AI4PEOPLE'S 7 AI GLOBAL FRAMEWORKS

AI IS NOT MERELY
ANOTHER UTILITY THAT NEEDS
TO BE REGULATED
ONLY ONCE IT IS MATURE.

IT IS A POWERFUL FORCE
THAT IS RESHAPING
OUR LIVES, OUR INTERACTIONS,
AND OUR ENVIRONMENTS.

Luciano Floridi
2018 Chairman, Scientific Committee AI4People, Professor of Philosophy and Ethics of Information and Director of the Digital Ethics Lab at Oxford University.

AI4PEOPLE'S 7 AI GLOBAL FRAMEWORKS

Following its past work on AI ethics (with the "*AI4People's Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations*") and on AI governance (with the "*AI4People Report on Good AI Governance: 14 Priority Actions, a S.M.A.R.T. Model of Governance, and a Regulatory Toolbox*"), in 2020 AI4People has identified seven strategic sectors (Automotive, Banking & Finance, Energy, Healthcare, Insurance, Legal Service Industry, Media & Technology) for the deployment of ethical AI, appointing 7 different committees to analyze how can trustworthy AI be implemented in these sectors: the *AI4People's 7 AI Global Frameworks* are the result of this effort.

7

MEDIA & TECHNOLOGY

AI in Media & Technology Sector: Opportunities, Risks, Requirements and Recommendationss

Authors

Jo Pierson
Professor of Media, Innovation and Technology at Vrije Universiteit Brussel, Belgium

Stephen Cory Robinson
Senior Lecturer/Assistant Professor in Communication Design at Linköping University, Norrkoping, Sweden

Paula Boddington
Senior Research Fellow, New College of the Humanities London, UK

Patrice Chazerand
Director at DIGITALEUROPE

Aphra Kerr
Professor of Sociology at Maynooth University and Maynooth lead of the ADAPT Centre for Digital Media Technology, Ireland

Stefania Milan
Associate Professor of New Media and Digital Culture, University of Amsterdam

Fons Verbeek
Full Professor in Bio-Imaging and Bio-Informatics, Leiden Insitute of Advanced Computer Science

Cornelia Kutterer
Senior Director, Rule of Law & Responsible Tech, European Government Affairs at Microsoft

Evdoxia Nerantzi
European Government Affairs at Microsoft

Elizabeth Crossick
Head of Government Relations at RELX

ABSTRACT

As AI systems increasingly pervade modern society and lead to manifold and diverse consequences, the development of internationally recognized and industry-specific frameworks focusing on legal and ethical principles is crucial. This report aims at (a) understanding how the 7 Key Requirements for Trustworthy AI impact the Media and Technology sector (MTS) and at (b) putting forward guidelines to ensure compliance with the 7 Key Requirements.

The report identifies four application areas of AI MTS, i.e. automating data capture and processing, automating content generation, automating content mediation and automating communication. Subsequently, the 7 Key Requirements are discussed within each of the four identified themes. Ultimately, recommendations are made to ensure that AI development and adoption in Media and Technology sector is compliant with the 7 Key Requirements. Three clusters of recommendations are proposed: (1) addressing data power and positive obligations, (2) empowerment by design and risk assessments and (3) cooperative responsibility and stakeholder engagements.

Keywords:

Artificial Intelligence, Media and Technology Sector, Trustworthy AI

1. Introduction

AI systems are increasingly pervasive in the individual, organisational, and institutional layers of modern society. Laying the foundations for a "Good AI Society", the multi-stakeholder initiative AI4People initiated the development of internationally recognized and industry-specific frameworks, considering ethics principles.[1] [2] This report examines the large-scale deployment of AI (understood as intelligent and/or autonomous systems) in the Media and Technology sector (MTS). Within this sector, the report lays out four central themes: *automating data capture and processing, automating content generation, automating content mediation*, and *automating communication*. For each of these themes, the report identifies overarching opportunities and risks stemming from the use of AI.

In this report, the Media and Technology Committee – chaired by Jo Pierson – puts forward guidelines to ensure compliance with the 7 Key Requirements for Trustworthy AI. The 7 Key Requirements for Trustworthy AI were originally developed by the European Commission's High-Level Expert Group on Artificial Intelligence[3] and include:

1. Human agency and oversight: Allowing humans to make informed decisions and ensuring human oversight mechanisms;
2. Technical robustness and safety: Ensuring resilient and secure AI systems, a fall back plan, accuracy, reliability and reproducibility;
3. Privacy and data governance: Respecting privacy and ensuring protection, governance, quality of and access to data;
4. Transparency: Ensuring transparent, explainable, and traceable AI models;
5. Diversity, non-discrimination and fairness: Ensuring accessibility to all while diminishing prejudice, discrimination, and unfair bias;
6. Societal and environmental well-being: Ensuring sustainable and environmentally friendly AI systems and considering social and societal impact;
7. Accountability: Ensuring responsibility and accountability of AI systems and their outcomes and adequate redress.

This report is structured in three main sections. After the introduction, Section 2 delineates the Media and Technology sector based on Garnham's framework of mediation and identifies the application areas of AI. This section also provides the

1 All co-authors of this paper constitute the AI4People-Automotive Committee. Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., Vayen, E. (2018). AI4People: An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. Minds and Machines 28, 689–707.
2 Hagendorff, T. (2020). The ethics of AI ethics. An evaluation of guidelines. Minds and Machines, 30: 99–120.
3 HLEG (2019). Ethics guidelines for trustworthy AI. Brussels: Independent High-Level Expert Group on Artificial Intelligence.

definition of AI used throughout this report. Section 3 summarizes the state-of-the-art in European AI governance. Section 4 is divided into two parts. First, the 7 Key Requirements are discussed within the four identified themes of the Media and Technology sector. Second, recommendations are made to ensure that AI development and adoption is compliant with the 7 Key Requirements in the Media and Technology sector.

2.
Conceptual framework for AI in Media and Technology Sector

a. Definition Media and Technology Sector

This section delineates the Media and Technology sector and identifies the application areas of AI. This is no simple matter given the broad field and fast evolution of the Media and Technology sector due to constant innovation. To begin with, MTS involves every form of technologically supported interaction and communication within an ecosystem where they intersect with specific dynamics, i.e. personalization algorithms. This refers to digital media, i.e. digitised traditional content media, and digital platforms which act as socio-technological intermediating architectures and infrastructures enabling and steering interaction and communication between users through collection and circulation of data.[4] These data are collected, processed and used in MTS for many purposes, among which automated personalisation of (recommendations for) content (e.g. news) and advertising (e.g. targeted advertisements). We observe how especially apps are taking an increasingly prominent place in the MTS, as a large amount of digital media communication today happens via apps, while being embedded within a wider ecosystem.

To determine the essential dimensions of the MTS and to situate AI, we adopt the three main components of Garnham's concept of mediation.[5]

- The first dimension includes **human agents** (*human intermediaries*) which refer to people themselves being mediators, e.g. 'gate-keepers' in (citizen) journalism and news production.
- The second dimension includes **content** (*systems of symbolic representation*) in the form of language and symbols, i.e. how humans produce ('encode') text and consume ('decode') text and what happens to the meaning when it is transported and mediated through languages and cultures.
- The third dimension, which includes **technological systems** (*technological tools in media systems*), prevails when it comes to AI applications in the MTS. The dimension refers to the role and meaning of media systems and related technologies.

4 Helberger, N., Pierson, J., & Poell, T. (2018). Governing online platforms: From contested to cooperative responsibility. The information society, 34(1), 1-14.
5 Garnham, N. (2000). Emancipation, the media, and modernity: arguments about the media and social theory. New York: Oxford University Press.

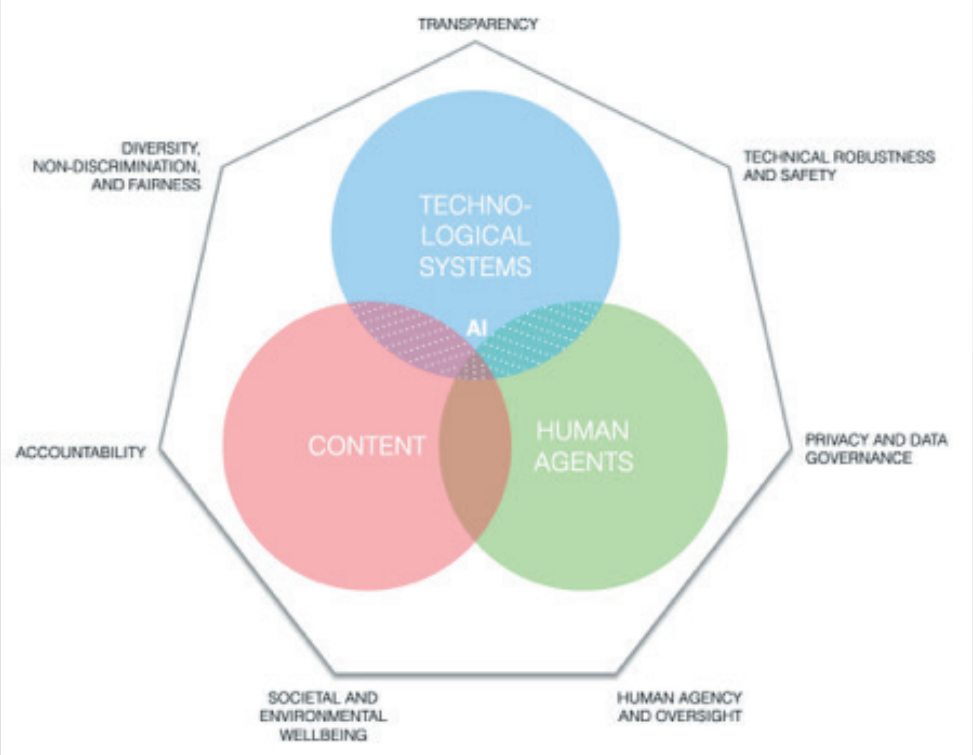

Figure 1: 'Trustworthy AI' heptagon in the Media and Technology sector (own figure)

Within the proposed frame, we identify the following examples of application areas of AI in the MTS which this report examines in the light of the 7 Key Requirements of 'Trustworthy AI':

Human agents +Technological systems	• Automating tools for journalists (Twitter analysis) • Data journalism tools • Newsletter and Customer Relationship Management (CRM) tools • Social media advertising • Etc.
Content + Technological systems	• Computational journalism and robot journalism » Augmenting journalistic practices (high level of AI autonomy): information sharing and gathering, content generation (e.g. sports and financial reporting), revision and distribution • Chatbots, cobots, robots • Deepfakes production and diffusion based on AI • Search engines (algorithm-based technologies) • Smart speakers, voice assistants, new forms of communication (VR/AR) » Speech and face recognition systems » Image analysis software • Marketing automation, programmatic advertising (real-time bidding) and online behavioural advertising • Etc.
Human agents + Content + Technological systems	• Digital media » Journalistic practices (low or medial level of AI autonomy): information sharing and gathering, content generation (e.g. sports reporting), revision and distribution • Digital intermediaries » Digital platforms • General-purpose social media platforms, e.g. Facebook, Twitter • Specific-purpose platforms, e.g. Craigslist, Upwork • News via social media by journalists, citizen journalists and people • Messaging services • Personalisation algorithms (e.g. based on inferential predictive analytics) • Video games • Etc.

Table 1: AI application areas in the MTS

We see the MTS is highly relevant and even exemplary for discussing opportunities, risks and requirements for Trustworthy AI. This is related to several factors. The sector is more directly user-facing compared to other sectors such as energy or automotive sector, with, for example, social media platforms being essential for social interaction and information sharing. This means that people might peg the confidence they should have in digital technology to how much they can trust social media platforms. However, at the same time, the MTS offers the opportunity to provide AI with a promising front office, by realistically framing doom stories and possibly showcasing the advantages of cutting-edge technology. In that way, people can learn the ropes of empowerment in an environment which is more familiar, or less forbidding than anything related to health or mobility. Consequently, the MTS lends itself very well for analysing and discussing Key Requirements for Trustworthy AI in Europe.

b. Definition AI

The AI HLEG (2019) defines Artificial Intelligence as human designed systems which are implemented in the digital or physical environment in the form of software-based systems or possibly hardware devices. Being given a certain goal, AI collects data and assesses the information based on reasoned decision-making in order to suggest relevant actions to achieve the goal. This process is guided by a set of symbolic rules or a numeric model as well as the ability of AI to learn from their environment and previous outputs.[6] The COM (2018) on Artificial Intelligence in Europe emphasizes the intelligent and to a certain extent autonomous behaviour of AI systems.[7]

In fact, AI can be defined as machines that acquire cognitive capabilities such as learning, taking decisions, communicating and interacting based on digital data. Machine learning (ML) and algorithms are two essential features of this process. An algorithm is a software which processes input, i.e. data, based on described rules and selects the relevant information for the user. Moreover, AI is capable of prediction-making, decision-making and problem-solving.[8]

6 HLEG (2019). A Definition of AI: Main Capabilities and Disciplines. Brussels: Independent High-Level Expert Group on Artificial Intelligence.
7 COM (2018). Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions. Artificial Intelligence for Europe. European Commission, 237 final.
8 Just, N., & Latzer, M. (2017). Governance by algorithms: reality construction by algorithmic selection on the Internet. Media, culture & society, 39(2), 238-258.

Examining the level of autonomy of AI, Boucher (2019) differentiates between two waves in AI development.[9] The first wave, called symbolic artificial intelligence, is rather human-centered, which means that even though the AI system performs tasks autonomously, the decision-making process is still guided by humans (*human in the loop*). The system's intelligence stems from the encoding of human expertise and, hence, makes the process and output more comprehensible for humans. In the second wave, called data-driven machine learning, algorithms gain more autonomy and become rather independent from human expertise as they train themselves from data and statistics (from *human-over-the-loop* to *human-out-of-the-loop*). Striking a balance between data-driven and human-centred expertise and assistance is important especially within the scope of the MTS sector as automatisation processes increasingly penetrate journalism and communication activities, a core feature of European democratic processes.

Given that AI systems make recommendations and provide normative solutions, the notion of trust is important to be examined.[10] According to the AI HLEG, trustworthiness should represent a "prerequisite for people and society to develop, deploy and use AI".[11] Hence, the MTS needs to be continuously vigilant that AI systems stay trustworthy even after having been developed, implemented and/or used. People should not be "nudged" or forced to use systems they do not trust or that do not adhere to the 7 Key Requirements. In this light, the next section briefly examines how this has been tackled by the EU to date and what this in particular means for AI applications in the MTS.

9 Boucher, P. (2019). How artificial intelligence works. Brussels: European Parliament Research Service.ù
10 Ferrario, A., Loi, M., & Viganò, E. (2019). In AI We Trust Incrementally: A Multi-layer Model of Trust to Analyze Human-Artificial Intelligence Interactions. Philosophy & Technology. doi: 10.1007/s13347-019-00378-3
11 HLEG (2019). Ethics guidelines for trustworthy AI. Brussels: Independent High-Level Expert Group on Artificial Intelligence.

3. European AI Governance for MTS

The EU's long-standing emphasis on democratic values and the rule of law also shapes its technology governance approach. This is especially relevant for the MTS where automated-decision making technologies and algorithm-dependent processes are becoming increasingly essential. We take a closer look at how AI governance is taking form in the EU, with a focus on MTS-related issues.
For the purpose of this report the High-Level Expert Group on Artificial Intelligence (AI HLEG)[12] was an important initiative, being appointed by the European Commission in June 2018. Since then, the AI HLEG's work has been considered as substantial in defining a "European" governance approach centred around the concepts of "ethical" and "trustworthy" AI. The AI HLEG bases its considerations on three key requirements for AI: legal (i.e. AI should comply with the law); ethical (i.e. AI should fulfil ethical principles); and robust (i.e. AI should be built safely and on the highest quality standards). In July 2020, the AI HLEG published their final Assessment List for Trustworthy Artificial Intelligence (ALTAI) for all relevant stakeholders, particularly those involved in developing and deploying AI systems, to self-assess compliance of specific AI use cases with the 7 Key Requirements for Trustworthy AI.

The European Commission (EC) also incorporated the AI HLEG recommendations in their latest White Paper on Artificial Intelligence.[13] The document sets forth to promote and develop AI based on European values, following a regulatory and investment-based approach. The EC refers to the MTS particularly in the context of protecting fundamental human rights and ensuring legal certainty.[14]

Specifically, the EC highlights the use and potential impact of AI (1) for information selection and content moderation by online intermediaries; (2) in tracing people's daily habits; and (3) in creating information asymmetries by which citizens might be left powerless. The EC is particularly concerned about some potential AI systems' features, such as "opacity ('black box-effect'), complexity, unpredictability and partially autonomous behaviour",[15] in overseeing and enforcing the existing EU legal fundamental rights framework. This may be the reason for the introduction of specific rules for 'high-risk' AI systems in a possible forthcoming EU regulatory framework for

12 European Commission. (2019). High Level Expert Group on Artificial Intelligence. Retrieved on May 20, 2020, from https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence
13 European Commission. (2020). White Paper on Artificial Intelligence. A European approach to excellence and trust. Retrieved on March 20, 2020, from https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf
14 European Commission. (2020). White Paper on Artificial Intelligence. A European approach to excellence and trust. Retrieved on March 20, 2020, from https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf
15 European Commission. (2020). White Paper on Artificial Intelligence. A European approach to excellence and trust. Retrieved on March 20, 2020, from https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf, 12

AI systems. An AI system could be considered as 'high-risk' if "both the sector and the intended use involve significant risks"[16], particularly if safety, consumer rights or fundamental rights are at stake. If an AI system meets the 'high-risk' criteria, compliance with strict requirements and oversight would be mandatory. The EC explicitly sets out the protection of the following EU rights:

- Fundamental rights: Free expression; political freedoms; personal data protection; privacy protection; non-discrimination.
- Legal certainty: Safety; liability; cybersecurity.

These fundamental EU rights are relevant for the MTS since information, communication and mediation activities are all intrinsically linked and somewhat a prerequisite for democracy and the rule of law in the EU. Particularly relevant for the MTS is that the White Paper specifically mentions "online intermediaries" and their responsibility in adequately safeguarding the abovementioned rights as required by EU legislation. Further, the EC underlines that citizens should clearly be aware about their interactions "with an AI system, and not a human being."[17] According to the specific context in which the AI application operates, the EC emphasises "objective, concise and easily understandable" information provision. Next to the White Paper on Artificial Intelligence, the European Commission provides an interpretation of the existing safety and liability framework specifically for Artificial Intelligence, the Internet of Things and robotics.[18] Further, the documents apply in addition to key requirements for protecting data subjects and their data, as set out by the EU data protection legislation (GDPR).

In October 2020, the European Parliament released two legislative initiatives to develop an ethics framework for AI and a civil-oriented liability framework for AI causing damage. The first initiative calls for a legal framework outlining the ethical principles and legal obligations for AI following guiding principles such as human-centric and human-made AI, safety, transparency and accountability, safeguards against bias and discrimination, right to redress, social and environmental responsibility, and respect for privacy and data protection.[19] The second initiative encourages the development of a civil-oriented liability framework, calling for liability of humans

16 European Commission. (2020). White Paper on Artificial Intelligence. A European approach to excellence and trust. Retrieved on March 20, 2020, from https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf, 17

17 European Commission. (2020). White Paper on Artificial Intelligence. A European approach to excellence and trust. Retrieved on March 20, 2020, from https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf, 20

18 European Commission. (2020). Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics. Retrieved on March 20, 2020, from https://eur-lex.europa.eu/legal-content/en/TXT/?qid=1593079180383&uri=CELEX%3A52020DC0064.

19 García del Blanco, I. (2020). Report with recommendations to the Commission on a framework of ethical aspects of artificial intelligence, robotics and related technologies. Retrieved on November 1, 2020, from https://www.europarl.europa.eu/doceo/document/A-9-2020-0186_EN.html.

when operating with high-risk AI activity.[20] Moreover, the European Parliament published a report on intellectual property rights. The report urges to distinguish between AI-assisted human creations and AI-generated creations and encourages an effective intellectual property rights system (IPR) as well as safeguards for the EU's patent system to protect innovative developers.[21]

Considering the indicated European AI policy initiatives, this report contributes to the EU AI governance process by establishing an ethical framework for AI applications in the MTS.

4.
Research questions

The Media and Technology committee consists of 12 members representing different stakeholders of academia and media and technology industry. The goal is to establish a concerted perspective on the meaning and significance of the HLEG 7 Key Requirements for Trustworthy AI in relation to the MTS. For this, the committee held regular gatherings to discuss implications of the requirements in their respective expertise and industry. The multistakeholder procedure for developing the main research questions, consecutive outcomes and the final report was organised as follows:

- Discussing 7 Key Requirements and specific cases based on committee members' expertise and practical experience;
- Asking members to submit case studies containing best and worst practice use cases of AI in the MTS;
- Members submitted their cases;
- Discussion of the submitted cases;
- Members provided additional information, literature and explanation on cases;
- Committee Chair and Advisors scanned all cases for keywords and overlapping issues and best/worst practices;
- Committee Chair and Advisors grouped submitted AI case studies into four categories (themes);
- Committee Chair and Advisors cross-combined four themes with 7 Key Requirements;

20 Voss, A. (2020). Report with recommendations to the Commission on a civil liability regime for artificial intelligence. Retrieved on November 1, 2020, from https://www.europarl.europa.eu/doceo/document/A-9-2020-0178_EN.html.
21 Séjourné, S. (2020). Report on intellectual property rights for the development of artificial intelligence technologies. Retrieved on November 01, 2020, from https://www.europarl.europa.eu/doceo/document/A-9-2020-0176_EN.html.

- Committee Chair and Advisors identified tensions within the four themes;
- Cross-combination of four themes with 7 Key Requirements was first discussed in-depth, after which the view of the committee was further validated and visualised through an online form and interviews among members;
- Members proposed recommendations to ensure compliance with the 7 Key Requirements within the four main themes of the MTS;
- Committee Chair and Advisors identified three prevailing recommendation clusters.

I.
How do the 7 Key Requirements impact the Media and Technology sector?

AI technologies are used in various MTS areas and for various purposes. Given the related manifold and diverse consequences of AI, the analysis and discussion of the 7 Key Requirements for trustworthy AI in MTS is structured according to four main *MTS AI application and use themes*:

a) Automating data capture and processing;
b) Automating content generation;
c) Automating content mediation;
d) Automating communication.

The four themes are mapped in line with the typical (big) data life cycle of data capture, processing and interpretation, preparation and creation, and usage.[22] The four MTS AI application and use themes aim (1) to be mutually inclusive and (2) to largely capture all relevant cases which fall under the MTS in the scope of this report. This is made tangible in the following paragraphs by focusing on concrete examples, when discussing the Key Requirements for each individual theme. This approach allows for a content-based discussion instead of discussing various cases and impacts under each key requirement. This bottom-up, practically oriented methodology also allows to discuss tensions between the 7 Key Requirements.

22 Jagadish, H. V., Gehrke, J., Labrinidis, A., Papakonstantinou, Y., Patel, J. M., Ramakrishnan, R., & Shahabi, C. (2014). Big data and its technical challenges. Communications of the ACM, 57(7), 86-94.

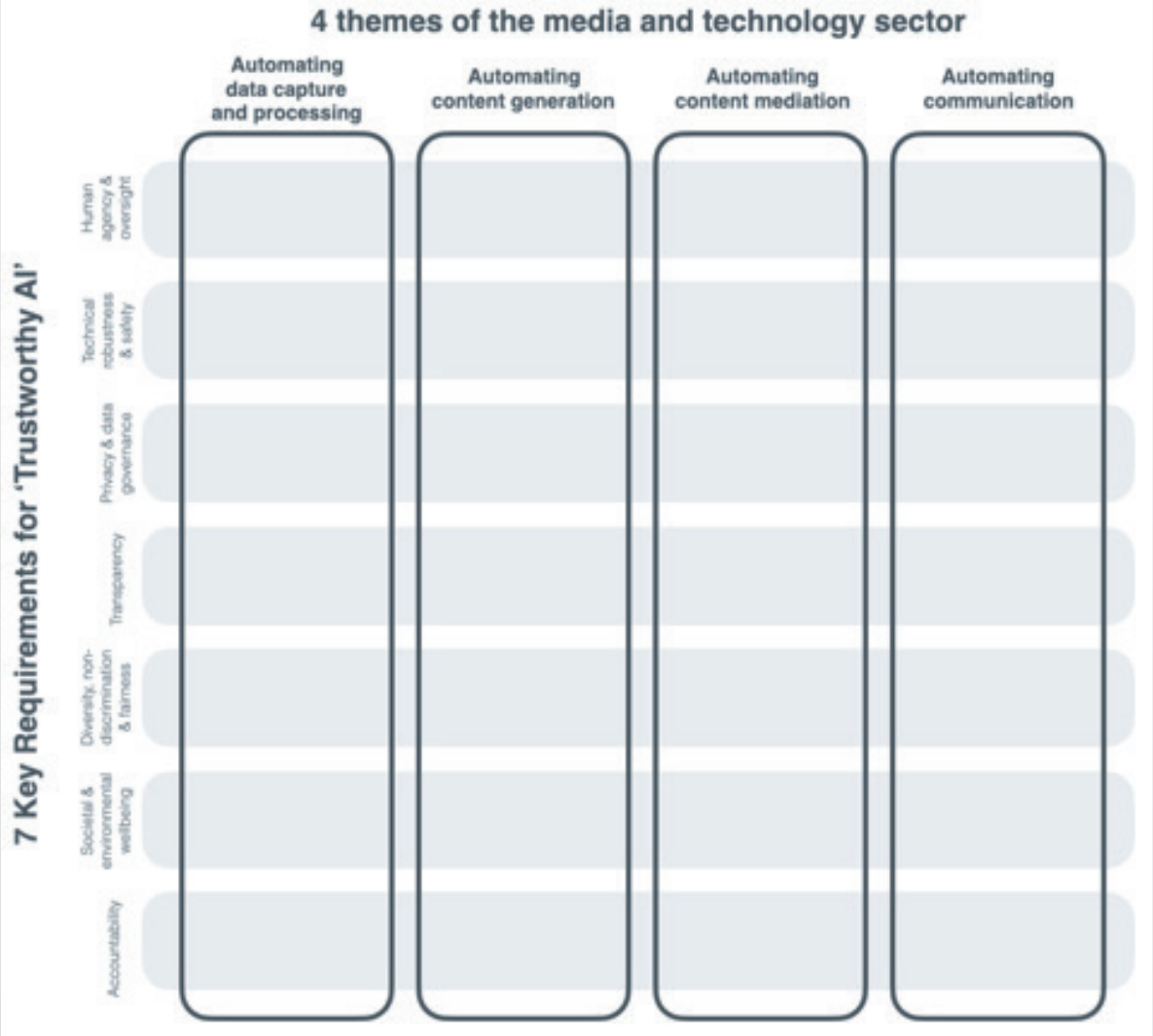

Figure 2: The 7 Key Requirements for "Trustworthy AI" in relation to the four themes of the Media and Technology Sector.

a. Theme 1: Automating data capture and processing

The first theme, **automating data capture and processing**, entails a variety of AI technologies concerned with the systematic capture and processing of data in the MTS. This typically includes data capture and processing by digital media, platforms and websites for reasons of personalisation, profiling, inferential predictive analytics, targeted advertising, etc. However, this type of automation also includes emotional AI in the form of facial and voice recognition systems as well as GPS/location tracking,

contact tracing apps, and VR/AR headsets.[23] Within the 'Trustworthy AI' heptagon (Fig. 1) this theme is concerned with the diminishing aspect of humans as agents against augmenting impacts of technological systems on content.

The principles *human agency and oversight* as well as *privacy and data governance* have a high impact on automated data capture and processing. First, the EU's legislative framework, in particular data and consumer protection standards, protects individuals' fundamental rights to make informed and independent choices. This also applies in relation to automated data capture and processing AI systems: *human agency and oversight* as well as *privacy and data governance* demand that citizens should always be able to decide if and how they choose to use a certain service or be inadvertently tracked by it. In case of using a MTS service, there is the right to decide on what and how much data will be collected, what it would be used for, where it would originate from, and how it would be shared. Given the advertising-driven business model for a significant part of MTS, special attention is needed on how data capturing and processing takes shape with regard to adtech and marketing automation. This is particularly relevant for online behavioural advertising (OBA), where internet users' behavioural data (website visits, clicks, mouse movements, etc.) and metadata (browser type, location, IP address, etc.) are collected and processed to create profiles used to personalise ads and to improve conversion rates. Recent events have shown that especially automated advertising systems of real-time bidding (RTB) have been capturing and processing in possibly prohibited and unethical ways.[24] RTB in ad auctions is the system by which advertisers bid on the possibility of instant targeted advertising to website visitors by using personal data that is collected through tracking and is shared with all bidders. Even advertisers who do not win the auction receive personal data in order to ascertain their interest in the auction. Some advertisers are reported to participate in the auctions merely to enrich their data sets. The targeting is based on profiles of users built via the extensive and persistent tracking of online and possibly offline activities (e.g. via cookies or pixels). The profiles contain categories of users' past behaviour, but also inferred preferences and affinities, being often sensitive categories protected by the GDPR. For example, Google and several data brokers have been accused of violating EU's data protection rules by harvesting and processing people's personal data to build detailed online profiles, including information on sexual orientation, health status and religious beliefs.[25] Additionally, the Norwegian consumer council investigated the data traffic from popular mobile apps. This revealed a number

23 For example immersive mixed reality headset (i.e. Microsoft HoloLens).
24 Information Commissioner's Office (2019). Update report into adtech and real time bidding, 20 June 2019. Retrieved on November 13, 2020, from https://ico.org.uk/media/about-the-ico/documents/2615156/adtech-real-time-bidding-report-201906.pdf
25 Scott, M., Manacourt, V. (2020). Google and data brokers accused of illegally collecting people's data: report. in: POLITICO, 21 September 2020. Retrieved on November 12, 2020, from https://www.politico.eu/article/google-and-data-brokers-accused-of-illegally-collecting-data-report/amp/

of serious privacy infringements and a large amount of illegal data sharing and processing.[26] Academics and data protection practitioners have made proposals to address these type of privacy infringements. Wachter and Mittelstadt suggest introducing the "right to reasonable inferences" by which meaningful control and choice over inferences and profiles are granted to data subjects.[27] This would be particularly relevant for high-risk inferences that are privacy invasive or reputation damaging and have low verifiability in the sense of being predictive or opinion-based.[28] Envisaged as an ex-ante mechanism to provide justification for the reasonability of an inference, disclosing relevance of the data in question, relevance of the inferences drawn, accuracy and statistical reliability of the methods used, these disclosures should be accompanied by an ex-post mechanism enabling inferences to be challenged. This right should close the gap both of explainability and accountability.

In addition, given the importance of being able to collect and process as much as possible (personal) data for optimising personalisation of content and advertising, special attention is needed to safeguard a level playing field in MTS. Although all players in the media and advertising ecosystem are affected by the GDPR, larger players may be more resilient to regulatory interventions. In case smaller competitors drop away, the consolidation of personal data in fewer hands might also increase, and perversely, negatively affect people's rights and freedoms overall. For that reason, various initiatives have been taken, especially in smaller media markets, to pool data and to process them for the benefit of different (competitive) companies at once.[29]

Automated data capture and processing also takes place in other types of applications (in work, health, leisure time) as well as devices (VR/AR headsets[30]). Especially if emotion-reading and -inferring AI systems were to be adapted on a large scale for partially abled people, the option to not use or be subjected to such AI systems should always be available for the person. Moreover, individuals should be aware if they are being systematically tracked, such as by websites, platforms, apps and cameras, and for which purpose, based on an opt-in regime in line with the GDPR. However,

26 ForbrukerRådet. (2020). Out of control: How consumers are exploited by the online advertising industry. Report by the Norwegian Consumer Council.
27 Pop Stefanija, A. (2019, July 7-11). Algorithmic selfie: on the right to assess algorithmic identity and exercise right of access". Madrid, Spain: IAMCR 2019 Conference.
28 Wachter, S., & Mittelstadt, B. (2019). A right to reasonable inferences: Re-thinking data protection law in the age of big data and AI. Colum. Bus. L. Rev., 494.
29 van Zeeland, I., Ranaivoson, H., Hendrickx, J., Pierson, J., Van den Broeck, W. & van der Bank, J. (2019). Salvaging European media diversity while protecting personal data. Brussels, Belgium: SMIT Policy Brief #23, Report for Chair 'Data Protection on the Ground' (Media Sector).
30 As such, Microsoft HoloLens - immersive mixed reality headset - can help people who are blind and with low vision learn who is where in their social environment.
Roach, J. (2020). Using AI, people who are blind are able to find familiar faces in a room. Retrieved on May 2, 2020, from https://news.microsoft.com/innovation-stories/project-tokyo/?utm_source=pre-amp

there can also conditions where certain types of capture are required for overriding purposes (e.g. tracking potential terrorists, in case of a substantiated suspicion). The key principle to safeguard *human agency and oversight* also implies that citizens must be aware if their information or face is being recorded, especially if personal data is tracked. Giving their consent in context of AI applications often involves a weighing of benefits and harms of not opting in, resulting in a rather reluctant agreement than genuine willingness. [31] [32]
Given that tracking of data has become an essential part of many platforms and services in MTS, not opting into the conditions often leads to substantial disadvantages for users. A swift implementation of automated systems for data capture and processing during emergency situations may however lead to lower standards of accuracy and ethical oversight (e.g. in the case of COVID-19 contact tracing apps).[33] In addition, the wide adoption of tracking apps may lead to chilling effects in surveillance of free movement and/or individual behaviour.

Moreover, users should have agency over their data when they visit MTS websites or use apps which record certain information on purpose. Users' given consent needs to be purpose-limited and context-specific. The key principle *transparency* would enable increased *human agency and oversight* if AI uses systematic tracking and tracing features. Transparent and explainable automated data capture AI can build human oversight and trust in technologies. However, a challenge to becoming transparent poses the complex nature of AI itself, such as machine learning. It seems to be a difficult matter to agree on what a 'transparent' explanation of the system must contain and to come to an understanding of what sort of information is enough for all types of individuals subject to the data capture and processing. Linking this to consent, the specific requirements of transparency needed to obtain genuine consent could vary from domain to domain.

The principle of *diversity, non-discrimination and fairness* would impact the MTS in a way that algorithmic data capture and processing AI should not discriminate and/or be biased, and promote a stated conception of fairness. Conceptions of what is a fair

31 Apps for coronavirus contact tracing could trace the spread of disease, to understand infection pathways for risk individuals and communities, and could help in delivering resources to where needed. However, the same technology could be used for wider surveillance of populations and for very punitive consequences in some societies, especially in combination with other technology such as facial recognition to monitor citizens' behaviour. Surveillance measures may outlast the need.
Prasso, S. (2020). Corona Virus surveillance helps, but the programs are hard to stop. Retrieved on April 20, 2020, from https://www.bloomberg.com/news/articles/2020-04-06/coronavirus-surveillance-helps-but-the-programs-are-hard-to-stop
32 Gershgorn, D. (2020). We mapped how the Coronavirus is driving new surveillance programs around the world. Retrieved on April 20, 2020 from https://onezero.medium.com/the-pandemic-is-a-trojan-horse-for-surveillance-programs-around-the-world-887fa6f12ec9
33 Van Zeeland, I. & Pierson, J. (2020). Contact tracing apps and solutionism. Position statement for the Future of Privacy Forum's "Privacy & Pandemics: Responsible Uses of Technology and Health Data During Times of Crisis - An International Tech and Data Conference".

distribution of anything in society differs. Subsequently, what a system developer may deem as 'fair' should be explicitly stated, as well as promoted already in the data capture stage (e.g. do not only process data on young people's news preferences if this is further used for providing recommendations to seniors). In relation to automating data capture, the link between *technical robustness and safety* is paramount because automated data capture systems need to fully represent all potential users and other individuals who will be affected by the systems' outcomes, and not only a certain (biased) part of its dataset. In the MTS, those key requirements are particularly relevant for content-related platforms, like search engines (e.g. Google, Bing), social network sites (e.g. Facebook, Twitter) and video sharing services (e.g YouTube, Vimeo). Equally important are to uphold the key requirements *diversity, non-discrimination and fairness* to avoid the simplistic classification of emotions, which could result in unwanted social sorting. More generally, cultural norms in emotions are not yet fully researched, and psychological research would be suited to inform technology developers about the social norms behind public display of emotions. Furthermore, if the AI system could appropriately adopt cultural norms, it would require consideration if reinforcing certain cultural norms is desirable or not. It is, more fundamentally, worth considering which ethically legitimate purposes could be served by processing data on human emotions at scale.

Technical robustness and safety are also highly important to not market any AI systems for which the impact is not well-researched, and which are not yet fully developed, based on the precautionary principle. The risk is to release it too early. As such, the 'misreading' of emotions could create serious damage to both users and corporate reputation. This is also why the auditing of automated and algorithmic data capture and processing AI systems is key: *accountability* enables a more comprehensive assessment of the purpose, development and deployment of automated data capture and processing AI and additionally enhance *transparency*. Finally, *accountability* can further be improved through multi-stakeholder deliberation, maintenance and oversight, where also citizens and civil society organisations are represented in meaningful way.

Prospectively, automating data capture and processing technology could allow more immersive interactions with surrounding environments by means of Virtual Reality (VR) or Augmented Reality (AR) technologies.[34] Large amounts of data could determine the large-scale nudging by "recommendations", as such for online maps, services or products. Automating data capture and processing could enable targeted consumer choice but likewise decrease *human agency and oversight*. More and better datasets by automating data capture and processing could create powerful nudges based

34 Pollock, D. (2019). Digital billboards open-up advertising to blockchain, artificial intelligence, and cryptocurrency. Retrieved on April 20, 2020, form https://www.forbes.com/sites/darrynpollock/2019/04/18/delving-into-digital-advertising-as-blockchain-cryptocurrency-iot-ar-and-ai-enter-the-frame/

on emotional appeals which one is unable to rationally and cognitively process. This is why *human agency and oversight* are key requirements as long as AI technologies for widespread automating data capture and processing progress. At the same time, prediction based on automating data capture and processing requires considering *accountability*. People should be able to know who is providing the information and what database the prediction is based on, such as models of other people or past behaviour. Several principles come together, as *accountability* towards users and supervisory authorities, to enable *effective independent oversight*, requires *transparency* for data sets to determine whether the captured and processed data indeed supports *diversity, non-discrimination, and fairness*.

b. Theme 2: Automating content generation

The second theme, **automating content generation**, refers to online content produced either fully by automated systems or partly in combination with human agents. Examples of common AI uses in content generation are text-based news reporting apps (based on user preferences)[35] and translation tools, and - in a malign way - disinformation and deepfakes on online platforms. The question remains how much of this type of content is fully automated. The automated element is perhaps more prevalent in the diffusion and amplification of the content rather than the production of it. In addition, an emerging AI application area are creative industries, such as the music and games industry, and creative AI/computing.[36] Considering the 'Trustworthy AI' heptagon (Fig. 1), strong links between technological systems and content become evident. In sum, increasing automation in content generation may provoke an imbalance disfavouring the role of human agents in content generation.

As human agents like journalists play a major role in providing trustworthy information, the aspects of *human oversight*, *accountability*, and *technical robustness* are highly important. AI-driven tools are already employed in journalistic content generation, which relates to the principle of human agency and oversight. In data journalism, for instance, AI helps to identify patterns in large datasets. AI-driven tools can suggest titles and photos, help to find a new topic angle, and produce draft versions of articles. Automated systems assist the journalist in writing the story, but the journalist is still the main storyteller.[37] Thus, a high level of editorial input and human oversight remains; at the same time, publishing articles becomes more efficient. The increasing

35 For example Google News, Apple News, Reddit, Digg, and Flipboard.
36 Amato, G., Behrmann, M., Bimbot, F., Caramiaux, B., Falchi, F., Garcia, A., & Koenitz, H. (2019). AI in the media and creative industries. arXiv preprint arXiv:1905.04175.
37 Willens, M. (2019). Forbes is building more AI tools for its reporters. Retrieved on March 4, 2020, from https://digiday.com/media/forbes-built-a-robot-to-pre-write-articles-for-its-contributors/.

pace and efficiency of news production triggered by automation can put pressure on smaller newsrooms which usually do not dispose large datasets and robust AI-systems.[38] In some news genres, especially those that are rather fact-based, the automation in news generation is higher. For instance, specific natural language processing tools can generate sports articles and financial reporting,[39] while recent projects even involve video reporting[40] Higher automation of content generation can eventually lead to transitions in working opportunities and possible job loss, impacting *societal well-being*.[41] [42] In addition, content produced by AI systems is often not flagged as such to the user and this, hence, links to the importance of *transparency*.

Technical robustness of AI-driven tools in content generation is essential to manage large amounts of data. Data journalism requires robust AI systems to analyse data correctly and to extract "relevant" information. A key issue is as the definition of 'relevant' information today. New forms of news/ information coupled with commercial pressures on the internet are shaping what is presented as 'news' and how it is presented, e.g. clickbait (content whose main purpose is to attract attention and encourage visitors to click on a link to a particular web page). How and what information AI systems extract can, ultimately, shape how the reader understands the information, e.g. positive or negative attitude toward a subject. This can be linked to the principle of *diversity, non-discrimination, and fairness*.

Another example of the significance of technically robust AI systems is in preserving practices of European cultural and architectural heritage. AI systems are capable of digitising high volumes of information which is stored in physical form in archives and museum;[43] for instance, IVOW's "Culturally Sensitive Deep Learning model" can create captions for photos generated by natural language processing algorithms.[44]

38 Helberger, N., Eskens, S. J., van Drunen, M. Z., Bastian, M. B., & Möller, J. E. (2019). Implications of AI-driven tools in the media for freedom of expression.
39 Peiser, J. (2019). The Rise of the Robot Reporter (Published 2019). Retrieved on November 13, 2020, from https://www.nytimes.com/2019/02/05/business/media/artificial-intelligence-journalism-robots.html
40 Chandler, S. (2020). Reuters uses AI to prototype first ever automated video reports. Retrieved on May 10, 2020, from https://www.forbes.com/sites/simonchandler/2020/02/07/reuters-uses-ai-to-prototype-first-ever-automated-video-reports/#7eb6a99f7a2a
41 Lindén, C.-G., Tuulonen, H. (Eds.) (2019). News Automation. The rewards, risks and realities of 'machine journalism'. Frankfurt: WAN-IFRA. Retrieved November 22, 2020, from http://immersiveautomation.com/wp-content/uploads/2019/06/WAN-IFRA_News_Automation-FINAL.pdf. sws
42 Srnicek, N. (2017). Platform Capitalism. Polity Press.
43 Ibaraki, S. (2019). Artificial Intelligence For Good: Preserving Our Cultural Heritage. Retrieved on March 6, 2020, from https://www.forbes.com/sites/cognitiveworld/2019/03/28/artificial-intelligence-for-good-preserving-our-cultural-heritage/#200a70094e96https://www.forbes.com/sites/cognitiveworld/2019/03/28/artificial-intelligence-for-good-preserving-our-cultural-heritage/#200a70094e96
44 IVOW. (2020).An AI and Storytelling Startup. Retrieved on May 10, 2020, from https://www.ivow.ai

Technical robustness is also highly relevant in producing and translating texts. Automated translation risks replication biases (e.g. stereotypes, gender and racial biases) and errors from training datasets. This can affect the principle of *diversity, non-discrimination, and fairness*. Given the linguistic diversity in the European Union, robust automated translation is highly important. It preserves linguistic and cultural plurality. For example the ADAPT research center[45] in Ireland aims to develop data sets and intelligent models that automatically translate online content for native speakers of low-resource languages, and make important content available to people in their language of choice. Projects have focused on developing resources for Irish, Serbian, Basque, and non-European languages including Hindi. Their approach is to employ both, AI and human, rather than fully automated systems.[46]

On social media platforms, deepfakes generated by AI-driven tools grow in popularity. These formats simulate a speech or an action, usually of a public persona (such as politicians, celebrities and actors), where the generated content does not correspond to reality but reveals striking resemblance. This is highly problematic because such false information is often generated without the knowledge of the individuals in question and viewers may be unaware the video was tampered with. This applies to the principle of *human agency* and *societal wellbeing*. It can foster the spread of contentious content like 'fake news', disinformation, hate speech and harmful content. One of the most popular videos that went viral in 2019 portrays Marc Zuckerberg claiming to conquer the world. A recent study shows that 72 percent of people reading an AI-generated news story thought it was credible.[47] Another example is the Chinese app Zao which allows people to seamlessly swap themselves into famous movie scenes.[48] Generating deepfakes and producing disinformation challenges media integrity. In addition, it can severely harm individuals through inappropriate and false representation as well as harassment, for example by malign actions like revenge-porn, affecting not just public figures, but also regular, common people. Forms of redress to tackle these issues seem to be underrepresented or do not guarantee general accessibility to citizens. Hence, it can also be linked to the principle of *diversity, non-discrimination, and fairness*.

45 Transforming Global Content. (2020). Retrieved on May 5, 2020, from https://www.adaptcentre.ie/research/transforming-global-content/

46 For example, they have developed a high-quality Irish-English system called Tapadóir to translate documents into Irish for the Irish government. From 2021 all European documents will also have to be translated into Irish and much of this will be done using these automated systems supplemented by Irish language native speakers and translators.

47 Leibowicz, C. (2019). On AI & Media Integrity: Insights from the Deepfake Detection Challenge. Retrieved on April 20, 2020, from https://www.partnershiponai.org/on-ai-media-integrity-insights-from-the-deepfake-detection-challenge/

48 Kambhampati, S. (2019). Perception won't be reality, once AI can manipulate what we see. Retrieved on April 20, 2020, from https://thehill.com/opinion/cybersecurity/470826-perception-wont-be-reality-once-ai-can-manipulate-what-we-see

In the music sector, deploying AI-driven tools links to the principles of *human agency*, *societal wellbeing*, and *diversity*, *non-discrimination*, and *fairness*. While AI may be beneficial for musicians as it could enhance music education and composition, it also causes concerns about, for instance, replacing human creativity and removing the personal aspect of music creation. Furthermore, the human agency in question may affect *societal wellbeing* in hampering the development of human talent. This can result in reducing opportunities for live music and can produce a cycle by which music is generated and experienced online and remotely, with an impact on human social life.[49] Ultimately, the music sector does not represent an urgent human demand to be complemented by AI systems, to justify replacing human labour.

c. Theme 3: Automating content mediation

The third theme, **automating content mediation**, involves automated filtering systems in the distribution and moderation of online content and advertising. AI technologies in content distribution occur in the form of recommender systems for entertainment and social media content, online news aggregators, and programmatic advertising (including RTB) which provide user-specific and context-conform content. A further set of AI systems is employed to moderate content to detect and tackle contentious content like fake news, mis- and disinformation[50], and harmful content.[51] Linking this to the three components in the 'Trustworthy AI' heptagon (Fig. 1) reveals that *online content* is increasingly processed by technological systems either fully automated or assisting human agents.

Employing automated filtering systems in online content and advertising mediation tasks requires a careful consideration of the principles *diversity*, *non-discrimination*, and *fairness* and *human agency and oversight*. For example, we observe how years after the initial research into discrimination in online employment ads, higher salary positions are still advertised to predominantly (assumed) male users.[52] As AI technologies somehow occupy the new role of traditional gatekeepers and doing agenda-setting in the online sphere, they can also co-determine what people see or not see as well as what content users can generate online. This could affect freedom of expression, media diversity and plurality of voices.[53] In the case of algorithmic content distribution, it can constrain access to a diversity of information and create 'filter bubbles' leading to 'echo

49 Castro, A. (2019). We've been warned about AI and music for over 50 years, but no one's prepared. Retrieved on May 1, 2020, from https://www.theverge.com/2019/4/17/18299563/ai-algorithm-music-law-copyright-human
50 EPRS (2019). Regulating disinformation with artificial intelligence.
EPRS (2019). Automated tackling of disinformation.
51 Lacoma, T. (2020). League of Legends Survey Reveals Nearly Every Player Has Been Harassed. Retrieved on May 1, 2020 from https://screenrant.com/league-legends-survey-harassment-toxicity-riot-games-everyone/
52 Datta, A., Tschantz, M.C., Datta, A. (2015). Automated Experiments on Ad Privacy Settings: A Tale of Opacity, Choice, and Discrimination. arXiv. Retrieved on November 12, 2020 from https://arxiv.org/pdf/1408.6491.pdf

chambers', i.e. personalised content. Especially online platform recommender systems tend to magnify hyperactive users' interests and content, while passive users' interests and content become more invisible.[54] [55] Hence, political microtargeting and opinion formation could become subject to (un)intentional algorithmic manipulation. Furthermore, the datasets as well as developed standards about fairness and non-discrimination for algorithmic filtering systems might contain bias and could leverage discrimination and social sorting. With regards to diversity, non-discrimination, and fairness in the media sector, media recommendation algorithms may worsen the position of smaller countries and their cultural values in media creation.

These issues raise the importance of *human agency and oversight* in online content mediation. Algorithmic filtering systems constrain human agency as users are hampered in choosing which content they receive or if they want to be exposed to algorithmic recommendations at all. Furthermore, human oversight is crucial in detecting and tackling disinformation and harmful content, also in relation to programmatic advertising with advertisers being worried about brand safety with their ads being placed besides contentious content on digital platforms. Take, for instance, the 'infodemic' or large circulation of disinformation and misleading ads during the Covid-19 pandemic (e.g. drinking more water would cure an individual from the disease).[57] [58] [59] Especially in the context of health crises, correct information and reliable sources are particularly important and the lack of it can have severe, even fatal consequences. This case reveals the importance of human oversight in fact-checking the content by professionals, such as health advice. Nevertheless, algorithmic filtering systems are required to master the high volume and fast-paced production of online content. When it comes to content moderation, AI systems are crucial assistants for augmenting human agents in their demanding work of evaluating harmful content such as child abuse, racism, and harassment. This has immediate effects on the physical, mental and *societal well-being* of human content moderators. AI systems can facilitate and support content moderation for humans[60] by flagging harmful content, blurring out areas that are particularly harmful, or engaging in 'visual question answering', i.e. humans' moderations can ask questions to the AI tool about the content without

53 Helberger, N. Karppinen, K., D'Acunto, L. (2018). Exposure diversity as a design principle for recommender systems. In: Information, Communication & Society, 21:2, 191-207.
54 Content Personalisation Network. (2020). Retrieved on May 1, 2020, from https://www.projectcpn.eu
55 Papakyriakopoulos, O., Serrano, J.C.M., Hegelich, S. (2020). Political communication on social media: A tale of hyperactive users and bias in recommender systems. Online Social Networks and Media, 15. https://doi.org/10.1016/j.osnem.2019.100058
56 WFA and platforms make major progress to address harmful content. (2020). Retrieved on November 13, 2020, from https://wfanet.org/knowledge/item/2020/09/23/WFA-and-platforms-make-major-progress-to-address-harmful-content
57 Stolton, S. (2020). EU Rapid Alert System used amid coronavirus disinformation campaign. Retrieved on May 1, 2020, from https://www.euractiv.com/section/digital/news/eu-alert-triggered-after-coronavirus-disinformation-campaign/
58 Mozilla Insights. (2020). When Content Moderation Hurts. Retrieved on May 4, 2020, from https://foundation.mozilla.org/en/blog/when-content-moderation-hurts
59 Ofcom. (2020). Half of UK adults exposed to false claims about coronavirus. Retrieved on May 1, 2020, from https://www.ofcom.org.uk/about-ofcom/latest/features-and-news/half-of-uk-adults-exposed-to-false-claims-about-coronavirus
60 Ofcom. (2019). Use of AI in Online Content Moderation. Cambridge Consultants.

actually seeing it. The actual efficiency of these techniques also depends on the human response time to review the proposed content. Other AI-driven methods to tackle malicious online behaviour are to address the online audience directly in community management. Such AI 'nudging' techniques involve notifications or comments by chatbots that make the user aware that the post contains harmful content, or the technology can cause a short delay in the posting process which could encourage the user to rethink his or her message.[61] AI systems can also provide alternative, more positively expressed content suggestions which still resemble the original message. In both instances, the human agent, namely content moderator or user, takes the ultimate decision.

The complexity of online content challenges the *technical robustness* of AI systems in content moderation. First, AI systems face limitations due to the large variety of content formats, such as text, image, video, and audio which can also appear in a combination of different formats, such as in GIFs, memes, and emojis in combination with text. Advanced content types such as deepfakes and live video streams represent a considerable challenge for human and algorithmic content moderation.[62] Second, content moderation often requires evaluation beyond the content: it must take into account contextual understanding, e.g. societal, cultural, historical, and political aspects, and 'metadata', i.e. surrounding online information such as the number of followers and platform activities. Third, the variety of languages and nuances, e.g. sarcasm, represent challenges. These points create a challenge for both, algorithmic systems as well as human agents. However, users have higher expectations and less tolerance for mistakes in AI rather than human performance.[63] A poor algorithmic performance can have direct impact on the trust of humans in machines. To increase the *technical robustness*, training the AI systems requires large, suitable, and high-quality diverse datasets and constant updating, which is, however, challenged by complex contextual nuances. In particular, smaller newsrooms face difficulties in keeping up with big tech companies. Data and trained engineers for machine learning tend to be underrepresented and/or being insufficiently diverse, e.g. on gender, cultural background. In addition, training AI systems substantially affects *environmental well-being*. An AI training process is highly energy intensive and, hence, incurs considerable environmental costs. This leads to significant sustainability issues. [64] [65]

61 Statt, N. (2020). Twitter tests a warning message that tells users to rethink offensive replies. Retrieved on May 5, 2020, from https://www.theverge.com/2020/5/5/21248201/twitter-reply-warning-harmful-language-revise-tweet-moderation
62 Ofcom. (2019). Use of AI in Online Content Moderation. Cambridge Consultants.
63 Ofcom. (2019). Use of AI in Online Content Moderation. Cambridge Consultants.
64 Hao, K. (2019). Training a single AI model can emit as much carbon as five cars in their lifetimes. Deep learning has a terrible carbon footprint. Retrieved on May 31, 2020, from https://www.technologyreview.com/2019/06/06/239031/training-a-single-ai-model-can-emit-as-much-carbon-as-five-cars-in-their-lifetimes/
65 Matheson, R. (2020). Reducing the carbon footpring of artificial intelligence. MIT system cuts the energy required for training and running neutral networks. Retrieved on May 31, 2020, from http://news.mit.edu/2020/artificial-intelligence-ai-carbon-footprint-0423. bon footpring o

Transparency and *accountability* are two major principles to trace algorithmic decision-making and to counter potential abuse. First, the highly complex architecture of well performing AI moderating tools makes it difficult to analyse and reveal their decisions-making process.[66] Algorithmic standards that are not clearly defined and articulated can result in leaving 'negative' content online and/or removing 'appropriate' content. In this regard, it must be transparent who is to what extent accountable for algorithmic decisions and judgments and to whom it must be disclosed, e.g. general public, certain sectors, human agencies and/or oversight bodies. At the same time, transparency of algorithmic moderating tools towards users can increase the relation of trust between humans and machines. Second, a lack of transparency as well as the algorithmic system itself can be abused in political online campaigns during election periods, such as it often remains unclear who is paying for it, how much is being spent, and how audiences are segmented and targeted, e.g. through ads and chatbots.[67] An abuse of algorithmic filtering systems could further result in censorship which would violate democratic principles. In this regard Tracking Exposed[68] and Algorithms Exposed (ALEX)[69] introduced open-source software as algorithmic auditing methods to tackle the consequences of personalisation algorithms on social media and shopping platforms. Their goal is to empower both advanced users and low-skill users in the data extraction and enhance data literacy.

d. Theme 4: Automating communication

The fourth theme, **automating communication**, includes all forms of interaction and communicative actions and infrastructure enabled by AI. As such, chatbots, smart speakers, voice assistants, automated marketing communication belong to this theme. Everything from AI systems that simulate a proper conversation as well as encoding and decoding conversational messages and data from users falls under this theme. Referring to the 'Trustworthy AI' heptagon (Fig. 1), it is expected that the AI technology in this theme further diminishes aspect of human agents and is in favour of content generated by technological systems.

The most important key requirements for the automated communication theme are *human agency and oversight*, *diversity*, *non-discrimination and fairness* as well as *transparency*. First, transparency would require all AI-empowered communication channels to lay open or make auditable to specialists much of their data infrastructure and thus also how information and output is compiled. *Transparency* would also enable users to understand better how their conversation data is being used and evaluated. Especially in the field of automated marketing and communication, users can fall prey

66 EPRS. (2019). Understanding algorithmic decision-making: Opportunities and challenges.
67 Transparent Referendum Initiative. (2020). Retrieved on May 1, 2020, from http://tref.ie/
68 Tracking Exposed. (2020). Retrieved on May 1, 2020, https://tracking.exposed
69 Algorithms Exposed. (2020). Retrieved on May 4, 2020, https://algorithms.exposed

to misleading messages or biased information, which could be avoided if *transparency* in marketing practices would be made mandatory. More open and transparent automated communication technologies would ultimately give users greater reassurance while open infrastructures and datasets could enable research and generate public interest value. As such, open-source anonymous data may benefit for example the AI-driven development of translation services for low-resource languages. The *transparency* principle is therefore closely linked to maintaining *privacy and data governance*. At the same time, the trade-off by enhanced transparency could result in a backlash against data protection civil society groups advocating for better protection of aggregated datasets. In any case, ensuring multi-stakeholder governance of data and robust privacy measures is also relevant in relation to the data collection and purposes around voice and emotional AI, since it must be clear to users how the information is stored and used in a long-term perspective.

Referring to *diversity, non-discrimination and fairness* in automated communication AI systems, open unbiased datasets would not only favour users' communication experience but are also key to not distort a certain conversation or flow of information between humans and machines. Further, considering the EU's linguistic diversity, automated communication systems can already discriminate or disadvantage certain linguistic minorities. Finally, users should be able to choose whether they want to interact with a chatbot or with a human being, reflected in the principle *human agency and oversight*. This also links to the principle of *accountability* as far as imprecise or wrong information given by a consumer-oriented chatbot, e.g. for a bank, can cause harm or damage.[70] As such, the redressing of automated decisions by chatbots should be considered when discussing accountability in automated communication.

Also, *technical robustness* is relevant in that regard because the key principle would encourage more testing and development of automated communication systems prior to market them as a solution, by the creators to the potential customers. This would allow for a better user experience as well as more trust in automated AI-enabled communications.

The societal and environmental wellbeing principle also demands emphasis on the overall decision whether it is suitable, viable, sensible, considerate, and sustainable to adapt automated communication AI for a certain case. As such, beneficial cases include automated descriptions of visual content by using object recognition technology for the blind and vision-loss community.[71] Likewise, automated communication AI tools develop datasets and intelligent models that automatically translate online content

70 However, this also applies to wrong information from a human employee, and in both cases the bank is liable anyway.
71 Facebook automated alternative text. (2016). Retrieved on April 20, 2020, from https://www.facebook.com/accessibility/videos/1082033931840331/

for native speakers of low-resource languages[72], thereby making important content accessible to linguistically diverse communities. However, deploying already existing data risks replicating biases and errors from training datasets, e.g. stereotypes, gender and racial biases, especially for fully automated translation AI interfaces. The fact that employment opportunities for translators are significantly diminished by automated communication AI technologies threatens the *societal wellbeing* principle, according to which automated communication AI companies were required to mitigate the impact of their technologies on the traditional job sector.

Ultimately, automated communication should enhance human work which is achieved if the communication flows are still subject to human oversight. Delegating the decision to the AI system without human oversight should be avoided.

To improve the validity and emphasise the significance of the 7 Key Requirements in the four themes identified above, the figure below represents the view of the committee. The members assessed the significance of each requirement for Trustworthy AI within the context of the four themes of the MTS, i.e. automating data capture and processing, automating content generation, automating content mediation and automating communication. While the colours dark red and light red indicate a higher significance for the corresponding theme, orange, yellow and white reveal slightly lower significance. The view of the committee reveals that each of the 7 Key Requirements has a high significance throughout the MTS. This figure identifies *human agency and oversight*, *transparency*, and *accountability* as prevailing requirements throughout the MTS.

72 As such, an Irish research centre deploys automated translation and natural language processing AI for low-resource languages to preserve linguistic and cultural plurality in the EU: ADAPT center. Transforming Global Content. (2020). Retrieved on April 20, 2020, from https://www.adaptcentre.ie/research/transforming-global-content/

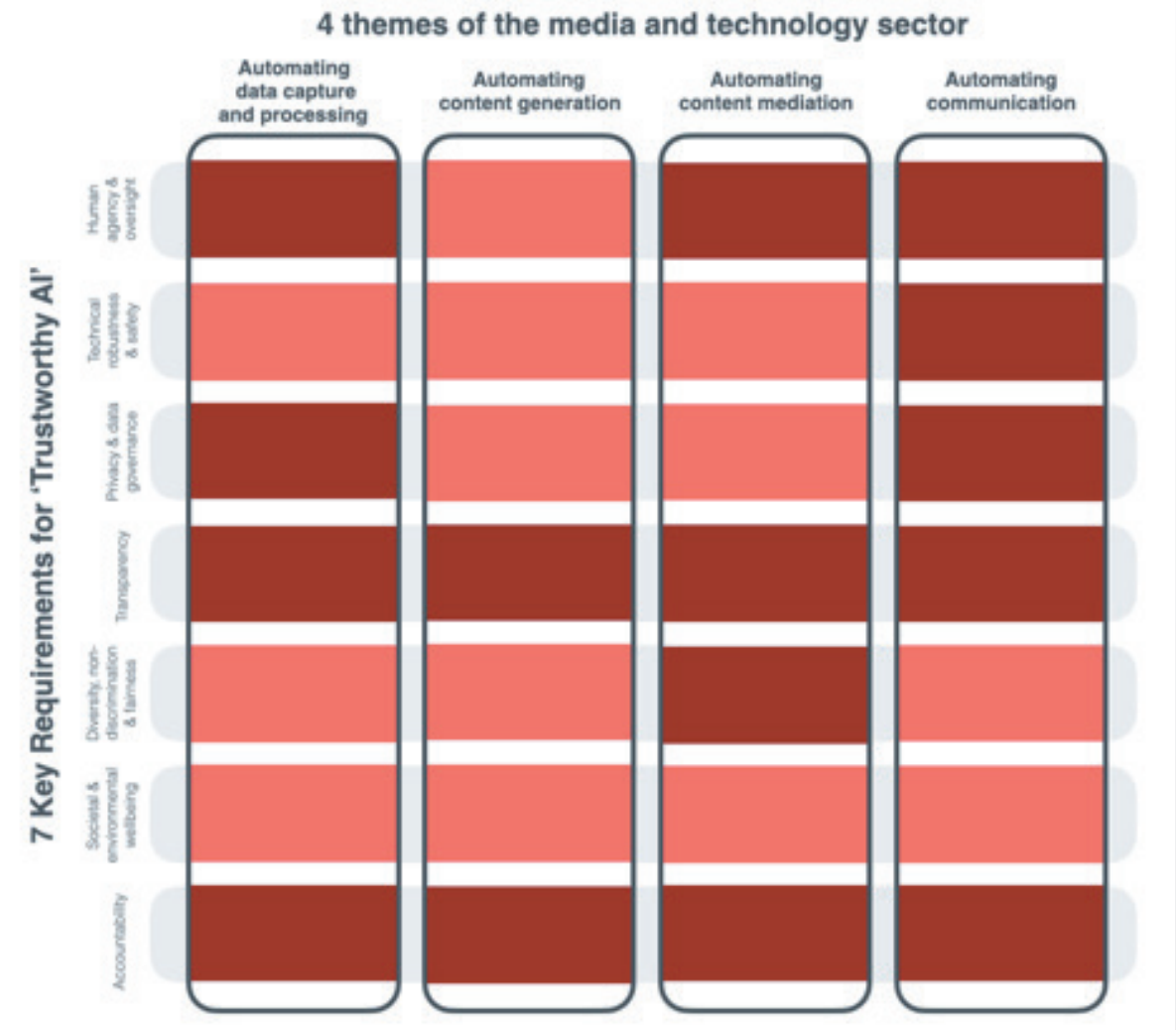

Figure 3: View of the committee on the significance of the 7 Key Requirements for 'Trustworthy AI' in relation to the four themes of the Media and Technology Sector

e. Possible tensions among the 7 Key Requirements for Trustworthy AI

Technical robustness and safety; Diversity, non-discrimination and fairness; Societal and environmental well-being

Possible **tensions** could arise between *technical robustness and safety*; *diversity, non-discrimination and fairness*; *societal and environmental well-being* because the large-scale implementation of AI tools such as holo-lenses for blind people as indicated in theme 1 (*automating data capture and processing*) still requires more extensive research and testing in order to be deployed on a large scale. Given the costs of development, it seems very hard to achieve non-discrimination in early adoption. In order to avoid longer term discrimination, care should be taken to ensure that products being developed

are trialled on diverse groups. In addition, currently, such advanced AI systems are not yet available for most of the blind population, especially for socioeconomically disadvantaged populations.

Technical robustness and safety; Human agency and oversight; Accountability; Transparency

Maximizing efficiency through *technical robustness* of AI systems in content generation can create **tensions** within the key requirements. Technical robust AI systems in data analysis can reduce *human agency and oversight*, *accountability*, and *transparency*. Furthermore, technical robust AI systems can become so efficient and well-advanced that they, ultimately, replace humans in tasks for which it is not necessary or desirable. This is at odds with *diversity, non-discrimination, and fairness and societal wellbeing*.

Transparency; Technical robustness and safety; Privacy and data governance; Societal and environmental wellbeing

Particularly in algorithmic content mediation, **tensions** can appear between *transparency*, *technical robustness*, *privacy and data governance*, and *environmental wellbeing*. A good moderating performance of AI systems might be based on a complex design of AI systems which, eventually, hampers explainability and transparency. Moreover, large datasets including a lot of user information are applied to increase the accuracy and efficiency of algorithmic systems. These data sets contain comprehensive information such as location, consumer preferences, political interests, education and workplace, relationship status, etc., which underlines once again the importance of privacy protection and data governance. Moreover, improving the accuracy of AI operations through well-trained system occurs at the expense of *environmental wellbeing*.

II. What must the Media and Technology sector do to be compliant with the 7 Key Requirements?

This section sets out guidelines for the implementation of AI in the Media and Technology sector. Specifically, it recommends how to adhere to the 7 Key Requirements, the *'Trustworthy AI' heptagon*, within the four identified MTS themes *automating data capture and processing, automating content generation, automating mediation, and automating communication*. Three clusters of recommendations are proposed: addressing data power and positive obligations (oriented mainly at people), empowerment by design and risk assessments (oriented mainly at infrastructure) and cooperative responsibility and stakeholder engagement (oriented mainly at stakeholders).

a) Addressing data power and positive obligations

Key requirements: *Privacy and data governance*; *Human agency and oversight*; *Transparency*

Aforementioned issues of consent are legitimate, particularly regarding the theme of *automating data capture and processing*. Do customers know when their personal data is being collected by AI-enabled systems? This relates to furthering 'data literacy' and 'data agency', which means stimulating awareness, building attitudes, enhancing capabilities and adjusting behaviour among users regarding (personal) data collection, processing and (re)use in the area of digital media and technologies.[73] However, at the same time, it should be avoided to put too much of the burden on the shoulders of relatively powerless citizens. It is first and foremost the task of data controllers to meaningfully explain what is happening with the data. Some users may never be fully digitally literate, yet data controllers also need to make clear to them what is going on. This requires more investigation into explaining well and meaningfully the data capturing, processing and (re)use. This could also mean a positive obligation for AI-driven business to conduct such research on an ongoing basis, as has been suggested in the past by WP29 in their guidelines on valid consent, which were recently updated by the European Data Protection Board.[74]

Positive data obligations also enable citizens to act with agency in the face of data power.[75] Automated data collection by AI systems happens in the background, particularly in remote biometric identification datasets and emotion detection AI. This raises, for example, the question if people should be able to decide if and how their emotions can be tracked, profiled, and re-used for specific purposes in order to avoid potentially harmful effects. For instance, Spotify's data analytics team conducts studies into musical preferences to profile users, not only to present them with better musical advice. One of Spotify's data analytics goals is to target advertising at users depending on the mood they are in, which is a play at manipulation using people's unconscious vulnerabilities.[76]

The meaningful, intentional and informed consent might erode in the presence of AI in the MTS. Users should, therefore, be informed when their volunteered, observed or inferred personal data is being used to train machine learning algorithms, and based on that decide whether to opt in, which could be described as **positive obligations**.

73 Pierson, J. (forthcoming) Media and Communication Studies, Privacy and Public Values: Future Challenges. In: González-Fuster, G., van Brakel, R. and De Hert, P. (eds.) Research Handbook on Privacy and Data Protection Law: Values, Norms and Global Politics, Cheltenham: Edward Elgar Publishing.
74 EDPB (2020). Guidelines 05/2020 on consent under Regulation 2016/679, adopted on May 4, 2020.
75 Kennedy, H., Poell, T. & van Dijck, J. (2015) Data and agency. In: Big Data & Society, July-December, 1-7.
76 See e.g. https://mitpress.mit.edu/books/spotify-teardown

Therefore, **the Committee recommends ensuring clear and strong consent (opt-in) and transparency obligations** for algorithmic training and testing with user data in MTS. This can be operationalised by for example setting-up algorithmic registries, as done by the cities of Amsterdam and Helsinki.[77] On top of providing understandable and easily accessible information on *automating data capture and processing* to users, the latter must also be able to contact a human to provide further information about the aforementioned aspects and users must be guaranteed satisfactory and effective remedies if they have been negatively affected by decisions of AI systems.[78] **The Committee, therefore, recommends responsive redress mechanisms.**

Disclosure of personal data should be a human-consented transaction, not one enticed or (unconsciously) demanded by technology. Data minimisation by design as required by the GDPR should be clearly implemented and enforced in the MTS. Companies should be obliged to undergo regular data reviews to ensure they are not 'casting their nets' farther than necessary. Lastly, data anonymisation or at least pseudonymisation by design should become a key principle. More research investments by the MTS sector are needed in this field. Pseudonymising data is not only favourable for users, but further mitigates risks arising from data breaches, systemic surveillance and cybercrime.

Explainability is a complex, nuanced problem, considering the variety of European citizens. Research and funding for increasing AI transparency and explainability should be pursued and prioritized. This should be combined with (co-)regulatory efforts for establishing more transparency from digital platforms vis-à-vis independent regulators, on matters like internal processes for handling harmful and illegal content through algorithms and AI. In that way we can better address and regulate the behaviour of platform-specific architectural amplifiers of contentious content, e.g. in recommendation engines, search engine features (such as autocomplete), features like 'trending', and other mechanisms that predict what we want to see next. This approach fits in with suggestions being made on ex ante principles-based co-regulatory approaches for addressing online harms as a key operational objective of digital platforms, in a way which is reflective of their reach, their technical architecture, their resources, and the risk such content is likely to pose.[79] Hence, **the Committee recommends strengthening research, process-based (co-)regulation and oversight on AI transparency and explainability**, especially with regards to architectural elements for algorithmic amplification.

77 Moltzau, A. (2020). Algorithm Registries in Amsterdam and Helsinki. Retrieved on November 13, 2020, from https://alexmoltzau.medium.com/algorithm-registries-in-amsterdam-and-helsinki-c1364b70ca6

78 Fanni, R., Steinkogler, V. E., Zampedri, G., & Pierson, J. (2020, September). Active Human Agency in Artificial Intelligence Mediation. In Proceedings of the 6th EAI International Conference on Smart Objects and Technologies for Social Good (pp. 84-89).

79 Vermeulen, M. (2019). Online Content: To Regulate or not to Regulate-Is that the Question?. Vermeulen, Mathias, Online content: to regulate or not-is that the question.

Anticipatory data management policy should be a future priority in EU legislation. Privacy is a moving target, and new categories of personal data will be utilized, collected and created. Therefore, it is imperative that GDPR and the ePrivacy directive update consider emerging sensitive AI-related personal identifiers, whether emotional data or even predicted behaviour AI systems foresee an individual taking.

Individual consent decisions will not prevent all types of societal harms stemming from abusive uses of automated personal data processing. While individuals may consent to the use of information about e.g. their emotions, political affiliation, health or sexual orientation, this may have large-scale effects beyond a single citizen, for which individual choices cannot bear responsibility. Political microtargeting offers an example: individual users may consent to the use of data about their political preferences and emotional states on a platform, but in aggregated form, data on attitudes and emotions linked to political preferences may be used to automatically manipulate voting behaviour of other citizens with potentially major societal effects, as the Cambridge Analytica scandal has illustrated.[80] Prevention of such malignant applications of automated data processing cannot rest on an individual's shoulders and should be addressed with regulation based on an interdisciplinary, multi-stakeholder engagement to uphold public values.

The Committee recommends multi-stakeholder processes for investigating how predictive analytics, sentiment analysis and emotional AI threaten the integrity and autonomy of digital media users, especially in online behavioural advertising and synthetic content production. This approach is in line with the remit of Art. 22 GDPR ('Automated individual decision-making, including profiling').

b) Empowerment by design and risk assessments

Key requirements: *Human agency and oversight*; *Transparency*; *Diversity, non-discrimination and fairness*; *Societal and environmental well-being*; *Technical robustness and safety*

AI technologies being used for activities like profiling, content personalisation and targeted advertising can pose threats to *human agency*, to *transparency*, to *diversity, non-discrimination and fairness*, to *societal well-being*, and to *technical robustness and safety*.

Therefore, it is important that comprehensive solutions are being investigated and developed to address these threats. This fits in with the idea of 'empowerment by design', i.e. building infrastructures and systems in such a way that (organised) citizens have agency to safeguard and strengthen their fundamental rights and the public interest.[81]

80 The Guardian, The Cambridge Analytica Files. Retrieved on November 13, 2020, from https://www.theguardian.com/news/series/cambridge-analytica-files

81 Pierson, J. and Milan, S. (2017) Empowerment by design: Configuring the agency of citizens and activists in digital infrastructure. Presentation at Communication Policy & Technology section for IAMCR Conference 'Transforming Culture, Politics & Communication: New media, new territories, new discourses', 17 July 2017, Cartagena, Colombia.

The targeted advertising industry in MTS is a complex and multi-sided market with a multitude of actors, many of whom intermediaries, such as networks of third parties with tracking technology, intermediary data brokers, and exchanges all competing in the market of RTB and automated auctions.[82] Sensitive information about individuals can be inferred and used, e.g. ethnicity, gender, sexual orientation, religious beliefs, for online behavioural advertising and affinity profiling, i.e. grouping people according to their assumed interests rather than their personal traits. Several scholars and digital rights organisations have made suggestions for empowering consumers in case of illegal or unethical automated capturing and processing of their personal data. Hence **the Committee recommends investigating comprehensive solutions for addressing legal and ethical risks of automated decision-making and profiling, like the 'right to reasonable inferences'**.

Besides issues of profiling in digital marketing, AI is also used in emotion detection and sentiment analysis in MTS. This can have positive uses, but it also bears risks to manipulating human behaviour. These systems could powerfully 'nudge' people into taking certain behavioural actions; used to infer belief and attitude; and incentivise use or concealment of certain emotional expressions. Emotion detection could likewise exacerbate existing biases specifically for vulnerable groups of the society. A set of actions could help to mitigate the risks posed by emotion detection AI. First, users should have to opt-in if any of their data is being used to detect emotions. The consent by users should be mandatory for MTS business, as required by EU data protection law. However, consenting to the data collection does not suffice, as the issue lies with how the results of data analysis are applied, e.g. avoiding that citizens are manipulated at scale. The (dynamic) consent should be reviewed and renewed on a recurring basis with full disclosure over the purpose and scope of the emotion AI implementation areas, and only for sound reasons such as health or safety. Those developing sentiment analysis and emotion detection AI need to be urged to full transparency and public discussion with relevant experts such as sociologists, psychologists, anthropologists, media scholars and psychiatrists. Overall, **the Committee recommends designing an EU-wide, dynamic, and mandatory high-risk assessment scheme** for AI systems detecting sentiments from their users, leading to empowerment by design for citizens and society.

82 Binns, R., Zhao, J., Kleek, M. V., & Shadbolt, N. (2018). Measuring Third-party Tracker Power Across Web and Mobile. ACM Trans. Internet Technol., 18(4), 52:1–52:22. https:// doi.org/10.1145/3176246

More largely, high-risks assessment schemes also need to consider the value of the AI-enabled system(s) against the risks. The latter also refers to minimising unintentional and unexpected harm, and preventing unacceptable harm, which is related to the principle of technical robustness and safety. Simply put, the value of the service enabled/provided must outweigh the risk of the data collected. Thus, a theoretical continuum exists where risks associated with disclosure of personal data and reward or value of received product or service are balanced cognitively.[83] [84] This applies in scenarios where humans interact with AI systems, such as in the first theme. The AI HLEG Assessment List for Trustworthy Artificial Intelligence (ALTAI) already provides a tool to self-assess compliance of specific AI use cases with the 7 Key Requirements for Trustworthy AI. The Committee recommends that the EU-wide, dynamic, and mandatory high-risk assessment scheme should be coherent with the ALTAI, specifically focussing on the potential risks and societal impacts arising in the MTS.[85]

c) Cooperative responsibility and stakeholder engagement

Key requirements: *Accountability*; *Societal and environmental well-being*; *Diversity, non-discrimination, and fairness*; *Transparency*

Many concerns that arise in this sector can only be tackled by means and resources beyond the sector. For instance, social media has enabled targeted harassment of private individuals, which may evade current attempts to regulate, and savvy abusers can readily avoid penalty. *Accountability* issues can arise if companies fail to catch up with technology, if the technology or service provided is ineffective, or if services available only to people with plenty of resources. Yet, effective legal remedies against abusive individuals could be one way of helping to prevent blanket social media policies which may have more draconian effects on freedom of expression. Targeted online harassment of individuals needs to be taken more seriously especially considering the EU fundamental human rights framework and legal obligations. These policies should consider the context since it is vital to communication and hence, a policy that works in one context in social media could be disastrous in another.

ICT blurred borders between media production, consumption and literacy. The most effective way to secure *societal and environmental well-being* should be a shared responsibility between civil society (users), industry (platforms) and governments (education remit). This type of 'cooperative responsibility' requires that digital media platforms, policy makers, users and possible other actors develop a division of labour

83 Robinson, C. (2017). Disclosure of personal data in ecommerce: A cross-national comparison of Estonia and the United States. Telematics and Informatics, 34(2), 569-582.
84 Petronio, S. (2002). Boundaries of privacy: Dialectics of disclosure. Suny Press.
85 European Commission. (2020). Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment. Retrieved on July 17, 2020, from https://ec.europa.eu/digital-single-market/en/news/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment.

on how to manage their responsibility for their role regarding public values.[86] The EU preliminary principle demands that the MTS can only be 'compliant' in presence of an oversight body including a transparent system of compliance, an appeal (redress) and a complaints procedure. Any such system would also have to acknowledge and interface somehow with legacy governance structures in the MTS. Given the legal obligations in the EU, **the Committee recommends setting up an advisory body with all relevant stakeholders involved for feedback and evidence on EU technology policy**.

Providing an outlook, the New European Media Initiative (NEM)[87] is a key European technology platform organisation for the MTS, that – since Framework Programme 7 – is intensely involved in the EU research, thereby driving the future of digital experience. In their "Vision Paper 2030 – Towards a future media ecosystem", NEM aims to unite the MTS with European core values, drivers and goals. Acting ethical, transparent and accountable, being human-centric and sustainable, and encouraging an empowered and critical society are the main ambitions.[88] **In line with the 7 Key Requirements for AI in the MTS, the Committee recommends fostering exchanges and best practices with other institutions, network organisations and multi-stakeholder initiatives**, for example, NEM, Forum on Information & Democracy,[89] Re-Imagine Europe,[90] and the Council of Europe.

Furthermore, impacts on human creativity and societal wellbeing in the media and creative industry could be serious, e.g. in case of music creation by AI. Remedies could, for example, include tax channelled to live music venues/music schools, regulators to remove barriers to live music performance, encouragement of music tuition at all levels of schooling, and open provision of software to educational establishments. This also includes broader support for public service media and creative industry to safeguard creativity and wellbeing. Therefore, **the Committee recommends allocating funding to the most severely impacted creative media industries in the EU**, especially on cultural and public service/information grounds.

The MTS and online intermediaries in particular should be encouraged more to set up an appropriate architecture for empowering users. More standardised methodologies and deliberation fora to facilitate ongoing exchange with the specific user community should be put in place. Also, media production cycles such as designing websites (access, monitoring and dissemination) should involve multiple stakeholders. Likewise, the same stakeholders should be taught the essentials of *diversity*, *non-*

86 Helberger, N., Pierson, J. and Poell, T. (2018) Governing online platforms: from contested to cooperative responsibility. In: The Information Society, 34 (1), 1-14.
87 https://nem-initiative.org
88 Adzic, J., D'Andria, F., Behrmann, M. Boi, S., Castillo, P., Clarke, J., Danet, P-Y., Delaere, S., Fernandez, S. Hrasnica, H. Lippold, S., Matton, M., Menéndez, J.M., De Rosa, S. (2020) NEM Vision 2030: Towards a future media ecosystem, NEM – New European Media, April 2020, 18.
89 https://informationdemocracy.org
90 https://reimagine-europa.eu

discrimination, fairness and human rights, as the ISFE-Council of Europe guidelines to online game developers did.[91] **The Committee recommends incentivizing and developing educational trajectories, guidelines, training, materials and tools** for professional and technical staff (e.g. via online courses or curriculum changes in higher education) to better understand and engage with EU fundamental human rights and the principle of trustworthy human-centered AI.

Example: The PEGI Case

To present the recommendations in applied context, the Pan European Game Information (PEGI) System demonstrates how a voluntary regulatory system can work in practice. The system recommends content and age policies for video games. It is pan-European, interacts with other regional systems in Asia and North America, and sits on top of national governance systems. PEGI is advised by national councils and an expert advisory board made up of representatives (e.g. academics, parent bodies, film rating bodies) from around Europe. These all meet with PEGI staff face to face once a year and online in between the annual meetings. The committee member names are published online, which provides transparency. PEGI and its North American and Asian equivalents are working together to develop an International Age Rating Coalition (IACR).

PEGI is a system that results in information notices on the back of physical boxed media products and now also in the online app and other stores. Publishers fill out a questionnaire and send it to PEGI before a game is released. PEGI can refuse to give a rating to a game, ask for clarifications and it can increase or decrease a rating on appeal. It also takes complaints directly from the general public.

The system works reasonably well in terms of a high level of accountability, but it also has weaknesses. Some online platforms do not participate. How games are rated and on what grounds can be opaque to those outside of the organisation. Further, the system does not have legislative backing and thus cannot take punitive actions against game companies like the game rating systems for example in Germany and the UK do. Thus, while under national legislation it is illegal to sell an over 18 game to a minor in the UK, this is a matter of national legislation. The system is highly focused on protecting children but less on negative impacts or procedures for adults or other vulnerable populations. Further, it is unclear what impact the system has in practice in terms of purchasing behaviour and game playing. PEGI is a co-regulatory system, with a focus on 'educating' consumers but especially protecting children. For a critical discussion see Felini (2015).

91 DG of Human Rights and Legal Affairs. (2008). Human rights guidelines for online game providers. Developed by the Council of Europe in co-operation with the Interactive Software Federation in Europe. Retrieved on June 1, 2020, from https://rm.coe.int/16805a39d3.

Any system that might emerge may want to consider the rather stronger role and stance taken in some countries in relation to the 'traditional media' industries including for example the Press Councils and Press Ombudsman in Ireland which operates to oversee both print and online only news media[92] and the communications regulation bodies like Ofcom in the UK which oversee telecoms and broadcast media.[93] Any governance system might also need to work with established worker unions like the National Union of Journalists, both in terms of training and educating journalists, and in terms of whistleblowing and worker rights. In sum, **the Committee recommends strengthening workers' rights and public interest values in the media as new AI systems evolve and emerge**.

Public information campaigns and initiatives about the functioning and possible risks of new AI initiatives should be promoted. As such, the Media Literacy Initiative[94] involves public, commercial and not for profit/community organisations to counter mis- and disinformation around Covid-19 and is running across online and traditional media channels.[95] Similar information and public communication initiatives are taken at European level including of course Safer Internet Day.[96] **The Committee recommends extending existing publicly supported media, data and AI literacy programmes to include information and public awareness of AI applications, services and impacts**.

5.
Conclusion

Artificial intelligence systems have a substantial impact on various areas of the European media and technology sector (MTS). This report identified four themes of AI applications in the MTS: *automating data capture and processing, automating content generation, automating content mediation, and automating communication*. This report analysed the core opportunities and risks of AI applications within these proposed themes. The 7 Key Requirements for Trustworthy AI developed by the European Commission High-Level Expert Group on AI were at the centre of discussion. The report addresses its recommendations to all stakeholders involved in the development, deployment, use, and governance of AI systems in the MTS.

92 Press Council of Ireland. Office of the Press Ombudsman (2020). Retrieved on June 1, 2020, from https://www.presscouncil.ie/.
93 Ofcom. (2020). TV, radio and on-demand. Retrieved on June 1, 2020, from https://www.ofcom.org.uk/tv-radio-and-on-demand.
94 Be smart media. An Initiative of Media Literacy Ireland. (2020). Members. Retrieved on June 1, 2020, from https://www.bemediasmart.ie/members.
95 Be smart media. An Initiative of Media Literacy Ireland. (2020). About. Retrieved on June 1, 2020, from https://www.bemediasmart.ie/about.
96 Be smart media. An Initiative of Media Literacy Ireland. (2020). Members. Retrieved on June 1, 2020, from https://www.bemediasmart.ie/members.

Recommendation cluster 1: Addressing data power and positive obligations

- Ensuring clear and strong consent (opt-in) and transparency obligations for algorithmic training and testing with user data in MTS.
- Establishing responsive redress mechanisms, so that users can contact humans to provide understandable and easily accessible information on automating data capture and processing, and have satisfactory and effective remedies when negatively affected by AI decisions.
- Strengthening research, process-based (co-)regulation and oversight on AI transparency and explainability, especially with regards architectural elements for algorithmic amplification.
- Ensuring a multi-stakeholder process for investigating how predictive analytics, sentiment analysis and emotional AI threaten the integrity and autonomy of digital media users, especially in online behavioural advertising and synthetic content production.

Recommendation cluster 2: Empowerment by design and risk assessments

- Investigating comprehensive solutions for addressing legal and ethical risks of automated decision-making and profiling, like the "right to reasonable inferences".
- Designing an EU-wide, dynamic, and mandatory high-risk assessment scheme for AI systems detecting sentiments from their users, leading to empowerment by design for citizens and society.
- The EU-wide, dynamic, and mandatory high-risk assessment scheme should be coherent with the ALTAI, specifically focussing on the potential risks and societal impacts arising in the MTS.

Recommendation cluster 3: Cooperative responsibility and stakeholder engagement

- Setting up an advisory body with all relevant stakeholders involved for feedback and evidence on EU technology policy.
- Fostering exchanges and best practices with other institutions, network organisations and multi-stakeholder initiatives.
- Allocating funding to the most severely impacted creative media industries in the EU, especially on cultural and public service/information grounds.
- Incentivizing and developing educational trajectories, guidelines, training, materials and tools for professional and technical staff to better understand and engage with EU fundamental human rights and the principle of trustworthy human-centered AI.
- Facilitating and strengthening workers' rights and public interest values in the media as new AI systems evolve and emerge.

- Extending existing publicly supported media, data and AI literacy programmes to include information and public awareness of AI applications, services and impacts.

The report concludes by emphasising the involvement of public, private, scientific and civil society stakeholders in order to achieve a holistic AI governance framework across the EU.

This report and especially the proposed recommendations aim to tackle concerns that arise due to the proliferation of AI systems in the MTS, thereby ensuring an ethical and sustainable AI implementation throughout this sector. As such, public, private and civil society organisations representing the media and technology sector in Europe, as well as other institutions in Europe are encouraged to consult this report and actively implement the proposed recommendations.

Acknowledgements
We thank Valerie Eveline Steinkogler and Rosanna Fanni for their assistance with the development of the project and for contributing to the draft report, as well as Giulia Zampedri for the support in the finalisation of the report. In addition, we are also grateful to Ana Pop Stefanija and Ine van Zeeland for their revisions and input, in their capacity as PhD researchers for respectively the FWO Research Project DELICIOS (Delegation of Decision-Making to Autonomous Agents in Socio-Technical Systems) (https://coast.uni.lu/delicios) and the VUB Research Chair 'Data Protection on the Ground' (www.dataprotectionontheground.be).