1.1 Exploring the tensions of AI regulation

At each workshop the participants were asked to discuss nine 'provocative' statements, three for each of the identified challenge areas (see methodology section 2.1.2 for more detail). Across the three workshops, there were similarities in how the participants reacted to the statements,

but also differences, which will be summarised here. The full discussions of the workshops can be found online for workshop 1 and workshop 2, whereas the short report for workshop 3 will be published in early 2024.

For the **media landscape**, the participants generally expressed concerns with the statement *'Generative AI should increasingly be used to write news'.* Across all three workshops, there was a solution-oriented focus arguing for the need to ensure the technology was only used when appropriate and transparently via disclosure and that humans needed to remain in the loop. In the second workshop (with technologists), there was an emphasis on how this had already and would continue to become more and more common.

Equally, the statement '*Editorial content from media organisations should never be removed by private platforms if not illega*l' produced quite similar responses with a clear focus on the need to ensure healthy debate, which was particularly highlighted by the media professionals (workshop one). However, they also all raised the issues of defining 'illegal' and other forms of problematic content.

For the last statement *'Media organisations should simply use easily accessible AI solutions (like OpenAI) rather than open source in-house development'*, there was clear agreement of the risk this would pose to media independence and in terms of producing AI monopolies. In the last workshop (with legal and media researchers), participants highlighted the need for alternatives to commercial AI infrastructures to ensure media independence, while in the first (with media professionals) highlighted the need for sustainable funding to provide media a chance to build their own systems.

For the **research landscape**, *'API privatisation like on Twitter (now X) will become standard practice and research access will be lost'* was at once seen as a reality by all participants but also highly problematic and there was an insistence on the need for access for researchers and civil society to ensure accountability mechanisms. In the second workshop (technologists), they also highlighted how it is not only the privatisation that is the issue but also the costs associated with it.

The statement *'Training data becomes a problematic trade-off - either you cannot get the data you need due to strict regulation or you have to use illegal/unethical datasets'* was seen more as a fact than controversial by the participants in the second workshop, while the participants in the first workshop found it simplistic and argued for the need for regulation to ensure the production of ethical datasets.

There was a unanimous understanding of the third statement *'AI development/research should never rely on funding from big tech (e.g. Google DNI, Meta)'* as all agreed that this funding would be necessary in the future as well and hard to avoid. However, there were also arguments for the need to better regulate and make such funding transparent as well as providing other alternative funding schemes and better conditions for public-private collaborations.

For the **regulatory and policy landscape**, there was quite a large dissensus among the stakeholder groups around the statement *'Big tech wants strict regulation of AI to eliminate small competitors, such as small media organisations or start-ups'*, where the participants in the

last workshop (legal and media scholars) agreed this was the case, while the participants of the first workshop (media professionals) did not agree with this. The participants of the second workshop (technologists) were more in the middle.

Oppositely, the statement *'The use of AI should always be transparently disclosed - this should be a strict regulatory demand'* was probably the one that produced the most consensus, where all participants found this a vital condition. However, there were concerns about how to find the right balance of disclosure in practice in terms of when full disclosure was needed and how to provide meaningful disclosure for the audiences.

The last statement, *'All applications of AI in media should be high risk in the AI Act'* was seen as too simplistic and that it would be unproductive with such excessive regulation. However, participants in the first workshop (media professionals) did find it important to produce string regulation to preserve media independence, while participants in the second and third workshops (technologists & legal and media scholars) argued for the need to consider the application and impact to ensure a well-functioning regulation, but also that there is a need to include more considerations around media in the AI Act.

5.1.2 Exploring the Wishes for AI Regulation

In the workshops the participants were also asked to provide three wishes for future regulation on AI and Media and as a group rank these wishes in terms of, which would be most important to prioritise. Generally, wishes for **data access, more transparency and better governance** were seen across the workshops.

Interestingly, the participants in the first workshop (media professionals) particularly raised the need for more **sustainable funding**, as being a core condition for the future of AI and media. In the second workshop (technologists), there was a specific focus on facilitating more **open-source models and datasets**, but also this was the only workshop where the question of the **environment** was raised and the need to regulate the effects of AI.

In the last workshop (legal and media scholars), there were also different wishes raised relating particularly to **consumer rights and protection**, particularly in relation to threats of manipulation.

The prioritised wish lists can be found in the short reports but also feed into the policy recommendations discussed in section 6.

5.1.3 Results in brief

The short reports produced based on the workshops can be found online for workshop 1 and workshop 2 (Figure 18), whereas the short report for workshop 3 will be published in early 2024. The findings have been disseminated back to the workshop participants of the first workshop via email and on social media.

Figure 18: Screenshots of the short reports outlining the insights from the policy workshops.

6 Policy discussion

In the initial whitepaper, several policy recommendations were made regarding the need for (1) Regulation that promotes and fosters responsible AI practices in the media sector, rather than attempt to constrain the use, (2) Good practices and policies of 'diversity by design', (3) Increased focus on global AI divides and their consequences, (4) More research and policies addressing potential displacement patterns resulting from AI, (5) Good practices and policies regarding disclosure of AI systems for the media sector, and (6) More transparency in moderation systems and AI fact checking systems.

This section revisits the policy recommendations in light of the findings from the industry and policy workshops. The aim is to highlight specific challenges that were identified across the different workshops, and which clearly highlight where there is a need for action by policymakers to support the responsible development and use of AI in media.

6.1 Six cross-cutting policy needs

This final whitepaper highlights six cross-cutting policy needs which were deemed to be the most urgent and important based on their impact on the media industry and how strongly and often they were brought up by the consulted stakeholders.

6.1.1 Impose and support good practices for transparency

Transparency was a topic that was discussed in all workshops and there was a strong consensus that good transparency practices will be essential to enable the responsible use of AI, but also that there was a need to provide guidance and requirements of transparency. The following concrete policy needs were highlighted:

Policies should support ***internal transparency*** by helping to minimise the current 'intelligibility gap'[6] that often hinders fruitful and interdisciplinary collaborations within organisations that develop and use AI for media.

Policies should support and require ***external transparency*** directed toward end users, for example, via clear regulation on the appropriate level of disclosure required when using AI in the media sector that also takes into account considerations of the user experience and emerging research on what forms and levels of explanations are helpful to users.

Policies should require increased ***provider transparency*** directed at third-party providers of AI solutions for media that will enable the buyers of these solutions to (1) thoroughly assess whether the solutions were developed in ways that align with the organisation's ethical standards and (2) enable them to be transparent towards their end-users about the workings of the AI solution.

[6] See Jones et al. (2022): https://doi.org/10.1080/21670811.2022.2145328

6.1.2 Support research in and of AI solutions

The importance of stimulating the research both in responsible AI solutions and of existing deployed solutions was viewed as uniquely important to secure future solutions of AI and to hold irresponsible actors accountable for problematic usages of AI. The following concrete policy needs were highlighted:

Policies that require the providers of AI solutions to ***give access to data sets and system APIs*** for research and investigative purposes to enable, for example, civil society actors to hold the AI providers accountable for the workings of their systems or to better train models based on scraped data from, for example, large platforms. While this was a clear need expressed across the workshops, there have been concerns about how simply providing access could lead to 'open washing' as it might not provide meaningful insights into the workings of the AI model.

Policies should actively support the production of ***open and shareable datasets***, for the purpose of developing robust and responsible AI systems, as there currently is a lack of ethical open-source datasets and current privacy legislation (e.g., the GDPR) is also making it difficult to share datasets.

Policies that enable ***long-term and sustainable funding schemes*** that support the development of AI solutions that do not pursue short-term goals, but rather take into account long-term societal interest and monitor potential harms.

6.1.3 Stimulating responsible development of AI

In a similar vein, the importance of stimulating the responsible development of AI solutions was a core need amongst the different stakeholders, particularly the organisations who wish to develop AI applications in-house. The following concrete policy needs were highlighted:

Policies should support the ***upskilling and guidelines on responsible practices*** to help organisations build AI solutions in-house and improve AI literacy.

Policies that enable ***sustainable funding schemes*** that support societal interest and have a long-term focus to ensure that the developed tools are not 'quick fixes'. Existing studies have illustrated the issues with short-term funding for producing AI solutions that in fact move into production and produce the hoped value for society[7].

Policies that stimulate ***public-private collaborations*** to enable solutions to be developed with the support of researchers who can provide the latest knowledge on both the technological developments and potential impacts of these technologies and where public organisations can inform the work of future research.

[7] See e.g., De-Lima Santos et al. (2023): https://www.cogitatiopress.com/mediaandcommunication/article/view/6400

6.1.4 Mitigating AI divides

One of the challenges that was highlighted across the workshops was the inequity in access to AI solutions across geographic locations and between small and large organisations, which have also previously been documented[8]. The following concrete policy needs were highlighted:

Policies supporting ***organisational capacity building*** particularly for small organisations (e.g., local news organisations) or organisations that are based in geographies that due to historical developments (e.g., later adoption and access to the internet) are experiencing barriers to engaging with AI.

Policies supporting increased ***local and regional collaborations*** to limit the costs of developing AI by joining forces in areas where languages might be similar and solutions could be transferable with limited extra costs.

Policies supporting the ***development of open datasets and AI solutions for low-resource languages*** to counteract the fact that most AI solutions are developed for and work best in English and enable local actors to develop local AI solutions.

6.1.5 Mitigating power imbalances

Another core challenge faced by many of the organisations engaging in the development of AI is the current power imbalances, where the power of AI is in the hands of the few, which delimits the possibilities of making local AI solutions. The following concrete policy needs were highlighted:

Policies supporting ***alternative funding schemes*** that allow organisations developing AI to be independent of funding from big tech companies to sustain a critical distance and ensure the funding is steered by public interest as opposed to commercial interests.

Policies supporting ***the investment in*** **and** ***development of public and open-source critical AI infrastructures and models*** (e.g., public LLMs, computing infrastructures and hardware) to ensure a diverse landscape of AI developments and put in place dynamics to ensure big tech cannot eliminate the competition by buying, for example, competing start-ups[9].

Policies that enable organisations to have more ***negotiation and bargaining power*** with third-party providers and be able to make demands about, for example, data ownership and allow for tweaking models to suit the local conditions.

[8] See e.g., Beckett (2019): https://blogs.lse.ac.uk/polis/2019/11/18/new-powers-new-responsibilities/
[9] See e.g., Widder et al. (2023): https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4543807

6.1.6 Global and societally focused policies

A last challenge often mentioned is related to the currently scattered legal landscape and the need for not only clearer legislation but also policies that have a global, ecological and societal focus. The following concrete policy needs were highlighted:

Policies that support ***a global and collective approach*** to AI that both help to level the playing field and counteract the growing power imbalances emerging from the US and China and also make the legal landscape more cohesive. However, it will also be important to monitor both the positive and negative effects that emerge from the EU's legislative work on AI[10].

Policies that focus more on ***societal risks***, as opposed to individual risks and rights, to ensure that long-term harms to society (e.g., polarisation) are always considered and accounted for when building AI solutions. Here there are specific societal risks that need to be considered including (1) the potential displacement of workers, (2) the effects of copyright infringements and the risks it produce for intellectual property and (3) the ecological risks of AI on society.

6.2 Future impact of recommendations

The recommendations will be disseminated in their current forms to relevant stakeholder groups via the AI Media Observatory and other media channels as well as via future engagements with stakeholder groups (e.g., via future public consultations).

They will furthermore feed into the final policy recommendations (D2.6) to be delivered in August 2024 and be part of the discussion at the future policy event organised within the consortium (current effort led by KUL), which is aimed to directly feed into ongoing legislative discussions in the EU.

[10] See e.g., Goffi (2021): https://www.aiethicsjournal.org/10-47289-aiej20210716-1

7 Conclusions

This whitepaper concludes and summarises the work under T2.4, which included the initial literature review of the social, economic and political impacts of AI (D2.2) and the workshops that have been conducted to further qualify this work, which has been presented in this whitepaper. Lastly, it included the development of the AI Media Observatory, which launched earlier this year. The whitepaper has summarised the approaches used in the different workshops, provided an overview of the core findings and described the efforts to disseminate the work conducted to varying stakeholder groups.

Concluding, the whitepaper provides insights into some of the concrete AI-related challenges faced by the media sector, but also delivers good practices and suggestions for how to provide better conditions for the responsible use and development of AI for media.

The final discussion of the whitepaper, furthermore, highlights some of the policy gaps that remain dominant to support the responsible development of AI for media.

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