Artificial intelligence in the audiovisual sector

IRIS *Special*

A publication
of the European Audiovisual Observatory

European Audiovisual Observatory

COUNCIL OF EUROPE
CONSEIL DE L'EUROPE

IRIS *Special* 2020-2
Artificial intelligence in the audiovisual sector
European Audiovisual Observatory, Strasbourg 2020
ISBN 978-92-871-8806-9 (print edition)

Director of publication – Susanne Nikoltchev, Executive Director
Editorial supervision – Maja Cappello, Head of Department for legal information
Editorial team – Francisco Javier Cabrera Blázquez, Sophie Valais, Legal Analysts
European Audiovisual Observatory

Authors (in alphabetical order)
Mira Burri, Sarah Eskens, Kelsey Farish, Giancarlo Frosio, Riccardo Guidotti, Atte Jääskeläinen, Andrea Pin, Justina Raižytė

Translation
France Courrèges, Julie Mamou, Marco Polo Sarl, Nathalie Sturlèse, Stefan Pooth, Erwin Rohwer, Sonja Schmidt, Ulrike Welsch

Proofreading
Anthony Mills, Catherine Koleda, Gianna Iacino

Editorial assistant – Sabine Bouajaja
Press and Public Relations – Alison Hindhaugh, alison.hindhaugh@coe.int
European Audiovisual Observatory

Publisher
European Audiovisual Observatory
76, allée de la Robertsau
F-67000 Strasbourg, France
Tél. : +33 (0)3 90 21 60 00
Fax : +33 (0)3 90 21 60 19
iris.obs@coe.int
www.obs.coe.int

Cover layout – ALTRAN, France

Please quote this publication as:
Cappello M. (ed.), *Artificial intelligence in the audiovisual sector*, IRIS S*pecial*, European Audiovisual Observatory, Strasbourg, 2020

Artificial intelligence in the audiovisual sector

Mira Burri, Sarah Eskens, Kelsey Farish, Giancarlo Frosio, Riccardo Guidotti, Atte Jääskeläinen, Andrea Pin, Justina Raižytė

European Audiovisual Observatory

COUNCIL OF EUROPE
CONSEIL DE L'EUROPE

Foreword

According to Elon Musk, the founder of SpaceX and CEO of Tesla, "we should be very careful about artificial intelligence," it may be "our biggest existential threat." Sounds scary, doesn't it? And yet, everybody is talking about it, and more and more companies are using it. It is the future, or so they say.

But what is AI? It is certainly not the dystopian vision served up by Hollywood in so many films, from *2001, Space Odyssey* to *Blade Runner* or *Terminator*.

Not yet, at least.

I am afraid that the answer to the question "What is AI?" is much more prosaic than all that: in the end, AI is mostly computers being computers; software code gulping tons of data and using all this raw information according to predefined instructions.

Of course, the potential is awe-inspiring. Medicine, economics, transportation, energy … you name it. The practical applications are seemingly limitless. However, as with any other technical development, it is not without risks. AI also has a dark side, of course, but probably not as dramatic as Elon Musk would have us believe.

Not yet, at least.

In the audiovisual industry, as in other sectors, the increasing use of artificial intelligence is likely to herald a paradigm shift, as it can transform the entire value chain: from content production, programming and advertising, to consumer expectations and behaviours due to the abundance of offers and devices and the personalisation of content. On the dark side, though, AI can contribute to the proliferation of "fake news", and it raises issues regarding users' right to information, media diversity and pluralism, and data protection, to name but a few.

The European Audiovisual Observatory has decided to take a closer look at these effects by publishing this report, following a workshop we organised in December 2019, to discuss the opportunities and challenges raised by AI in the audiovisual sector, particularly in the journalistic field and in the film sector. More information about the interesting event, including a summary of the discussions and links to the participants' presentations, is available here:

https://www.obs.coe.int/en/web/observatoire/-/workshop-artificial-intelligence-in-the-audiovisual-industry-

This report, conceived, shaped and coordinated by the European Audiovisual Observatory's legal department during the difficult months of lockdown, explores different issues requiring analysis from a regulatory standpoint, and is divided into three parts.

The first is devoted to general umbrella issues and opens with **Chapter 1**, written by IT specialist Riccardo Guidotti (University of Pisa), who explains what AI is and delves in particular into the AI black box problem, that is to say the lack of transparency in how AI systems operate and make decisions, as well as into how explainable AI could be made possible. Two overview chapters follow this technical introduction: in **Chapter 2**, Andrea Pin (University of Padua) offers an explanation of the regulatory problems thrown up by the collection and use of the stuff AI dreams are made of: Big Data. **Chapter 3**, written by Sarah Eskens (University of Amsterdam), provides an overview of issues relating to the

impact of AI on freedom of expression and information, including the legal framework for AI use by the media and the effects on the freedom of expression rights of others.

The second part of the publication presents specific fields of media law and policy where AI may have a profound impact in the future. First comes cultural diversity in the algorithmic age. Mira Burri (University of Lucerne) discusses in **Chapter 4** how, from news personalisation to recommendation algorithms on video on demand services, AI appears to hold the key to our information needs and entertainment wishes, what effect this has, and whether there is a need for regulation. Other tricky legal questions are dealt with by Giancarlo Frosio (Center for International Intellectual Property Studies at the University of Strasbourg) in **Chapter 5**: if machines can "create" works, can they be copyright holders? Or can a person or company be the copyright holder of a work created by a machine? In **Chapter 6**, Justina Raižytė (European Advertising Standards Alliance) explains how AI offers a new world of possibilities for advertisers and, in theory, can be more convenient for the customer, but also raises important privacy issues. In **Chapter 7**, Kelsey Farish (law firm DAC Beachcroft LLP, London) takes us on a tour of personality rights issues - ghost acting, life and *post-mortem* personality rights, and infringement issues (notably deepfakes).

In the third part of the publication, Atte Jääskeläinen (LUT University and London School of Economics and Political Sciences) presents in **Chapter 8** what are, in his view, the main regulatory challenges raised by AI in the audiovisual sector, focusing on possible fields of regulation along with potential risks.

The introductory texts and the concluding remarks, authored by Francisco Javier Cabrera Blázquez, senior legal analyst at the European Audiovisual Observatory, aim at putting all these diverse legal and policy issues in perspective

To this brilliant set of authors go my warmest thanks for having made this report so rich. To our readers, I can just say: enjoy the read!

Strasbourg, December 2020

Maja Cappello
IRIS Coordinator
Head of the Department for Legal Information
European Audiovisual Observatory

The Council of Europe is addressing AI in the Human Rights and other specific contexts.

We invite you to visit https://www.coe.int/en/web/artificial-intelligence/home for more information about the work of the Council of Europe's Ad hoc Committee on Artificial Intelligence (CAHAI).

Table of contents

Figures

Tables

List of abbreviations

AGI	Artificial general intelligence
AI	Artificial intelligence
AI4SG	AI for social good
ANN	Artificial neural networks
AVMSD	Audiovisual Media Services Directive
CAN	Creative Adversarial Network
CDPA	Copyright Designs and Patents Act
CGI	Computer-generated
CJEU	Court of Justice of the European Union
CPU	Traditional processors
DNN	Deep neural networks
DSP	Demand-side-platform
EASA	European Advertising Standards Alliance
ECHR	European Convention on Human Rights
ECtHR	European Court on Human Rights
EDPB	European Data Protection Board
EGE	European Group on Ethics in Science and New Technologies
EPG	Electronic programme guides
EPRS	European Parliamentary Research Service
GAN	Generative adversarial networks
GDPR	General Data Protection Regulation
GPS	General Problem Solver"
GPU	Graphics processor units
IAB	Interactive Advertising Bureau
IP	Intellectual Property
IPR	Intellectual Property Rights
LT	Logic Theorist (first reasoning program)
NLG	Natural language generation
NLP	Natural language processing
PSB	Public service broadcasters

ROI	Return on investment
RTB	Real-time bidding
SRO	Advertising self-regulatory organisations
SSP	Supply-side platform
SVM	Support vector machines
VFX	Visual special effects
WMFH	Work-made-for-hire
XAI	Explainable AI

The black box

As mentioned in the foreword of this publication, AI is both a fascinating and scary development. Its current achievements and its potential are awe-inspiring indeed, and the different contributions of this publication bear witness to the many ways AI can revolutionise (or is already revolutionising) the audiovisual sector. AI machines can write music and lyrics, tell you what to watch and read next, and they can even (virtually) bring the dead back from the grave! Which is maybe why AI, like any other disruptive technological discovery of the past, provokes feelings of fearfulness. This is only natural. It is human nature to fear what one can neither comprehend nor control. That is why the most pressing problem to be solved in the AI regulatory field appears to be the so-called "black box problem". As explained by ***Riccardo Guidotti*** *in his contribution to this publication, "black-box models are tools used by AI to accomplish a task for which either the logic of the decision process is not accessible, or it is accessible but not human-understandable". In other words, it is a machine taking decisions over humans' lives without human oversight or awareness of the reasons behind those decisions. The problem is, according to Guidotti, "not only the lack of transparency but also possible biases inherited by the black boxes from prejudices and artifacts hidden in the training data used by the obscure machine learning models of the AI systems". Indeed, one of the main issues with the use of algorithms today is transparency. If, as they say, an algorithm is like a cooking recipe, the algorithms used by certain companies must be like the Coca-Cola formula, the best kept recipe secret in the world. But it is also true that many people deal with algorithms the way they deal with certain foods: as long as they like what they are eating, they don't really care about the recipe, and in most cases, they actually prefer not to know the ingredients. Anyway, at least in extreme cases, there is plenty to be scared about. Hence the calls from experts to have AI systems whose workings and results are explainable.*

1. Artificial intelligence and explainability

Riccardo Guidotti, University of Pisa

Artificial Intelligence is nowadays one of the most important scientific and technological areas, with a huge socio-economic impact and pervasive adoption in every field of the modern information society. High-profile applications based on artificial intelligence include voice assistants (e.g. Siri and Alexa), autonomous vehicles (e.g. self-driving cars, drones, cleaning robots), medical diagnosis, spam filtering, and image recognition. Artificial intelligence systems achieve their impressive performance in emulating human behaviour mainly through obscure machine learning models. These models are generally based on deep neural networks that hide the logic of their internal processes.

The lack of transparency on how these models make decisions is a key ethical issue and a limitation to their adoption in socially sensitive and safety-critical contexts. Indeed, the problem is not only the lack of transparency but also the possible biases inherited by black-box models from artifacts and preconceptions hidden in the training data. In addition, artificial intelligence can be used for creating synthetic realistic contents. Artificial intelligence is profoundly changing the media and entertainment industries, from personalised recommendations to content creation, underpinned by monetisation.

1.1. What is artificial intelligence?

Artificial intelligence (AI) is the "intelligence" shown by machines or by any technology or software in performing an activity.[1] The term "artificial" is used to distinguish it from the "natural" or "biological" intelligence displayed by humans. AI is a field of research in computer science that tries to understand the heart of intelligence and to produce intelligent machines that reason and respond, simulating human intelligence. The study of AI is historically considered to be the study of "intelligent agents" perceiving an *environment* and performing *actions* that maximise their chances of successfully achieving

[1] Russell, S. and Norvig, P., *Artificial intelligence: a modern approach*. Pearson.

a predefined target.[2] The theories and technologies related to AI have become more and more mature since its birth, and the application fields have been expanding.

1.1.1. A short history of artificial Intelligence

The term "artificial Intelligence" was proposed by John McCarthy during a workshop at Dartmouth University in 1956[3] to distinguish AI from *cybernetics*.[4] The workshop is recognised as the moment in which AI was born.

1.1.1.1. The early years of AI

The early years of AI (1952 – 1969) were full of successes limited to the primitive computers of the time and to the belief that computers were no more than powerful calculators only able to do maths. Allen Newell and Herbert Simon, after "Logic Theorist" (LT), the first reasoning program, designed the "General Problem Solver" (GPS), which, differently from LT, was designed to imitate human problem-solving behaviours. Thanks to GPS, Newell and Simon formulated the famous *physical symbol system hypothesis*, which states that any system exhibiting intelligence must operate by manipulating symbols. In 1958, John McCarthy at MIT defined the high-level programming language Lisp which was used until the 1990s as the dominant AI language.[5] In the 1960s, there were many successful new research directions[6] in AI.

1.1.1.2. The first AI winter

From 1970 to 1980, AI faced the so-called "AI winter" in which the research had a consistent slowdown.[7] The researchers' promises of progress in AI didn't hold up because of the technical limits imposed by the computers used to realise AI programs that were able to solve only "toy" problems.[8] There was not enough processing speed or memory to achieve anything really useful. Logic-based AI systems introduced by McCarthy implementing deduction programs were not able to solve real problems, as they required

[2] Poole, D., Mackworth, A., and Goebel, R., *Computational Intelligence*. Pearson.
[3] Crevier, D. (1993). *AI: the tumultuous history of the search for artificial intelligence*, Basic Books, Inc.
[4] AI and cybernetics are two different but interconnected research fields based on the same principle of binary logic. However, while AI is about creating machines that mimic human intelligence and that can behave like humans, cybernetics is based on a constructivist vision of the world, and it focuses on human-machine interactions: how a system processes information, responds to it and changes accordingly. Thus, the differences between AI and cybernetics are not just semantical but rather conceptual.
[5] Reilly, E. D., *Milestones in computer science and information technology*, Greenwood Publishing Group.
[6] McCorduck, P. and Cfe, C., *Machines who think: A personal inquiry into the history and prospects of artificial intelligence*. CRC Press.
[7] Russell, S. and Norvig, P., *op.cit.*
[8] Crevier, D., *op.cit.*

a huge number of steps to prove very simple theorems.[9] Also, many AI programs practically need enormous amounts of data. Unfortunately, no one in that period had, or was able to collect, a database large enough. The 1970s saw the creation of the successful logic programming language Prolog,[10] a more fruitful approach to logic for AI that permitted tractable computation. Critics of the logical approach started a debate between the need to have machines that think like people versus the need for machines that can solve problems independently from how people do. Consequently, the agencies that funded AI research became disappointed with the lack of progress and cut off almost all funding for research. Simultaneously, research on neural networks was interrupted for almost 10 years after the book *Perceptrons* was published in 1969. A perceptron[11] is a primitive form of neural network, and nowadays, neural networks are a vital part of modern AI systems. An (artificial) neural network is a machine learning model inspired by biological neural networks and composed of artificial neurons. It receives an input, combines the input with the neurons' internal state, and produces output using an activation function. The inputs are data, such as tables, images, or documents, and the output is a classification. A neural network learns how to return output based on a certain input from an annotated training dataset.

1.1.1.3. The boom of AI

In the 1980s "knowledge" became the focus of AI research, and many companies started to adopt forms of AI called "expert systems". An expert system is an algorithm that, by exploiting a given knowledge represented with "if–then" rules, mimes the decision-making ability of a human expert.[12] An expert system is formed by a "knowledge base" that represents facts and rules, and by an "inference engine" that applies the rules to the known facts to deduce new facts. Expert systems were among the first successful AI software programmes adopted in business companies. Researchers realised that the power of expert systems came from the knowledge they contained and that "... intelligence might be based on the ability to use large amounts of diverse knowledge in different ways".[13] This injection of confidence in AI pushed lenders to invest again in AI research. In parallel, there was a 'revival' of neural networks. Hinton and Rumelhart made popular "backpropagation",[14] an effective method for training neural networks. This training method made effective the usage of artificial neural networks (ANNs), machine learning systems inspired by the biological neural networks of human brains.[15] ANNs "learn" from examples contained in a dataset of knowledge how to assess a task, but without requiring existing task-specific rules. For instance, they can recognise if an image

[9] McCorduck, P. and Cfe, C., *op.cit.*
[10] Crevier, D., *op.cit.*
[11] Tan, P.-N. et al., *Introduction to data mining*. Pearson Education India.
[12] Jackson, P., *Introduction to expert systems*. Addison-Wesley Longman Publishing Co., Inc.
[13] McCorduck, P. and Cfe, C., *op.cit.*
[14] Rumelhart, D., Hinton, G. & Williams, R, "Learning representations by back-propagating errors", *Nature* 323, 533–536 (1986), https://doi.org/10.1038/323533a0.
[15] Tan, P.-N. et al., *op.cit.*

contains a pedestrian or a car by learning from images labelled with their content and without any prior knowledge of the objects studied.

1.1.1.4. The second winter of AI

Investments in research in AI went up and down during "the second winter of AI" (1987 – 1993). Desktop computers not requiring any form of AI from Apple and IBM were slowly augmenting power and speed, and in 1987 they became more effective than the expensive Lisp and Prolog machines. However, despite criticisms from some investors and governments, AI kept pushing forward. In these years the concept of "intelligent agents" was finalised thanks to economists' definition of a "rational agent". An intelligent agent is a system that takes actions that maximise the chances of success with respect to a predefined goal. In addition, AI became a "rigorous" scientific discipline because AI researchers increased the usage of sophisticated mathematical tools for developing AI programs. For instance, probability and decision theory were brought into AI by Judea Pearl's book.[16] However, despite these evident steps forward, AI as a theoretical academic research field received little attention because algorithms originally developed for AI began to be exploited as parts of larger systems in the technology industry, such as data mining, medical diagnosis, speech recognition, search engines, banking software, industrial robotics, etc.

1.1.1.5. Big data, deep learning and AI

Despite the aforementioned advances, the real turning point was mostly due to the enormous increase in the power of computers by the 1990s. Very famous examples of successes due to these technological advancements in AI are Deep Blue[17] and Watson.[18] The IBM Deep Blue was the first chess-playing AI system to win against a world chess champion, Garry Kasparov,[19] in 1997. In 2011, IBM's question-answering system Watson beat the champions of "Jeopardy!", a TV quiz show, by a significant margin. In addition, starting from 2010, on top of the advances in computer power, AI entered a new era thanks to technological progress in terms of storage capability, the ease of accessing big data, and advanced machine learning techniques like deep neural networks.

- "Big data" identifies a huge collection of data that cannot be stored, managed and processed using conventional software. The era of big data originated from two main flows:
    - *(i)* the industrial sectors storing information ranging from the log of activities to purchases of clients;

[16] Pearl, J. (1988), *Probabilistic reasoning in intelligent systems*.
[17] Available at https://www.ibm.com/ibm/history/ibm100/us/en/icons/deepblue/.
[18] Available at https://www.ibm.com/ibm/history/ibm100/us/en/icons/watson/.
[19] Russell, S. and Norvig, P., *op.cit.*

  - *(ii)* the widespread collection by smartphones and mobile devices of personal information of users from various sources such as online posts on social networks, emails, mobility traces, health records, etc.
- Deep neural networks (DNNs) are models realised as an evolution of traditional ANN by composition of many processing layers (deep). Deep learning is the branch of machine learning that studies DNNs. DNNs can be applied for assessing tasks that are much more complex than those that can be solved with ANN, such as image recognition, speech recognition, natural language processing, etc. However, the recent popularity of DNNs is mainly due to novel computer graphics processing units (GPU). GPUs allow a marked acceleration in the learning process of DNNs and their efficient execution, as compared to traditional processors (CPUs). Unfortunately, as discussed in the next chapter, DNNs suffer from a profound drawback: the lack of interpretability.[20]

1.1.2. Different approaches for artificial intelligence

Historically four different notions in terms of dealing with AI have been recognised,[21] with respect to two dimensions:

1. observing the artificial way of *thinking* versus observing artificial *behaviour*;
2. modelling *humans* or modelling an *ideal* standard (called rationality).

Hence, the four different notions are "thinking humanly", "thinking rationally", "acting humanly", and "acting rationally". Different researchers with different approaches have aligned with these four notions. As a consequence, research on AI has been divided into subfields that often fail to communicate with each other. These sub-fields can be differentiated with respect to philosophical variances and notions, the objectives of reaching particular goals, and the usage of certain technical methods.

Concerning the philosophical differences, we can recognise the human-centred approach and the rationalist approach. The human-centred approach suggests that AI should simulate natural intelligence. On the other hand, a rationalist approach involves a combination of mathematics and engineering, and suggests that human biology is irrelevant. Under this vision, either AI can be designed through simple, elegant principles such as logic or optimisation, or it requires solving many distinct and complex problems. Regarding the different challenges in AI, the general problem of creating an intelligence has been divided into sub-problems that consist of specific capabilities that an intelligent system should have. The principal sub-problems are machine learning, planning, reasoning, problem-solving, representing knowledge, perception, robotics, natural

[20] Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., and Pedreschi, D. (2018), "A survey of methods for explaining black box models", *ACM computing surveys (CSUR)*, 51(5):1–42.
[21] Russell, S. and Norvig, P., *op.cit.*

language processing, and social intelligence.[22] Each sub-problem corresponds to a subfield of study on computer science.

Finally, in the history of AI, we recognise a broad set of methods belonging to three different categories:

1. "Cybernetics" explores the connections between neurobiology and information theory and tries to design machines that use electronic networks to display rudimentary intelligence;[23]
2. "Symbolic AI" is based on the assumption that through the manipulation of symbols, it is possible to model many aspects of the human intelligence.[24]
3. "Statistical learning-based AI" relies on strong mathematical approaches.

Well-known methods used in AI systems are: "logic", used for knowledge representation and for problem-solving; "probabilistic methods", used in reasoning, planning, learning, perception, and robotics; "search and optimisation methods", used for planning and for robotics; "machine learning methods" such as decision tree classifiers, support vector machines; and "deep neural networks", used to address almost every challenge. A drawback of some of these powerful statistical learning methods is that they are not interpretable, that is to say a human cannot understand the logic of these systems in making decisions.

1.1.3. Applications of artificial intelligence

What can AI do today, and in which fields it is applied? A complete answer to this question is not easy, as nowadays AI is applied to a plethora of areas and tasks. In the following section, we briefly report some AI applications that may be remarkable or interesting for readers of this publication.

- **Robotic vehicles.** Self-driving autonomous vehicles have been made possible thanks to the advancements in AI. Distinct AI components incorporated in systems such as collision prevention, lane changing, braking, etc. contribute to the overall functioning of autonomous cars. AI companies involved with robotic vehicles are Tesla, Google, and Apple.[25]
- **Healthcare.** AI in healthcare is used to support doctors. For instance, AI systems can be used for disease diagnosis, analysing the relationship between treatments and outcomes, discovering issues related to dosage, supporting surgeons during

[22] Poole, D., Mackworth, A., and Goebel, R., *op.cit.*
[23] Weiner, N., *Cybernetics (or control and communication in the animal and the machine)*, Cambridge (Massachusetts).
[24] Haugeland, J., *Artificial intelligence: the very idea.*
[25] CBInsights, *33 Corporations working on autonomous vehicles.* Retrieved on 16 March 2017.

operations, supporting radiologists in interpreting images, and creating new drugs.[26]

- **Marketing, economics and finance.** Companies and financial institutions were the first adopters of AI systems for market analysis, churn prediction, price forecasting, stocks supervision, portfolio management, algorithmic trading, etc. Also, AI is effectively used to reduce fraud and financial crimes.
- **Media.** The analysis of media content such as TV programmes, advertisements, movies and videos can be demanded of specific AI applications. The typical usage refers to face or object recognition, automatic subtitling, recognition of relevant scenes, and to summarise content. Media analysis based on AI allows the creation of descriptive keywords for a media item in order to simplify media searches. Another application consists of monitoring the suitability of media content or automatically detecting appropriate/inappropriate logos and products related to advertisements.
- **News and publishing.** Nowadays, many companies are using AI techniques to generate news and reports automatically. Through AI, companies are also capable of writing text. An example of an application is the generation of personalised recaps for sports events.[27] Another application turns structured data into comments in natural language.
- **Music.** AI has allowed, to an extent, the emulation of human-like composition, and it helps humans play music or sing.[28] Computer accompaniment technologies are able to listen and follow a human performer so that they can play in synchrony. Interactive composition technologies allow AI to respond with a music composition to the performance of a live musician. Finally, projects like Google Magenta, Sony Flow Machines, or IBM Watson Beat are able to compose music in any style after analysing large databases of songs. Other AI applications for music also cover music marketing and listening.
- **Deepfakes.** Deepfakes are synthetic media contents created through AI techniques which appear real to humans.[29] Generally, they are images or videos in which a person is replaced with someone else with deep learning methods. The main methods used to create deepfakes involve the training of generative approaches such as generative adversarial networks (GAN)[30] or autoencoders.[31] Even though deepfakes can be used for comedic purposes, they are better known as hoaxes, "fake news", celebrity pornographic videos, and financial frauds. Consequently, both governments and industries work to develop AI tools to detect and limit

[26] Coiera, E., *Guide to Medical Informatics, the Internet and Telemedicine*, Chapman & Hall, Ltd., GBR, 1st edition.
[27] Available at https://www.barrons.com/articles/big-data-and-yahoos-quest-for-mass-personalization-1377938511.
[28] Roads, C., "Research in music and artificial intelligence", *ACM Computing Surveys (CSUR)*, 17(2):163–190.
[29] Kietzmann, J., Lee, L. W., McCarthy, I. P., and Kietzmann, T. C., "Deepfakes: Trick or treat?", Business *Horizons*, 63(2):135–146.
[30] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y., "Generative adversarial nets", in *Advances in neural information processing systems*, pages 2672–2680.
[31] Makhzani, A., Shlens, J., Jaitly, N., Goodfellow, I., and Frey, B., "Adversarial autoencoders", *arXiv preprint arXiv*:1511.05644.

them. The reason is that, in the foreseeable future, AI will probably be able not only to create realistic images and videos, but full media content such as movies, TV series, and TV programmes like reality shows and quizzes.

1.2. What is explainable artificial intelligence?

Nowadays, AI systems are not only able to simulate the information process of human thinking and learning, but can also exceed human intelligence in resolving some tasks. This is possible because artificial intelligence is not human intelligence and, due to the widespread adoption of complex methods such as deep learning, AI does not act like human intelligence, or at least acts using a decision process that is not always human-understandable. Indeed, the last decade has witnessed the rise of what Frank Pasquale calls the "black-box society",[32] where AI systems adopt obscure decision-making models to carry on their decision processes. This choice is driven by high performance in terms of accuracy[33] achieved by these black-box models. Examples include neural networks and deep neural networks, support vector machines (SVMs), and ensemble classifiers, but also compositions of expert systems, data mining, and hard-coded software that "hide" the logic of their internal decision processes from humans.[34] Thus, black-box models are tools used by AI to accomplish a task for which either the logic of the decision process is not accessible, or it is accessible but not human-understandable.

The lack of explanations of how these black-box models make decisions poses a problem for their adoption in safety-critical contexts and socially sensitive domains such as healthcare and law. The problem is not only the lack of transparency but also possible biases inherited by the black-boxes from prejudices and artifacts hidden in the training data used by the obscure machine learning models of the AI systems. Indeed, machine learning algorithms build models after a learning phase that is made possible by big data coming from logs of business processes and from the digital traces that people leave behind while performing daily activities (e.g. purchases, movements, posts in social networks, etc.). This huge amount of data might contain human biases and prejudices. Hence, decision models whose learning is drawn from them may inherit such biases, possibly leading to unfair and wrong decisions. Consequently, the research in explainable AI (XAI) has recently garnered much attention.[35]

---

[32] Pasquale, F., *The black box society*, Harvard University Press.
[33] Tan, P.-N. et al., *op.cit.*
[34] The interested reader can find details about neural networks, SVMs, and ensemble classifiers in Tan, P.-N. et al., *op.cit.*
[35] Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., and Pedreschi, D. (2018), *op.cit.* Miller, T., "Explanation in artificial intelligence: Insights from the social sciences", *Artificial Intelligence*, 267:1–38. Adadi, A. and Berrada, M., *Peeking inside the black-box: A survey on explainable artificial intelligence (xai)*. IEEE Access, 6:52138–52160.

Moreover, the General Data Protection Regulation (GDPR)[36] introduces a right of explanation for all individuals, to obtain "meaningful explanations of the logic involved" when automated decision making takes place. Despite conflicting opinions among legal scholars regarding the real scope of these clauses,[37] there is a common agreement that the implementation of such a principle is imperative and that it represents today a huge open scientific challenge.

XAI is at the heart of a responsible science across multiple industry sectors and scientific disciplines. How can companies trust their AI products without understanding the rationale of their machine learning components? In turn, how can users trust AI services? It will be impossible to increase the trust of people in AI without explaining the rationale followed by obscure models.

1.2.1. Motivations for XAI

Besides theoretical, ethical, and legal motivations behind the need for explainable AI, there are real cases in which discrimination or errors could have been avoided if the AI had not been obscure. Having access to the reasons for AI decisions is particularly crucial in safety-critical AI systems like self-driving cars and medicine, where a possible wrong decision could even lead to the death of people. For example, in the case of a self-driving Uber car that knocked down and killed a pedestrian in Tempe, Arizona, in 2018, the use of interpretable models would have helped Uber understand the reasons behind the decision, and manage their responsibilities.

Another inherent risk of black-box components used by AI systems is the possibility of making wrong decisions learned from spurious correlations or artifacts in the training data. For instance, Ribeiro et al.[38] show that a classifier trained to recognise wolves and husky dogs was basing its predictions regarding distinguishing a wolf solely on the presence of snow in the background. The AI made this choice because all the training images with wolves had snow in the background. In another example, in 2016, the AI software used by Amazon to determine the areas of the United States to which Amazon would offer free same-day delivery, unintentionally restricted minority

[36] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). Available at https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1532348683434&uri=CELEX:02016R0679-20160504.

[37] Malgieri, G. and Comandé, G., "Why a right to legibility of automated decision-making exists in the General Data Protection Regulation", *International Data Privacy Law*, 7(4):243–265. Goodman, B. and Flaxman, S., "EU regulations on algorithmic decisionmaking and a 'right to explanation'", in *ICML workshop on human interpretability in machine learning (WHI 2016)*, New York, NY. http://arxiv.org/abs/1606.08813 v1. Wachter, S., Mittelstadt, B., and Floridi, L., "Why a right to explanation of automated decision-making does not exist in the general data protection regulation", *International Data Privacy Law*, 7(2):76–99.

[38] Ribeiro, M. T., Singh, S., and Guestrin, C. (2016), "Why should I trust you?: Explaining the predictions of any classifier", in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1135–1144. ACM.

neighbourhoods from participating in the programme (often when every surrounding neighbourhood was allowed)[39]. More recently, the journalists of ProPublica showed that the COMPAS score, a predictive model for the "risk of crime recidivism" (proprietary secret of Northpointe), has a strong ethnic bias. Indeed, according to this score, a black person who did not re-offend was classified as "high risk" twice as often as whites who did not re-offend. On the other hand, white repeat offenders were classified as "low risk" twice as often as black repeat offenders.[40]

1.2.2. The dimensions of interpretability

To "interpret" means to give or provide meaning or to explain and present in understandable terms certain concepts.[41] Therefore, AI "interpretability" is defined as the ability to "explain" or to provide meaning with regard to decisions, in terms understandable to a human.[42] This definition assumes that the concepts composing an explanation are self-contained and do not need further explanations. Basically, an explanation is an "interface" between a human and an AI, and it is at the same time both human-understandable and an accurate proxy of the AI. We can identify a set of "dimensions" to analyse AI systems' interpretability that, in turn, reflect on existing different types of explanations.[43]

1.2.2.1. Black-box explanation vs. explanation by design

We distinguish between black-box explanation and explanation by design. In the first case, the idea is to couple an AI with a black-box model with an explanation method able to interpret the black-box decisions. In the second case, the strategy is to substitute the obscure model with a transparent model in which the decision process is accessible by design. More in detail, the black-box explanation idea is to maintain the high performance of the obscure model used by the AI and to use a technique from XAI to retrieve the explanations.[44] This kind of approach is the most frequent one nowadays in the XAI research field. On the other hand, the "explanation by design" consists of directly designing a transparent model which is interpretable, and of substituting the black-box

[39] Available at http://www.techinsider.io/how-algorithms-can-be-racist-2016-4.
[40] Available at http://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.
[41] Available at https://www.merriam-webster.com/.
[42] Doshi-Velez, F. and Kim, B., "Towards a rigorous science of interpretable machine learning", *arXiv preprint arXiv*:1702.08608. Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R., et al., "Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai", *Information Fusion*, 58:82–115.
[43] Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., and Pedreschi, D. (2018), *op.cit.*
[44] Craven, M. and Shavlik, J. W., "Extracting tree-structured representations of trained networks", in *Advances in neural information processing systems*, pages 24–30. Ribeiro, M. T., Singh, S., and Guestrin, C. (2016), *op.cit.* Lundberg, S. M. and Lee, S.-I., "A unified approach to interpreting model predictions", in *Advances in neural information processing systems*, pages 4765–4774.

component in the AI system with the new interpretable model.[45] In the literature, there are various models recognised as being interpretable. Examples are "decision tree", "decision rules", and "linear models".[46] These models are considered easily understandable and interpretable for humans. However, nearly all of them sacrifice performance in favour of interpretability. In addition, they cannot be applied effectively to data types such as images or text, but only to tabular, relational data, in other words tables.

1.2.2.2. Global vs. local explanations

We distinguish between a global or local explanation depending on whether the explanation allows understanding of the whole logic of a model used by an AI system, or whether it refers to a specific case, that is to say only a single decision is interpretable. A "global" explanation consists in providing a way to interpret any possible decision of a black-box model. Generally, the black-box behaviour is approximated with a transparent model trained to mimic the black-box behaviour and also to be human-understandable. In other words, the interpretable model approximating the black-box provides a global interpretation. Global explanations are quite difficult to achieve and, up to now, can be provided only for AI working on tabular data. A local explanation consists in retrieving the reasons for the "outcome" returned by a black-box model relative to the decision for a specific instance. In this case, it is not required to explain the whole logic underlying the AI, but only the reason for the prediction with regard to a specific input instance. Hence, an interpretable model is used to approximate the AI black-box behaviour only in the "neighbourhood" of the instance analysed, in other words with respect only to similar instances. The idea is that in such a neighbourhood, it is easier to approximate the AI with a simple and understandable interpretable model. Several local explanation approaches are analysed in the following sections.

1.2.2.3. Interpretable models for explaining AI

In the following section, we briefly describe the interpretable models most frequently adopted to explain obscure AI systems or to replace black-box components.

- A "decision tree" exploits a graph structured like a tree and composed of internal nodes representing tests on features or attributes (e.g. whether a variable has a value lower than, equal to or greater than a threshold), and leaf nodes representing a decision. Each branch represents a possible outcome.[47] The paths from the root to the leaves represent the classification rules. The most common

[45] Rudin, C., "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead", *Nature Machine Intelligence*, 1(5):206–215. Rudin, C. and Radin, J., "Why are we using black box models in ai when we don't need to? a lesson from an explainable ai competition", *Harvard Data Science Review*, 1(2).

[46] Freitas, A. A., "Comprehensible classification models: a position paper", *ACM SIGKDD explorations newsletter*, 15(1):1–10.

[47] Quinlan, J. R., *C4.5: Programs for Machine Learning*. Elsevier.

rules are "if-then rules", where the "if" clause is a combination of conditions on the input variables. If the clause is verified, the 'then' part reveals the AI action.

- For a "list of rules", given an ordered set of rules, the AI returns as the decision the output of the first rule that is verified.[48]
- Finally, "linear models" allow visualisation of features importance: both the sign and the magnitude of the contribution of the attributes for a given prediction.[49] If the sign of an attribute-value is positive, then it contributes by increasing the model's output, otherwise, it decreases it. Higher magnitudes of attribute-values indicate a higher influence over the prediction of the model.

1.2.2.4. Desiderata of interpretability

Since interpretable models are required to retrieve explanations, some desiderata should be taken into account when adopting them,[50] in order to increase the trust in a given model.

- "Interpretability" consists in evaluating to what extent a given explanation is human-understandable. An approach often used for measuring the interpretability is the "complexity" of the interpretable surrogate model. The complexity is generally estimated with the 'size' of the interpretable model. For example, the complexity of a rule can be measured with the number of clauses in the condition; for linear models, it is possible to count the number of non-zero weights, while for decision trees it is the depth of the tree.
- "Fidelity" consists in evaluating to what extent the interpretable surrogate model is able to accurately 'imitate', either globally or locally, the decision of the AI. The fidelity can be practically measured in terms of Accuracy score, F1-score, etc.[51] with respect to the decisions taken by the black-box model. Moreover, an interpretable model should satisfy other important general desiderata: for instance, having a high accuracy in terms of evaluating the ability of the interpretable surrogate model to take decisions relating to unprecedented instances.
- "Fairness" and "privacy" are fundamental desiderata to guarantee the protection of groups against discrimination,[52] and to ensure that the interpretable model does not reveal sensitive information.[53]

---

[48] Yin, X. and Han, J., "Cpar: Classification based on predictive association rules", in *Proceedings of the 2003 SIAM International Conference on Data Mining*, pages 331–335. SIAM.
[49] Ribeiro, M. T., Singh, S., and Guestrin, C. (2016), *op.cit.*
[50] Freitas, A. A., *op.cit.*
[51] Tan, P.-N. et al., *op.cit.*
[52] Romei, A. and Ruggieri, S., "A multidisciplinary survey on discrimination analysis", *The Knowledge Engineering Review*, 29(5):582–638.
[53] Aldeen, Y. A. A. S., Salleh, M., and Razzaque, M. A., *A comprehensive review on privacy preserving data mining*. SpringerPlus, 4(1):694.

- "Usability" is another property that can influence the trust in a model: for example, an interactive explanation can be more useful than a textual and fixed explanation.

1.2.2.5. Model-specific vs. model-agnostic explainers

We distinguish between model-specific or model-agnostic explanation methods depending on whether the technique adopted to retrieve the explanation acts on a particular model adopted by an AI system, or can be used on any type of AI. The most used approach to explain AI black-boxes is known as 'reverse engineering'. The term stems from the fact that the explanation is retrieved by observing what happens to the output, that is to say the AI decision, when changing the input in a controlled way.

- An explanation method is 'model-specific', or not generalisable,[54] if it can be used to interpret only particular types of black-box models. For example, if an explanation approach is designed to interpret a random forest[55] and internally uses a concept of distance between trees, then such an approach cannot be used to explain the predictions of a neural network.
- On the other hand, an explanation method is 'model-agnostic', or generalisable, when it can be used independently from the black-box model being explained: the AI's internal characteristics are not exploited to build the interpretable model approximating the black-box behaviour.

1.2.2.6. User background

Varying levels of background knowledge and diverse experiences in various tasks are tied to different notions and requirements for the usage of explanations. Domain experts can be able to understand complex explanations, while common users require simple and effective clarifications. Indeed, the meaningfulness and usefulness of an explanation depends on the stakeholder.[56] For instance, taking as an example the aforementioned COMPAS case, a specific explanation for a score may make sense to a judge who wants to understand and double-check the suggestion of the AI support system and possibly discover that it is biased against black people. On the other hand, the same explanation is not useful to a prisoner who cannot change the reality of being black. However, the prisoner can find useful, and therefore meaningful to him, the suggestion that when he is older he will have a lower risk of recidivism and house arrest will be granted more easily.

---

[54] Martens, D., Baesens, B., Van Gestel, T., and Vanthienen, J., "Comprehensible credit scoring models using rule extraction from support vector machines", *European journal of operational research*, 183(3):1466–1476.
[55] Tan, P.-N. et al., *op.cit.*
[56] Bhatt, U., Xiang, A., Sharma, S., Weller, A., Taly, A., Jia, Y., Ghosh, J., Puri, R., Moura, J. M., and Eckersley, P., "Explainable machine learning in deployment", in *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pages 648–657.

1.2.2.7. Time limitations

The time that the user is allowed to spend on understanding an explanation or is available to do so is a crucial aspect. Obviously, the time availability of a user is strictly related to the scenario where the predictive model has to be used. In some contexts where the user needs to quickly take the decision, for example surgery or in the event of an imminent disaster, it is preferable to have an explanation that is simple and effective. In contexts, though, where the decision time is not a constraint, such as during a procedure to release a loan, one might prefer a more complex and exhaustive explanation.

1.2.3. Different explanations and how to read them

The emerging field of XAI is giving birth to a broad set of alternatives for explaining the black-box components of AI systems. Indeed, it is not possible to define a unique type of explanation that is suitable for every application. The following sections illustrate the most used types of explanations.

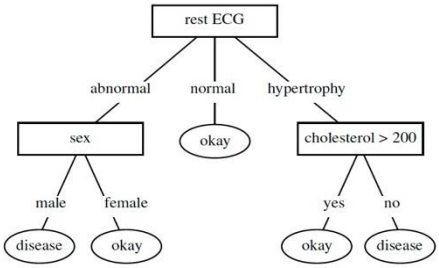
1.2.3.1. Global explanations

1.2.3.1.1. Tree-based explanations

Approximating an obscure AI component with a tree was one of the first approaches introduced.[57] The TREPAN method is able to represent all the possible decisions taken by a neural network acting on tabular data through a single decision tree. TREPAN builds a decision tree approximating the concepts represented by the networks by maximising a gain ratio[58] calculated on the fidelity of the tree with respect to the decision of the neural network. TREPAN results allow to globally explore a neural network through a tree structure that, starting from a root, shows for every path the conditions driving the decision process of the AI system.

[57] Craven, M. and Shavlik, J. W., *op.cit.*
[58] Tan, P.-N. et al., *op.cit.*

Figure 1. Example of global tree-based explanations returned by TREPAN



1.2.3.1.2. List of rules

As previously mentioned, an alternative to explaining black-box classifiers is to directly design transparent models for the AI systems. The CORELS method[59] builds a list of rules and provides an optimal solution for tabular data. An example of list of rules is reported in Figure 2. The rules are read one after the other, and the AI takes the decision of the first rule for which the conditions are verified.

Figure 2. Example of list of rules explanations returned by CORELS

```
if (age=23-25) and (priors=2-3) then predict yes
else if (age = 18 - 20) then predict yes
else if (sex = male) and (age = 21 - 22) then predict yes
else if (priors > 3) then predict yes
else predict no
```

1.2.3.2. Local explanations

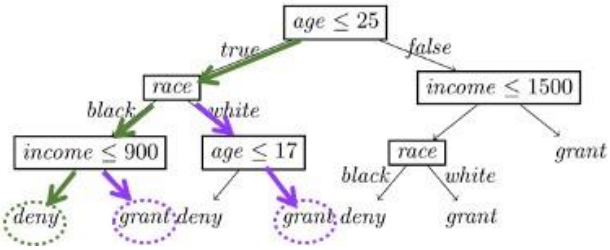
The above explanations are global explanations. However, when the obscure AI models to explain are too complicated, it is better to adopt a local XAI method and separately retrieve the reasons for the decisions for the various instances. Thus, nowadays, research on XAI is focusing more on local explanations. The most representative local explanations are described in the following sections.

[59] Angelino, E., Larus-Stone, N., Alabi, D., Seltzer, M., and Rudin, C., "Learning certifiably optimal rule lists", in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 35–44. ACM.

1.2.3.2.1. Rule-based explanations

In if-then rule explanations under the prism "if conditions, then consequent", the "consequent" corresponds to the decision of the AI, while the "conditions" explain the "factual reasons" for the "consequent". For example, the explanation for the denial of a request by a customer of a loan with "age=22, race=black, and income=800" from a bank that uses an AI could be the factual rule "if age≤25 and race=black and income≤900 then deny". The LORE method builds a local decision tree in the neighbourhood of the instance analysed,[60] then extracts from the tree a single rule revealing the reasons for the decision with regard to the specific instance (see the green path in Figure 3). ANCHOR[61] is another XAI approach for locally explaining AI through decision rules referred to as anchors. An anchor contains a set of features with the values that are fundamental for obtaining a certain decision.

Figure 3. Example of factual and counter-factual rule-based explanation returned by LORE



1.2.3.2.2. Features importance

Local explanations can also be returned in the form of features importance. Figure 4 shows the features importance returned by LIME[62] with positive and negative contributions towards the black-box outcome and assigning their importance. LIME adopts a linear model as an interpretable local surrogate and returns the importance of the features as an explanation exploiting the regression's coefficients. Figure 5 shows the feature importance returned by SHAP.[63] SHAP provides the local unique additive feature importance for a specific record. The higher a Shaply value, the higher the contribution of the feature. Under appropriate settings, LIME and SHAP can also be used to explain AI systems working on text.

---

[60] Guidotti, R., Monreale, A., Giannotti, F., Pedreschi, D., Ruggieri, S., and Turini, F. (2019a), "Factual and counterfactual explanations for black box decision making", *IEEE Intelligent Systems*.
[61] Ribeiro, M. T., Singh, S., and Guestrin, C. (2018), "Anchors: High-precision model-agnostic explanations", in *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence (AAAI)*.
[62] Ribeiro, M. T., Singh, S., and Guestrin, C. (2016), *op.cit.*
[63] Lundberg, S. M. and Lee, S.-I., *op.cit.*

Figure 4. Example of explanation based on features importance by LIME

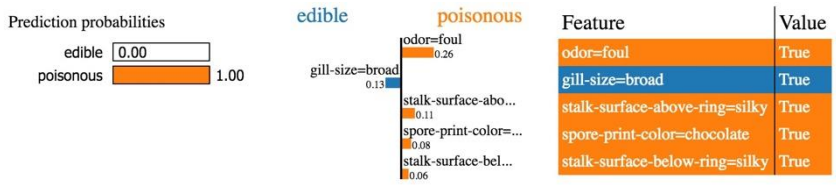


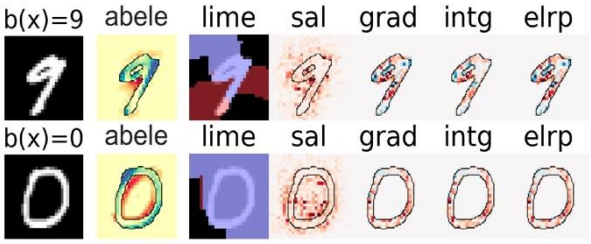
Figure 5. Example of explanation based on features importance by SHAP

1.2.3.2.3. Saliency maps

In image processing, typical explanations consist of "saliency maps", in other words images that show the positive (or negative) contribution of each pixel to the black-box outcome. Saliency maps are efficiently built to locally explain DNN models by gradient and perturbation-based attribution methods. These XAI approaches search the most important pixels of the image such that it maximises the probability that the AI returns the same answer without considering irrelevant pixels. Under appropriate image transformations that exploit the concept of "super-pixels", methods such as LORE and LIME can also be employed to explain AI working on images. The method ABELE[64] uses generative models to return a saliency map that highlights the contiguous areas that can be varied maintaining the same decision from the black-box used by the AI. Figure 6 is a comparison of saliency maps for classification of the handwritten digits 9 and 0 under the explanation methods ABELE,[65] LIME,[66] SALiency,[67] GRADInput,[68] INTGrad,[69] ELRP.[70]

[64] Guidotti, R., Monreale, A., Matwin, S., and Pedreschi, D. (2019b), "Black box explanation by learning image exemplars in the latent feature space", in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 189–205. Springer.
[65] Guidotti, R., Monreale, A., Matwin, S., and Pedreschi, D. (2019b), *op.cit.*
[66] Ribeiro, M. T., Singh, S., and Guestrin, C. (2016), *op.cit.*
[67] Simonyan, K., Vedaldi, A., and Zisserman, A, "Deep inside convolutional networks: Visualising image classification models and saliency maps", *arXiv preprint arXiv*:1312.6034.
[68] Shrikumar, A. et al., "Not just a black box: Learning important features through propagating activation differences", *arXiv*:1605.01713.
[69] Sundararajan, M. et al., "Axiomatic attribution for dnn", in *ICML*. JMLR. Tan, P.-N. et al.
[70] Bach, S., Binder, A., et al., "On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation", *PloS one*, 10(7):e0130140.

Figure 6. **Example of saliency maps returned by different explanation methods. The first column contains the image analysed and the label assigned by the black-box model b of the AI system.**

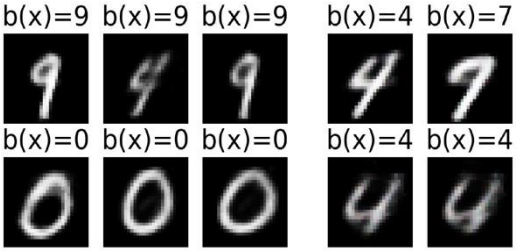


1.2.3.2.4. Prototype-based explanations

An explanation based on "prototypes" returns specimens similar to the instance analysed, which makes clear the reasons for the AI system's decision. Prototypes are used as a foundation of representation of a category, or a concept.[71] Prototype-based explanations can refer to tabular data, images, and text. In Li et al.[72] and Chen et al.,[73] image prototypes are used as the foundation of the concept for interpretability.[74] Kim et al.[75] discuss the concept of "counter-prototypes" for tabular data, in other words prototypes showing what should be different, to obtain another decision. Exemplars and counter-exemplars are used by ABELE[76] to augment the usability of the explanation based on a saliency map. Exemplars (left) and counter-exemplars (right) for 9 and 0 are shown in Figure 7.

[71] Frixione, M. and Lieto, A., "Prototypes vs exemplars in concept representation", in *KEOD*, pages 226–232.
[72] Li, O., Liu, H., Chen, C., and Rudin, C., "Deep learning for case-based reasoning through prototypes: A neural network that explains its predictions", in *Thirty-second AAAI conference on artificial intelligence*.
[73] Chen, C., Li, O., Barnett, A., Su, J., and Rudin, C., "This looks like that: deep learning for interpretable image recognition", *arXiv*:1806.10574.
[74] Bien, J. and Tibshirani, R., "Prototype selection for interpretable classification", *The Annals of Applied Statistics*, 5(4):2403–2424.
[75] Kim, B., Koyejo, O. O., and Khanna, R., "Examples are not enough, learn to criticize! criticism for interpretability", in *Advances In Neural Information Processing Systems*, pages 2280–2288.
[76] Guidotti, R., Monreale, A., Matwin, S., and Pedreschi, D. (2019b), *op.cit.*

Figure 7. Example of exemplars (left) and counter-exemplars (right) explanation returned by ABELE. On top of each (counter-)exemplar is reported the label assigned by the black-box model b of the AI system.



1.2.3.2.5. Counterfactual explanations

A "counterfactual" explanation shows what would have to be different, to change the decision of the black-box model. The importance of counterfactuals is that they help people in reasoning on the cause-effect relations between observed features and classification outcomes.[77] While factual, direct explanations such as decision rules, and features importance, are crucial for understanding the reasons for a certain outcome, a counterfactual reveals what should change in a given instance, to obtain a different classification outcome.[78] The aforementioned LORE method[79] provides, in addition to a factual explanation rule, a set of *counterfactual rules*. With respect to Figure 3, the set of counterfactual rules is highlighted in purple and shows "if income ≥ 900 then grant, or if race = white then grant", clarifying which changes would reverse the decision. The ABELE explanation method[80] proposes counter-exemplar images highlighting the similarities and differences between same-class and other-class instances.

1.3. AI and XAI in the media field

AI technologies are transforming and reinventing the media industry and its marketing, especially to facilitate the monetisation of content and to provide final users with super-personalised services and advertising. In particular, there is a wide usage of AI applications in cinema, television, radio, the written press, and advertising. According to

[77] Byrne, R. M., "Counterfactuals in explainable artificial intelligence (xai): evidence from human reasoning", in *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*, IJCAI-19, pages 6276–6282. Apicella, A., Isgrò, F., Prevete, R., and Tamburrini, G., "Contrastive explanations to classification systems using sparse dictionaries" in *International Conference on Image Analysis and Processing*, pages 207–218. Springer.
[78] Wachter, S., Mittelstadt, B., and Floridi, L., *op.cit.*
[79] Guidotti, R., Monreale, A., Giannotti, F., Pedreschi, D., Ruggieri, S., and Turini, F. (2019a), *op.cit.*
[80] Guidotti, R., Monreale, A., Matwin, S., and Pedreschi, D. (2019b), *op.cit.*

Chan-Olmsted,[81] we can recognise two characteristics to distinguish AI applications in the media field:

- Some applications are more relevant to media audiences, the "demand side", others focus more on internal strategies of media providers, the "supply side". At the same time, some are applicable to both groups, for example audience engagement, augmented experience, and message optimisation.
- Some AI applications refer to "content creators", while others are more relevant to "content distributors" for content analysis and discovery. The companies most active, by far, in the adoption of AI technologies are online news services of companies such as the *New York Times* and video on demand services such as Netflix or Prime Video. Examples of such AI applications are recommendation, personalisation, social network monitoring and listening, emotional tracking and accessibility, video creation and post-production, information verification, predictive success analytics, customer relations, automated drafting, and voice assistants.

1.3.1. AI applications and explainability

Tech companies like Amazon, Netflix, Facebook, and Google are leading AI expansion in the media sector. For instance, the 'recommender systems' of Amazon Prime, Netflix, and Spotify are based on AI methods.[82] A recent survey shows that the most common way in which new media are exploiting AI is to improve recommendation services. Another application for AI could be to reinvent the media-audience connection, that is to say AI might be used to understand audience sentiments, preferences and social conversations. This would make possible the matching of audience interest in real-time to deliver a better consumption experience through personalised media contents. Finally, AI might help media companies identify new business opportunities: storylines or characters might be created based on users' preferences and tastes, opinions on social networks, conversations, etc..[83] In the following section, we discuss some specific applications of AI in the media industry.

1.3.1.1. Recommendation

The most notable use of AI in the media field is for content recommendation. The aim of a 'recommender system' is to predict the 'rating' or 'preference' a user would give to certain content, with respect to others. Recommender systems usually make use of 'collaborative

[81] Chan-Olmsted, S. M., "A review of artificial intelligence adoptions in the media industry", *International Journal on Media Management*, 21(3-4):193– 215.
[82] Chan-Olmsted, S. M., *op.cit.*
[83] Kietzmann, J., Paschen, J., and Treen, E., "Artificial intelligence in advertising: How marketers can leverage artificial intelligence along the consumer journey", *Journal of Advertising Research*, 58(3):263–267.

filtering' and 'content-based filtering', as well as other systems such as knowledge-based systems.[84] Recommender systems have been widely adopted in many fields, but the media field is the one that better fits their usage. The idea of these approaches is to model a user's past behaviour with the media content previously 'selected' or with numerical ratings given to those contents. Besides, recommender systems can also consider similar behaviours made by other users.

The widespread usage of recommender systems and the need to gain trust in AI systems from users implies that each user must have access to explainable recommendations.[85] In other words, the recommendations must be not only accurate and useful but also understandable. The most relevant types of explainable recommendations for media are user-based explanations, feature-based explanations, and item-based explanations:

- For user-based explanations the explanation can be something like: "This content is recommended to you because similar users have selected it before", and it is composed of a set of (anonymised) similar users together with the contents they have selected.
- A feature-based explanation would reveal: "This content is recommended to you because it is described by these features (e.g. features related to topics, actors, music, etc.) that you like", and offers the features according to the ratings you have assigned to them.
- Finally, an item-based explanation would say: "This content is recommended to you because it is similar to these other contents you have liked before".[86]

Early recommendation models, such as item/user-based models, are transparent and explainable. Achieving greater transparency has been recognised as a crucial aspect in raising trustworthiness, effectiveness, persuasiveness, efficiency, and satisfaction in the final user.[87] Recent advances in AI and the use of DNN have helped improve precision in recommendation, but have completely erased transparency because of the use of complex, obscure models such as DNN. The lack of explainability in recommender systems in the media industry can lead to many problems. Without letting the users know why specific results are provided, the system may be less effective in nudging the users toward a particular content, which may further decrease the system's trustworthiness.

---

[84] Manning, C. D., Raghavan, P., and Schütze, H., *Introduction to information retrieval,* Cambridge university press.

[85] Zhang, Y. and Chen, X, "Explainable recommendation: A survey and new perspectives", *arXiv preprint arXiv*:1804.11192.

[86] The interested reader can find in Zhang, Y. and Chen, X., *op.cit.*, other relevant types of explainable recommendations.

[87] Tintarev, N. and Masthoff, J., "A survey of explanations in recommender systems", in *2007 IEEE 23rd international conference on data engineering workshop*, pages 801–810. IEEE.

1.3.1.2. Personalisation and customisation

Personalisation in the proposal and curation of media contents is a fundamental aspect addressed in the media industry through AI. Indeed, AI systems in the media industry excel in precisely tailoring content distribution strategy thanks to recommender systems. For instance, AI systems can analyse trends on social networks, to identify the best content to broadcast. Another application is to analyse audiences, to automatically generate titles/summaries/illustrations with keywords that guarantee higher content visibility. In addition, AI can automate media content generation, and curation, regularly update theme-based playlists, and profile users to make customised recommendations. In this way, the media content proposed to each user can be different and tailored to each user's profile, the journey/commuting of the user, or when and where the media service is used. Other applications are relative to engaging the user with the right content, proposed in the proper format at the right moment in a completely personalised way. It is like having a personal editor for each individual to curate the perfect reading experience.

1.3.1.3. Content creation

As previously discussed, one of the most recent uses of AI is for creating news, music, and videos. In particular, we can use the term 'robot journalism' or 'automated journalism'. In this case, AI systems use natural language generation algorithms to turn data and knowledge into news stories, images, and videos. For instance, AI systems can easily write articles that are relatively boring for humans, such as those on weather or financial reports, based on previous articles and available data. With respect to videos, by exploiting image recognition, AI can produce coherent video montages. Most of the major editing software publishers have already added automatic video processing functions to save editors time. Other software programmes like Gingalab[88] adopt AI to create automated 'best of' videos based on pre-defined editorial lines (e.g. humour, tension, focus on a protagonist, etc.). In September 2018, the BBC broadcast a programme entirely created by a robot.[89]

1.3.1.4. Fake content detection

The weak point of this incredible achievement of AI that is 'creation' is the deepfake phenomenon.[90] Luckily, although AI can generate fake media content, it can also contribute to detecting fake content. Indeed, AI can be a crucial asset in countering misinformation because the same technology used to fabricate a fake can be exploited to detect it. Through extensive analytical capabilities and machine learning algorithms, AI can partially automate the verification of media content like news, images, and videos.

[88] Available at https://gingalab.com/.
[89] Available at https://www.bbc.co.uk/programmes/b0bhwk3p#:~:text=Made%20by%20Machine%3A%20When%20AI%20Met%20the%20Archive,Documentary.
[90] For a definition of deepfakes see 1.1.3 above.

The main problem is that the quality of the detection comes from the experience of the AI which is translated into the availability of data sources for discriminating between real and fake contents. Besides, such a source of information has to be generated by humans who manually annotate media content as fake or not. This manual step can create bias in the data because humans may not be able to verify all the media content necessary to train a fully working AI system, and may have to rely on their feelings about what is real and what is fake. XAI can be crucial in this phase of training for two reasons: First, users of the AI for fake content detection want to be sure that the logic followed in recognising the fake content is human-understandable. The expectation is something like: "This news is fake because sentences are too repetitive and the images displayed are taken from existing websites."; Second, AI systems must not rely on a biased dataset in providing suggestions. If all the real news comes from the same source, the explanation could reveal something like: "This news is fake because it is not being shared by the *New York Times*."

1.3.1.5. Further applications

AI and XAI can be used for many other applications in the media industry. In the following section, we name some of them without entering into details. AI can be used as a tool to improve conversations on the Internet, in other words to recognise hate speech, discrimination, trolls, etc. In this case, too, it is vital to access the reasons for which inadequate posts are recognised as such. AI for voice recognition is a basic for many modern services and vocal assistants like Amazon's Alexa, Google Home, or Apple's Siri, which are present in every smart device. They exploit AI and natural language processing to answer our questions and fulfil our orders. Finally, it is worth mentioning that AI has strategic implications for monetising and predicting the success of media content. At the same time, concerning media ethics, XAI becomes crucial for communicating with the audience, in a transparent way, the logic adopted by the AI systems interacting with the users or making decisions for them. Certain questions could arise, and through XAI, users can possibly have these questions answered: For example, what is the right proportion between personalisation and content discovery? What level of recommendation do we want? Why is this media content considered real? Under the GDPR, the first step of the media industry is to clearly reveal which contents are recommended/created by an AI.

1.3.2. VOD services in practice

VoD services have transformed the way we watch media content ranging from TV series and comedy shows to movies and cartoons. These services algorithmically adapt the users' experience through heavy personalisation that is based on a large set of metadata (including genre, categories, cast, and release date), on user behaviour data (such as searching, browsing, rating, and device type), but also on the rows selected for the homepage, the titles selected for those rows, the visuals for each movie, the movies in the playlist, etc. The AI system adopted evolves, constantly collecting the personal data of

each user, and always offers a customised visualisation of options on which the user is most likely to click, depending on their use and context. The final goal is to find the best combination of contents that can satisfy users instead of contents simply corresponding to the most users. Algorithms thus underpin creativity and diversity, rather than standardisation. These remarkable features are offered based on the AI recommendation systems collecting the data of millions of users watching and rating the content on these platforms.

Nobody knows exactly how these recommendation system works. VoD services usually provide a description of their recommendations system in plain language, but they do not reveal details of their decision-making. In this sense, these AI recommendation systems are black-box models par excellence. Theoretically, a user may not be interested in how recommendations are happening because they are just going to relax in front of some enjoyable media content. However, such recommendations may not be entirely personal, but channelled by marketing strategies or even worse by bias in the data used for the machine learning models. When supported by the application of AI using obscure recommenders based on machine-learning models, decisions or predictions rest on the learning obtained by automated processes, and the available or selected training dataset may not represent the population it was designed to assess. For instance, statistics based on people avoiding movies with Asian heroes could result in discrimination against this category of film and wrongly rate it with a low score for a population that might nonetheless be interested in this kind of media content. Thus, the use of XAI to understand the training dataset and analyse how the data affect the results for different populations is crucial in identifying bias. For any of those services, a global explanation could describe how the algorithm behaves in general. For instance, we could discover that the AI recommender will not suggest a three-hour movie just before midnight on a weekday. On the other hand, a local explanation could describe how the AI behaves for a specific individual. For example, if the customer under analysis generally watches VoD content from 12:00 to 14:00 in her lunch break, at work, then the service in question will not suggest a three-hour movie in this time slot. This is because the service in question may have inferred that this is the best course of action based on routine, even though the three-hour movie perfectly fits the user's interests. On the other hand, perhaps the user does not want the service in question to exploit this type of personal information in making recommendations. So detailed explanations would help VoD services gain more trust from their users. Theoretically, in every application in the media field using AI, suggestions should be based on unbiased recommendations and a trusted relationship between the service and its users.

1.4. Conclusion

Artificial Intelligence cannot be the final solution for any application, and especially in the media field, it needs to be attached to a human being, both when creating and checking media content, but also when watching recommended media content. Indeed, AI is fundamental on the demand side, on the content access side, and for monetisation. AI

has a great potential for the social good in helping navigate masses of content, by optimising searches and personalised recommendations, and by preventing manipulation. With the appropriate XAI tools and degree of trust from the audience and vendors, AI would effectively boost the media industry and all its related sectors and applications.

Big data

The obtaining and using of personal data by third parties, whether provided willingly or inadvertently by the users, can also have a very intrusive effect on their personal lives. Moreover, there are situations in which the state or private parties require insight into a user's life that goes beyond what a user is prepared to accept. In his contribution to this publication, ***Andrea Pin*** *states that the "vast deployment of AI nowadays requires that the media sphere become aware of its unique role and that the media sector should strive to use AI in a lawful, ethical, and robust way". A matter of special concern is the appropriate role of media platforms in managing their contents. Debates are ongoing on the extent to which they should "go beyond a merely passive role to pursue the worthwhile ethical goal for media platforms to patrol their content". In these cases, AI's lack of humanity is precisely one of its biggest drawbacks. Filtering algorithms are extremely efficient in addressing and removing potential harmful content, but they cannot match humans in making nuanced decisions on complex legal areas.*[91]

[91] Barker A., Murphy H., "YouTube reverts to human moderators in fight against misinformation", *Financial Times*, 20 September 2020, https://www.ft.com/content/e54737c5-8488-4e66-b087-d1ad426ac9fa

2. The stuff AI dreams are made of – big data

Andrea Pin, Associate Professor of Comparative Public Law, University of Padua

2.1. Introduction

It is commonly said that big data is the oil of the AI revolution.[92] Since data science and technological engineering joined forces, a massive flow of information has flooded the globe, affecting how we live and understand politics, the economy and culture. Thanks to AI's capabilities, the phenomenon of big data has had an enormous, and probably enduring, impact on how individuals and groups make plans, obtain information about themselves and the world, entertain themselves, and socialise.

Nowadays' computers are technologically capacious. Their algorithms are extremely sophisticated. Their neural networks replicate the intellectual processing of human beings and enable them to make complex analyses. By processing big data, firms can anticipate customers' choices and preferences at such an early stage that they can predict what customers want even before *they* do. Thanks to big data, business processes are moving from a "reactive" to a "proactive" approach.[93]

The Internet is playing a fundamental role within this scenario. As individuals use the Internet to share information, even about themselves and their lives, practically without interruption, the web gathers the raw materials from which AI will draw inferences, make guesses, and find out responses to queries. Oxford philosopher Luciano Floridi coined the concept of "onlife" to describe how frequently and unconsciously human beings transition between the real world and the online world.[94]

This phenomenon is escalating. In 2023 it is estimated there will be more than five billion Internet users and 3,6 devices per capita, and 70% of world population will

[92] Pan S. B., "Get to know me: Protecting privacy and autonomy under big data's penetrating gaze", *Harvard Journal of Law and Technology* 30, 2016, p. 239, https://jolt.law.harvard.edu/assets/articlePDFs/v30/30HarvJLTech239.pdf; Surden H., "Artificial intelligence and law: An overview", *Georgia State University Law Review* 35, 2019, pp. 1311 and 1315.
[93] Microsoft Dynamics 365, *Delivering personalized experiences in times of change*, 2007, p. 3, https://www.hso.com/wp-content/uploads/2020/03/Digitally-transforming-customer-experiences-ebook.pdf.
[94] Floridi L., "Soft ethics and the governance of the digital", *Philosophy & Technology* 31, 1, 2018, p. 1.

have mobile connectivity.[95] The more the world is connected, the more big data will be produced. It is not by chance that one of the most hotly currently debated issues is the introduction of 5G networks, since they can provide considerable informational advantage to their owners.

The media field and industry are big players in this scenario. Their job has always consisted in collecting, processing, and disseminating information. Thanks to big data, now they can profile their audience and learn what it expects, how to couch news or to tell a story, or what would be a good finale for a certain movie. Big data allows customisation of the offering through identification of potential news-readers, or movie-goers, as "computers are more accurate than humans at predicting from 'digital footprints' personality traits [or] political attitudes".[96]

The novelty brought about by big data is also changing the media landscape. "… [D]igital TV/movies/music and a myriad of online distribution models have been challenging incumbent distributors (CDs, cable) for years … Online publishers are mining consumer signals from what they read, where they are, the social signals they send –for example what articles they share, what topics are trending on Facebook and Twitter – to serve up personalised, relevant content while not being too repetitive and predictable, thus automating and surpassing what human editors can do".[97] Traditional media now compete in generating news with non-professional information providers that sift through the web searching for news or bloggers that share their views on social media platforms within which distribution and consumption of content are virtually indistinguishable.[98]

This chapter addresses the most relevant legal ramifications of such a global shift in the media world. It touches upon the crucial issue of privacy protection. It then deals with the potential discriminations and bias that a big data-driven strategy can run into and considers the risks of misinformation, polarisation of politics, and the media field becoming a mass surveillance system. Later on, the chapter casts a bird's eye view at how media markets and strategies are changing in light of big data dynamics. Finally, it briefly addresses the debates on the correct regulatory approach to big data.

Overall, the need to regulate AI has gained much traction throughout the years. Although technologies are global and know no border, the regulatory purpose, approach,

---

[95] Cisco, *Cisco Annual International Report (2018-2023) White Paper*, 9 March 2020, https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html?fbclid=IwAR31-e732ws1p1cIW5PYHQjVOJkPSzV0dGt3sq_qkX_P8wb9Q4Yn0Ez0a0Y.

[96] European Data Protection Supervisor, Opinion 7/2015 Meeting the challenges of big data, 19 November 2015, p. 16, https://edps.europa.eu/sites/edp/files/publication/15-11-19_big_data_en.pdf.

[97] Byers A., "Big data, big economic impact", 10, 2015, https://kb.osu.edu/bitstream/handle/1811/75420/ISJLP_V10N3_757.pdf?sequence=1&isAllowed=y.
See also Bruckner M. A., "The promise and perils of algorithmic lenders' use of big Data", *Chicago-Kent Law Review* 93, 2018, p. 8, https://scholarship.kentlaw.iit.edu/cklawreview/vol93/iss1/1/ or Ambrose M. L., "Lessons from the Avalanche of Numbers: Big Data in Historical Perspective", *ISJLP*, 11, 2015, p. 213, ("Netflix predicts our movies").

[98] Perritt H. H. Jr., "Technologies of storytelling: New models for movies", *Virginia Sports & Entertainment Law Journal*, 10, 2010, p. 153, http://blogs.kentlaw.iit.edu/perrittseminar/files/2016/07/perritt-technologies-of-storytelling-Westlaw_Document_05_56_44.pdf.

and scheme of the big legal players within this scenario – the United States, the European Union and China – diverge deeply. The US approach is committed to ensuring that markets within which AI is massively deployed remain open and efficient; the EU's paramount concern seems to consist in ensuring that the dignity of the individual is respected; China is mostly preoccupied with social peace, stability, and the ordered development of its economy. Each of these approaches accords big data a specific legal treatment.

2.2. Privacy as the big data gatekeeper

Concerns proliferate that big data-driven tools may integrate in a pervasive system of mass surveillance and manipulation. One of the main safeguards against this threat is privacy. Many countries and supranational legal systems have put in place regulations that limit and monitor what and how information is collected and processed, also with the purpose of constraining big data analytics and preventing social disruption. In this respect, privacy laws serve as a shield against big data's overreach.

2.2.1. The United States of America

The Western world is split in its understanding and protection of privacy. The approaches of the United States and the European Union are far from aligned. Despite its historical sensitiveness to privacy, the United States lacks comprehensive regulation of the collection and gathering of information on the web. Several legal regimes coexist, each regulating a specific sector, without any comprehensive nationwide regulation.[99] The US approach, however, usually sees information as a new, huge market, with positive ramifications for the national economy. While certain states have started implementing pieces of legislation that protect and regulate privacy, with California in a leading position, the collection and gathering of personal data is largely allowed and even promoted. A quite general legal baseline is that the subjects who confer their data should be merely *aware* that their information will be processed in various ways, including for profiling and the trading of their preferences. Since most of the protagonists of the AI-based global industry are based in the US, such a favourable regulatory scheme allows them to fully exploit the advantages of the new oil of data.

[99] Houser K. A. & Voss W. G., "The end of Google and Facebook or a new paradigm in data privacy", *Richmond Journal of Law and Technology*, 25, 2018, p. 18, https://jolt.richmond.edu/files/2018/11/Houser_Voss-FE.pdf.

2.2.2. The European Union

Privacy protection within the European Union is based on the General Data Protection Regulation (GDPR),[100] which was adopted on 27 April 2016 and became applicable as of 25 May 2018. The GDPR itself is the peak of a longer process that has enhanced the protection of personal data over the decades, and represents a very different journey from that of the United States. Although the European Union is committed to making it "easier for business and public authorities to access high quality data to boost growth and create value",[101] the European Union's overall attitude rests on a rejection of the commodification of personal data.[102] The GDPR's legal baseline is that a subject must give his/her *consent* to data processing.[103] Consent itself must be unambiguous, freely given, and well informed:[104] the subject must be given the details about the scope and the purpose of the processing.[105] The GDPR's protection covers EU citizens as well as any other natural persons' data, as long as the processing takes place within the EU. In other words, it protects anyone within its territories.[106]

The gap between the US and the European approaches has created a rift in the exchange of data across the Atlantic. The GDPR is very conservative as to the sharing of information gathered within the European Union, and requires that any data transfer outside EU borders comply with EU standards.[107] The EU regulatory philosophy has been perceived to be so protective of privacy that many non-EU citizens tend to prefer EU-based companies over entities not subject to the jurisdiction of the European Union. Conformance with the GDPR has therefore become a reputation asset for companies working in the field of AI even outside the European Union, pushing them to implement privacy protection rules spontaneously.[108]

---

[100] Consolidated text: Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance), https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:02016R0679-20160504.

[101] European Data Protection Supervisor, Opinion 3/2020 on the European strategy for data, 16 June 2020, p. 4, https://edps.europa.eu/sites/edp/files/publication/20-06-16_opinion_data_strategy_en.pdf. See also Council of the European Union, Shaping Europe's Digital Future – Council Conclusions, 9 June 2020, https://data.consilium.europa.eu/doc/document/ST-8711-2020-INIT/en/pdf.

[102] European Data Protection Board, Guidelines 2/2019 on the processing of personal data under Article 6(1)(b) GDPR in the context of the provision of online services to data subjects, Version 2.0, 8 October 2019, No. 54, https://edpb.europa.eu/sites/edpb/files/files/file1/edpb_guidelines-art_6-1-b-adopted_after_public_consultation_en.pdf

[103] Art. 6 GDPR.

[104] Manheim K. & Kaplan L., "Artificial intelligence: Risks to privacy and democracy", *Yale Journal of Law & Technology*, 106, 2019, p. 1069, https://yjolt.org/sites/default/files/21_yale_j.l._tech._106_0.pdf.

[105] Art. 6, par. 4, and 7, GDPR.

[106] European Data Protection Supervisor, Opinion 3/2018 EDPS Opinion on online manipulation and personal data, 19 March 2018, p. 14, https://edps.europa.eu/sites/edp/files/publication/18-03-19_online_manipulation_en.pdf.

[107] Art. 45 GDPR.

[108] Moerel L. & Lyon C., "Commoditization of data is the problem, not the solution – Why placing a price tag on personal information may harm rather than protect consumer privacy, *Future of Privacy Forum*, 24 June

Such a high level of privacy protection from the GDPR comes, however, at a cost. The companies' need to obtain consent from the Internet users who visit their websites translates into a plethora of repetitious, and sometimes obscure, requests for consent that traditionally pop up as soon as a webpage is displayed.[109] This phenomenon has flooded the Internet to the extent that most users simply click "yes" and keep navigating the website without paying attention to how their information is collected, processed and disseminated.[110] This course of action is certainly risky but understandable. Some have made the estimation that a normal person – not a skilled lawyer or a maniacally meticulous Internet user – would waste 76 working days per year reading all the privacy warnings that pop up while he/she is online.[111] Too much privacy protection can be counter-productive: individuals may give away all the protection by consenting in too superficial a manner, thereby allowing massive harvesting of their information.

Moreover, the potentials of big data analysis can weaken the privacy protection accorded by the GDPR on many fronts. First, the GDPR imposes fewer restrictions on anonymised data, as anonymisation is supposed to protect privacy. Thanks to increasing AI capabilities, however, "it is becoming ever easier to infer a person's identity by combining allegedly 'anonymous' data with other datasets including publicly available information for example on social media"[112] … "The bigger and the more comprehensive" a data collection, the more likely it is that an individual whose data has been anonymised will be re-identified.[113]

On top of this, EU privacy rules require that individuals be given detailed information regarding the purpose and scope of the processing of the data they confer. Through neural networks and deep learning, AI-based systems draw inferences that even software developers cannot fully anticipate. This very capacity of big data jeopardises how EU privacy regulation is construed. As big data processing returns results that cannot be fully foreseen, it is extremely difficult to provide individuals with a detailed picture of what their information will be used for.[114]

2020, https://fpf.org/2020/06/24/commoditization-of-data-is-the-problem-not-the-solution-why-placing-a-price-tag-on-personal-information-may-harm-rather-than-protect-consumer-privacy.

[109] European Data Protection Supervisor, Opinion 7/2015 Meeting the challenges of big data, *op. cit.*, p. 11.

[110] Tsesis A., "Marketplace of ideas, privacy, and the digital audience", *Notre Dame Law Review*, 94, 2019, p. 1590, https://scholarship.law.nd.edu/cgi/viewcontent.cgi?article=4845&context=ndlr.

[111] Hartzog W., *Privacy's blueprint,* Harvard University Press, 2018.

[112] European Data Protection Supervisor, Opinion 4/2015. Towards a new digital ethics, September 11, 2015, p. 6, https://edps.europa.eu/sites/edp/files/publication/15-09-11_data_ethics_en.pdf.

[113] European Data Protection Supervisor, Opinion 7/2015, "Meeting the challenges of big data", *op. cit.*, p. 15.

[114] AGCM, AGCOM, and Garante per la protezione dei dati personali, Indagine conoscitiva sui *Big Data*, p. 25-26, https://www.agcom.it/documents/10179/17633816/Documento+generico+10-02-2020+1581346981452/39c08bbe-1c02-43dc-bb8e-6d1cc9ec0fcf?version=1.0. The document explains how "dynamic consent" is taking off as a viable option within the EU privacy regulatory scheme. This concept understands consent as a gradual process, during which the subject can be contacted more than once to ask whether he or she consents to a certain usage of his or her information.

2.2.3. China

Chinese public and private institutions draw massive amounts of data from a wealth of sources to profile individuals with the highest degree of accuracy. Collecting and processing personal data about the Chinese population is instrumental to China's grand civic plan, which foresees the implementation of a wide-ranging surveillance and monitoring scheme that exploits AI to profile and predict individuals' and groups' behaviours.[115] The overall goal of this plan consists in the construction of a pervasive social credit system – an AI-based mechanism that gathers information from personal records, smartphones, and mass-surveillance systems, and then ranks individuals and accords them privileges and rights based on their previous conduct.[116]

In China, public institutions are trying to make everyone's life transparent, and not private. To this end , they partner with Chinese private firms. A handful of big tech companies such as WeChat and Alibaba thus operate as digital hubs for the lives of Chinese citizens.[117] The Chinese are encouraged to use the same mobile app for a wide array of activities – from reserving a taxi to paying for a restaurant, socialising or interacting with a public administration. A huge amount of information about anyone is thus gathered and passed over to public institutions for profiling.[118]

2.2.4. Three different approaches?

Odd as it may seem, some have speculated that a similar social credit system is already in place also in the private sector of the United States.[119] Private companies don't merely profile their clients to make them loyal. They also sell the information about them to other companies. Personal preferences and purchase habits are thus matched to better profile users, anticipate their decisions, and nudge them.[120] A bank or an insurance

---

[115] State Council, Notice of the State Council Issuing the New Generation of Artificial Intelligence Development Plan, No. 358 July 2017, pp. 2-5, and 18-21, https://flia.org/notice-state-council-issuing-new-generation-artificial-intelligence-development-plan.

[116] State Council, Notice concerning Issuance of the Planning Outline for the Construction of a Social Credit System (2014-2020), No. 21, 14 June 2014, https://chinacopyrightandmedia.wordpress.com/2014/06/14/planning-outline-for-the-construction-of-a-social-credit-system-2014-2020.

[117] Pieranni S., *Red Mirror*, Laterza, 2020, pp. 22-23.

[118] Ibid, pp. 40 and 115.

[119] Baker L. C., "Next generation law: Data-driven governance and accountability-based regulatory systems in the West, and social credit regimes in China", *Southern California Interdisciplinary Law Journal*, 28, 2018, pp. 170-171, https://lcbackerblog.blogspot.com/2019/05/just-published-next-generation-law-data.html.

[120] The European Parliament has recently called on the European Commission to "ban platforms from displaying micro-targeted advertisements": European Parliament, Resolution of 18 June 2020 on competition policy – annual report 2019, https://www.europarl.europa.eu/doceo/document/TA-9-2020-0158_EN.html. According to Morozov E., "Digital socialism?", *New Left Review*, 116/117, March-June 2019, p. 62, https://newleftreview.org/issues/II116/articles/evgeny-morozov-digital-socialism, "Amazon got a patent on 'anticipatory shipping' – allowing it to ship products to us before we even know we want them".

company can accurately assess an individual's financial risk based on a variety of information, ranging from his/her education, his/her lifestyle, or the places and people he/she visits. A political party can assess the political inclination of an individual based on the movies he/she watches, the media channels he/she prefers, or his/her family records.

It should be of little or no surprise that the overall US approach to data protection overlooks the negative potential of such a private accumulation of personal data. The US culture of rights has traditionally focused on keeping public powers under check. This approach is still lively, and keeps the US attention focused on the threats of public powers, whereas Europe has always been more attentive to private companies' capacity to violate fundamental rights.[121] The paradoxical result is that the US is the global hub for big data innovation, but does not see the big data threat to fundamental rights the way Europe appears to do.

Such different approaches to privacy have powerful consequences for the ordinary lives of citizens and media companies alike. As will become apparent below, the exploitation of AI-based technologies transforms media corporations into more than information givers. They can become information gatherers and participate in profiling individuals.

2.3. Big data bias and discrimination

Although one would not expect software to be biased, one of the biggest challenges for data-driven technologies is their discriminatory potential. The gathering, processing, and dissemination of information can incorporate, embed and amplify prejudices. The most famous example probably is the Microsoft chatbot Tay. In 2016, Microsoft created a Facebook profile for innovative software capable of interacting on the media platform with other Facebook users by gathering information from the web, identifying trends, and exchanging opinions accordingly.

In the span of 16 hours, the Facebook account was opened and then shut down, after its creators realised it was engaging in sexist and racist posts.[122] The software developers certainly did not provide their bot with the set of prejudices it later displayed on the web. Its makers simply used the web itself to teach the bot, which evidently found racism and sexism to be widespread and attention-drawing. Tay shaped its language and

[121] As to the European attentiveness to private companies' harmful potential, see European Data Protection Officer, Opinion 8/2016 EDPS Opinion on coherent enforcement of fundamental rights in the age of big data, 23 September 2016, p. 5, https://edps.europa.eu/sites/edp/files/publication/16-09-23_bigdata_opinion_en.pdf. *See* also Pollicino O., "L'"autunno caldo' della Corte di giustizia in tema di tutela dei diritti fondamentali in rete e le sfide del costituzionalismo alle prese con i nuovi poteri privati in ambito digitale", *Federalismi*, 15 October 2019, https://www.federalismi.it/nv14/editoriale.cfm?eid=533.

[122] "Microsoft 'deeply sorry' for racist and sexist tweets by AI chatbot", *The Guardian*, 26 March 2016, https://www.theguardian.com/technology/2016/mar/26/microsoft-deeply-sorry-for-offensive-tweets-by-ai-chatbot.

themes based on the training it was subject to. It learned and adopted prejudices on its own.

Tay's ephemeral life explains the importance of training for AI. AI-based systems require a lot of data in order to learn. The more information they gather, the more capable they become of making inferences and choices. Unfortunately, big datasets to train algorithms are often unavailable, so software programmers often exploit what is already available on the web. This choice is extremely problematic, because human beings cannot fully supervise the learning process, and AI can take unforeseen or even unwelcome directions. It can draw and incorporate biases from society, boosting them with its activity.[123]

Unbalanced datasets can unintentionally create biases, as the case of facial recognition exemplifies. Western AI systems of face recognition often fail to correctly identify non-Caucasian individuals because other ethnic groups appear on the web less often than Caucasians, while AI software developed in China suffers from the reverse problem.[124] As a result, there is a higher probability that, say, in Western countries an African individual is mistaken for someone else than a Caucasian is. Media systems that incorporate big data-based processes therefore face a formidable challenge, as by exploiting AI they may incorporate prejudices and social imbalances.

Fighting discrimination is very difficult in the field of big data and neural networks because of the dangers of "proxy discrimination".[125] Proxy discrimination is a private or public policy that includes a requisite or factor that is facially neutral but actually embeds a discriminatory tradition, practice, or belief. For example, in socially or territorially divided societies, the zip code or the housing price can serve as a proxy discrimination for insurance policies or zoning, as it may deprioritise some ethnicities while preferring others. Even if software developers expressly prohibit AI from considering ethnicity while making inferences, other factors can serve as proxies for discrimination.[126] Within a given society, big data-driven market strategies, political campaigns, or welfare providers can – even involuntarily – isolate and systematically discriminate worse-off groups by proxy.

---

[123] Stevenson M. T. & Doleac J. L., *Algorithmic Risk Assessment in the Hands of Humans*, Institute of Labor Economics, 1 December 2019, p. 1, http://ftp.iza.org/dp12853.pdf; Bruckner M. A., *op. cit*, p. 25.

[124] Grother P., Ngan M., Hanaoka K., "Face recognition vendor test (FRVT) Part III. Demographic effects", National Institute of Standards and Technology Interagency 8280, December 2019, https://doi.org/10.6028/NIST.IR.8280.

[125] Prince A. E. R. & Schwarcz D., "Proxy discrimination in the age of artificial intelligence and big data" *Iowa Law Review* 105, 2020, p. 1260, https://ilr.law.uiowa.edu/print/volume-105-issue-3/proxy-discrimination-in-the-age-of-artificial-intelligence-and-big-data.

[126] Idem.

2.4. Informing the people: Media, misinformation, and illegal content

AI is a powerful media tool. It can discover facts, detect preferences, profile users and anticipate social trends. In a few words, it can provide people with more of what they want to receive. Customising media offerings through big data has a price, though.

AI is a very good tool for the pre-selection of content that media users may find of interest. Given the overflow of information, AI's capacity to profile a user can predict his/her interests in a piece of information, making the media's work more effective and the user's experience more enjoyable. However, AI exploitation may make media users unaware of the fact that their horizons are narrowing – that the type of information they receive may not portray reality accurately, but only the "reality" of what AI understands their interests to be.

Feeding users with more of what they already prefer, know, or are interested in, tends to create social bubbles. Big data technologies can filter information depending on what a media user supposedly likes or believes. Instead of widening the horizon of users, AI is thus able to boost individuals' intellectual selectiveness. A user-friendly news industry may lose sight of its purpose of providing society with broad perspectives, fully informed news and challenging viewpoints.

Big data-driven media strategies can thus unwillingly trigger the creation of informational bubbles. There is the additional risk, however, that a bubble is generated intentionally. Big tech companies can profile users and information to boost or hinder the spread of certain information depending on their market strategies or agendas.[127]

Big data also pits traditional media against social media. Social media exploit the strong protection normally accorded to freedom of speech, and live off their continuous presence on the web and their capacity to feed the audience with more news.[128] They therefore offer a cheap and easily accessible alternative to professional media operators and outlets. Such asymmetric competition has triggered a dangerous "race to the bottom" in the field of news providers.[129] In order to avoid losing the audience, traditional media try to keep up with the speed of non-professional services such as blogs, often at the expense of accuracy.[130]

AI-based media platforms' bubbles often participate in spreading "fake news". A plague in today's news industry, according to some statistics "fake news" is capable of

---

[127] Singer H., "How Washington should regulate Facebook", *Forbes*, 18 October 2017, https://www.forbes.com/sites/washingtonbytes/2017/10/18/what-to-do-about-facebook.

[128] Shefa M. C., "First Amendment 2.0: Revisiting Marsh and the quasi-public forum in the age of social media", *University of Hawaii Law Review*, 41, 2018, p. 160.

[129] AGCM, AGCOM, and Garante per la protezione dei dati personali, Indagine conoscitiva sui Big Data, *op. cit.*, p. 30.

[130] European Data Protection Supervisor, Opinion 3/2018 EDPS Opinion on online manipulation and personal data, *op. cit.*, p. 13 ("There is evidence that … concentration and elimination of local journalism facilitates the spread of disinformation").

reaching more people and more quickly than curated, fact-checked information,[131] giving life to what Cass Sunstein has called "cybercascades".[132] The bubble system aggravates the process, as it filters out facts and different viewpoints, thereby reinforcing deeply held viewpoints and even prejudices.

Big data-driven strategies are calling into question the historical role that the media system and freedom of speech have played in democratic regimes. Instead of broadening horizons, challenging viewpoints, exposing biases and making society progress, contemporary media platforms run the risk of mutually insulating social groups and reinforcing deeply held opinions. Traditionally, liberal constitutionalism values and protects freedom of speech greatly because different viewpoints make societies progress through the free exchange of opinions. Contrarily, big data technologies are capable of creating "echo chambers",[133] which expel dissent and gravitate around unchallenged beliefs. Opinions that challenge deeply seated worldviews are ejected from a bubble and will probably find their place within another bubble, which offers virtually no exchange outside itself.[134] Big data can thus narrow perspectives and immunise prejudices from the benefits of freedom of speech.

Private and public institutions have grown aware of the distortions that big data can cause to media and broader society. For example, Twitter recently created a contentious fact-checker tool with the purpose of detecting "fake news" or tweets that harm identifiable groups.[135] The EU's *Code of Practice on Disinformation*[136] has urged a comprehensive consideration of the phenomenon, emphasising that "all stakeholders have roles to play in countering the spread of disinformation". A list of signatories to the code that includes Facebook, Google, Mozilla, TikTok and Twitter has thus promised to "[d]ilute the visibility of disinformation by improving the findability of trustworthy content", and to "facilitate content discovery and access to different news sources representing alternative viewpoints". Overall, many are calling for regulation of the deployment of AI in a way that would bring Internet service providers closer to the "traditional media responsibility standards".[137]

EU policies especially target terrorist content, child sexual abuse material, racism, and xenophobic and hate speech,[138] which are usually topics of great concern for today's

---

[131] Idem.

[132] Sunstein C. R., "*#republic: Divided democracy in the age of Social Media*", Princeton University Press, 2017, p. 57.

[133] Sasahara K. et al., "*On the inevitability of online echo chambers*", https://arxiv.org/abs/1905.03919.

[134] Jones R. L., "Can you have too much of a good thing: The modern marketplace of ideas", *Missouri Law Review*, 83, 2018, p. 987, https://scholarship.law.missouri.edu/mlr/vol83/iss4/8/.

[135] Pham S., "Twitter says it labels tweets to provide 'context, not fact-checking'", *CNN Business*, https://edition.cnn.com/2020/06/03/tech/twitter-enforcement-policy/index.html.

[136] EU Code of Practice on Disinformation, https://ec.europa.eu/digital-single-market/en/news/code-practice-disinformation.

[137] European Data Protection Supervisor, Opinion 3/2018 EDPS Opinion on online manipulation and personal data, *op. cit.*, p. 16.

[138] Policy Department for Economic, Scientific and Quality of Life Policies, "Online platforms' moderation of illegal content online", June 2020, p. 9, https://www.europarl.europa.eu/RegData/etudes/STUD/2020/652718/IPOL_STU(2020)652718_EN.pdf.

social media. In fact, given the massive inflow of data, filtering information before deciding whether to host it is technically unrealistic. Online platforms thus normally blend two different schemes: on the one hand, they adopt a "notice-and-takedown" system - anyone can complain that a specific display of content is in breach of the law and have the medial platform make an assessment; on the other hand, most platforms adopt big data-based filtering systems that sift through the materials automatically and pervasively, making decisions on what should be concealed from the public.[139] Most platforms have an additional safeguard against such automated decisions, allowing individuals to challenge a software decision to remove some material.[140]

Within the US and the EU, which has "one of the most comprehensive regulatory frameworks for tracking illegal content online",[141] service providers enjoy broad liability exemptions. Such exemptions aim to preserve their positive role in connecting people and disseminating information.[142] EU law has reinforced this rule by prohibiting its member states from imposing general obligations on hosting platforms to monitor the material they host.[143] The scenario is in flux, however.[144] In interpreting the Directive on electronic commerce, the Court of Justice of the European Union has stated that service providers that do not simply passively display materials are expected to do more than simply review and remove materials when necessary once they are requested to do so.[145] In fact, the court stated, a judicial order of removal extends "to information, the content of which, whilst essentially conveying the same message [to which the judicial order refers], is worded slightly differently, because of the words used or their combination,

[139] Ibid, p. 45 .
[140] Ibid, p. 10.
[141] Ibid, p. 66.
[142] For the United States, see Title 47, Section 230 of the Communication Decency Act, https://www.fcc.gov/general/telecommunications-act-1996; For the EU, see Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market ('Directive on electronic commerce'), https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32000L0031&from=EN, Art. 14: "1. Where an information society service is provided that consists of the storage of information provided by a recipient of the service, Member States shall ensure that the service provider is not liable for the information stored at the request of a recipient of the service, on condition that: (a) the provider does not have actual knowledge of illegal activity or information and, as regards claims for damages, is not aware of facts or circumstances from which the illegal activity or information is apparent; or (b) the provider, upon obtaining such knowledge or awareness, acts expeditiously to remove or to disable access to the information." As for the protection of minors, see Directive 2010/13/EU of the European Parliament and of the Council of 10 March 2010 on the coordination of certain provisions laid down by law, regulation or administrative action in Member States concerning the provision of audiovisual media services (Audiovisual Media Services Directive; codified version; text with EEA relevance). A consolidated version including the amendments introduced in 2018 is available at https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:02010L0013-20181218.
[143] Policy Department for Economic, Scientific and Quality of Life Policies, *op. cit.*, p. 21.
[144] Nunziato D. C., "The marketplace of ideas online", *Notre Dame Law Review*, 94, 2019, p. 1521, https://scholarship.law.nd.edu/cgi/viewcontent.cgi?article=4844&context=ndlr.
[145] C-324/09, *L'Oréal et al. v. eBay International AG*, paras. 113-115, http://curia.europa.eu/juris/document/document.jsf?text=&docid=107261&pageIndex=0&doclang=en&mode=lst&dir=&occ=first&part=1&cid=12642628.

compared with the information whose content was declared to be illegal".[146] Some have criticised this sensible principle because it would result in the "Good Samaritan paradox": the more a platform is committed to patrolling the information it publishes, the more it becomes liable. There are concerns that such a judicial approach would encourage providers to remain passive and limit their monitoring activity in order to avoid liability risks.[147] It is now a matter of debate whether the EU should revise its policy and imitate the US approach, which has preserved the liability exemption for platforms, as this would encourage them to become more proactive, or whether this would jeopardise the protection of individuals and groups.[148]

In the context of illegal materials posted on online platforms, AI can certainly play an important role. Given the huge amount of data exchanged and the tendency to create bubbles within which media users hardly find information they do not like or viewpoints they disagree with, illegal materials may not be detected by human beings for a long time. Developing AI-based systems that filter content may therefore become advisable or even necessary. AI and big data are not just part of the problem – they can be part of the solution. Obviously, AI-based monitoring should not become a form of automated censorship. Providers may exploit AI systems to filter out materials that are simply controversial, thereby insulating the public sphere from minoritarian opinions or information that many would find hard to engage with . This risk should be kept in check.

2.5. Big data politics and the political bubble[149]

Democracies need a sound public sphere to survive and flourish.[150] The existence and exchange of alternative worldviews and political opinions is crucial for their survival. More generally, within democracies "people should be exposed to materials that they would not have chosen in advance",[151] as one of the benefits historically associated with democracies is that "biases are filtered out in the large republic".[152]

Social media have flooded contemporary politics. Legal academia and courts have responded by slowly but steadily developing the classical idea of public forums to

[146] C-18/18, *Eva Glawischnig-Piesczek v. Facebook Ireland Ltd*., par. 41, http://curia.europa.eu/juris/document/document.jsf?text=&docid=218621&pageIndex=0&doclang=en&mode=lst&dir=&occ=first&part=1&cid=12642666.
[147] Policy Department for Economic, Scientific and Quality of Life Policies, *op. cit*., p. 20; Policy Department Economic and Scientific Policy, "Liability of Online Service Providers for Copyrighted Content – Regulatory Action Needed?", January 2018, p. 10, https://www.europarl.europa.eu/RegData/etudes/IDAN/2017/614207/IPOL_IDA(2017)614207_EN.pdf.
[148] Ibid, p. 67.
[149] For a different viewpoint on the filter-bubble/echo chamber issue see chapter 5 of this publication.
[150] Wischmeyer T., "Making social media an instrument of democracy", *European Law Journal*, 25, 2019, p. 172, https://onlinelibrary.wiley.com/doi/abs/10.1111/eulj.12312.
[151] Sunstein C. R., *op. cit.* p. 6.
[152] McGinnis J. O., *Accelerating Democracy*, Princeton University Press, 2013, p. 127.

incorporate also social media sites that are privately owned.[153] Because of their pervasive social role and their pivotal importance in providing the public with news feeds and political opinions, the US Supreme Court has dubbed social media sites as "the modern public square".[154] They are so essential to social and political life – the court has argued – that they must be accessible to the general public.[155] Since 2001, US courts have also "treated computers and Internet access as 'virtually indispensable in the modern world of communications and information gathering'."[156]

Social media are not universally accessible places within which everybody is welcomed and able to make an argument, however. Big data analysis allows social media to segment the public sphere in self-referential bubbles.[157] Even the media platforms that do not intentionally filter information, still tailor their news feeds to their users' needs and choices, therefore creating informational bubbles. Such bubbles are capable of dividing public opinion into impenetrable, homogenous spheres of influence.[158]

The creation of homogenous, partisan, non-conversational echo chambers is no substitute for democratic pluralism[159] and can even threaten it.[160] The scandal of Cambridge Analytica, which allegedly harvested data of Facebook users without their consent to develop "psychographic profiles" and then target selected individuals to nudge their voting behaviours,[161] is just one example of how big data can affect politics.[162] And there is wider evidence of the deployment of big data-fed bots to influence political agendas.[163]

Harvard Law Professor Cass Sunstein has explored the impact of AI-based social media platforms in the political sphere in his acclaimed volume *#Republic*.[164] Sunstein has persuasively shown AI's capacity to generate informational clusters and polarise politics. Political campaigns can target well-profiled users, exposing them to certain opinions or facts while silencing or downplaying the statements of political opponents or facts that

---

[153] Nunziato D. C., *op. cit.*, p. 3.
[154] *Packingam v. North Carolina* 582 U.S. ___ (2017), https://www.supremecourt.gov/opinions/16pdf/15-1194_08l1.pdf.
[155] Ibid.
[156] Shefa M. C., *op. cit.*, p. 164.
[157] Sunstein C. R., *op. cit.*
[158] Sasahara K. et al., *op. cit.*
[159] Wischmeyer T., *op. cit.*, p. 173-174.
[160] Manheim K. & Kaplan L., *op. cit.* , p. 109.
[161] Ibid, p. 139.
[162] For more examples drawn from various countries, see Gurumurthy A. and Bharthur D., "Democracy and the algorithmic turn", *Sur International Journal of Human Rights*, 27, 2018, pp. 43-44, https://sur.conectas.org/en/democracy-and-the-algorithmic-turn, and Tenove C., Buffie J., McKay S. and Moscrop D., *Digital threats to democratic elections: how foreign actors use digital techniques to undermine democracy*, January 2018, passim, https://democracy2017.sites.olt.ubc.ca/files/2018/01/DigitalThreats_Report-FINAL.pdf.
[163] When the Federal Communication Commission considered repealing some rules regulating the Internet in 2017, 21 out of 22 million commentsthe Commission received on its website were fake news (Manheim K. & Kaplan L., *op. cit.*, p. 145.)
[164] Sunstein C. R., op. cit.

would call into question their own platform and agenda.[165] AI thus splinters the public sphere into homogenous environments which hardly interact together. Successful politicians often go to extremes to galvanise their supporters and reinforce the bubble system.

Big data politics often blurs the line between personal and institutional capacity. Many political figures prefer using their personal social media profiles rather than institutional profiles also to communicate with the general public on institutional matters. By using their personal profiles, they force the public – which would normally follow institutional media pages and profiles - into their sphere of supporters.

Some legal systems have deployed countermeasures to fight this privatisation of the public sphere into separate media echo chambers. The US experience provides the most telling example of this development. Many public figures – including President Donald Trump – who have used personal websites for institutional purposes have blocked individuals making critical comments about their posts, therefore walling them out from their briefing activity to citizens.[166] Some citizens thus ejected from the audience sued the politicians - and won in court. Judges considered the structure of media platforms and how politicians were using them, and concluded that such platforms had to be considered public places that should remain open to everyone. Politicians could still "mute" their followers, thereby preventing them from engaging in a conversation within their own profile, but not "block" them, as this would have prevented some citizens from being informed on matters of public interest.[167]

2.6. Media as surveillance watchdogs?

Big data analysis has been instrumental to the development of artificial face recognition techniques. Thanks to AI capabilities, software can peruse and compare an enormous amount of images, to find matches. Differently from old-fashioned close-circuit cameras, which human agents scrutinise looking for matches, today's computer vision has the capacity to process images almost instantly. In a 2019 decision, a Welsh court dealt with artificial face recognition.[168] The software that the Welsh police had deployed at several public events was able to process up to 40 faces per second. The total figure is impressive: in roughly 50 deployments, the software processed roughly 500 000 individuals – one out of six of the total population of Wales. AI can become a powerful tool of mass surveillance, as has already happened in countries such as China, where a

[165] Mor N., "No Longer Private: On Human Rights and the Public Facet of Social Network Sites", *Hofstra Law Review* 47 (2018), p. 669, https://www.hofstralawreview.org/wp-content/uploads/2019/04/bb.7.mor_.pdf (6 August 2020).
[166] *Ibidem*, p. 42 ff.
[167] *Knight First Amendment Inst. at Columbia Univ. v. Trump* 302 F. Supp. 3d 541 (SDNY 2018), https://digitalcommons.law.scu.edu/cgi/viewcontent.cgi?article=2780&context=historical (6 August 2020).
[168] *(Bridges) v. The Chief Constable of South Wales Police et al.*, [2019] EWHC 2341, https://www.judiciary.uk/wp-content/uploads/2019/09/bridges-swp-judgment-Final03-09-19-1.pdf.

project of a systematic AI-based surveillance system, with more than half a billion of cameras deployed, is ongoing.[169]

Face recognition cuts across a variety of issues seen above. First, face recognition techniques are a matter of privacy. They process human faces – not just of those in a database, but of everyone. In fact, in order to exclude someone from the group of persons of interest, a software must process their face first. According to the European legal culture, such a massive privacy intrusion must be properly justified. As the European Court of Human Rights has repeatedly insisted, public interests do not override privacy concerns – on the contrary, they require a preliminary assessment of the expected benefits and costs to ensure that any deployment is proportionate to the task.[170]

Second, face recognition techniques runs the risk of being biased. As noted above, "false positives" – wrong matches – are more frequent in ethnic groups that are underrepresented in the training materials.[171] False positives often have practical consequences, as they may reinforce racial prejudices and nudge public institutions, such as police patrols, to target ethnic minorities for which software returns more false positives.[172]

Third, face recognition can be misleading on a variety of grounds. Some software programmes are able to exploit the immense AI capabilities by using live and recorded images coming from any Internet source.[173] Such technology can exploit the media industry to gather more materials and increase its database. A debate is ongoing on the pros and cons of developing or adopting software that sifts through the web to find matches of people, as has happened in many local police agencies of the U.S. to track down suspects. Such a huge dataset draws on a variety of materials that can be spurious, incorporate bias,[174] and transform any single bit of social life or media broadcast into a record.

---

[169] Carter W. M., "Big Brother facial recognition needs ethical regulations", *The Conversation*, 22 July 2018, https://theconversation.com/big-brother-facial-recognition-needs-ethical-regulations-99983.

[170] *Lopez Ribalda and others v. Spain* (apps. No. 1874/13 and 8567/13: http://hudoc.echr.coe.int/fre?i=001-197098); *Gorlov and others v. Russia* (app. no. 27057/06; 56443/09; 25147/14: http://hudoc.echr.coe.int/spa?i=001-194247); *Antovic and Mirkovic v. Montenegro* (app. no. 70838/13: http://hudoc.echr.coe.int/fre?i=001-178904); *Bărbulescu v Romania* (app. no. 61496/08: http://hudoc.echr.coe.int/spa?i=001-177082).

[171] Buolamwini J. & Gebru T., "Gender shades: Intersectional accuracy disparities in commercial gender classification" *Proceedings of Machine Learning Research* 81, 2018, pp. 1 and 15, http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf.

[172] Fung B. and Metz R., "This may be America's first known wrongful arrest involving facial recognition", 24 June 2020, *CNN Business*, https://edition.cnn.com/2020/06/24/tech/aclu-mistaken-facial-recognition/index.html.

[173] Hill K., "The secretive company that might end privacy as we know it", *New York Times*, 18 January 2020, https://www.nytimes.com/2020/01/18/technology/clearview-privacy-facial-recognition.html; Ducklin P., "Clearview AI facial recognition sued again – this time by ACLU", *Naked Security*, 29 May 2020, https://nakedsecurity.sophos.com/2020/05/29/clearview-ai-facial-recogition-sued-again-this-time-by-aclu.

[174] Geiger R. S. et al., "Garbage in, garbage out? Do machine learning application papers in social computing report where human-labeled training data comes from?", https://arxiv.org/abs/1912.08320.

It is no surprise that IBM,[175] Microsoft[176] and Amazon[177] have recently issued statements that they will not offer their face recognition technologies to the police anymore. Many US states are considering banning artificial face recognition or have already implemented legislation that limits or prohibits it.[178] There is therefore a growing consensus in Western countries that even public interests cannot justify pervasive mass surveillance systems that exploit the web.

2.7. The media market: Big data-driven market strategies

Big data has revolutionised the universe of media. Many players in the media industry now depend on big tech companies to better connect with their audiences.[179] In fact, gathering and processing huge amounts of data in a fruitful way requires capabilities that few own. The pool of companies that can harvest big data is very limited, and the majority of market players rely on this pool to better understand who their clients are, what type of market strategy they should implement or how to gain more visibility. Some big tech companies in the field, such as Amazon, even produce media content themselves. Thanks to their technological capabilities, big tech companies thus now operate either (or both) as media makers and as mediators between the media industry and its consumers.

The Court of Justice of the European Union's landmark Google Spain case[180] encapsulates the paramount role that big tech companies now play in the news field and their resistance to the laws governing it. When an individual complained that a Google search of his name returned a list of results at the top of which was a very old newspaper item about him that could still ruin his reputation, Google's first line of defence was that it did not handle personal data; it only connected searches with results.[181] In other words, Google made the argument that it was not responsible for what it made available through Google search. The court responded with a historical judgement, showing its awareness of the unique role of Google in Internet searches. It found that Google was responsible for how it ranked its answers to a query, as it could resurrect long forgotten pieces of information that would not have been accessible to the general public otherwise.

---

[175] Krishna A., "IBM CEO's Letter to Congress on Racial Justice Reform", 8 June 2020, https://www.ibm.com/blogs/policy/facial-recognition-susset-racial-justice-reforms/.

[176] Greene J. Microsoft won't sell police its technology, following similar moves by Amazon and IBM", *The Washington Post*, 11 June 2020, https://www.washingtonpost.com/technology/2020/06/11/microsoft-facial-recognition/.

[177] Hao K., "The two-year fight to stop Amazon from selling face recognition to the police", *MIT Technology Review*, 12 June 2020, https://www.technologyreview.com/2020/06/12/1003482/amazon-stopped-selling-police-face-recognition-fight. See also Hartzog W., *op. cit.*, p. 76-77.

[178] See the Illinois Biometric Information Privacy Act, https://www.termsfeed.com/blog/bipa/.

[179] Tsesis T., *op. cit.*, p. 1589.

[180] *Google Spain SL et al. v. Agencia Española de Protección de Datos*, C-131/12, https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:62012CJ0131&from=EN.

[181] Ibid., para. 22.

Big tech companies do not simply populate the media market. They deeply affect its dynamics, too. Their unique ability to profile the market entraps their users in a "lock in" phenomenon and generates a quasi-market monopoly.[182] They are so pervasive and indispensable that those who do not want to use them often have to leave the market altogether. Many Internet users know that "visiting a single website results typically in the disclosure of browsing behaviour to over 100 third parties who seek to limit their own legal liability by means of dense 'privacy policies' which can run to hundreds of pages", but they cannot avoid visiting the same websites time and again.[183] The few companies that exploit the potentials of big data may patrol their territories even further by engaging in "killer acquisitions", through which they purchase innovative start-ups to either mine the data they have collected[184] or protect their dominant position.[185] In Frank Pasquale's words, like "Pharaoh trying to kill off the baby Moses", big tech companies can deny their rivals "the chance to scale".[186]

The simultaneous presence of more than one company that uses big data does not ensure that a market is competitive.[187] Big data can help the development of market strategies, including pricing, that benefit the competitors, not the customers. There is evidence that algorithms of different companies can maximise pricing through an implicit collusive strategy, simply by processing information about the market itself.[188] An algorithm can suggest a company raise prices because it predicts that its competitors will decide to do the same. Thanks to user profiling and clustering, they can also "segment … the market" and charge each user according to their willingness to pay. These practices create the "maximum revenue [for firms] but no consumer welfare".[189] Such a data-driven market strategy is usually not punishable, as there is no collusion, but has the benefits that normally attach to collusive behaviours.[190]

---

[182] AGCM, AGCOM, and Garante per la protezione dei dati personali, Indagine conoscitiva sui Big Data, *op. cit*., p. 26 and 78.

[183] European Data Protection Supervisor, Opinion 3/2018 EDPS Opinion on online manipulation and personal data, *op. cit*., p. 7.

[184] Zuboff S., *The Age of Surveillance Capitalism*, Profile Books, 2019, pp. 102-103.

[185] AGCM, AGCOM, and Garante per la protezione dei dati personali, "Indagine conoscitiva sui Big Data", *op. cit*., p. 81, See also Hughes C., *op. cit*.

[186] Pasquale F., *The Black Box Society*, Harvard University Press, 2015, p. 67.

[187] European Data Protection Supervisor, Opinion 3/2020 on the European strategy for data, op. cit., p. 8 (where it is warned against the creation or reinforcement of "situations of data oligopoly").

[188] Den Boer A. V., "Dynamic pricing and learning: Historical origins, current research, and new directions", *Surveys in operations research and management science*, 20, 2015, p. 1, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2334429); AGCM, AGCOM, and Garante per la protezione dei dati personali, "Indagine conoscitiva sui Big Data", op. cit.

[189] European Data Protection Officer, Opinion 8/2016 EDPS Opinion on coherent enforcement of fundamental rights in the age of big data, op. cit., p. 6.

[190] Harrington, J. E. Jr., "Developing competition law for collusion by autonomous artificial agents", *Journal of Competition Law & Economics*, 14, 2019, pp. 349-351, https://academic.oup.com/jcle/article-abstract/14/3/331/5292366?redirectedFrom=fulltext.

2.8. Regulatory approaches to AI-based systems

Many have voiced the need for new regulatory schemes in order to ensure that AI is utilised in a way that respects the rule of law, fundamental rights and ethical values. Big tech companies have long resisted public efforts to regulate the field,[191] but now appear to have come to terms with the necessity of constraining AI, although they push for company self-regulation rather than state rules.

Most constraints, however, do not aim to depress the utilisation of AI; in fact, they are expected to boost its role by making it more trustworthy and reliable.[192] There is wide consensus, in fact, that AI needs to be "lawful" (law-compliant), "ethical" (committed to respecting ethical principles and values) and "robust" (technologically and sociologically safe), in order to successfully integrate with human societies.[193]

Debates often emphasise that big data analyses need a new approach to legal regulation. Traditional tools may not be sufficient to ensure that the world of big data respects basic human values. Because of AI's black box structure and large-scale effects, legal sanctions are hardly capable of constraining big data-based technologies and strategies. Lawsuits may arrive late, when one's reputation or a company is in ruins, and liabilities may be hard to locate. AI needs to incorporate legal values within its data processing, in order to make sure that it protects them while it is operating.

Because of the wealth of information it gathers, its pervasive deployment and its capacity to replace human operators with robots, AI also poses ethical questions. *Digital ethics* is a new frontier for AI regulation and has drawn considerable attention especially in the US, in Canada and in Europe, where ethical codes have mushroomed.[194] As a field, digital ethics covers a wealth of topics, including "moral problems relating to *data* and *information* ..., *algorithms* ... and corresponding *practices* and *infrastructures*",[195] in a way that cuts across different disciplines and perspectives. Albeit extremely lively, the situation is magmatic at the moment, also because of the difficulties in drawing lines between the legal and the ethical components of AI regulation.[196]

---

[191] Zuboff S., op. cit., p. 105.

[192] Van Dijk N. & Casiraghi S., "The ethification of privacy and data protection law in the European Union: The case of artificial intelligence", *Brussels Privacy Hub*, 6, 22, May 2020, p. 5, https://brusselsprivacyhub.eu/publications/BPH-Working-Paper-VOL6-N22.pdf.

[193] High-Level Expert Group on Artificial Intelligence, Ethics Guidelines for Trustworthy AI, p. 2, https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai. See N. van Dijk & S. Casiraghi, *op. cit.*, p. 14.

[194] Jobin A., Ienca M. and Vayena E., "The global landscape of AI ethics guidelines", *Nature Machine Intelligence*, 1, 2019, pp. 393-395, https://www.nature.com/articles/s42256-019-0088-2.

[195] Floridi L., *op. cit.*, p. 3.

[196] For example, see the Council of Europe's *Recommendation CM/Rec(2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems*, 8 April 2020, https://search.coe.int/cm/pages/result_details.aspx?objectid=09000016809e1154, which showcases the variety of regulatory layers necessary for the development of sound AI-based systems.

2.9. Conclusion

Big data is a big reason for the societal, economic, and political success of AI. Processing vast amounts of data is crucial for big tech companies. It has not been just a blessing, however, and it requires people working in the field to take action to ensure that AI is beneficial to human beings.[197] Chris Hughes, co-founder of Facebook, has warned that the digitalisation of the economy may contribute to what he perceives to be "a decline in entrepreneurship, stalled productivity growth, and higher prices and fewer choices for consumers".[198] The stakes are so high that a member of the National Assembly, the lower house of the French Parliament, has even submitted a proposal to entrench a *Charter of artificial intelligence and of algorithms* within the preamble of the French constitution, to better protect human rights.[199]

AI maximises people engagement. Eliciting "as much response as possible from as many people as possible" is a key factor of success, as it provides feedback and allows companies to adjust their business plans and models to their customers in real time.[200] Political players and social influencers exploit this phenomenon by triggering emotional responses from their potential audience. Big data politics and economy place media at the centre stage, as they spread news, gather information, process emotions, and connect social spheres.

Big data aggrandises the role of the media for contemporary societies. Companies, politicians, influencers and other political figures exploit big data to market their ideas, agendas and opinions, as well as to shape their audiences.[201] Internet platforms allow legacy media to spread their content and generate new competition between traditional and new outlets.

Media players can also play a negative role. Through profiling the "thinking patterns and psychological makeup," they can deliberately misinform and mislead an audience.[202] Moreover, in countries where few media players operate, or where there are only or almost exclusively state-run social media,[203] a political regime can effectively control the news and also how people react to it, by disseminating fabricated favourable feedback and insulating unfavourable comments.[204] Within the scenario generated by big

---

[197] See the Asilomar Principles, developed in conjunction with the 2017 Asilomar conference. *Future of Life Institute*, https://futureoflife.org/ai-principles.

[198] See also Hughes C., *op. cit.*

[199] http://www.assemblee-nationale.fr/dyn/15/textes/l15b2585_proposition-loi.

[200] Akin Unver H., "Artificial intelligence, authoritarianism and the future of political systems", Centre for Economics and Foreign Policy Studies, July 2019, p. 3, https://edam.org.tr/wp-content/uploads/2018/07/AKIN-Artificial-Intelligence_Bosch-3.pdf.

[201] Idem.

[202] European Data Protection Supervisor, Opinion 4/2015. Towards a new digital ethics, *op. cit.*, p. 7.

[203] Pasquale F., *op. cit.*, p. 10, notes that "the distinction between state and market is fading" because of massive AI deployment in strategic sectors of public and private interest.

[204] Akin Unver H., *op. cit.*, p. 8. See also Meaker M., "How governments use the Internet to crush online dissent", *The Correspondent*, 27 November 2019, https://thecorrespondent.com/142/how-governments-use-the-internet-to-crush-online-dissent/18607103196-db0c0dab.

data, media can discharge a critical role in protecting democracy, equality, minority groups and open societies - or in undermining them.[205]

Finally, mass surveillance can have a chilling effect on creativity and innovation. Despite earlier expectations that AI would simply boost inventiveness,[206] some have detected "a tendency to discourage or penalise spontaneity, experimentation or deviation from the statistical 'norm', and to reward conformist behaviour".[207]

The vast deployment of AI nowadays requires that the media sphere become aware of its unique role. The media sector should strive to use AI in a lawful, ethical, and robust way. Thanks to their connecting role, the media could encourage the wider world of AI-based businesses to embrace the same values and become lawful, ethical, and robust. In particular, an ethical commitment may encourage media platforms to go beyond a merely passive role. While many regulations limit providers' legal liability for the content they host,[208] and more burdens imposed on media have not succeeded in encouraging more policing, it can still be a worthwhile ethical goal for media platforms to patrol their content.[209]

[205] High-Level Expert Group on Artificial Intelligence, "Ethics guidelines for trustworthy AI, *op. cit.*, p. 11.
[206] Perritt, H. H., Jr., *op. cit.*, p. 107.
[207] European Data Protection Supervisor, Opinion 4/2015. Towards a new digital ethics, *op. cit.*, p. 9. See also Pan S. B., *op. cit.*, p. 257 ("The goal of big data is to generalize") and Pasquale F., *op. cit.*, p. 188.
[208] Perritt H. H., Jr., *op. cit.*, p. 149.
[209] ERGA2020 Subgroup 1 – Enforcement, ERGA Position Paper on the Digital Services Act, p. 6, https://nellyo.files.wordpress.com/2020/06/erga_sg1_dsa_position-paper_adopted-1.pdf.

Freedom of expression, diversity and pluralism

One specific issue of concern raised by the use of AI relates very particularly to the media field: diversity and pluralism. And yet, for those who are old enough to remember the times when TV channels in a given country could be counted on the fingers of one hand and newspapers were called papers for a reason, the problem of diversity (at least in quantitative terms) might seem a bit exaggerated. Nowadays, there are scores and scores of TV channels and any newspaper on the globe is only one click away. It could actually be said that the only thing preventing anybody today from getting all the information in the world is not algorithms but rather paywalls. But precisely because the information offering is so overwhelmingly broad, people look for filters. And as mentioned before, filtering is one thing that AI does very well. Video on demand or news services can carry out this news personalisation for any Internet user, based on his or her personal viewing, reading history or other preferences. This has a downside: the so-called filter bubbles that occur when algorithms filter out "facts and different viewpoints, thereby reinforcing deeply held viewpoints and even prejudices".[210] The existence and effects of such filter bubbles are, however, not something everybody agrees upon. In her contribution to this publication, ***Mira Burri****, while acknowledging some of the precarious implications of tailored media on diversity and the need to pay attention to the power of platforms, voices also doubts about their direct link with a fragmentation of the public discourse and possible polarization of views.[211] Even promoters of the filter bubble thesis admit that they cannot prove its existence in real life[212] and that the empirical evidence of these bubbles is so far scarce.[213]* ***Sarah Eskens****, in her contribution to this publication, notes: "[T]he current challenge for news media and public authorities is to develop journalistic codes of ethics, self-regulatory standards, and possibly government regulation to contain the risks of AI for freedom of expression, while enabling AI to contribute to public debate, media pluralism, the free flow of information, and other societal goals".*

---

[210] See Andrea Pin's contribution to this publication.

[211] For other critical views on this matter see e.g. Bruns A., "It's Not the Technology, Stupid: How the 'Echo Chamber' and 'Filter Bubble' Metaphors Have Failed Us", http://snurb.info/node/2526.

[212] Zuiderveen Borgesius, F., Trilling, D., Moeller, J., Bodó, B., de Vreese, C. H., & Helberger, N., "Should we worry about filter bubbles?" Internet Policy Review, 5(1). https://doi.org/10.14763/2016.1.401.

[213] Helberger N., Eskens S., van Drunen M., Bastian M., Moeller J., Implications of AI-driven tools in the media for freedom of expression, https://rm.coe.int/cyprus-2020-ai-and-freedom-of-expression/168097fa82.

3. Implications of the use of artificial intelligence by news media for freedom of expression

Sarah Eskens, University of Amsterdam

3.1. Introduction

News media are increasingly using artificial intelligence in their businesses. In 2018, the Reuters Institute for the Study of Journalism surveyed almost 200 leaders in journalism. Almost three quarters of the leaders surveyed said they were already using AI in their organisation.[214]

The use of AI creates opportunities for news media and may help them fulfil their democratic role. A report by the European Broadcasting Union also underlined the opportunities of AI for public service journalism.[215] Accordingly, the use of AI by news media may fall within the scope of the protection of freedom of expression for the media, which is important considering the push to regulate AI in various domains. At the same time, the use of AI by news media may affect the extent to which other participants in public debate can exercise their freedom of expression rights. For example, news organisations can use AI to automatically moderate comments on their websites. If automated moderation is biased towards, for example, general American English, then certain voices in public debate might not be heard.

As lawmakers are discussing the need to regulate AI, the question arises to what extent the use of AI by news media can be regulated and how media freedom should be balanced with other rights and interests associated with the use of AI by news media. The use of AI by news media shapes our information environment and may have significant effects on open debate, media pluralism and diversity, the free flow of information, and

[214] Newman N., "Journalism, media, and technology trends and predictions 2018". Reuters Institute for the Study of Journalism, p. 29, https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2018-01/RISJ%20Trends%20and%20Predictions%202018%20NN.pdf.

[215] European Broadcasting Union, "The next newsroom: Unlocking the power of AI for public service journalism", https://www.ebu.ch/publications/news-report-2019.

other public values attached to the institution of the news media.[216] The questions raised by the use of AI by news media are unique due to the democratic role of the news media. This chapter therefore focuses on the use of AI by news media and not on the use by other types of media, such as entertainment media, which requires a different balancing of interests.[217]

This chapter is set up as follows: it describes the general framework for the protection of freedom of expression; it discusses to what extent the use of AI falls under media freedom, which kind of news actors that use AI can benefit from media freedom, and what duties and obligations news media have when they use AI; thereafter, it describes how certain applications of AI in the news media can limit the freedom of expression rights of other participants in public debate, including news users and citizens or politicians who make themselves heard via the media. The purpose is not to offer an exhaustive overview of all the risks of the use of AI by the news media. Rather, the aim is to illustrate the risks of AI for freedom of expression in order to discuss the substance of the human rights of various participants to public debate. Finally, the chapter analyses what kind of obligations states have regarding freedom of expression in the face of the use of AI by news media. But before providing these analyses, it briefly sets out what goals news media have when they use AI.

3.2. AI applications for news media

Similarly to its definition in other studies about AI and the news media, for the purpose of this chapter, artificial intelligence is loosely defined as "a collection of ideas, technologies, and techniques that relate to a computer system's capacity to perform tasks normally requiring human intelligence".[218] AI is thus an umbrella term that refers to various digital technologies, including, among others, machine learning, image recognition, natural language processing, and natural language generation. The other chapters in this publication showcase the variety of applications for media that are considered to be AI.

In this chapter, four goals are distinguished for which news media can use AI: newsgathering; news production; news distribution; and moderation of reader

[216] Council of Europe, Recommendation CM/Rec(2018)1 of the Committee of Ministers to member States on media pluralism and transparency of media ownership, https://search.coe.int/cm/Pages/result_details.aspx?ObjectId=0900001680790e13#_ftn1;
Council of Europe, Recommendation CM/Rec(2015)6 of the Committee of Ministers to member States on free, transboundary flow of information on the Internet, https://search.coe.int/cm/Pages/result_details.aspx?ObjectID=09000016805c3f20.
[217] See Chapter 4 of this publication.
[218] Beckett C., "New powers, new responsibilities: A global survey of journalism and artificial intelligence", London School of Economics and Political Science, p. 16, https://blogs.lse.ac.uk/polis/2019/11/18/new-powers-new-responsibilities/.

comments.[219] The first three goals relate to classic journalistic processes, and the fourth goal relates to the fact that online news media sometimes allow user comments on their websites. As remarked in a recent report by the Parliamentary Assembly, most of the discussions about the use of AI in online communication processes has focused on content moderation, while the manner in which AI shapes the online information environment is equally important.[220]

To begin the journalistic process, news media can use AI for newsgathering. This includes the use of AI to find information and newsworthy events, generate story ideas, and monitor events or issues. Once journalists have gathered information on potential stories, they can use AI for the production of news. This includes the use of AI for writing news items (sometimes called "automated journalism"),[221] creating images and videos, fact-checking information, or repurposing content for new audiences. In the final step of the journalistic process, news media can use AI for the distribution of news. This includes the use of AI for providing personalised recommendations, finding new audiences, marketing the news brand, and selling subscriptions. The use of personalisation often has a dual goal. Personalisation helps news media better serve their users and, especially for public service media, fulfil their public remit.[222] Personalisation also helps news media retain subscribers, increase user engagement, and consequently generate sales and advertising revenues.[223]

As news media have opened their online platforms for reader comments, they can use AI to more effectively moderate these comments. For instance, The *New York Times* implemented a system that uses machine learning to prioritise comments for moderation and automatically approve comments.[224] The *New York Times* moderates almost 12 000 comments per day and the automated system allows the comment section to be open for longer and approve comments faster.

Other research about the use of AI in the news media sector discusses the use of AI for comment moderation as part of news distribution.[225] However, automated comment moderation brings specific risks for the freedom of expression rights of the people who are commenting on news stories and engaging in public debate. For the use of AI for other distribution goals, such as personalisation and marketing, news users are

[219] Beckett C., Ibid, p. 20.
[220] Parliamentary Assembly, Report: Need for democratic governance of artificial intelligence (24 September 2020), para. 18-19, https://pace.coe.int/en/files/28742/html.
[221] Dörr K.N., "Mapping the field of algorithmic journalism", *Digital Journalism* 4(6), pp. 700–722, https://doi.org/10.1080/21670811.2015.1096748.
[222] Van den Bulck H. and Moe H., "Public service media, universality and personalisation through algorithms: Mapping strategies and exploring dilemmas", *Media, Culture & Society* 40(6), pp. 875–92, https://doi.org/10.1177/0163443717734407.
[223] Bodó B., "Selling news to audiences: A qualitative inquiry into the emerging logics of algorithmic news personalization in European quality news media", *Digital Journalism* 7(8), pp. 1054–75, https://doi.org/10.1080/21670811.2019.1624185.
[224] Etim B., "The Times sharply increases articles open for comments, using Google's technology", *The New York Times*, https://www.nytimes.com/2017/06/13/insider/have-a-comment-leave-a-comment.html.
[225] Beckett C., Ibid, p. 28.

addressees of the communication but they are not themselves active speakers. This chapter therefore considers AI for comment moderation on news platforms as a separate category. After this overview of the various uses of AI in the news media, the next section discusses to what extent the use of AI is protected by media freedom.

3.3. The use of AI by news media as an element of media freedom

In Europe, the European Convention on Human Rights (ECHR) provides the legal basis for the human right to freedom of expression. Article 10(1) ECHR provides that everyone has the right to freedom of expression, which includes freedom to hold opinions and to receive and impart information and ideas without interference by public authorities.

The European Court on Human Rights (ECtHR) was set up to ensure that states comply with their obligations under the ECHR. The ECtHR has produced a huge body of case law in which it interprets and develops the right to freedom of expression. In one of its first cases on freedom of expression, the ECtHR affirmed that freedom of expression is one of the foundations for a democratic society and for the development of every person.[226]

3.3.1. Democratic role of the news media

Journalism scholars have distinguished several democratic roles for news media.[227] The media are a source of information for democratic debate, by providing citizens information on politics and current affairs. Furthermore, the media function as the "fourth estate" by critically scrutinising the exercise of power by government, businesses, and other powerful actors. The media also are a mediator between citizens and politicians because they facilitate the existence of a public space in which citizens and politicians can communicate via letters, op-eds, broadcasted studio debates, and contributions to news articles.

The ECtHR has affirmed these various democratic roles of the media. The ECtHR has considered that the news media have the task to distribute information and be a public watchdog.[228] In this respect, the ECtHR has determined that freedom of expression protects both the gathering and publication of information,[229] and both the content of

[226] ECtHR, *Handyside v. the United Kingdom* [1976], 5493/72, para. 49, http://hudoc.echr.coe.int/eng?i=001-57499.
[227] McNair B., "Journalism and democracy" in T. Hanitzsch and K. Wahl-Jorgensen (eds.) *The Handbook of Journalism Studies*. Routledge, pp. 237–49.
[228] ECtHR, *Barthold v. Germany* [1985], 8734/79, para. 58, http://hudoc.echr.coe.int/eng?i=001-57432.
[229] ECtHR, *Dammann v. Switzerland* [2006], 77551/01, para. 52, http://hudoc.echr.coe.int/eng?i=001-75174.

communication and the technical means for the distribution and reception of information.[230] Additionally, the ECtHR has found that one task of the news media includes the creation of forums for public debate.[231]

The use of AI by news media fits within their democratic roles as protected by the right to freedom of expression. News media can use AI to gather information on new issues, for example by using AI to analyse big data. Furthermore, news media can use AI to distribute relevant information to different citizens, depending on each individual's personal interests and information needs. News media can also use AI to watch and monitor the behaviour of large corporations or the implementation of public policies. Finally, news media can use AI to improve the forum for public debate by automatically moderating reader comments.

Because of their democratic roles, the news media receive special freedom of expression protection. The ECtHR has held that freedom of expression is of particular importance as far as the news media are concerned.[232] In the case of the news media, the ECtHR therefore speaks of "freedom of the press",[233] which is also called media freedom.[234] The ECtHR has found that public authorities have a smaller margin of appreciation to decide if there is a pressing social need to interfere with media freedom,[235] compared to the margin of appreciation that public authorities have when they interfere with the freedom of expression of other types of speakers. Furthermore, the ECtHR has determined that media freedom protects the news media against influence from powerful economic or political groups in society, and ensures their editorial freedom.[236] As part of the gathering of information, media freedom protects journalistic sources,[237] and media may have a right to access information held by public authorities.[238] To the extent that the use of AI falls under media freedom, public authorities are thus limited in the regulation of AI. The next section discusses which actors can enjoy media freedom.

3.3.2. Beneficiaries of media freedom

These days, the news media environment is formed by a complex network of different actors, including news publishers, news users, and online intermediaries. Legacy news

[230] ECtHR, *Autotronic AG v. Switzerland* [1990], 12726/87, para. 47, http://hudoc.echr.coe.int/eng?i=001-57630.
[231] ECtHR, *Társaság a Szabadságjogokért v. Hungary* [2009], 37374/05, para. 27, http://hudoc.echr.coe.int/eng?i=001-92171.
[232] ECtHR, *The Sunday Times v. the United Kingdom (No. 1)* [1979], 6538/74, para. 65, http://hudoc.echr.coe.int/eng?i=001-57584.
[233] ECtHR, *The Sunday Times v. the United Kingdom (No. 1)*, para. 66.
[234] Oster J., *Media Freedom as a Fundamental Right*. Cambridge University Press, p. 48.
[235] ECtHR, *Busuioc v. Moldova* [2004], 61513/00, para. 65, http://hudoc.echr.coe.int/eng?i=001-67745.
[236] ECtHR, *Manole and Others v. Moldova* [2009], 13936/02, para. 98, http://hudoc.echr.coe.int/eng?i=001-94075.
[237] ECtHR [GC], *Goodwin v. the United Kingdom* [1996], 17488/90, para. 39, http://hudoc.echr.coe.int/eng?i=001-57974.
[238] Ibid..

media and digital-born news media gather original information and publish news articles via their own offline and online news outlets. Before the Internet, news publishers reached their audiences directly on the whole, when people bought a certain newspaper or tuned in to a certain radio or television channel. On the Internet, news publishers can still reach their audiences directly, when people browse to news websites or use apps of news publishers. But people are accessing and finding news increasingly via social media, search engines, and news aggregators.[239] These platforms function as intermediaries between news publishers and news users. Rather obviously, both traditional news media and digital-born news media qualify for media freedom. But the involvement of online intermediaries in the news environment raises the question to what extent these intermediaries can also rely on media freedom when they use AI.

The ECtHR has determined that various actors can fulfil the democratic roles that the media traditionally perform. The ECtHR remarked that there is a strong public interest in enabling campaign groups to contribute to public debate by distributing information on matters of public interest.[240] The ECtHR has therefore analysed the conduct of public authorities with regard to campaign groups in the light of media freedom (but it has also held that campaign groups are expected to meet certain duties and responsibilities typically reserved for the media; see next section). In one case, the ECtHR considered that the creation of forums for public debate is not limited to professional news media and that non-governmental organisations may also fulfil that role.[241] The ECtHR has therefore characterised NGOs as "social watchdogs". The activities of civil society organisations may thus warrant similar ECHR protection as that afforded to the news media.[242] These judgements of the ECtHR could provide a basis to build on to also recognise online intermediaries as actors comparable to the news media, depending on the societal role they play.

As discussed in the previous section, freedom of expression law recognises three democratic roles for the news media: providing information to the public; creating a forum for public debate; and acting as a watchdog. Online intermediaries can fulfil in particular two of these roles, aided by AI. Online intermediaries can increase the accessibility and findability of information via personalised news feeds and easy access to a range of news publishers. Furthermore, online intermediaries can create forums for public debate by allowing news publishers, politicians, and citizens to post content on their platforms in public and private groups and affording different forms of engagement with online content, including 'liking', commenting, and forwarding content. Online intermediaries can thus play roles similar to the news media and may have, just like the media, a gatekeeping and agenda-setting function. The Committee of Ministers of the Council of Europe has also remarked that online intermediaries may "exert forms of

[239] Newman N. et al., "Digital News Report 2020", Reuters Institute for the Study of Journalism, , pp. 11–12, https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2020-06/DNR_2020_FINAL.pdf.
[240] ECtHR, *Steel and Morris v. the United Kingdom* [2005], 68416/01, para. 89, http://hudoc.echr.coe.int/eng?i=001-68224.
[241] ECtHR, Társaság a Szabadságjogokért v. Hungary, para. 27.
[242] Ibid.

control which influence users' access to information online in ways comparable to media, or they may perform other functions that resemble those of publishers".[243]

On the basis of these freedom of expression principles, one could argue that online intermediaries may qualify for media freedom when they are using AI to perform democratic roles similar to those of the news media. At the least, online intermediaries may qualify for "normal" freedom of expression rights when they are making news more easily accessible through news feeds and search results. In that regard, Van Hoboken distinguishes the production of "information about information" by search engines, such as when they publish search results, from the referencing to information elsewhere. Van Hoboken concludes that the publication of search results by a search engine is protected under Article 10 ECHR.[244] In a similar manner, one could argue that the AI-driven selection, ranking, and personalisation of news feeds by social media and news aggregators is the production of "information about information" and deserves freedom of expression protection.

When traditional news media, digital-born news media, and ultimately online intermediaries, exercise their media freedom or freedom of expression, they also assume certain duties and obligations. The next section discusses these duties and obligations.

3.3.3. Duties and responsibilities and journalistic codes of ethics

While the first paragraph of Article 10 ECHR guarantees the human right to freedom of expression, the second paragraph lays down that the exercise of freedom of expression "carries with it duties and responsibilities" and may therefore be subject to restrictions as are prescribed by law and are necessary in a democratic society for a legitimate aim. In other words, Article 10 ECHR contains a mechanism to ensure that people and organisations exercising freedom of expression do so in a responsible manner. For the purpose of this chapter, the focus is on duties and responsibilities and not on the conditions under which interference with the right to freedom of expression may be justified. The question is what are the duties and responsibilities that come with freedom of expression and what do these duties and responsibilities mean for the use of AI by the news media.

Various actors have duties and responsibilities when they participate in or contribute to the exercise of freedom of expression. The ECtHR has held that in addition to speakers or authors themselves, persons or organisations providing other authors a

[243] Council of Europe, Recommendation CM/Rec(2018)2 of the Committee of Ministers to member States on the roles and responsibilities of internet intermediaries, para. 5, https://search.coe.int/cm/Pages/result_details.aspx?ObjectID=0900001680790e14.

[244] Van Hoboken J. V. J., "Search engine freedom: On the implications of the right to freedom of expression for the legal governance of Web search engines", University of Amsterdam, p. 182, http://hdl.handle.net/11245/1.392066.

medium or platform, such as publishers[245] or Internet news portals,[246] take on duties and responsibilities regarding the publication and distribution of third-party content.

The ECtHR has determined that the scope of an actor's duties and responsibilities depends on various factors. First of all, someone's duties and responsibilities depend on their situation and the technical means they use for communication.[247] As news media have a special democratic role, their duties and responsibilities also assume a special significance. The ECtHR has stipulated that the duties and responsibilities of news media are specifically important when their work might undermine the rights of others.[248] Furthermore, the potential impact of the medium forms a factor to determine the scope of duties and responsibilities.[249] The more impactful the medium, the more weight the duties and responsibilities of a news media actor retain. Still, the ECtHR has held that the duties and responsibilities of Internet news portals may differ to some degree from those of traditional publishers as regards third-party content.[250] In a similar manner, the Committee of Ministers of the Council of Europe has recommended that the duties and responsibilities of online intermediaries should, given the multiple roles they play, be determined with respect to the specific services and roles they perform.[251]

The ECtHR has held that the duties and responsibilities of news media mean they should act "in good faith in order to provide accurate and reliable information in accordance with the ethics of journalism".[252] Journalist codes of ethics existed already long before the ECHR introduced the idea that the exercise of freedom of expression comes with duties and responsibilities. In the 1920s, the American Society of Newspaper Editors adopted the Canons of Journalism, which is seen as one of the first codes of ethics for the news media.[253] But in the years thereafter, criticism about the corporate press created pressure to subject the news media to government regulation.[254] In response, the news media developed codes of ethics, press councils, ombudsmen and other forms of self-regulation to prevent regulation by the government.[255]

In Europe, the idea that the news media should not be regulated holds mainly for the printed press. European countries regulate audiovisual media in several ways, most notably through the EU Audiovisual Media Services Directive. The manner in which the

---

[245] ECtHR, *Éditions Plon v. France* [2004], 58148/00, para. 50, http://hudoc.echr.coe.int/eng?i=001-61760; ECtHR, *Chauvy and Others v. France* [2004], 64915/01, para. 79, http://hudoc.echr.coe.int/eng?i=001-61861.
[246] ECtHR, *Magyar Tartalomszolgáltatók Egyesülete and Index.hu Zrt v. Hungary* [2016], 22947/13, para. 62, http://hudoc.echr.coe.int/eng?i=001-160314; ECtHR [GC], *Delfi AS v. Estonia* [2015], 64569/09, para. 113, http://hudoc.echr.coe.int/eng?i=001-155105.
[247] ECtHR, *Handyside v. the United Kingdom*, para. 49.
[248] ECtHR [GC], *Bladet Tromsø and Stensaas v. Norway* [1999], 21980/93, para. 65, http://hudoc.echr.coe.int/eng?i=001-58369.
[249] ECtHR [GC], *Jersild v. Denmark* [1994], 15890/89, para. 31, http://hudoc.echr.coe.int/eng?i=001-57891.
[250] ECtHR [GC], *Delfi AS v. Estonia*, para. 113.
[251] Recommendation CM/Rec(2018)2 of the Committee of Ministers to Member States on the roles and responsibilities of Internet intermediaries, para. 11.
[252] ECtHR [GC], *Bladet Tromsø and Stensaas v. Norway*, para. 65.
[253] Ward S.J.A., *The invention of journalism ethics: The path to objectivity and beyond*. MQUP, pp. 236–37.
[254] Ward S.J.A., Ibid, p. 244.
[255] Ward S.J.A., Ibid, p. 245.

ECtHR has interpreted the notion of duties and responsibilities now also creates a normative legal basis for a social responsibility theory of the printed press, under which the printed press should commit to self-regulation and codes of ethics for their profession.

The question is if current journalistic codes of ethics are fit to deal with the use of AI by news media. In 2001, journalism scholars concluded that the Internet creates new ethical issues for journalists while traditional journalistic codes of ethics provided insufficient guidance for the conduct of news media in the online environment.[256] More than a decade later, research into journalistic codes of ethics has found that the majority of codes still do not include rules for online journalism and digital media.[257] Helberger and Bastian therefore call for the development of algorithmic journalistic ethics, to guide the news media in their use of AI for the production, publication, and distribution of news.[258] Similarly, Dörr and Hollnhuchner, two communication science scholars, conclude that media organisations should adopt ethical codes for algorithmic journalism.[259]

Another question is to what extent online intermediaries that use AI to shape the online environment for freedom of expression should follow codes of ethics. Social media and search engines have long tried to escape responsibility for the manner in which they select and prioritise information by arguing they are not media. In an interview, the CEO of Facebook stressed that Facebook is "a social network" and that he prefers that term over "social media" because the notion of a social network focuses on the "people part" of the platform and less on the content part of it.[260] If social media are indeed not journalistic entities, then they do not have to follow journalistic codes of ethics.

It is apparent, then, that Article 10 ECHR as interpreted in case law of the ECtHR provides a normative legal basis for duties and responsibilities for actors contributing to freedom of expression, regardless of whether or not they are "real" news media organisations. At the same time, it also becomes clear that the duties and responsibilities of online intermediaries may differ from those of traditional news media. This means that online intermediaries cannot be obliged to follow journalistic codes of ethics, although freedom of expression principles make clear that online intermediaries have responsibilities when they use AI to regulate expression on their platforms and exercise their own freedom of expression rights. If online intermediaries do not develop adequate codes and other instruments of self-regulation, then governments may justifiably regulate the manner in which online intermediaries exercise their freedom of expression while

---

[256] Deuze M. and Yeshua D., "Online journalists face new ethical dilemmas: Lessons from the Netherlands", *Journal of Mass Media Ethics* 16(4), pp. 273–92, https://doi.org/10.1207/S15327728JMME1604_03.

[257] Díaz-Campo J. and Segado-Boj F., "Journalism ethics in a digital environment: How journalistic codes of ethics have been adapted to the Internet and ICTs in countries around the world", *Telematics and Informatics* 32(4), pp. 735–44, https://doi.org/10.1016/j.tele.2015.03.004.

[258] Helberger N. and Bastian M., "AI, algorithms and journalistic ethics", presented at the Future of Journalism conference, Cardiff, 2019.

[259] Dörr K. N., "Mapping the field of algorithmic journalism", *Digital Journalism* 4(6), pp. 700–722, https://doi.org/10.1080/21670811.2015.1096748.

[260] Swisher K., "Zuckerberg: The Recode interview", *Recode*, https://www.recode.net/2018/7/18/17575156/mark-zuckerberg-interview-facebook-recode-kara-swisher.

using AI. A similar line of thinking is also evident in the way in which the European Commission approaches online intermediaries. The European Commission threatens to regulate online intermediaries if they do not adhere to self-regulatory codes for dis- and misinformation. A similar approach could be taken regarding the use of AI by online intermediaries to shape the news environment.

After analysing the freedom of expression principles for the use of AI by news media and other actors playing a role in the online news environment, the next question is how the use of AI by news media affects the freedom of expression rights of other participants in public debate.

3.4. Implications of AI for the freedom of expression rights of news users and other participants in public debate

The use of AI by news media for newsgathering, production, distribution, and moderation presents various risks for the freedom of expression rights of news users and other stakeholders in the news media environment. It is important to note that the use of AI by news media may also enhance the right to freedom of expression and the right to receive information of news users and participants in public debate. From a regulatory perspective, the question is thus how to contain the risks while allowing AI to have a positive effect on the news media environment and public debate.

In the newsgathering stage, news media can use AI to identify trends and facts in big data. Used in this way, AI can uncover original stories in big data that could not be seen by the human eye. However, the use of AI for newsgathering depends on the availability of (public) datasets. Events or societal issues that do not come with a large dataset may remain invisible to the gaze of the automated story discovery system.[261] The voices of the people implicated by events and stories that do not generate big data may stay out of the focus of data-driven news media and these voices may thereby remain unheard in the public debate.

In the news production stage, news media can use AI to generate texts and images, verify and fact-check information, automatically translate, write posts for social media, or tailor mass-produced stories to specific audiences. The use of AI for automated content production can lead to unlawful output that infringes the rights of others, such as hate speech, defamatory content, or copyright infringement. The question then arises who is accountable or liable for this unlawful content: the news organisation that decided to deploy the AI tool, the developer of the AI tool, or the AI itself?

[261] Hansen M. et al., "Artificial intelligence: Practice and implications for journalism", Tow Center for Digital Journalism, p. 17, https://doi.org/10.7916/D8X92PRD.

Legal scholars have argued that it is possible to hold AI agents accountable.[262] American legal scholars have also argued that the First Amendment, which guarantees the constitutional right to free speech, should protect automated speech.[263] However, from a positive law perspective, AI cannot have legal personhood and thus cannot be accountable or liable in European jurisdictions. In 2017, the European Parliament called on the European Commission to consider the creation of a specific legal status ("electronic person") for robots.[264] In its Communication about AI for Europe, the Commission did not mention such an electronic personality.[265] This omission suggests that so far, the European Commission does not intent to consider legal personhood for robots or AI entities under EU law.[266]

It would appear most appropriate to hold the news media organisation that decides to use AI accountable for unlawful content produced by AI. Reversely, if public authorities want to censor automatically generated content or bots that contribute to public debate, then these tools and their output could be protected both via the freedom of expression rights of the news media organisation as well as via the right to receive information of news users.[267]

In the news distribution stage, news media can use AI to personalise the news offering for each individual news user, which may engage the news users' right to receive information. The human right to freedom of expression as protected by Article 10 ECHR includes the right to receive information. The ECtHR has determined that the media have the task to provide information and ideas on issues of public interest and "the public also has a right to receive them".[268] In addition, the public has a right to be properly informed.[269] The right to receive information entails that the public should have access through the media to diverse information.[270] At the same time, news users do not have a subjective right to receive information from the media.[271] Still, the use of AI by news media can affect the enjoyment of the right to receive information, such as when personalisation decreases the diversity of information that people have access to. A

[262] Hage J., "Theoretical foundations for the responsibility of autonomous agents", *Artificial Intelligence and Law* 25(3), pp. 255–71, https://doi.org/10.1007/s10506-017-9208-7.
[263] Collins R.K.L. and Skover D. M., *Robotica: Speech rights & artificial intelligence*. Cambridge University Press.
[264] European Parliament, Report with recommendations to the Commission on civil law rules on robotics, para. 59, https://www.europarl.europa.eu/doceo/document/A-8-2017-0005_EN.html.
[265] European Commission, Communication from the Commission: Artificial intelligence for Europe, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52018DC0237.
[266] For an in-depth discussion about the legal status of AI concerning copyright law see Chapter 5 of this publication.
[267] Kaminski M.E., "Authorship, disrupted: AI authors in copyright and First Amendment law", *U.C. Davis Law Review* 51(2), pp. 589–616.
[268] ECtHR, *The Sunday Times v. the United Kingdom (No. 1)*, para. 65.
[269] ECtHR, *The Sunday Times v. the United Kingdom (No. 1)*, para. 66.
[270] ECtHR, *Manole and Others v. Moldova*, para. 100.
[271] Eskens S., Helberger N. and Moeller J., "Challenged by news personalisation: Five perspectives on the right to receive information", *Journal of Media Law* 9(2), pp. 259–84, https://doi.org/10.1080/17577632.2017.1387353.

limitation on the right to receive information caused by the conduct of news media may give rise to positive obligations for states (see next section).

When online intermediaries use personalisation, they may also implicate the freedom of expression rights of news media or citizen journalists posting news stories on their platforms to reach a wider audience. As one of the first in Europe, the German government therefore developed legal safeguards for media freedom and the visibility of news organisations in the face of personalisation on social media and search engines. The new German *Medienstaatsvertrag* provides that online intermediaries "may not unfairly disadvantage (directly or indirectly) or treat differently providers of journalistic editorial content to the extent that the intermediary has potentially a significant influence on their visibility".[272] This legal provision is a novelty for European media law, and although it has to be seen how the law works out in practice, digital rights organisations state that the new German legislation has "important symbolic value" and that its core goals are "laudable".[273]

Finally, news media can use AI to moderate user comments on their websites. This may engage the right to freedom of expression of people who post comments. Research shows that AI systems are more likely to classify social media posts in African American English as offensive compared to posts in general American English.[274] If AI-driven content moderation is biased against certain societal groups, then this may lead to unequal chances to communicate.

Other technical limitations for automated content analysis arise from the difficulties automated content moderation systems have in understanding the context of a reader comment, the lack of natural language processing tools trained in the domain in which they will be applied, and the underrepresentation of certain groups of speakers in the training data.[275] These technical limitations may lead to false positives and false negatives in the reviewing of comments. As Llansó and colleagues remark, false positives can put a burden on individuals' freedom of expression, while false negatives "can result in a failure to address hate speech, harassment, and other objectionable content that may

[272] Helberger N., Leerssen P. and van Drunen M., "Germany proposes Europe's first diversity rules for social media platforms", *Media@LSE*, https://blogs.lse.ac.uk/medialse/2019/05/29/germany-proposes-europes-first-diversity-rules-for-social-media-platforms/.

[273] Nelson, M., "Germany's new media treaty demands that platforms explain algorithms and stop discriminating. Can it deliver?", *AlgorithmWatch*, https://algorithmwatch.org/en/new-media-treaty-germany/.

[274] Sap M. et al., "The risk of racial bias in hate speech detection", in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 2019, pp. 1668–78, https://www.aclweb.org/anthology/P19-1163. The Parliamentary Assembly has highlighted that "[t]he use of biased datasets, or datasets that reflect historical bias, prejudice or discrimination, is a major cause of discrimination in AI"; see Parliamentary Assembly, Report: Preventing discrimination caused by the use of artificial intelligence (29 September 2020), para. 43, https://pace.coe.int/en/files/28715/html. The above example shows that these risks are equally present in the media field.

[275] Llansó E. et al., "Artificial Intelligence, content moderation, and freedom of expression", Institute for Information Law, pp. 7–8, https://www.ivir.nl/publicaties/download/AI-Llanso-Van-Hoboken-Feb-2020.pdf.

create a chilling effect on some individuals' and groups' willingness to participate online".[276]

In principle, the right to freedom of expression does not give people the right to speak on private platforms. The ECtHR has determined that freedom of expression "does not bestow any freedom of forum for the exercise of that right".[277] Still, the ECtHR also stated that if the bar on access to private property prevents any effective exercise of freedom of expression or destroys the essence of freedom of expression, then states may have a positive obligation to protect the enjoyment of freedom of expression rights by regulating property rights.[278] Furthermore, as discussed in the previous sections, when news media or online intermediaries are moderating comments, they are essentially shaping the forum for public debate. The creation of a forum for public debate is a democratic role, which comes with duties and responsibilities. Some news media have accepted these responsibilities regarding the moderation of comments. For instance, the *New York Times* investigated the way it automatically moderates user comments following research about how automated content moderation may discriminate, or reinforce biases.[279]

Following this discussion of freedom of expression principles, rights, and duties and responsibilities for multiple stakeholders in the news media environment, the final question is what obligations states have regarding freedom of expression in the face of the use of AI by news media.

3.5. Obligations of states regarding media freedom

The rights and freedoms in the ECHR are formulated as negative rights. The provisions prohibit public authorities from interfering with the rights and freedoms of individuals. For example, in the case of freedom of expression, Article 10(1) ECHR provides that everyone has the right to freedom of expression "without interference by public authority". The human rights in the ECHR thus contain negative obligations for states.

Over the years, the ECtHR has accepted that human rights in the ECHR may also give rise to positive obligations for states. In the 1960s, the ECtHR for the first time accepted the idea that states may have positive obligations under certain ECHR rights.[280] It took some years before the ECtHR read positive obligations in the right to freedom of expression. But in the 2000s, the ECtHR found that the right to freedom of expression may contain positive obligations for states, even in the sphere of relations between

[276] Llansó E. et al., Ibid p. 9.
[277] ECtHR, *Appleby and Others v. the United Kingdom* [2003], 44306/98, para. 47, http://hudoc.echr.coe.int/eng?i=001-61080.
[278] ECtHR, *Appleby and Others v. the United Kingdom*, para. 47.
[279] Salganik M.J. and Lee R.C., "To apply machine learning responsibly, we use it in moderation", *NYT Open*, https://open.nytimes.com/to-apply-machine-learning-responsibly-we-use-it-in-moderation-d001f49e0644.
[280] ECtHR, *Case 'relating to certain aspects of the laws on the use of languages in education in Belgium' v. Belgium* [1968], para. 27, http://hudoc.echr.coe.int/eng?i=001-57525.

individuals.[281] For example, the ECtHR held that states have a positive obligation to ensure that the public has access through news media to impartial, accurate, and diverse information, and that journalists can impart this information.[282] Additionally, the ECtHR observed that states have a positive obligation to adopt a solid legislative and administrative framework to guarantee pluralism in the audiovisual media market.[283]

In the case of *Dink v. Turkey*, the ECtHR formulated a particularly strong positive obligation for states. The ECtHR found that states are required to create a favourable environment for participation in the public debate by all the persons concerned, enabling them to express their opinions and ideas without fear.[284] The ECtHR repeated this statement in the case of *Khadija Ismayilova*.[285] These two cases concerned attacks and harassment of journalists. In the case of *Dink*, the ECtHR found that the state had a positive obligation to protect a journalist against attacks by people who felt insulted by his publications. In the case of *Khadija Ismayilova*, the ECtHR held that the state had a positive obligation to more effectively investigate intrusions into the private life of a journalist. McGonagle argues that the notion of a favourable environment has great potential.[286] Still, it is an open question how far the positive obligation of states to ensure an enabling environment for freedom of expression reaches. From a positive law perspective, it currently does not guard against the risks of AI for freedom of expression.

Although the ECtHR's recent reiteration of the requirement to ensure a favourable environment may not extend to the use of AI by the news media, the foregoing shows that states do have positive obligations that may be relevant regarding AI. When the use of AI by the news media diminishes the diversity of information that people receive, then the positive obligations of states may be engaged. More concretely, states may have a positive obligation to ensure that news users receive diverse information through AI-driven online news media in the event that personalisation and automated content moderation become so pervasive that they reduce the diversity of news media content that people receive. In its guidelines on media pluralism and transparency of media ownership, the Committee of Ministers of the Council of Europe also stresses that states should make efforts to ensure that "the broadest possible diversity of media content, including general interest content," is accessible to everyone.[287] These guidelines apply to the use of AI by online news media as well.

---

[281] ECtHR, *Özgür Gündem v. Turkey* [2000], 23144/93, para. 43, http://hudoc.echr.coe.int/eng?i=001-58508; ECtHR, *Fuentes Bobo v. Spain* [2000], 39293/98, para. 38, http://hudoc.echr.coe.int/eng?i=001-63608.
[282] ECtHR, *Manole and Others v. Moldova*, para. 100.
[283] ECtHR [GC], *Centro Europa 7 S.r.l. and Di Stefano v. Italy* [2012], 38433/09, para. 134, http://hudoc.echr.coe.int/eng?i=001-111399.
[284] ECtHR, *Dink v. Turkey* [2010], 2668/07, 6102/08, 30079/08, 7072/09, 7124/09, para. 137, http://hudoc.echr.coe.int/eng?i=001-100383.
[285] ECtHR, *Khadija Ismayilova v. Azerbaijan* [2019], 65286/13, 57270/14, para. 158, http://hudoc.echr.coe.int/eng?i=001-188993.
[286] McGonagle T., "Positive obligations concerning freedom of expression: Mere potential or real power?", in Andreotti O. (ed.) *Journalism at risk: Threats, challenges, and perspectives*, Council of Europe, pp. 9–35.
[287] Council of Europe, Appendix to Recommendation CM/Rec(2018)1, Guidelines on media pluralism and transparency of media ownership.

Furthermore, on a societal level, the use of AI by the news media and other actors in the news media environment may threaten media pluralism. Large news media and online intermediaries have access to more and better user data and more powerful AI technologies, which gives them a competitive advantage over local news media. The uptake of AI in the news industry for various stages of the journalistic process can push these smaller and local news media out of the market, which risks decreasing media pluralism. States have a positive obligation to ensure media pluralism amidst the growing popularity of AI for the news business by, for example, creating a level playing field for online news media to use AI and data-driven technologies.

3.6. Conclusion

This chapter analysed the implication of the use of AI by news media for the human right to freedom of expression, as protected by Article 10 of the European Convention on Human Rights. Four goals in the pursuit of which news media can use AI were analysed: newsgathering; news production; news distribution; and moderation of reader comments. The use of AI by news media falls within the democratic roles of the media as recognised and affirmed by the ECtHR: distributing information to the public; acting as a public watchdog; and creating a forum for public debate.

Because of their democratic role, the news media enjoy media freedom. Media freedom protects the use of AI to gather, publish, distribute, and receive information, as freedom of expression protects both the content and the technical means for communication.

Online intermediaries can fulfil roles similar to the democratic roles of the media when they augment the accessibility of information and enable public debates on their platforms. The selection, ranking, and prioritising of news by online intermediaries may therefore qualify for freedom of expression or even media freedom.

News media and other actors exercising or contributing to freedom of expression by using AI also assume duties and responsibilities. For the news media, these duties and responsibilities are spelled out in journalistic codes of ethics and enforced through various self-regulatory instruments. However, current journalistic codes of ethics do not contain guidance for the use of AI. Some scholars are therefore calling for the development of algorithmic journalistic ethics. The concept of duties and responsibilities also provides a normative legal basis to require online intermediaries to develop adequate self-regulatory instruments for the use of AI when they contribute to the exercising of freedom of expression. If journalistic codes of ethics and other self-regulatory instruments continue to stay behind on the realities of AI in the news media environment, then states have a legal justification to regulate this domain when AI has significant effects on the freedom of expression rights of news users and other participants in public debate.

The use of AI by news media presents, among others, the following risks for freedom of expression: Events and news stories that do not generate big data may be

overlooked by the algorithmic eyes of automated news-gathering systems and the voices of people implicated by these stories may therefore go unheard; automated journalism may produce unlawful content that violates the rights and dignity of other people; the personalised distribution of news may affect the right of news users to receive diverse information, which is an inherent part of freedom of expression; finally, automated moderation of user comments may be biased against minority groups, which could lead to unequal chances to communicate and participate in public debate.

States may have positive obligations to ensure that everyone can effectively enjoy their right to freedom of expression in the face of AI. States have a positive obligation to ensure that news users receive diverse news and to create a favourable environment for freedom of expression. To the extent that the use of AI by large news corporations and online intermediaries threatens the competitive viability of smaller players, states may have positive obligations to create a level playing field for the use of AI by news media.

These conclusions resonate with calls from human rights organisations and digital rights organisations regarding the implications of AI for freedom of expression. The UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression recommends that states "create a policy and legislative environment conducive to a diverse, pluralistic information environment", which includes "taking measures to ensure a competitive field in the artificial intelligence domain".[288] The OSCE Representative on Freedom of the Media[289], as well as Privacy International and Article 19,[290] also point to various threats related to AI for freedom of expression.

The current challenge for news media and public authorities is to develop journalistic codes of ethics, self-regulatory standards, and possibly government regulation to contain the risks of AI for freedom of expression, while enabling AI to contribute to public debate, media pluralism, the free flow of information, and other societal goals. The principles embedded within Article 10 ECHR provide concrete guidance for public and private bodies when they take up this challenge.

[288] Kaye D., "Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression", United Nations, para. 64, https://freedex.org/wp-content/blogs.dir/2015/files/2018/10/AI-and-FOE-GA.pdf.

[289] OSCE Representative on Freedom of the Media, Artificial intelligence & freedom of expression, https://www.osce.org/representative-on-freedom-of-media/447829?download=true.

[290] Privacy International and Article 19, Privacy and freedom of expression in the age of artificial Intelligence, http://privacyinternational.org/report/1752/privacy-and-freedom-expression-age-artificial-intelligence.

4. Cultural diversity policy in the age of AI

Mira Burri, University of Lucerne

4.1. Introduction

Diversity of content is essential to a vibrant public discourse, to cultural and social inclusion and to cohesion. Cultural diversity has accordingly been long defined as a regulatory objective in national media and cultural policies, especially in Europe, and the mandate to protect and foster it has only been strengthened after the 2005 UNESCO Convention on Cultural Diversity.[291] While cultural diversity has remained a key public policy objective, despite widely differing implementations in national policies on the ground, as the technological environment has profoundly changed, some fundamental questions have remained unanswered. Two critical questions need to be asked in this sense: Firstly, to what extent the affordances of the digital medium have enabled, as well as challenged, diversity online – both in terms of availability of diverse content and its actual consumption; secondly, how suited are cultural policy toolkits, as now applied, to actually address and foster engagement with culturally diverse content. This contribution will show that the answers to these questions are not simple and that policy-makers may need to engage in complex trade-offs, as well as be more innovative in the implementation of their cultural policies – by governing through intermediaries and through technologies.

This contribution will look at the affordances of digital media and artificial intelligence (AI) in particular and their implications for content policies; it will not engage however in the broader discussions about creativity in the age of AI,[292] nor does it look at diversity as embedded in AI to reduce biases in its decision-making[293] or diversity in the AI industry.[294]

---

[291] See e.g. Burri M., "The UNESCO Convention on Cultural Diversity: An appraisal five years after its entry into force." *International Journal of Cultural Property* 20, 4, pp. 357–380, November 2013.
[292] See in this context, Kulesz O., "Culture, platforms and machines: The Impact of Artificial Intelligence on the Diversity of Cultural Expressions", report for UNESCO, DCE/18/12.IGC/INF.4, 2018.
[293] Melendez C., "In AI, Diversity Is a Business Imperative", *The Forbes*, 14 November 2019.
[294] See e.g. Paul K., "'Disastrous' Lack of Diversity in AI Industry Perpetuates Bias", *The Guardian*, 17 April 2019.

4.2. Understanding the changed environment of content creation, distribution, use and re-use

The transformations in the digital networked environment epitomised by the societal penetration of the Internet have been multi-faceted and over the years their effects have been captured, although not without contention, by a host of excellent studies.[295] The centrality of data and the predominance of multiple-sided markets for data, as well as the rise of AI, have created a new level of complexity, and it is essential to understand well the contemporary dynamics around content, so as to be able to design adequate cultural toolkits. We focus on those specific developments that may be critical for the pursuit of cultural diversity objectives in this new space and are particularly interested in the changed ways content is produced, distributed, accessed, and consumed online, as well as in the related modifications in the patterns of user experience and participation, whenever these can be identified. This chapter uses the changing role of intermediaries as critical gatekeepers as an entry point to this complex discussion.

4.2.1. Understanding the new intermediaries

There have been assumptions, some of them backed by evidence, that the digital environment would bring about abundance, diversity and empowerment of users not possible under the conditions of analogue media.[296] One of the core elements supporting these positive accounts is that intermediaries do not exist in cyberspace and one can freely choose any content at any time. Yet, as contemporary digital media practice shows, this claim may be flawed. In fact, it may be that there *are* various intermediaries with different types of control over the choices we *make* and over the possibility for choices we *see*. We do not discuss here the physical intermediaries, such as network operators or Internet service suppliers (although these can be very important[297]), but focus on those gatekeepers existing at the applications and the content levels – the so-called "choice intermediaries"[298] or "new editors",[299] which often also employ AI technologies.

---

[295] See e.g. Benkler Y., *The wealth of networks: How social production transforms markets and freedom*. New Haven: Yale University Press, 2006); Sunstein C. R., *Republic.com 2.0*. Princeton: Princeton University Press, 2007.

[296] Benkler Y., Weinberger D., *Everything is miscellaneous: The power of the new digital disorder*. New York: Henry Holt, 2007; Jenkins H., *Convergence culture: Where old and new media collide*. New York: New York University Press, 2008.

[297] Benkler Y. (2006).

[298] Helberger N., "Diversity label: Exploring the potential and limits of a transparency approach to media diversity." *Journal of Information Policy* 1. pp. 337–369, 2011; Helberger N., "Diversity by design", *Journal of Information Policy*, pp. 441–469, 2011.

[299] Miel P. and Farris R., *News and information as digital media come of age*, Cambridge: The Berkman Center for Internet and Society, 2008, at p. 27.

To understand the new media space, it may be helpful to compare it to the functioning of legacy media. Conventionally, in the offline/analogue world, editorial roles were concentrated under the roof of a single institution. Editorial choices were based on a certain, limited pool of materials, and editorial products were finite, bounded by the limitations inherent to each medium, such as the pages of a printed newspaper or the length of a broadcast. The targeted audience was also typically addressed in a certain rhythm, which influenced the breadth and depth of the content – for example daily newspapers or a weekly edition. The editorial decisions made as to the content and the format reached the entire audience of any given publication or programme in the same way – they were not tailored to a particular user. Depending on the format, there was also a certain balance between local, national and international topics, which were presented in a contextualised and trustworthy manner. These were the key editorial functions of broadcasters and other legacy media, which were in many jurisdictions also under a specific mandate to feature local and national content; there were commonly mechanisms in place to supervise the fulfilment of certain content quantity and quality requirements. In the European Union, for instance, there are, in addition to the obligation to carry a majority of European works on audiovisual channels, requirements and obligations at the national level. Overall, these relatively neatly defined editorial functions had important consequences for the production and distribution of knowledge.[300] They also supported the conviction, which underlies almost all national media policies, that diversity in supply will be reflected in diversity of consumption.

The picture is strikingly different now, as digital media forms remove these analogue limitations and provoke "fundamental shifts in the composition and consumption of media products".[301] The "new editors" are multiple, disintegrated and distributed.[302] The "new editors" are AI-driven and it is ultimately algorithms[303] that define the new media space.

Aggregation is the first such editor and refers to the process of assembling different types of content in a tailored fashion and constantly updating it. This sort of personalised editor is offered on different platforms, for different types of content – be it news, entertainment or gossip. It automatically generates information tailored to a particular user profile and/or previous experience in a seemingly seamless and incessant manner. The mechanism driving this content feed is usually an algorithm that is specific

[300] Weinberger D., *Too big to know*, New York: Basic Books, 2012.
[301] Miel and Farris at p. 27. See also Kleis Nielsen R., Gorwa R., and de Cock Buning M., *What can be done? Digital media policy options for strengthening European democracy*, Oxford: Reuters Institute Report, 2019.
[302] Ibid.; also Latzer M., Hollnhuchner K., Just N., and Saurwein F., "The economics of algorithmic selection on the Internet" in Bauer J. M. and Latzer M. (eds.), *Handbook on the economics of the Internet*. Cheltenham: Edward Elgar. pp. 395–425, 2016.
[303] For a comprehensive definition of algorithms, see Latzer M., and Just N., "Governance by and of algorithms on the Internet: Impact and consequences", in *Oxford Research Encyclopedia, Communication*, Oxford: Oxford University Press, 2020.

to the platform (be it Facebook or Instagram for example) and may discriminate between different types of content.[304]

Social bookmarking has also become increasingly important as a mechanism of giving prominence to content. Here the crowd acts as an editor through different ranking and bookmarking systems, such as Reddit, Technorati or Del.icio.us. With the wide adoption of Twitter and Instagram, in particular by younger generations, the use of hashtags as a type of metadata tag, allowing users to create dynamic, user-generated indexes, has increased. These mechanisms can not only tailor media consumption but also succeed in commanding the attention of large groups.[305] This may be true for political campaigns but also for mobilising consumer attention in marketing campaigns.

And finally, as a digital intermediary, *search* is nowadays absolutely essential. It is often the starting point for the majority of online experiences and is the most significant driver of traffic to most websites. Without being indexed and searchable on the net, content is plainly rendered non-existent.[306] Search is again typically driven by proprietary algorithms and the business is highly concentrated around very few providers, with Google clearly distancing itself from its competitors.

4.2.2. Implications of AI-driven editorial agents

Through all these different mechanisms, the network functions as a multi-channel editor and an important intermediary in the content value chain – it replaces in fact the role of traditional media as a "general interest intermediary".[307] On the positive side, it has been suggested that "the networked media environment as a virtual social mind [...] produces something richer, more representative, and more open to ideas than the top-down mass media model of the past".[308] While we should not underestimate the affordances of digital platforms and the processes of communication, participation and engagement that they enable, at least so far, there is profound uncertainty and indeed increasing doubt as to the

[304] The algorithms often combine different mechanisms and are driven by different factors: (1) general popularity of the item among all users is the simplest approach, where all users get the same recommendation, which ultimately results in popular items becoming even more popular and the disappearing of unpopular items; (2) semantic filtering recommends items that match the currently used item or items previously used by the same user on a number of pre-defined criteria (such as topics, the author or source of an article); (3) collaborative filtering or social information filtering is an automated 'word-of-mouth' recommendations generator – items are recommended to a user based upon values assigned by other people with similar taste. These methods are usually applied in hybrid forms, including also other methods like weighing items by recency or pushing content that has specific features such as paid content. Platforms have also over the years accumulated large amounts of valuable data based on past behavior and can additionally apply user data such as age or location to calibrate the content feed. See Bozdag E., "Bias in algorithmic filtering and personalization", in *Ethics and Information Technology* 15. Pp. 209–227, 2013.

[305] Miel and Farris, at p. 30.

[306] Council of Europe, Draft Recommendation on the Protection of Human Rights with Regard to Search Engines, Strasbourg, 11 March 2010.

[307] Sunstein (2007).

[308] Miel and Farris, at p. 30.

ability of this self-organising mechanism to reliably identify salient information.[309] There is also a dose of scepticism as to its impact on the diversity and quality of content, and on users' capabilities to find and access content that is diverse and trustworthy.

Thinking about the societal functions of the media and the goal of cultural diversity in the context of this chapter's discussion, it could be that this complex environment presents certain dangers of reduced diversity and fragmentation of the public discourse.[310] First, we need to acknowledge the possible interferences with users' individual autonomy and freedom of choice. As Latzer et al. argue, while filtering reduces search and information costs and facilitates social orientation,[311] it can be "compromised by the production of social risks, among other things, threats to basic rights and liberties as well as impacts on the mediation of realities and people's future development".[312] The second worry in this context has to do with the impact of tailored media production and consumption. In the former sense, there has been a recent trend towards algorithmic content production, where algorithms drive decision-making in media organisations by predicting audiences' consumption patterns and preferences.[313] While in some areas this may be viewed as beneficial in giving the audiences what they want, in other areas, such as for news, this may be highly problematic, as local news and current affairs become tailored to the demographic, social and political variables of specific communities.[314] We should also be reminded of the so-called "content farms", which, based on search-engine data (such as popular search terms, ad word sales and actual available content) produce content rapidly and cheaply in order to meet that demand. Such creation of content is completely commodified and possibly harmful to any public interest function of the media, including in the cultural sphere.

In the sense of media consumption, the personalisation of the media diet, as based on a distinct profile or previous experience, "promotes content that is geographically close as well as socially and conceptually familiar"[315] … "This keeps users within familiar boundaries, feeding their curiosity with more of the same. When they are looking for new content or information, this reinforces existing opinions, gradually removing conflicting views."[316] One can of course state that this has been the case with legacy media as well, where people are naturally drawn to content they have liked in the past – the key difference in the current space is that users see *only this* content, and their active choice is so diminished or manipulated. Hoffman et al. argue that social media only

[309] Ibid.
[310] See e.g. Sunstein C. R. *Going to extremes: How like minds unite and divide*, Oxford: Oxford University Press, 2009.; Pariser E., *The filter bubble: What the Internet is hiding from you*, London: Viking, 2011.
[311] Latzer et al.
[312] Ibid., at pp. 29–30.
[313] Napoli P. M., "On automation in media industries: Integrating algorithmic media production into media industries scholarship" in *Media Industries Journal* 1. pp. 33–38, 2014; also Saurwein F., Just N., and Latzer M., "Governance of algorithms: Options and limitations', *info* 17. pp. 35–49, 2015.
[314] Napoli, ibid., at p. 34.
[315] Hoffman C. P., Lutz C., Meckel M., and Ranzini G., "Diversity by choice: Applying a social cognitive perspective to the role of public service media in the digital age", *International Journal of Communication*, 9, 2015, pp. 1360–1381.
[316] Ibid.

exacerbate this effect by combining two dimensions of homophily: similarity of peers and of content.[317] We should keep in mind in this context that despite a slight reduction in the use of social networking sites as an entry point to content and variations across countries,[318] they still are important gatekeepers. This reinforces the effect of homophily, as well as clearly illustrates the power of a few players and the deep impact of their decisions – for instance, when Facebook changed its algorithm in 2018 and downgraded news, this automatically led to less news consumption.[319]

The commercialisation of platforms and the radical increase in commercially or politically driven "fake news" should also be underscored.[320] Despite the slight shift towards reader payment models for news, it is worth remembering that the vast majority of online consumption still happens through free websites, largely supported by advertising. While some of the aggregated content is taken from legacy media,[321] which may disperse some of the conventional criticism that aggregators amplify the impact of unreliable non-traditional sources, it is still true that content is not made more abundant but has merely become more distributed – in this sense we do not have more and diverse content but simply more of the same. Still, it is fair to note that legacy media have responded to the technologically enabled aggregation and offer much more content online than in their print or broadcast versions. With specific regard to news, the Reuters Institute for the Study of Journalism found that private news organisations are making major investments in social media and report significant traffic, off-site reach, and/or additional digital subscribers.[322] While this may enable access to a variety of content over more platforms, also enticing young people, two drawbacks need to highlighted: the first relates to the almost full reliance of media organisations on Facebook, which brings a certain "platform risk" with it; the second is that private sector legacy news organisations' approaches to social media are strongly shaped by path-dependent business models oriented towards advertising and subscriptions, or a mix thereof.[323] This again may not lead to a sustainable offering of diverse local, regional and national content.[324] Overall,

---

[317] Ibid.

[318] For country analyses, see Reuters Institute for the Study of Journalism, *Digital News Report 2018*, Oxford, 2018.

[319] Reuters Institute for the Study of Journalism (2018); also Tucker J. A. et al. *Social media, political polarization, and political disinformation: A review of the scientific literature*, prepared for the Hewlett Foundation, March, 2018.

[320] Reuters Institute for the Study of Journalism (2018); European Commission, Tackling Online Disinformation: a European Approach, COM(2018) 236 final, 26 April 2018.

[321] Reuters Institute for the Study of Journalism (2018). Aggregators may be somewhat restricted by copyright, see Associated Press v. Meltwater U.S. Holdings, Inc., 931 F. Supp. 2d 537, 537 (S.D.N.Y. 2013) and newer initiatives in the field of EU copyright law.

[322] They identify three main strategic aims shaping the different ways in which news organisations approach social media: (1) driving on-site traffic through referrals; (2) driving off-site reach through native formats and distributed content; (3) driving digital subscription sales, often in part through advertising content on Facebook.

[323] Cornia A., Sehl A., Levy D. A., and Nielsen R. K., *Private sector news, social media distribution, and algorithm change*, Oxford: Reuters Institute for the Study of Journalism, 2018.

[324] A study of US local media has shown for instance that only about 17 percent of the news stories provided to a community were truly local – that is, about or having taken place within the municipality; fewer than half

despite increased amounts of content, there may be less local, regional and national content and real difficulties in finding it, because it is - or becomes - marginalised on online platforms.

With regard to search engines as intermediaries, it may be generally in the long-term interest of search providers to meet the needs of their users – both as consumers and as citizens. Research conducted by the UK's Ofcom suggests that demand for national public service content remains strong, and therefore it should continue to be in the interest of search providers to ensure that their results give due prominence to such content.[325] A recent comparative study also found that those who find news via search engines, on average use more sources of online news, are more likely to use both left- and right-leaning online news sources, and have more balanced news repertoires.[326] This said, and as earlier mentioned, search results are generated algorithmically and automatically assign relevance to certain information units. The automated selection is also prone to manipulation using a range of search engine optimisation techniques, whereby sponsored or other content gains more visibility and attracts more attention.[327]

In concluding this section, which offers only a snapshot of the complex contemporary media environment, one needs to stress its fluidity and the therewith related uncertainty as to its impact, as far the abundance and diversity of content and the conditions of free speech are concerned. On the one hand, there is a discourse in the literature that, under different labels such as "filter bubbles"[328] or "echo chambers",[329] highlights the risks of the current tailored media diet in leading towards a fragmentation of the public discourse and possible polarisation of views.[330] On the other hand, we are unsure to what extent this is true. A 2017 cross-country report found for instance that although search plays a major role in shaping opinion, it needs to be viewed in a context of multiple media and is not deterministic.[331] The study of "automated serendipity", which denotes a phenomenon whereby users are drawn to sources they would not have consulted otherwise, also reduces the fears of "echo chambers".[332] In the same context, it should be noted that the currently applied tools to track fragmentation tell us surprisingly

(43 percent) of the news stories were original – that is, actually produced by the local media outlet. See Napoli P. M., Weber M., McCollough K., and Wang Q., *Assessing local journalism: News deserts, journalism divides, and the determinants of the robustness of local news*. News Measures Research Project, August, 2018.

[325] Ofcom „Ofcom's Second Public Service Broadcasting Review, Phase Two: Preparing for the Digital Future", London: Ofcom, 2008, para. 5.60.

[326] The authors refer to a phenomenon of "automated serendipity", which leads people to sources they would not have used otherwise. See Fletcher R. and Nielsen R. K., "Automated serendipity" in *Digital Journalism* 6. pp. 976–989, 2018.

[327] See e.g. Bradshaw S., "Disinformation optimised: Gaming search engine algorithms to amplify junk news", in *Internet Policy Review* 8. pp. 1–24, 2019.

[328] Pariser (2011).

[329] Sunstein C. R., *Infotopia: How many minds produce knowledge*. Oxford: Oxford University Press, 2006.

[330] See also High Level Group on Media Freedom and Pluralism, *A free and pluralistic media to sustain European democracy*. Report prepared for the European Commission p. 27, January 2013.

[331] Dutton W. H. et al "Search and politics: The uses and impacts of search in Britain, France, Germany, Italy, Poland, Spain, and the United States", Quello Center working paper No 5 pp. 1-17, 18 May 2017.

[332] Fletcher and Kleis Nielsen (2018).

little about audience loyalties and how public attention moves *across* media.[333] Webster and Ksiazek find for instance little evidence that audiences are composed of devoted loyalists.[334] "Moreover, measures of exposure, no matter how precise, cannot tell us how content affects people. It may be that even modest periods of exposure to hate speech or otherwise obscure media have powerful effects on those who seek it out",[335] or it could be that the overall effect is balanced through other components in the media diet.[336] In this sense, we should not concentrate on snapshots but examine dynamics and track evolution over time.[337]

4.3. Possible avenues of action: New tools addressing and engaging digital intermediaries

In painting the picture of the transformed and transforming media landscape above, we observe the complexity of new "editorial" processes and the difficulty for individuals to navigate this potentially rich but distributed content space. We also identify some potential risks of tailored content consumption and polarisation of views in this environment, as the new intermediaries algorithmically drive supply and demand by selecting information elements and assigning relevance to them.[338] Against this backdrop, one can think of two viable channels for introducing cultural diversity measures: the first is in addressing the emergent environment and governing the algorithms as critical gatekeepers, since these have so far largely remained unregulated, especially for cultural policy purposes; the second is using the new intermediaries as a tool to promote cultural diversity exposure, in the sense of "governance through intermediaries".

4.3.1. Governance of algorithms

When speaking of governance of algorithms, there is a more generic, not necessarily cultural policy-related debate, which has to do with the observation that intermediaries, in particular those driven by algorithms, have gained a critical role in the online space and in this sense it is now governed *by* algorithms.[339] This discussion is closely related to that on the appropriate ways to address this new power – that is, the governance *of*

[333] Webster J.G. and Ksiazek T. B., "The dynamics of audience fragmentation: Public attention in an age of digital media", *Journal of Communication* 62, pp. 39–56, 2012.
[334] Ibid., at p. 40.
[335] Ibid., at p. 51.
[336] For very interesting findings, see Pew Research Center, "Political polarization and media habits: From Fox News to Facebook, how liberals and conservatives keep up with politics", Washington, DC: Pew Research, 2014.
[337] For newer trends in media consumption, see Reuters Institute for the Study of Journalism (2018).
[338] Saurwein et al. (2015), at p. 35.
[339] For an excellent analysis and review of the literature, see Saurwein et al. (2015).

algorithms.[340] Privacy protection questions have been particularly salient in this latter context[341] but also questions around copyright enforcement through intermediaries, as illustrated by the latest EU copyright reform and the discussion around Article 17 of the Copyright in the Digital Single Market Directive.[342] Latzer et al. identify nine categories of risk stemming from algorithmic selection that may need to be addressed: (1) manipulation; (2) diminishing variety, echo chambers and biases; (3) constraints on freedom of expression; (4) surveillance and threats to data protection and privacy; (5) social discrimination; (6) violation of intellectual property rights; (7) abuse of market power; (8) effects on cognitive capabilities; and (9) growing heteronomy, loss of human sovereignty and controllability of technology.[343] We were particularly interested in (2), (3) and (9) above, as immediately related to the core cultural diversity objectives pursued in the media domain.

Saurwein et al. provide a careful analysis of the different governance options that can address these risks, which, next to conventional command-and-control interventions, may involve regulation by market and various self- and co-regulatory solutions in between.[344] Yet, the authors also note that so far there have been hardly any tools designed to address the risks of bias, heteronomy and effects on cognitive capabilities.. It is indeed true in the specific setting of our discussion that most of these intermediary platforms will not fall under the regulatory scope of the current media regimes. Napoli has argued in this context that we should start approaching algorithms as a distinctive form of media institution.[345] He believes algorithms should be subject to restrictive types of regulation – that is, a ban on certain types of activities by the platform operators or content on these platforms, to protect privacy and counter graphic violence and hate speech. Napoli suggests that considering the crucial role these new intermediaries play, we ought to develop "affirmative approaches in the public interest",[346] as we have done for traditional electronic media built upon established media policy principles, such as plurality, diversity, and localism – prescribing for instance certain amounts or types of

[340] Saurwein et al. (2015).
[341] For instance with regard to the right to be forgotten, which is now enshrined in the EU data protection regime of the General Data Protection Regulation.
[342] See Directive 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC, OJ L (2019) 130/92, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32019L0790. See also Montagnani M. L. and Trapova A. Y., "Safe harbours in deep waters: A new emerging liability regime for Internet intermediaries in the digital single market", *International Journal of Law and Information Technology* 26. pp. 294–310, 2018; Senftleben M., "Institutionalized algorithmic enforcement – The pros and cons of the EU approach to UGC platform liability", *Florida International University Law Review* 14, 2020.
[343] Saurwein et al. (2015), at p. 37.
[344] Saurwein et al. (2015).
[345] Napoli P. M., "The algorithm as institution: Toward a theoretical framework for automated media production and consumption", paper presented at the Media in Transition Conference, Massachusetts Institute of Technology, May 2013; Napoli P. M., "Social media and the public interest: Governance of news platforms in the realm of individual and algorithmic gatekeepers. Media + the Public Interest Initiative Working Paper, 2014.
[346] Napoli (2014), ibid.

content.[347] This could address on the one hand the lack of diverse, trustworthy and local content, which despite the availability of more content online appears scarce, as well as increase the possibility for users to access such content.[348]

It should be noted in this context that the discussions on the regulation of algorithms, although relatively young, have rapidly evolved. In the age of big data, perils for personal data protection and "fake news" proliferation, the topic has gathered attention on the part of politicians and the broader public, and the need for action has been recognised.[349] However, the form of action is yet to be defined. So far, any action appears to be unfolding in the domains of self-regulation and soft regulatory approaches, as hard intervention may not only hinder platforms' and users' innovation but also defeat the very goal of promoting free speech in its active and passive forms. In this context, and in an attempt to design appropriate and forward-looking governance tools, one needs to carefully consider the experience so far in the field of fighting online disinformation. On the one hand, we must examine to what extent businesses have responded to the increased public awareness and users' demands for trust and quality, and to what extent different technical (for instance by adjusting algorithms) and other solutions (for instance working with users and other organisations) have effectively contributed towards the defined objective – in this case: constraining the amount and spread of "fake news".[350] Facebook's efforts, subsequent to the 2016 US presidential campaign, may provide a case in point. As a reaction to various accusations, Facebook endorsed a number of initiatives – for instance, it disseminated educational tools for information literacy, started the Facebook Journalism Project and joined the News Integrity Initiative with a number of academic and media partners focused on fostering engaged communities and more inclusive media, while seeking to better understand misinformation. Concretely and in order to reduce the spread of "fake news", Facebook entered into partnerships with about 40 third-party media organisations – such as Snopes, PolitiFact, the Associated Press, and FactCheck.org. They strive to fact-check shared news stories and identify them with a "disputed" label if they did not pass a fact-checking muster. Facebook also installed a "more info" button that lets users obtain additional context about articles in their news feeds.[351] Despite these wide efforts, research and anecdotal evidence suggest that Facebook's practices may still be insufficient to secure a 'healthy' media space.[352]

---

[347] Napoli P. M. "Social media and the public interest: Governance of news platforms in the realm of individual and algorithmic gatekeepers", *Telecommunications Policy* 39. pp. 751–760, 2015.

[348] See Napoli P. M. "Re-evaluating the long tail: Implications for audiovisual diversity on the Internet", in Albornoz L. A. and Garcia Leiva M. T. (eds.), *Audiovisual industries and diversity: Economics and policies in the digital era*. Abingdon: Routledge, 2019, at chapter 5; Napoli P. M., *Social media and the public interest: Media regulation in the disinformation age*, New York: Columbia University Press, 2019.

[349] See e.g. Balkin J., "Free speech in the algorithmic society: Big data, private governance, and new school speech regulation", *UC Davies Law Review* 51. pp. 1149–1210, 2018. There have also been multiple actions in the EU with regard to data protection and "fake news" – for instance, through the adoption of the GDPR. See also European Commission (2018).

[350] Reuters Institute for the Study of Journalism (2018).

[351] The 'additional information' button for news articles surfaced in the News Feed lets users click through to see: (1) background pulled from the Wikipedia page about the publisher; (2) other articles recently posted by

The role of governments, civil society and other user organisations should also be taken into account. A good recent example of coordinated efforts and working together on multiple fronts has been the European approach towards "fake news".[353] The European Commission subscribes to improving transparency of distributed information, its diversity and credibility and to an effort to fashion inclusive long-term solutions to this effect. Amongst other things, the Commission convened a multi-stakeholder forum to provide a framework for efficient cooperation amongst relevant stakeholders, including online platforms, the advertising industry and major advertisers, and to secure a commitment to coordinate and scale up efforts to tackle disinformation. The forum's first output was an EU-wide Code of Practice on Disinformation. Adopted in September 2018, the code sets out self-regulatory standards to fight disinformation; it aims at achieving the Commission's objectives by setting a wide range of commitments, from transparency in political advertising to the closure of fake accounts and demonetisation of purveyors of disinformation. The code also includes an annex identifying best practices that signatories pledge to apply to implement the code's commitments.[354] More decisive steps towards accountability, and even a move towards co-regulatory approaches, may be necessary, however,[355] as evidentiated by the acute problem of "fake news" around the Covid-19 pandemic.[356]

4.3.2. Governance through algorithms

Moving towards a more targeted cultural diversity toolkit, one may consider endorsing new forms of editorial intelligence,[357] as a sort of public interest mediation of the digital space that seeks to increase the visibility, discoverability and usability of discrete types of

---

the publisher; (3) a heat map of where in the world the article is being shared and which of the user's Facebook friends have shared it.

[352] Levin S., "They don't care: Facebook fact-checking in disarray as journalists push to cut ties", *The Guardian*, 13 December 2018. For a more in-depth analysis, see Saurwein F. and Spencer-Smith C., "Combating disinformation on social media: Multilevel governance and distributed accountability in Europe", *Digital Journalism*, 2020.

[353] European Commission (2018).

[354] For all documents, see https://ec.europa.eu/digital-single-market/en/news/code-practice-disinformation. The latest report of the European Regulatory Group for Audiovisual Media Services (ERGA) on disinformation shows some limits in the commitments taken by platforms within the code; see ERGA Report on Disinformation: Assessment of the Implementation of the Code of Practice, 2020, https://erga-online.eu/?p=732.

[355] Saurwein and Spencer-Smith (2020).

[356] See e.g. "Social media firms fail to act on Covid-19 fake news", *BBC News*, 4 June 2020.

[357] We do not address here media literacy policies, which can also be important from the user-centric perspective. See Helberger (2011), at p. 357; Burri M., "The global digital divide as impeded access to content", in Burri M., and Cottier T.(eds.), *Trade governance in the digital age*. Cambridge: Cambridge University Press. pp. 396–420, 2012; High Level Group on Media Freedom and Pluralism, "A Free and Pluralistic Media to Sustain European Democracy", report prepared for the European Commission, January 2013.

content.[358] We may also envision tools that incentivise exposure diversity – that is, the actual consumption of diverse content.[359]

This is not a completely new or exotic project. The European media framework, under the Audiovisual Media Services Directive (AVMSD)[360] includes a suggestion that the promotion of European works may relate to increasing the "prominence" of such works.[361] From the consultation of regulatory authorities in 2013,[362] it appears that many of them were in favour of prominence tools, while being sceptical about the promotion through asset share in catalogues. Many view this measure as the most efficient (also because it relates to actual higher consumption of European works) and the least burdensome for operators.[363] It has also been a standing practice in Europe that public service broadcasters (PSBs) have had the privilege to occupy the first slots in electronic programme guides (EPGs) and have so been given "due prominence".[364] Foster and Broughton show that EPGs have been an important tool for consumers finding and selecting programmes, and there is evidence that channels with slots near the top of each section of an EPG have had an advantage in viewers' selection over those further down.[365] "This approach [of "nudging" people towards the choices we hope they will make both in their own and society's wider interests] has so far worked reasonably well."[366] Recent evidence confirms that EPG

[358] Miel and Farris, at p. 3; also Goodman E. P., "Public media 2.0", in Schejter A. M. (ed.), *And communications for all: A public policy agenda for a new administration*. Lanham, MD: Lexington Books, pp. 263–280, 2009; Webster J. G., "User information regimes: How social media shape patterns of consumption", *Northwestern University Law Review* 104, pp. 593–612, 2010.

[359] See Helberger N., "Media diversity from the user's perspective: An introduction", *Journal of Information Policy* 1, pp. 241–245, 2011; Napoli P. M., "Exposure diversity reconsidered", *Journal of Information Policy* 1, pp. 246–259, 2011.

[360] The AVMS was last reviewed in 2018: see Directive (EU) 2018/1808 of the European Parliament and of the Council of 14 November 2018 amending Directive 2010/13/EU on the coordination of certain provisions laid down by law, regulation or administrative action in Member States concerning the provision of audiovisual media services (Audiovisual Media Services Directive) in view of changing market realities, OJ L (2018) 303/69.

[361] See the latest EU recommendations: Communication from the Commission Guidelines pursuant to Article 13(7) of the Audiovisual Media Services Directive on the calculation of the share of European works in on-demand catalogues and on the definition of low audience and low turnover 2020/C 223/03, OJ C (2020) 223/10.

[362] See e.g. European Audiovisual Observatory, IRIS Special: Video on Demand and the Promotion of European Works. Strasbourg: European Audiovisual Observatory, 2013.

[363] European Commission (2014), at p. 6.

[364] EPGs have been regulated at the EU level through the Access Directive (Directive 2002/19/EC of the European Parliament and of the Council of 7 March 2002 on access to, and interconnection of, electronic communications networks and associated facilities, OJ L 108/7, 24 April 2002, as amended by Directive 2009/140/EC of the European Parliament and of the Council of 25 November 2009, OJ L 337/37, 18 December 2009). The implementation of the directive differs – e.g. the British regulation allows a preferred treatment of PSB channels, while in Germany the regulation of EPGs is based on the equal treatment of public and commercial channels in EPG listings.

[365] Foster R. and Broughton T., "PSB prominence in a converged media world", report commissioned by the BBC, London: Communications Chambers at p. 12, 2012. Other factors that influence selection include having a memorable EPG channel number and being adjacent to another popular channel.

[366] Foster and Broughton, op.cit, pp. 13–14. This has been confirmed by a more recent report.

positioning is likely to have a significant impact on a channel's performance.[367] Such "nudging", albeit for commercial media, has also worked with the remote controls of SMART TVs, with the big online players, such as Netflix, YouTube and Google Play, appearing as buttons allowing direct access.

In a similar way, one may consider a deeper type of intervention that entails some sort of guidance for users as to the "relevant" and "quality" local, regional or national content, making sure they then consume the "right mix".[368] Two critical questions arise in this context – of awareness and of serendipity – i.e.: "Do people know about the full range of content opportunities available to them online, and how often do they stumble across content that they like but that they did not know existed?"[369] The UK's Ofcom has shown that barriers with respect to awareness and serendipity may be significant.[370]

One way of doing this is through the existing PSB systems.[371] One can first think of an updated variation of the EPG as a tool for enhancing the prominence of both the PSB brand and the local, regional and national content offering. Foster and Broughton see this "nudging" as a two-step process whereby viewers are attracted to the PSB channel or brand and then a range of techniques are used to "lead audiences to a wider range of content than they might otherwise have chosen for themselves".[372] The authors have justified the need for new legislation that will ensure prominence of PSB brands or individual service brands[373] on online platforms. Prominence requirements should apply to the core elements of any consumer interface, such as a channel grid or on-demand service menu and each PSB should expect to secure at least one icon/button on the first page of an on-demand guide or its equivalent.[374] The same rationale can be applied also for European works.[375]

---

[367] Ofcom, "EPG prominence: A report on the discoverability of PSB and local TV services", London: Ofcom, 2018. The data on the effect of prominence on VoD viewing are less comprehensive but suggest a similar correlation.

[368] Helberger (2011), at p. 346. Justifying also such an approach, see Sunstein C. R., "Television and the public interest", *California Law Review* 88. pp. 499–563.

[369] Ofcom "Ofcom's Second Public Service Broadcasting Review, Phase Two: Preparing for the Digital Future", London: Ofcom, 2008 at para. 3.95.

[370] Ofcom (2008) at para. 3.98.

[371] For a fully-fledged analysis, see Burri (2015).

[372] Foster and Broughton, *op.cit.*,. At p. 11.

[373] Foster and Broughton argue against prominence given to individual programmes, which, they argue, may fragment user experience and hurt the overall PSB brand.

[374] Foster and Broughton, *op.cit.*, p. 4.

[375] The amended AVMS contains in its Article 13 such a rule: "Member States shall ensure that media service providers of on-demand audiovisual media services under their jurisdiction secure at least a 30% share of European works in their catalogues and ensure prominence of those works'. Recital 34 explains further that: "...The labelling in metadata of audiovisual content that qualifies as a European work should be encouraged so that such metadata are available to media service providers. Prominence involves promoting European works through facilitating access to such works. Prominence can be ensured through various means such as a dedicated section for European works that is accessible from the service homepage, the possibility to search for European works in the search tool available as part of that service, the use of European works in campaigns of that service or a minimum percentage of European works promoted from that service's catalogue, for example by using banners or similar tools."

The second proposition (fostering serendipity) may help too – "in particular for introducing viewers to content they would not otherwise look for or challenging users' views and expanding their knowledge 'by chance'".[376] In this context, a scholars have stressed that "[s]erendipitous encounters might alleviate some concerns about restrictive coping strategies and a tendency in users to hide in their 'information cocoons',[377] and 'promote understanding' and open-mindedness, and thereby also advance democratic goals".[378] The digital space and different ways of analysing data and aggregating content do allow for the random delivery of different types of content, which can be displayed next to the "chosen by the viewer" content or in dedicated "less searched for", "less viewed" and other "less popular" "not-mainstream" lists. Also, since it appears that there is a great difference in the availability and discoverability of discrete genres of content (e.g. sports versus educational programmes), it may be appropriate to establish cross-genre linkages, so as to both highlight this type of content and to increase the chances of overall more diverse consumption.[379]

However, caution should be exercised with regard to these random offerings, as they can simply be ignored or can even disrupt a viewer's experience. Research has shown that there must be more to serendipitous encounters than just chance. Schönbach explains that in order to work and incentivise users, surprises must be "embedded in the familiar".[380] Helberger expounds further that "[i]n order to be able to make sense out of chance information exposure, the information must resonate with some prior knowledge, interest, or experience for the user".[381] Hoffman et al. argue along the same line: that we can speak of "diversity experience" only if users "perceive and digest this content according to their motivations, awareness, and capabilities".[382] Designing tools that work well for this purpose may be a difficult task that partly links to the theme of media literacy. Such tools may also be connected to certain algorithmic design functions as "empowerment nudges, which promote decision-making in the interests of citizens, as judged by themselves, without introducing further regulation or incentives or using any manipulative measures".[383]

---

[376] Ofcom (2008a). At paras 3.99–3.101.
[377] Helberger (2011a). At p. 454.
[378] Ibid., referring to Sunstein (2007). At pp. 27–28.
[379] For an experiment on fostering content diversity through recommendation systems, see Möller J. et al., "Do not blame it on the algorithm: An empirical assessment of multiple recommender systems and their impact on content diversity", *Information, Communication and Society*, 2018.
[380] Schönbach K. "The own in the foreign: Reliable surprise – An important function of the Media?", *Media, Culture and Society* 29. pp. 344–353, 2007.
[381] Helberger (2011a) at p. 462.
[382] Hoffman et al. (2015) argue that in order to experience diversity online, users must strive for diversity, be aware of the preconditions of diversity, and be able to ensure access to diversity.
[383] Hansen P. G. and Jespersen A. M. "Nudge and the manipulation of choice: A framework for the responsible use of the nudge approach to behaviour change in public policy", *European Journal of Risk Regulation* 1, pp. 3–28, 2013, at p. 24. For a fully-fledged analysis with regard to exposure diversity, see Helberger N., Karppinen K., and D'Acunto L., "Exposure diversity as a design principle for recommender systems", *Information, Communication and Society*, 2017.

Overall, there may be room to contemplate mechanisms that act as "good aggregators" and promote the visibility, availability and consumption of high quality and trusted local, national and regional content across various platforms. In the age of AI, it can be assumed that designing such smart editors is doable. The question of balancing between the virtue of the intervention and its possible side-effects intrinsic to such paternalistic actions remains and should be tackled carefully.[384]

4.4. Concluding remarks

It is evident that the media landscape has changed profoundly and is still in a state of flux. One discrete change that has been truly disruptive for media production, distribution and consumption lies in media's editorial functions. Digital platforms, such as Facebook and Google, have assumed, although to a different extent, key functions in content mediation and have so started to play a vital role in the realisation of critical public objectives, including in the cultural domain. As they impact on the availability of and access to local, national and regional content, these intermediaries may also impinge on the form and content of cultural exchanges, on democratic participation and civic engagement. In the last few years, awareness has risen as to the risks of algorithmic filtering and tailored media diets that may be severely restricted and/or commercialised. Labels such as "filter bubbles" or "echo chambers" have captured the attention of scholars and policy-makers alike. The jury is still out, however, on the real effects of the mediation through digital platforms and the causal link between types of media exposure and cultural, political and social engagement. This seems to be the case even with more straightforwardly "bad" content, such as "fake news". In this sense, two take-aways for policy makers may be highlighted:

(1) we need more data and independent research on the availability of different types of content, the consumption and engagement with that content, the participants involved in this process and the impact of these processes on individual and collective democratic and cultural performances;
(2) it is important to continue the dialogue between content creators, intermediaries, users, advertisers, and other stakeholders involved in the dynamics of the media space and underline the critical importance of culturally diverse media consumption in this dialogue. The heightened value attached to the availability of culturally diverse choices, the stress on trustworthiness and quality that users understand and appreciate, may very well incentivise platforms to deliver such options. There are indeed already steps in this direction – for instance, with regard to flagging or removing certain types of content or certain users, or with regard to more transparency as to the sources of content.

[384] Helberger N., "Merely facilitating or actively stimulating diverse media choices? Public service media at the crossroad", *International Journal of Communication* 9. pp. 1324–1340, 2015; Bodo B. et al. "Tackling the algorithmic control crisis – The technical, legal and ethical challenges of research into algorithmic agents", *Yale Journal of Law and Technology* 19. pp. 133–180, 2017.

Some form of additional action may still be needed. We sketch two possible avenues that may shape media consumption – governing *of* and *through* algorithms. As for the latter, one may consider some "good aggregators" that promote the visibility, availability and consumption of high quality and trusted local, national and regional content. While this may sound interventionist and like possible interference with user autonomy and free speech, as well as the freedom of platforms to conduct business, there may be ways to have a diversity-sensitive design that is not at odds with autonomous choices but indeed empowers users to make better informed choices. Technology is likely to permit many variations on the theme and policy-makers may need to keep an open mind here, and may experiment with public service media as curators of media experiences. Caution is still required and the pursuit of diversity objectives may not necessarily fit into the practical design of all recommender systems – in the case of search engines like Google for instance, where users actively search for answers, there may be a trade-off between accuracy and diversity.[385]

With regard to addressing the role of the intermediaries themselves and alleviating the risks of tailored and potentially distorted media consumption, there may be a need to act in the public interest. Yet, we cannot plainly blame the platform or the recommendation system and target all measures at them. As Helberger (2017) at al. note users also "play a role in the realization or erosion of public values on these platforms".[386] Indeed, we have a "problem of many hands" and there is a corresponding need to conceptualise a framework with the participation of, and different responsibilities for, all stakeholders – platforms, users, civil society and governments.[387] Multi-stakeholder mechanisms derived from Internet governance can be used as a model.[388] The experience gathered recently in the domain of tackling online disinformation in Europe can provide particularly helpful insights. The realisation of diverse content availability and informed and empowered user choices as core public values in the media space should be then the result of a dynamic interaction and deliberation between the stakeholders and may result in a palette of measures, such as codes of conduct, guidelines and principles, supervisory bodies of governmental or non-governmental character that ensure continuous and effective dialogue, or certain technological fixes.[389]

---

[385] Adomavicius G. and Kwon Y., "Maximizing aggregate recommendation diversity: A graph-theoretic approach", *Proceedings of the 1st International Workshop on Novelty and Diversity in Recommender Systems*. DiveRS 2011. Chicago, 2011.

[386] Helberger N., Pierson J. and Poell T., "Governing online platforms: From contested to cooperative responsibility", *The Information Society*, 2017, at p. 2.

[387] Ibid.

[388] See e.g. Marda V. and Milan S., "Wisdom of the crowd: Multistakeholder perspective on the fake news debate", *A Report by the Internet Policy Observatory at the Annenberg School, University of Pennsylvania*, 21 May 2018.

[389] Helberger et al. (2017).

Copyright

One of the biggest fears raised by AI is the replacement of humans by machines. People are increasingly worried that they will lose their jobs to robots, and this uneasiness has reached the audiovisual sector too. There are more and more examples of the creative intervention of AI in scriptwriting and music composition, just to name two aspects. This technobarbaric invasion into the creative realm is still of relatively low import, though, so that the fears of the destruction of creative jobs are most probably unwarranted, at least for the time being. And yet, the issue of the copyrightability of works made by machines has taken academia by storm. The question is quite pertinent: if we agree that machines can "create" works, can the creating machine be a copyright holder? Or can a person or a company be the copyright holder of a work created by a machine? The overview provided by ***Giancarlo Frosio*** *in his contribution to this publication answers a set of emerging legal questions concerning AI and creativity. AI can also be used against the enemies of creativity to find and remove copyright-infringing material on the Internet and hunt down pirates and industry leaks. However, depending on how algorithms are programmed, there is always the risk of false positives, which can have an impact on the freedom of expression of Internet users.*

5. Copyright - Is the machine an author?

Giancarlo Frosio, Center for International Intellectual Property Studies (CEIPI), University of Strasbourg *

5.1. Introduction

It is claimed that artificial intelligence (AI) is a fundamentally disruptive revolution for humankind.[390] Intelligent machines are coming in multiple shapes to serve diverse purposes, replacing humans potentially everywhere[391] with, predictably, both positive and negative externalities.[392] Apparently, AI shows potential for replacing even those activities that are more inherently human. Although so far most creatives still do not fear being replaced by robots,[393] actually, a major field where AI appears to be increasingly proficient is creativity. AI writes poems, novels and news articles, composes music, edits photographs, creates video games, and produces paintings and other artworks. Most creative industries will be substantially affected,[394] from the audiovisual sector[395] to music[396] and publishing.[397] The time of the A(I)uthor has already come.

---

*Associate professor, Center for International Intellectual Property Studies (CEIPI), University of Strasbourg; non-resident fellow, Stanford Law School Center for Internet and Society; faculty associate, NEXA Center for Internet and Society. I wish to thank my research assistant, Varnita Singh, for in-depth research and remarkable, critical assistance given in preparing this chapter.

[390] Floridi L., *The Forth Revolution: How the Infosphere is Reshaping Human Reality*, OUP, Oxford. Bughin J. et al. (2017), *Artificial Intelligence: The Next Digital Frontier?*, McKinsey Global Institute Discussion Paper, www.mckinsey.com/~/media/mckinsey/industries/advanced%20electronics/our%20insights/how%20artificial%20intelligence%20can%20deliver%20real%20value%20to%20companies/mgi-artificial-intelligence-discussion-paper.ashx. Elsevier, *Artificial Intelligence: How knowledge is created, transferred, and used - Trends in China, Europe, and the United States*, www.elsevier.com/research-intelligence/resource-library/ai-report. Ménière Y., Rudyk I. and Valdes J., *Patents and the Fourth Industrial Revolution: The Inventions behind Digital Transformation*, Munich, DE: European Patent Office.

[391] ITU, "Assessing the Economic Impact of Artificial Intelligence", Issue Paper No.1, International Telecommunications Union, Geneva, pp. 12-15, www.itu.int/dms_pub/itu-s/opb/gen/S-GEN-ISSUEPAPER-2018-1-PDF-E.pdf. Pricewaterhouse Coopers (PwC), *Sizing the prize: PwC's Global Artificial Intelligence Study: Exploiting the AI Revolution*, PwC, www.pwc.com/gx/en/issues/data-and-analytics/publications/artificial-intelligence-study.html.

[392] ITU, *op.cit.*, pp. 17-20.

[393] Pfeiffer A., *Pfeiffer Report: Creativity and technology in the in age of AI*, pp. 15, 29, https://www.pfeifferreport.com/wp-content/uploads/2018/10/Creativity-and-technology-in-the-age-of-AI.pdf.

[394] New European Media (2019), *AI in media and creative industries*, https://arxiv.org/ftp/arxiv/papers/1905/1905.04175.pdf. Pfeiffer A., *op.cit.*

In this context, the adaptation of the Intellectual Property (IP) system to AI-generated creativity and innovation (and the challenges that it brings about) is increasingly becoming a topic of critical interest.[398] A substantive corpus of literature dedicated to AI and IP is emerging.[399] Of course, existing IP regimes, including copyright law, trade secrets and patent law[400] can protect software on which AI technology is based.[401] However, the protection afforded to the software does not extend to the output possibly generated by the AI. Whether this protection is available is actually still an open question, based on the construction of the present copyright framework. A distinction should also be made between computer-assisted creativity, which is copyrightable as long as the user contribution is original, and computer-generated creativity proper, where a user's interaction with a computer prompts it to generate its own expression.[402] A report from the European Commission clearly presents the terms of this emerging quagmire:

> *Protection of AI-generated works […] seems to be […] problematic. In light of the humanist approach of copyright law, it is questionable that AI-generated works deserve copyright protection. […] While some copyright scholars clearly advocate for AI-generated works to be placed in the public domain, others have put forward a series of proposals aimed at*

---

[395] *Artificial Intelligence in the audiovisual industry, Summary of the EAO workshop*, Strasbourg, 17 December 2019, European Audiovisual Observatory, Strasbourg, 2019, https://rm.coe.int/summary-workshop-2019-bat-2/16809c992a. See also Baujard T., Tereszkiewicz R., de Swarte A., Tuovinen T., "Entering the new paradigm of artificial intelligence and series", study commissioned by the Council of Europe and Eurimages, December 2019, https://rm.coe.int/eurimages-entering-the-new-paradigm-051219/1680995331.

[396] Strum B. et al., "Artificial Intelligence and Music: Open Questions of Copyright Law and Engineering Praxis", Arts 8, pp. 115-129. BPI, *Music's smart future: How will AI Impact the Music Industry*, www.musictank.co.uk/wp-content/uploads/2018/03/bpi-ai-report.pdf.

[397] Lovrinovic C. and Volland H., *The future impact of artificial intelligence on the publishing industry*, Gould Finch and Frankfurter Buchmesse, available at https://bluesyemre.files.wordpress.com/2019/11/the-future-impact-of-artificial-intelligence-on-the-publishing-industry.pdf.

[398] Cubert J.A. and Bone R.G.A., "The law of intellectual property created by artificial intelligence" in Barfield W. and Pagallo U. (eds), *Research Handbook on the Law of Artificial Intelligence*, Edward Elgar, Cheltenham, pp. 411-427. OECD, Artificial Intelligence in Society, OECD Publishing, Paris, p. 104-105, https://doi.org/10.1787/eedfee77-en. WIPO (2019a), *Draft Issues Paper on Intellectual Property Policy and Artificial Intelligence*, WIPO, Geneva, https://www.wipo.int/meetings/en/doc_details.jsp?doc_id=470053. WIPO (2019b), *WIPO Technology Trends 2019 - Artificial Intelligence*, WIPO, Geneva, https://www.wipo.int/edocs/pubdocs/en/wipo_pub_1055.pdf.

[399] Iglesias M., Shamuilia S. and Anderberg A., *Intellectual Property and Artificial Intelligence: A Literature Review*, Publications Office of the European Union, Luxembourg, https://publications.jrc.ec.europa.eu/repository/bitstream/JRC119102/intellectual_property_and_artificial_intelligence_jrc_template_final.pdf.

[400] Whether protecting software as a computer-implemented invention or as such, depending on the jurisdiction.

[401] Calvin N. and Leung J., "Who owns artificial intelligence? A preliminary analysis of corporate intellectual property strategies and why they matter", *Future of Humanity Institute, University of Oxford*, https://www.fhi.ox.ac.uk/wp-content/uploads/Patents_-FHI-Working-Paper-Final-.pdf.

[402] *Payer Components South Africa Ltd v Bovic Gaskins* [1995] 33 IPR 407. Clark R. and Smyth S., *Intellectual Property Law in Ireland*, Butterworths, Dublin. Denicola R., "Ex Machina: Copyright Protection for Computer-Generated Works", Rutgers University Law Review 69, pp. 269-270.

ensuring a certain level of protection. With notable exceptions, these proposals [...] do not always sufficiently detail the possible elements underpinning such protection.[403]

In this very regard, this chapter is meant to answer a set of emerging legal questions within the AI-generated creativity conundrum. How does AI-generated creativity fit with traditional copyright theory and existing doctrines? In particular, which are the conditions for protection of creations generated by AI and deep neural networks under the main copyright regimes? Should legal personhood for AI be considered? Is AI an author according to traditional copyright standards? Can a machine be original? These questions —which are only a portion of the relevant questions related to AI-generated creativity — can be summarised in the single question of whether AI can be an A(I)uthor. Additionally, there are two other fundamental questions beyond the scope of this review, relating to the (Machine) Learner and the (A)Infringer. They refer to whether an AI can infringe copyright through the machine learning process and training that enables the AI to generate creativity and whether an AI can infringe copyright by creating an infringing output. In addition to genuine challenges related to standards for AI authorship, this chapter will finally consider the road ahead by reviewing policy options from different theoretical perspectives, such as personality theories and utilitarian/incentive theories of intellectual property.

5.2. Technology

The first book ever written by a computer was *The Policeman's Beard is Half Constructed: Computer Prose and Poetry by Racter*.[404] It was 1984 and Racter's prose was still rather obscure and unpolished. Since then, things have been changing. The quality of AI-generated creativity has improved dramatically, to the extent that a novel written by a machine made the first rounds of a literary competition in Japan, beating in the process thousands of human authors,[405] and *Sunspring*, a sci-fi film written entirely by an AI, placed top 10 in the Sci-Fi London annual film festival.[406] AIVA — as well as Amper or Melodrive — runs an AI that composes music, which is marketed to accompany audiovisual works, advertisements or video games.[407] The Z-Machines, a Japanese robot band, perform music changing the pace of their performance according to actions taken by their audience as well as people who access their website,[408] while Sony's Flow Machine can interact and

[403] Craglia M., *Artificial Intelligence: A European Perspective*, Publications Office of the European Union, Luxembourg, pp. 67-68.
[404] Racter, *The Policeman's Beard is Half Constructed: Computer Prose and Poetry by Racter - The First Book Ever Written by a Computer*, Warner Books, New York.
[405] Lewis D., *An AI-Written Novella Almost Won a Literary Prize*, Smithsonian Magazine, Washington, https://www.smithsonianmag.com/smart-news/ai-written-novella-almost-won-literary-prize-180958577.
[406] Craig C. and Kerr I., "The Death of the AI Author", *Osgoode Legal Studies Research Paper*, pp. 1-2.
[407] AIVA, https://www.aiva.ai.
[408] Bakare L., *Meet Z-Machines, Squarepusher's new robot band*, The Guardian, https://www.theguardian.com/music/2014/apr/04/squarepusher-z-machines-music-for-robots.

co-improvise with a human music performer.[409] Visual art, however, appears to be the creative field where AI performs best. An AI-generated "Portrait of Edmond de Belamy" was sold at Christie's for an astounding USD 432 500.[410]

Due to massive data availability, enhanced computational resources and novel deep-learning-based architectures, AI has experienced major breakthroughs over the past decade.[411] Tightly connected to these advancements, a fundamental development of AI-generated creativity has been caused by the advent of the Generative Adversarial Network (GAN).[412] This is quite a recent development. In June 2014, Ian Goodfellow published a paper entitled "Generative Adversarial Networks", and posted the code on GitHub under a BSD licence.[413] The paper describes a generative process that uses an adversarial model for machine learning. In this scenario, two neural networks compete against each other in a game. Given a training set, this technique learns to generate new data with the same statistics as the training set. This became a wildly popular method for training AI with large datasets. The technology further evolved into Creative Adversarial Network (CAN) systems, which build over GANs and "generate art by looking at art and learning about style; and become creative by increasing the arousal potential of the generated art by deviating from the learned styles".[414] GANs and CANs were deployed by the Paris-based Obvious arts collective to generate the "Portrait of Edmond de Belamy" and a series of generative images called "La Famille de Belamy".[415]

Like Google's Deep Mind, which generates and performs music or creates artworks, AI does so by listening to other music or analysing previous artworks online. Pindar Van Arman has been teaching an AI how to be creative for some time now. The project, called cloudpainter.com, provides an exemplification of the similarities between human and machine learning processes in order to create art.[416] Apparently, an AI would learn how to generate creativity through a multiple-step learning process, starting from technical exercises, such as completing connecting-dots images, to move later to experimentation, imitation and, finally, independent creation.

---

[409] Deltorn J.M. and Macrez F. (2018), "Authorship in the Age of Machine Learning and Artificial Intelligence", Center for International Intellectual Property Studies Research Paper No. 2018-10, 22-23, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3261329.

[410] Craig C. and Kerr I., *op.cit.*, 3-4.

[411] Goodfellow I., Bengio Y. and Courville A., *Deep Learning*, MIT Press, Cambridge.

[412] Svedman M., "Artificial Creativity: A case against copyright for AI-created visual work", *IP Theory* 9(1), pp. 3-4.

[413] Goodfellow I. et al., "Generative Adversarial Networks", *arXiv*, https://arxiv.org/abs/1406.2661.

[414] Elgammal A. et al., "CAN: Creative Adversarial Networks Generating "Art" by Learning About Styles and Deviating from Style Norms", pp. 1-22, https://arxiv.org/abs/1706.07068.

[415] Obvious AI & Art. Available at https://obvious-art.com.

[416] Cloud Painter. Available at https://www.cloudpainter.com.

5.3. Protection: Can AI-generated creativity be protected?

AI is transforming the way we create, and is impacting long-established copyright concepts and doctrines. In particular, genuine issues have been arising regarding the protectability of AI-generated creativity under the current copyright regime. This question can be answered by looking into three major conditions for protection and ownership of copyright works: (1) legal personality; (2) authorship; (3) originality.

5.3.1. Personality: Can a machine be a legal person?

A first relevant question would be whether machines can enjoy legal personality. Depending on the jurisdiction, cultural and religious belief and legal subjectivity, establishing the personality of machines may become a policy option. Japan always had a special relationship with robots and machines due to the Shinto beliefs that animal or human-like robots can be imagined to have a soul.[417] In October 2017, Sophia became the first robot to be granted citizenship by the Saudi Arabian government. The move was obviously a PR stunt. Nonetheless, it is an historical step into a possible assimilation of AI and humankind. Actually, one months later, in November 2017, Tokyo granted a chatbot official residence status in the Shibuya ward.[418]

The idea of legal personality of intelligent machines has been also supported by theoretical thinking. Nick Bostrom, for example, notes: "Machines capable of independent initiative and of making their own plans … are perhaps more appropriately viewed as persons than machines".[419] Authors have highlighted how there are no legal reasons or conceptual motives for denying the personhood of AI robots: the law should be entitled to grant personality on the grounds of rational choices and empirical evidence, rather than superstition and privileges.[420] Therefore, arguments have been made in favour of granting personhood to future hypothetical strong AIs that are autonomous (capable of making a decision without input action), intelligent (capable of self-programming and integrating information in a framework) and possess consciousness (capable of subjective experience).[421]

More strikingly, the European Parliament is considering the possibility of declaring AI and robots "electronic persons". In a resolution on civil law rules on robotics, the

---

[417] Holland-Minkley D., *God in the Machine: Perceptions and Portrayals of Mechanical Kami in Japanese Anime*, Master's Thesis, University of Pittsburgh.

[418] Cuthbertson, *Tokyo: Artificial Intelligence 'Boy' Shibuya Mirai Becomes World's First AI Bot to Be Granted Residency*, Newsweek, Washington, https://www.newsweek.com/tokyo-residency-artificial-intelligence-boy-shibuya-mirai-702382.

[419] Bostrom N., *Superintelligence: Paths, Dangers, Strategies*, OUP, Oxford.

[420] Solum L.B., "Legal Personhood for Artificial Intelligences", *North Carolina Law Review* 70, p. 1264.

[421] Zimmerman E., "Machine Minds: Frontiers in Legal Personhood", pp. 14-21, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2563965. See also Hubbard F.P., "Do Androids Dream? Personhood and Intelligent Artifacts", *Temple law Review 83*, pp. 406-474.

European Parliament wonders whether the ordinary rules on liability are sufficient or whether AI calls for new principles and rules.[422] Should the autonomous nature of robots be construed in the light of the existing legal categories or should a new category be created?[423] The resolution claims that "the more autonomous robots are, the less they can be considered simple tools in the hands of other actors (such as the manufacturer, the owner, the user, etc.)"[424] It is apparent to the European Parliament that EU legislation cannot fully address non-contractual liability for damages caused by autonomous AI. Traditional rules would still apply if the cause of the robot's act or omission can be traced back to a specific human agent such as the manufacturer, the operator, the owner or the user. Again, traditional liability rules still apply if the robot has malfunctioned or if the human agent could have foreseen and avoided the robot's harmful behaviour. But what if the cause of the robot's act or omission cannot be traced back to a specific human agent? What if there are no manufacturing defects? And the AI has not malfunctioned? And the injured person is unable to prove the actual damage, or the defect in the product or the causal relationship between damage and defect? What if, in fact, the AI has caused damages because it has actually acted autonomously according to its own programming and purpose? In this scenario, Directive 85/374/EEC on Product Liability should not apply. The resolution highlights that this makes the ordinary rules on liability insufficient and calls for new rules to clarify whether a machine can be held responsible for its acts or omissions.[425] Although the resolution recognises that "at least at the present stage the responsibility must lie with a human and not a robot", in the long run the it calls for: (1) an obligatory insurance scheme which takes into account all potential responsibilities in the chain;[426] (2) the creation of a specific legal status for robots, "so that at least the most sophisticated autonomous robots could be established as having the status of electronic persons responsible for making good any damage they may cause".[427]

The notion of an AI legal personality has been emerging in multiple discussions but so far the debate has been dominated by inconsistent, tinkering attempts at regulating a technology whose development is wholly unpredictable. Therefore, as has often happened, the discourse about granting legal personhood becomes a political issue with no rational basis. In this respect, Saudi Arabia granting citizenship to Sophia is redolent of the Roman emperor Caligula making his horse, Incitatus, a senator.[428]

---

[422] Cf. Vladeck D., "Machines without Principals: Liability Rules and Artificial Intelligence", *Washington Law Review* 89, pp. 117-150.

[423] European Parliament (2017), Civil Law Rules on Robotics: European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics, 2015/2103(INL), 16 February 2017, https://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_EN.pdf.

[424] European Parliament (2017), *op.cit.*

[425] European Parliament (2017), *op.cit.* See also European Commission, "Liability for Artificial Intelligence and other Emerging Technologies: Report from the Expert Group on Liability and New Technologies – New Technologies Formation", European Union, Brussels, https://ec.europa.eu/transparency/regexpert/index.cfm?do=groupDetail.groupMeetingDoc&docid=36608.

[426] European Parliament (2017), *op.cit.*

[427] European Parliament (2017), *op.cit.*

[428] Pagallo U., "Vital, Sophia, and Co.—The Quest for the Legal Personhood of Robots", *Information* 9(9), pp. 239-240.

Whether quasi-human or hyper-human AI will be coming, legal personality of machines is certainly unavailable under the present legal framework. Most likely it will be unavailable for the foreseeable future. Scholarship has been consistently stressing how any hypothesis of granting AI robots full legal personhood has to be discarded until fundamental technological changes may occur.[429] Pagallo highlights, among the normative arguments against legal personhood, the "missing something problem", according to which current AI robots lack most requisites that usually are associated with granting someone, or something, legal personhood: such artificial agents are not self-conscious, they do not possess human-like intentions, nor properly suffer.[430] Statistical analysis of different conditions for legal personhood set up by US case law, for example, would also show incompatibility between legal personhood and AI entities.[431] This empirical analysis proves that, to grant personhood, courts look at whether it is being granted directly or indirectly by a statute, if the artificial entity can sue and be sued, and finally if the entity is an aggregate of natural persons.[432]

These considerations serve also to set apart AI from corporations that are treated as a legal person. Unlike corporations, AI entities are neither "fictional" entities nor associations of natural persons.[433] Legal persons are formed by natural persons, who can ultimately exploit rights. In addition, although legal persons can own a copyright, that copyright originated from a work created by a natural person, who is the author of the work, which then fulfils both the requirements of authorship and originality. This would not be the case with an AI and an AI-generated work, as to be discussed in the next pages.

These arguments against AI's legal personality may have already been internalised by policy-makers, as the European Parliament's 2017 resolution could prove by excluding any form of AI legal personality at least in the short and mid-term. In addition, the European Parliament appears now to exclude AI's legal personality in specific connection to AI-generated creativity. In a recent "Draft report on intellectual property rights for the development of artificial intelligence technologies", the European Parliament noted, as part of a motion for a parliament resolution, that "the autonomisation of the creative process raises issues relating to the ownership of IPRs [but] considers, in this connection, that it would not be appropriate to seek to impart legal personality to AI technologies".[434] Rather than establishing the legal personality of machines, the policy challenge would be

[429] Banteka N., "Artificially Intelligent Persons", Houston Law Review 58 (forthcoming), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3552269. Mik E., "AI as Legal Person?" in Hilty R. and Liu K-C. (eds), *Artificial Intelligence and Intellectual Property,* Oxford University Press, England, forthcoming. Pagallo U., *op.cit.*, pp. 230-240.
[430] Pagallo U., *op.cit.*, pp. 237-238.
[431] Banteka N., *op. cit.*
[432] Banteka N., *op. cit.*, p. 50-52.
[433] Banteka N., *op. cit.,* p. 19.
[434] European Parliament (2020), Draft Report on intellectual property rights for the development of artificial intelligence technologies, 2020/2015(INI), 24 April 2020, https://europarl.europa.eu/doceo/document/JURI-PR-650527_EN.html.

to properly allocate accountability and liability for the activities of AI robots in cases of complex distributed responsibility, for example through contracts and business law.[435]

5.3.2. Authorship: Can a machine be an author?

Although, absent legal personality, AI cannot be vested with authorship or standing for enforcing rights on creativity that it might generate, it still remains relevant to consider whether that creativity is protectable under the present legal framework. Answering the broader question of whether AI-generated creativity is protectable under copyright law implies consideration of whether AI can be construed as an author according to traditional copyright standards. This boils down to whether the existence of a human being is an intrinsic requirement for authorship. Can an author be a machine or does it need to be human?

There is actually no definition in international treaties that can provide a definitive answer. However, it appears that textual reference to human creation in the Berne Convention[436] may exclude the possibility of construing AI as an author. For one thing, the term of "protection", linked to the life of the author, appears to rule out machines as authors (Berne Convention, Art. 7). Again, reference to the nationality—or residence—of the author seems to imply that the notion of authorship only applies to human agents (Berne Convention, Art. 3). Overall, it has been argued that "Berne's humanist cast" and its deference to idealist personality theories strongly support a "human-centred notion of authorship presently enshrined in the Berne Convention" that would exclude non-human authorship from Berne's scope.[437]

5.3.2.1. The European Union

A close review of EU law would most likely lead to similar conclusions.[438] Although there is no transversal definition in statutory law of the notion of authorship, an author is defined as a natural person, a group of persons or a legal person both by Art. 2(1) of the

[435] European Parliament (2017), *op.cit.* Pagallo U., *op.cit.*, pp. 239-240.
[436] Berne Convention for the Protection of Literary and Artistic Works (as amended on September 28, 1979), https://wipolex.wipo.int/en/treaties/textdetails/12214.
[437] Ginsburg J., "People Not Machines: Authorship and What It Means in the Berne Convention", International Review of Intellectual Property and Competition Law 49, pp.131-135. See also Aplin T. and Pasqualetto G., "Artificial Intelligence and Copyright Protection", §5.04, in Ballardini R., Kuoppamäki P. and Pitkänen O.(eds.), *Regulating Industrial Internet Through IPR, Data Protection and Competition Law*, Kluwer, Alphen aan den Rijn. Ricketson S., "The Need for Human Authorship - Australian Developments: Telstra Corp Ltd v Phone Directories Co Pty Ltd (Case Comme nt)", *E.I.P.R.* 34(1), p. 34. Ricketson S. (1991), "People or machines? The Berne Convention and the changing concept of authorship", *Columbia VLA Journal of Law & the Arts* 16, p. 34.
[438] Deltorn J.M. (2017), "Deep Creations: Intellectual Property and the Automata", *Frontiers in Digital Humanities*, p. 8, https://www.frontiersin.org/articles/10.3389/fdigh.2017.00003/full. Deltorn J.M. and Macrez F. (2018), *op.cit.*, p. 8.

Software Directive[439] and Art. 4(1) of the Database Directive[440]. In addition, Art. 2(1) of the Term Directive[441] provides that the principal director of a cinematographic and audiovisual work shall be considered its author or one of its authors. Actually, the *travaux préparatoires* of the Software and of the Database Directives were more straightforward in endorsing an anthropocentric vision of authorship by referring specifically to "the human author who creates the work" and "the natural person [who] will retain at least the unalienable rights to claim paternity of his work".[442] The original proposal for a Software Directive concluded: "[t]he human input as regards the creation of machine-generated programmes may be relatively modest, and will be increasingly modest in the future. Nevertheless, a human 'author' in the widest sense is always present, and must have the right to claim 'authorship' of the program".[443] In the Court of Justice of the European Union (CJEU) Painer case, Advocate General Verica Trstenjak stressed the same point by noting that "only human creations are therefore protected, which can also include those for which the person employs a technical aid, such as a camera".[444]

National legislation of EU member states confirms this approach. For example, Art. L.111-1 of the French Intellectual Property Code[445] requires copyrightable work to be the "creation of the mind". Art. 5 of the Spanish Copyright Act plainly states that "the author of a work is the natural person who creates it".[446] And, although Art. 7 of the German Copyright Act does not specifically limit authorship to natural persons, Art. 11 attaches authorship to a personality approach by protecting "the author in his intellectual and personal relationships to the work".[447]

In addition, EU law — as well as multiple national legislations (e.g. Dutch Copyright Act, Art. 4(1); French IP Code, Art. L113-1; Spanish Copyright Law, Art 6.1; Italian Copyright Law, Art. 8)[448] — endorses a human-centric approach when providing a

[439] Directive 2009/24/EC of the European Parliament and of the Council of 23 April 2009 on the legal protection of computer programs, O.J. L111/16.

[440] Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases, O.J. L077/20.

[441] Directive 2006/116/EC of the European Parliament and of the Council of 12 December 2006 on the term of protection of copyright and certain related rights, O.J. L372/12.

[442] Ramalho A., "Will Robots Rule the (Artistic) World? A Proposed Model for the Legal Status of Creations by Artificial Intelligence Systems", *Journal of Internet Law* 21, pp. 17-18.

[443] European Commission, Explanatory Memorandum to the proposal for a Software Directive, COM (88) 816 final, 17 March 1989, p. 21.

[444] Opinion of the AG Trstenjak (12 April 2011), *C-145/10 Eva-Maria Painer v. Standard VerlagsGmbH*, ECLI:EU:C:2011:239.

[445] Code de la propriété intellectuelle [Intellectual Property Code] 1912. Available at https://www.wipo.int/edocs/lexdocs/laws/en/fr/fr467en.pdf (France).

[446] The Intellectual Property Act 1996. Available at https://www.wipo.int/edocs/lexdocs/laws/en/es/es177en.pdf (Spain).

[447] Urheberrechtsgesetz – UrhG (Act on Copyright and Related Rights) 1965. Available at https://www.gesetze-im-internet.de/englisch_urhg/englisch_urhg.html (Germany).

[448] Auteurswet (Copyright Act) 1912. Available at https://wetten.overheid.nl/BWBR0001886/2012-01-01 (Netherlands); Code de la propriété intellectuelle [Intellectual Property Code] 1912. Available at https://www.wipo.int/edocs/lexdocs/laws/en/fr/fr467en.pdf (France); The Intellectual Property Act 1996. Available at https://www.wipo.int/edocs/lexdocs/laws/en/es/es177en.pdf (Spain); Law for the Protection of Copyright and Neighbouring Rights 1941. Available at

presumption of authorship for the *person* whose name is indicated in the work, in the absence of proof to the contrary (IP Enforcement Directive, Art. 5). Some national courts have even clarified that this presumption is only applicable to natural persons creating the work, and not to a legal person who might have obtained the economic rights.[449] In theory, this presumption of authorship could apply to AI-generated works, so that the person/s whose name/s is/are indicated in the work is/are regarded as the author/s. Of course, this solution provides no actual protection against infringement, given that the presumption can be rebutted by proving that the person named is not the author, but an AI is.

A brief overview of other major jurisdictions might lead to similar conclusions regarding the application of the notion of authorship to AI.

5.3.2.2. Australia

Australian law sets a quite clear statutory bar for non-human authors by defining an author as a "qualified person" in Section 32(1) of the Australian Copyright Act, who, Section 32(4) in turn, defines as an Australian citizen or a person resident in Australia.[450] Australian courts, then, link originality as a condition for protection of authorship. In *Acohs v. Ucorp*, involving subsistence of copyright in data sheets generated electronically, the court clarified that a work needs to "spring from the original efforts of a single human author".[451] The Phone Directories decision reinforces the point by noting that copyright "only subsists if it originates from an individual".[452] Finally, *IceTV v. Nine Network Australia* decided a case dealing with copyright for computer generation of weekly TV program schedules by concluding that only authors, thus persons according to the statutory definition, can be original.[453]

5.3.2.3. United States

The protection of products of computational creativity is not novel in the United States. Scholars started discussing possible protectability of computer-generated creativity in the

---

https://www.wipo.int/edocs/lexdocs/laws/en/it/it211en.pdf (Italy).

[449] *Herlitz PBS AG vs. Realister OÜ*, Estonian Supreme Court (7 February 2012) Case No. 3-2-1-155-11 (Estonia). See also Vasamae E., "Presumption of authorship: only natural persons", *Kluwer Copyright Blog*, Amsterdam, available at http://copyrightblog.kluweriplaw.com/2012/03/19/presumption-of-authorship-only-natural-persons/?doing_wp_cron=1594514535.1866068840026855468750.

[450] Copyright Act 1968. Available at https://www.legislation.gov.au/Details/C2019C00042(Australia).

[451] *Acohs Pty Ltd v. Ucorp Pty Ltd* (2010) 86 IPR 492 (AUS).

[452] *Phone Directories Co Pty v Telstra Corporation Ltd* (2010) 194 FCR 142 (AUS).See also McCutcheon J., "The Vanishing Author in Computer-Generated Works: A Critical Analysis of Recent Australian Case Law", *Melbourne University Law Review* 36(3), pp. 941-969, https://www.researchgate.net/publication/289409001_The_vanishing_author_in_computer-generated_works_A_critical_analysis_of_recent_Australian_case_law.

[453] *IceTV Pty Ltd v. Nine Network Australia Pty Ltd* [2009] HCA 14, 239 CLR 458 (AUS).

late 1960s.[454] The US Congress created a committee to determine whether computers or computer programmes can be authors whose output can be copyrighted. In 1978, the National Commission on New Technological Uses of Copyrighted Works (CONTU Commission) noted that computers were mere "inert tools of creation" which were not yet independently creating works. The CONTU Commission did not discuss copyright protection of automated works devoid of human authorship because it was considered too speculative at the time.[455] In 1986, the Congress Office of Technology Assessment (OTA) issued a report arguing that although computers were more than "inert tools of creation" the copyrightability of computer-generated works was undeterminable.[456]

The US Copyright Act does not have an express statutory definition of authorship, so that authors have initially argued that textually, the statute does not limit authorship to human authors.[457] However, both additional textual references and case law apparently exclude the possibility of construing non-human agents as authors under the statute. In particular, Section 101 of the Copyright Act[458] defines anonymous works as the "ones where no natural person is identified as an author", thus pointing at natural persons as potential authors. Also, there is a long-lasting understanding that the constitutional history of the word "copyright" would dispose in favour only humans as "authors".[459] U.S. courts have consistently supported this understanding. The Supreme Court has plainly stated that "[a]s a general rule, the author is [...] the *person* who translates an idea into a fixed, tangible expression entitled to copyright protection".[460] In *Feist v Rural*, the U.S. Supreme Court discusses at length the notion of authorship and author by reviewing the notion of originality, which would refer to inherently human features, such as "creative spark" or "intellectual production, of thought, and conception".[461] Earlier cases would support the same conclusion. The Trade-Mark Cases state that the copyright law only protects "the fruits of intellectual labor" that "are founded in the creative powers of the

[454] Milde K.F., "Can a Computer Be an "Author" or an "Inventor"?", Journal of the Patent Office Society 51, pp. 378-406.

[455] National Commission on New Technological Uses of Copyrighted Works (CONTU), *Final Report*, United States, p .44. See also Bridy A. (2012), "Coding Creativity: Copyright and the Artificially Intelligent Author", Stanford Technology Law Review 5, 22-24, para 53-60. Miller A.R., "Copyright Protection for Computer Programs, Databases, and Computer Generated Works: Is Anything New Since CONTU?", *Harvard Law Review* 977, pp. 977-1072.

[456] US Office of Technology Assessment, *Intellectual Property Rights in an Age of Electronics and Information*, United States.

[457] Bridy A. (2012), op.cit., 20, para 49. Denicola R. (2016), *op.cit.*, p. 275-283. Miller A.R., *op.cit.*, pp. 1042-1072; Samuelson P. (1986), "Allocating Ownership Rights in Computer-Generated Works", *University of Pittsburg Law Review* 47(4), pp. 1200-1204.

[458] The Copyright Act of 1976. Available at https://www.copyright.gov/title17/title17.pdf (US).

[459] Butler T., "Can a computer be an author? Copyright aspects of artificial intelligence", (Comm/Ent), A Journal of Communications and Entertainment Law 4(4), pp. 733-734. Clifford, R.D. (1996), "Intellectual Property in the Era of the Creative Computer Program: Will the True Creator Please Stand up", Tulane Law Review 71, 1682-1686. Kasap A., "Copyright and Creative Artificial Intelligence (AI) Systems: A Twenty-First Century Approach to Authorship of AI-Generated Works in the United States", *Wake Forest Intellectual Property Law Journal* 19(4), p. 358, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3597792. Milde K.F., *op.cit.*, pp. 391-392.

[460] *Community for Creative Non-Violence v. Reid*, 490 U.S. 730 (1989).

[461] *Feist Publications v. Rural Telephone Service*, 499 U.S. 340 (1991) (USA).

mind".[462] In the Burrow-Giles case, the US Supreme Court recalled that copyright law is limited to "original intellectual conceptions of the author".[463]

A recent case shed some further clarifications on the matter. This time, Naruto, a macaque monkey, came to the rescue. In *Naruto v. Slater*, two "monkey selfies" that received worldwide recognition were the subject of a dispute about whether animals can own copyright. The self-portraits were taken by seven-year-old crested macaque "Naruto" when wildlife photographer David Slater left his camera unattended on one of his visits to Indonesia. Shortly thereafter, Wikimedia Commons published the pictures on its website under the assumption that the monkey selfies have no human author and therefore belong to the public domain. Wikimedia consistently refused to take down the pictures. Changing earlier stances advertising the selfies as entirely taken by the monkeys with no human intervention, Slater later claimed the selfies were the result of his setting up the camera with the right angle, lighting, optimising settings and just luring the monkeys into pressing the camera button.

Although the question of whether the selfies belong to the public domain was not ultimately reviewed, courts had the opportunity to consider whether Naruto could be vested with a copyright for its selfie. In 2014, the monkey selfies were published in a book through Blurb Inc. which identified Slater and Wildlife Personalities Ltd as the copyright owners. In 2015, People for the Ethical Treatment of Animals (PETA) filed a complaint of copyright infringement as next friends and on behalf of Naruto against Slater, Wildlife Personalities Ltd and Blurb Inc. before the District Court, California. The District Court granted the motion to dismiss filed by the defendants on the basis that Naruto failed to establish statutory standing under the Copyright Act and noted: "If the humans purporting to act on Plaintiff's behalf wish for copyright to be among the areas of law where nonhuman animals have standing, they should make that dubious case to Congress – not the federal courts."[464] The decision was appealed and while the parties agreed to a settlement, the Court of Appeals declined to dismiss the appeal and affirmed the lower court decision. The majority found that while animals have Art. III standing to sue, animals do not have statutory standing under the Copyright Act.[465] The court relied on the Ninth Circuit decision in *Cetacean Community. v. Bush,* where it was held that animals have statutory standing only if the statute plainly states so.[466] Moreover, the terms 'children', 'grandchildren', 'legitimate', 'widow', and 'widower' used in the statute necessarily imply that the Copyright Act excludes animals that "do not marry and do not have heirs entitled to property by law".[467] The findings in the Naruto decision can easily be extended to any non-human and AI-generated creativity.

Meanwhile, the "Third Edition of the Compendium of U.S. Copyright Office Practices", which was published in December 2014 after the Naruto case started, provided

[462] *Trade-Mark Cases*, 100 U.S. 82 (1879) (USA).
[463] *Burrow-Giles Lithographic Co. v. Sarony*, 111 U.S. 53 (1884) (USA).
[464] *Naruto v. David Slater*, 15-cv-04324-WHO (N.D. Cal. 2016) (USA) ("Naruto 2016").
[465] *Naruto v. David Slater*, F.3d 418 (9th Cir. 2018) (USA) ("Naruto 2018").
[466] *Cetacean Community. v. Bush*, 386 F.3d 1169 (9th Cir. 2004) (USA).
[467] *Naruto v. David Slater*, F.3d 418 (9th Cir. 2018) (USA) ("Naruto 2018").

a non-binding expert guidance that excluded non-human authorship.[468] The compendium repeatedly refers to persons or human beings when discussing authorship. More specifically, under Section 306, "The Human Authorship Requirement" limits registration to "original intellectual conceptions of the author" created by a human being. As clarified under Section 313.2, "Works that Lack Human Authorship", works produced by nature, animal or plants and similarly, works created by a machine or by a mechanical process without intervention from a human author are not copyrightable. Making reference to the Trade-Mark Cases and Burrow-Giles, the Copyright Office concluded that it would refuse to register a claim if it determines that a human being did not create the work.[469]

5.3.2.4. China

In China, the questions of AI authorship and copyrightability of AI-generated works have been discussed by multiple courts. The Chinese position on AI authorship appears aligned with that of other jurisdictions, although it leaves some room for potential protection. In *Beijing Feilin Law Firm v Baidu Corporation*, affirming the requirement of human authors, the court denied copyright protection to works created solely by machines.[470] The matter involved a report published by the plaintiff — a Beijing-based law firm — on its official WeChat account. After an unidentifiable Internet user published the report online without permission, the plaintiff brought an infringement suit before the Beijing Internet Court. The report had been generated using Wolters Kluwer China Law & Reference —a legal information query software. While the plaintiff argued that the tool was used only for assistance, the defendants claimed that the entire report was generated by the software. The court agreed with the plaintiff. However, although the disputed report was found to be protected by Chinese copyright law, the court considered also the protectability of the report automatically generated by the software. In discussing protection of works exclusively generated by an AI, the court held that the notion of authorship requires the work to be created by a natural person. The court, however, came up with some interesting incentive analysis which rejects the conclusion that the work should be freely available in the public domain. Thus, the court believed that some sort of protection should be given to the user — not the software developer already rewarded with copyright over the software — in order to incentivise purchases of the software as well as generation and distribution of the works. Unfortunately, the judgement does not clarify which form this protection should take.

---

[468] U.S. Copyright Office, *Compendium of U.S. Copyright Office Practices*, 3rd edition, https://www.copyright.gov/comp3/comp-index.html

[469] U.S. Copyright Office, *op.cit.*, Section 313.2.

[470] *Beijing Feilin Law Firm v Baidu Corporation* (26 April 2019) Beijing Internet Court, (2018) Beijing 0491 Minchu No. 239. He K. (2020b), "Feilin v. Baidu: Beijing Internet Court tackles protection of AI/software-generated work and holds that copyright only vests in works by human authors", *The IPKat*, http://ipkitten.blogspot.com/2019/11/feilin-v-baidu-beijing-internet-court.html.
Chen M., "Beijing Internet Court denies copyright to works created solely by artificial intelligence", *Journal of Intellectual Property Law & Practice* 14(8), pp. 14-18.

In a later decision, *Shenzhen Tencent v. Yinxun*, the Nanshan District Court in Shenzhen basically confirmed the Beijing ruling.[471] The two decisions mirror each other insofar as the courts provided protection to the original contributions from human agents, rather than creativity exclusively AI-generated. The plaintiff Tencent Technology developed an AI writing assistant, Dreamwriter. In August 2018, the plaintiff published one of the AI-created works on its website, informing readers that the article had been written by Tencent's AI Dreamwriter. The defendant allegedly published the article on their website without the consent of the plaintiff. In a suit for infringement, the plaintiff argued that as authors of the article, they have exclusive rights under copyright law. They claimed that the article was generated under their supervision and they were responsible for the organisation and creation of the article as well as any liability arising thereof. In favour of the plaintiff, the court ruled that the article met the requirements of being an original literary work, as the content was a product of the input data, trigger conditions and arrangement of templates and resources selected by an operational group of the plaintiff. Since the expression of the article came from individual choices and arrangements made by the plaintiff, the AI-generated article was considered a work of legal entities under Article 11 of the Copyright Law and the defendant was held liable for infringement. However, although the court might have viewed the work as an integrated intellectual creation, deriving both from the contribution by the human team and the operation of Dreamwriter, the protectability granted apparently stems from the human team contribution, rather than any AI contribution.

5.3.3. Originality: Can a machine be original?

Besides the construction of the notion of authorship, also the notion of originality as a condition for copyright protection appears to preclude protection of AI-generated creativity. Textual references and case law construe originality through an anthropocentric model that emphasises self-consciousness. Originality is defined through a so-called personality approach that describes an original work as a representation of the personality of the author. The word 'author' itself bears this meaning on its face, as the most accredited etymology of the word would have it deriving from the ancient Greek *αὐτός*, which means 'self'.[472] This characterisation of originality builds upon idealist personality theories, according to which intellectual products are manifestations or extensions of the personalities of their creators.[473] Therefore, originality as a

[471] *Shenzhen Tencent v. Yinxun*, Nanshan District People's Court of Shenzhen, Guangdong Province [2019] No. 14010 (China), available at https://mp.weixin.qq.com/s/jjv7aYT5wDBIdTVWXV6rdQ. He K. (2020a), "Another decision on AI-generated work in China: Is it a Work of Legal Entities?", *The IPKat*, http://ipkitten.blogspot.com/2020/01/another-decision-on-ai-generated-work.html.

[472] Frosio G., *Reconciling Copyright with Cumulative Creativity: The Third Paradigm,* Edward Elgar, Cheltenham, p. 16.

[473] Fichte J., *Proof of the illegality of reprinting: a rationale and a parable*, Berlinische Monatsschrift 21, p. 447. Hegel G.H., *Philosophy of rights*, Thomas Knox, Clarendon Press, Oxford, para. 69. Kant I. (1785), *Von der Unrechtmäßigkeit des Büchernachdrucks* [On the injustice of counterfeitingbooks], Berlinische Monatsschrift 5,

representation of 'self' and self-consciousness would be, in theory, beyond the reach of machine-generated creativity. This construction of originality has been widely endorsed by the majority of jurisdictions. It has sidelined earlier approaches building upon Lockean fairness theories and endorsing "sweat of the brow" doctrines that rewarded "skills, labour and efforts" in creating intellectual work regardless of whether the work was representative of the personality of the author.[474]

In the European Union, three directives have vertically harmonised the notion of originality. According to Article 1(3) of the Software Directive, Article 6 of the Term Directive, and Article 3(1) of the Database Directive, a work is original if it is "the author's own intellectual creation".[475] Later, the CJEU 'horizontally' expanded originality to all copyright subject matters and further clarified the scope of the notion. In the Infopaq case, the CJEU noted that "[i]t is only through the choice, sequence and combination of those words that the author may express his creativity in an original manner and achieve a result that is an intellectual creation."[476] The Eva-Maria Painer decision further explained that a work — in that instance a portrait photograph — is original and can be protected, if it is: (1) an intellectual creation of the author; (2) reflects their personality; (3) expresses their free and creative choices in the production of that photograph.[477] By making those various choices, the author of a portrait photograph can stamp the work created with his 'personal touch'.[478] Finally, in the Football Dataco case, the CJEU rejected any remaining "sweat of the brow" doctrines and noted that significant labour and skill of the author cannot as such justify copyright protection, if that labour and that skill do not express any originality in the selection or arrangement.[479] Works produced merely based on technical rules or constraints lack the creative freedom required for authorship.

US jurisprudence has equally endorsed this personality approach to originality. Since early cases, such as *Burrow-Giles v. Sarony*, considering the copyrightability of a portrait photograph of Oscar Wilde, the U.S. Supreme Court has clarified that originality derives from the free creative choices of the author that imbue the work with his personality[480] "such as the final product duplicates his conceptions and visions" of what the work should be.[481] In particular, in the Burrow-Giles case, the court held photographs copyrightable because they could be traced from the photographer's "own original mental conception".[482] Later, in *Feist v. Rural*, the U.S. Supreme Court clearly stated that only

---

pp. 403–417. See also Fisher W., "Theories of intellectual property", in Munzer S. (ed.) *New essays in the legal and political theory of property*, Cambridge University Press, Cambridge, pp. 168–200.

[474] See e.g. *International News Service v. Associated Press*, 248 U.S. 215 (1918) (USA). *Jeweler's Circular Publishing Co. v. Keystone Publishing Co.*, 281 F. 83 (2nd Cir. 1922) (USA). See also Rahmatian A., "Originality in UK Copyright Law The Old "Skill and Labour" Doctrine Under Pressure", IIC 44, pp. 4–34.

[475] For a discussion see Rosati E., *Originality in EU Copyright - Full Harmonization through Case Law*, Edward Elgar, Cheltenham.

[476] *Infopaq International A/S v Danske Dagblades Forening*, C-5/08 (2009) ECLI:EU:C:2009:465.

[477] *Eva-Maria Painer v Standard VerlagsGmbH and Others*, C-145/10 (2011) ECLI:EU:C:2011:798.

[478] *Eva-Maria Painer v Standard VerlagsGmbH and Others*, C-145/10 (2011) ECLI:EU:C:2011:798.

[479] *Football Dataco Ltd and Others v Yahoo! UK Ltd and Others*, C-604/10 (2012) ECLI:EU:C:2012:115.

[480] *Burrow-Giles Lithographic Co. v. Sarony*, 111 U.S. 53 (1884) (USA).

[481] *Lindsay v. The Wrecked and Abandoned Vessel R.M.S. Titanic*, 52 U.S.P.Q.2d 1609 (S.D.N.Y. 1999) (USA).

[482] *Burrow-Giles Lithographic Co. v. Sarony*, 111 U.S. 53 (1884) (USA).

works with a minimum of creativity that represents the personality of the author can be original; labour and efforts alone in creating a work would not qualify for copyright protection.[483] In light of these systemic considerations, output such as computational shorthand[484] or listing of automatically numbered hardware parts created using software systems have been found to lack the originality for protection under copyright.[485]

Actually, Samuelson — and other authors — would argue there are no statutory limitations in the U.S. on treating a machine as an author as "[t]he copyright standard of originality is sufficiently low [so] that computer-generated works, even if found to be created solely by a machine, might seem able to qualify for protection."[486] I would argue that, after the Feist case, originality is not only a quantum question. For AI-generated creativity purposes, it is irrelevant whether the standard of originality is low or high. The standard the AI fails to reach is qualitative rather than quantitative. AI cannot express 'self'. The creativity that it generates cannot express the personality of the author because AI has none. In this regard, the United States joining the Berne Convention in 1988 and the Feist case in 1991 signal the crystallisation of a global, more harmonised view of copyright. This alignment of the United States with the European model also includes a construction of originality in personality theory terms.[487]

More recently, a few remaining — mainly common law — jurisdictions have been joining this personality approach to originality. This has been the case in Australia,[488] India,[489] and the United Kingdom,[490] which have finally rejected previous "labour, skill and efforts" approaches. Just a few countries still follow "sweat of the brow" doctrines and reject personality approaches to originality, including South Africa[491] and New Zealand.[492]

In sum, there appears to be an extremely consistent international construction of the notion of originality which emphasises an anthropocentric vision according to which a work is original if it is a representation of 'self', a representation of the personality of the author. Only if that inner attachment between the author and the work is present, is the originality requirement fulfilled, and protection granted. Of course, only a sentient self-conscious being would be capable of representing 'self' through a work. In turn, even if any possible textual anthropocentric construction of authorship is disregarded, absent the creator's self-consciousness, the originality requirement that lies in the representation of

---

[483] *Feist Publications v. Rural Telephone Service*, 499 U.S. 340 (1991) (USA).
[484] *Brief English Systems v. Owen*, 48 F.2d 555 (2d Cir. 1931) (USA).
[485] *Southco, Inc. v. Kanebridge Corporation*, 390 F.3d 276 (3d Cir. 2004) (USA).
[486] Samuelson P. (1986), op.cit., pp. 1199-1200. See also Brown N., "Artificial Authors: A case for copyright in computer-generated works", Columbia Science and Technology Law Review 9, pp. 24-27. Kaminski M., "Authorship, Disrupted: AI Authors in Copyright and First Amendment Law", *UC Davis Law Review* 51, p. 601. *Contra* see Clifford, R.D. (1996), *op.cit.*, pp. 1694-1695.
[487] Price M. E & Pollack M., The Author in Copyright: Notes for the Literary Critic, 10 Cardozo Arts & Entertainment Law Journal 703 (1992), pp. 717-720, https://larc.cardozo.yu.edu/faculty-articles/123.
[488] *IceTV Pty Ltd v. Nine Network Australia Pty Ltd* [2009] HCA 14, 239 CLR 458 (AUS).
[489] *Eastern Book Co. & Ors v. D.B. Modak & Anr* (2008) 1 SCC 1 (India).
[490] *Temple Island Collections v New English Teas* (No. 2) [2012] EWPCC 1. Rahmatian A., *op.cit.*, pp. 4-34
[491] *Appleton v. Harnischfeger Corp.* 1995 (2) SA 247 (AD) at 43–44 (SA).
[492] *Henkel KgaA v. Holdfast* [2006] NZSC 102, [2007] 1 NZLR 577 (NZ).

the personality of the author can never be fulfilled. Therefore, unless it can be claimed that machines have achieved self-consciousness, which might be the case for futuristic, hypothetical, strong AI, but not today,[493] AI-generated creativity cannot meet the originality requirement under the present legal framework.[494]

As some have argued, only a novel, perhaps a more formal, objective approach — as opposed to the existing, subjective approach — to the concept of originality would be able to include within the scope of copyright protection works created by creative robots as well as artworks generated by digital tools.[495] From this objective perspective, a judge should look at the final output per se, considering the field of art, the objective opinion of users, and similarity to other works, while disregarding the subjective intention of the author.[496] In this respect, the standard for originality in copyright should align itself more closely to the standard for novelty in patent law, which considers protectable subject matter from a social/historical perspective rather than an individual/subjective perspective.[497]

5.4. Policy options: Are incentives necessary?

Scholars and courts have raised the point that a legal system that does not grant protection to AI-generated creativity would create negative externalities from an 'incentive theory' perspective. Incentive theory or utilitarianism,[498] which is dominant in the United States and common law jurisdictions, is more removed from the humanity of its author than personality theories heavily influencing civil law jurisdictions.[499] This provides more room for argument in favour of non-human authorship and protectability of AI-generated creativity. According to the incentive theory approach, "providing financial incentives in order to encourage the growth and development of the AI industry and ensure the dissemination of AI-generated works is arguably the ultimate goal of assigning copyright to human authors".[500] Although a computer does not need an incentive to produce its output, the incentive may be useful for the person collaborating with the

[493] Zimmerman E., *op.cit.*, pp. 14-21.
[494] Clifford, R.D. (1996), *op.cit.*, 1694-1695. Deltorn J.M. (2017), *op.cit.*, p. 7. Deltorn J.M. and Macrez F. (2018), op.cit., p. 8. Gervais D.J., "The Machine as Author", *Iowa Law Review Vol. 105*, pp. 1-60. Mezei P., "From Leonardo to the Next Rembrandt – The Need for AI-Pessimism in the Age of Algorithms", https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3592187. Ramalho A., *op.cit.*, pp. 22-24.
[495] Yanisky-Ravid S. and Velez-Hernandez L.A., "Copyrightability of Artworks Produced by Creative Robots, Driven by Artificial Intelligence Systems and the Originality Requirement: The Formality-Objective Model", Minnesota Journal of Law, Science & Technology 19(1), pp. 40-48.
[496] Bonadio E. and McDonagh L., "Artificial Intelligence as Producer and Consumer of Copyright Works: Evaluating the Consequences of Algorithmic Creativity", Intellectual Property Quarterly 2, pp. 112-137, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3617197.
[497] Cf. Boden M., *The Creative Mind: Myths And Mechanisms*, Routledge, London, p. 32.
[498] Fisher W., *op.cit.*, pp. 177-180.
[499] Kaminski M., *op.cit.*, p. 599.
[500] Hristov K., "Artificial Intelligence and the Copyright Dilemma", *IDEA: The Intellectual Property Law Review* 57, p. 444. See also Brown N., *op.cit.*, 20-21.

computer.[501] In particular, authors argue that there should be some additional incentive to encourage industry to invest the time and money that it will take to teach machines to behave intelligently[502] or to reward users training and instructing AI to generate content.[503] As some argue, considerations of public policy under a utilitarian perspective would make it imperative for some form of protection to be given to AI-generated outputs whether copyright or unfair competition law protection or a sui generis protection,[504] as the process of creation — by human or computer — has no impact on its contribution to public welfare.[505]

However, most civil law jurisdictions may not be so fundamentally influenced by welfare and incentive considerations and may prefer to value systemic balance, thus rejecting any departure from the personality theory approach that shapes the civil law copyright perspective — and in fact the notion of originality in the large majority of jurisdictions. Therefore, although AI-generated creations may justify incentives to bolster innovation and commercialisation, the necessity of such incentives is questionable considering the impact they can have on human creations.[506] For example, considering the vast number of automated creations, granting protection for these works could devalue human authorship and existing jobs in the field and hamper creativity,[507] as it could discourage artists from publishing their creations due to the fear of infringing protected material[508] or clog the creative ecosystem with standardised and homogenised AI-generated outputs, impacting cultural diversity and identity politics.

The question to be determined is whether expansion of current copyright protection to computer-generated works is useful. The current legal framework may already provide enough protection through patent and copyright law to the underlying software, sui generis protection to databases or other legal mechanisms such as competition law to protect automated works without an extension of the existing copyright regime to non-human authors.[509] The questions should be investigated from a law and economics approach before any solutions are favoured.[510] Ginsburg and Budiardjo have stressed this point: '[w]e can conjure up a variety of scenarios supporting or

[501] Hristov K., *op.cit.*, 438-439. Miller A.R., *op.cit.*, 1067.
[502] Bridy A. (2012), *op.cit.*, 1-27. Butler T., *op.cit.*, p. 735. Farr E.H., "Copyrightability of Computer-Created Works", *Rutgers Computer & Technology Law Journal* 15, pp. 73-74. Kasap A., *op.cit.*, pp. 361-364; Abbott R., "I Think, Therefore I Invent: Creative Computers and the Future of Patent Law", 57 B.C.L. Rev. 1079 (2016), http://lawdigitalcommons.bc.edu/bclr/vol57/iss4/2. Milde K.F., *op.cit.*, p. 390.
[503] Brown N., *op.cit.*, p.37. Denicola R., *op.cit.*, p. 283. Ralston W.T., "Copyright in Computer-Composed Music: HAL Meets Handel", *Journal of the Copyright Society of the USA* 52, pp. 303-304. Samuelson p. (1986), *op.cit.*, pp. 1224-1228.
[504] Milde K.F., *op.cit.*, pp. 400-403.
[505] Butler T., *op.cit.*, p. 735. Denicola R., *op.cit.*, p. 273. Kaminski M., *op.cit.*, p. 599.
[506] Craglia M., *op.cit.*, pp. 67-68.
[507] Bonadio E. and McDonagh L., *op.cit.*
[508] Deltorn J.M. (2017), *op.cit.*
[509] Deltorn J.M. and Macrez F. (2018), *op.cit.*, p. 24.
[510] Craglia M., *op.cit.*, p. 68.

debunking the call for sui generis protection, but without empirical evidence, it would be imprudent (and premature) to seek to design a regime to cover authorless outputs".[511]

5.4.1. No protection: Public domain status of AI-generated works

As our earlier review of requirements for protection has suggested, the construction of the notion of legal personality, authorship and originality under the present copyright regime might exclude AI-generated creativity from copyright protection.[512] Relegating AI-generated creativity to the public domain would therefore be a possible policy option — and that most likely endorsed under the present legal framework.

With this approach, the ownership of copyright depends on the amount of human intervention. Mere data selection and classification by humans is insufficient to meet the 'originality' requirement; instead, actual and substantial human contribution to guide the AI system in creation is necessary for the granting of protection.[513] Only when there is substantial human input, and all creative choices are embedded in the computer code or users' instructions, would copyright be vested with the human author.[514] In this regard, four models of allocating authorship have been identified: (1) sole authorship to the user of the tool, if the designer of the tool does not contribute to the creative work generated; (2) sole authorship to the designers of the tool, if the operator plays no role in the output and the self-generative tool creates output based on the training and creative raw material provided by the designer; (3) joint authorship to the user and the programmer, when the outputs reflect the creative contributions of both designer and user; (4) authorless works — neither designer nor user contribute sufficient expression to form an original work of authorship.[515] In any event, if the creative output results both from human and machine choices, materials resulting from machine-made choices must be filtered out as is customary with public domain materials.[516] Only independently copyrightable human contributions will be protectable.

---

[511] Ginsburg J. and Budiardjo L.A., *op.cit.*, p. 448. See also Ginsburg J., *op.cit.*, pp. 131-135.
[512] Aplin T. and Pasqualetto G., *op.cit.*, §5.01-09. Clifford, R.D. (2018), "Creativity Revisited", *IDEA: The IP Law Review 59*, pp. 26-29. Clifford, R.D. (1996), *op.cit.*, 1700-1702. Huson G., "I, Copyright", *Santa Clara High Technology Law Journal* 35, pp. 72-78. Mezei P., *op.cit.* Palace V.M., "What if Artificial Intelligence Wrote This: Artificial Intelligence and Copyright Law", *Florida Law Review* 71(1), pp. 238-241. Saiz Garcia C., "Las obras creadas por sistemas de inteligencia artificial y su protección por el derecho de autor (AI Created Works and Their Protection Under Copyright Law)" InDret 1, pp. 38-39, https://ssrn.com/abstract=3365458. Gervais D.J., *op.cit.*, pp. 1-60. Ramalho A., *op.cit.*, pp. 22-24. Svedman M., *op.cit.*, pp. 1-22;
[513] Selvadurai N. and Matulionyte R., "Reconsidering creativity: copyright protection for works generated using artificial intelligence", *Journal of Intellectual Property Law & Practice* jpaa062, p. 539.
[514] Gervais D.J., *op.cit.*, pp. 51-60. Selvadurai N. and Matulionyte R., *op.cit.*, p. 538.
[515] Ginsburg J. and Budiardjo L.A., "Authors and Machines", *Berkeley Technology Law Journal* 34(2), pp. 404-445.
[516] Gervais D.J., *op.cit.*, p. 54.

An additional test has been proposed which would determine whether a work deserves protection depending on how much of the developers' meaning transmits to the final work and whether the user/developer could predict the output.[517] Such a test would involve implementing traditional copyright requirements granting protection only if the work is the product of the human authors' imagination and a conception of it. If this test is implemented, copyright law should prompt a shifting of the traditional burden of proof, so that the claimant must prove human authorship of certain AI-generated outputs by establishing that the output foreseeably includes a meaning or message that the author wishes to convey to his or her audience.[518]

5.4.2. Authorship and legal fictions: Should a human be the author?

In order to avoid AI-generated creativity falling in the public domain, and to grant necessary incentives to human agents involved with the AI creative process, proposals have been made — and legislation has been enacted — to set up a legal fiction, so that authorship of AI-generated works is conferred to the agents expending skills, labour and efforts to create, train or instruct the AI in the first place. This approach has also been termed the "fictional human author theory".[519]

This policy approach emerged quite early, when the creative potential and mechanics of machine learning and AI were wholly unknown. In the United Kingdom, the copyright protection of a computer-generated sequence for a lottery was discussed as early as 1985 in *Express Newspapers v. Liverpool Daily Post*. Justice Whitford assigned copyright protection for the automated output to the plaintiff and refused the notion that copyright in the work could be vested in the computer. The computer, he held, is a mere tool for creation; arguing that the computer is the author is similar to suggesting that in a written work, "it is the pen that is the author of the work rather than the person who drives the pen".[520] This position has been recently powerfully summarised by Dan Burk in terms of proximate cause, intent and volition, so that "[i]f there is an author, it is one or more of the humans who are sufficiently causally proximate to the production of the output. […] But the author is never the machine".[521]

---

[517] Boyden B., "Emergent Works", Columbia Journal of Law and the Arts 39, pp. 377-394.
[518] Boyden B., *op.cit.*, 393-394.
[519] Wu A.J., "From Video Games to Artificial Intelligence: Assigning Copyright Ownership to Works Generated by Increasingly Sophisticated Computer Programs", AIPLA Quarterly Journal, pp. 173-174.
[520] *Express Newspapers Plc v. Liverpool Daily Post & Echo Plc* [1985] 1 WLR 1089 (UK).
[521] Burk D.L., "Thirty-Six Views of Copyright Authorship, By Jackson Pollock", Houston Law Review 58, pp 1-38. See also Hedrick S.F., "I 'Think', Therefore I Create: Claiming Copyright in the Outputs of Algorithms", *NYU Journal of Intellectual Property & Entertainment Law* 8(2), pp. 324-375, and Grimmelmann J., "There is No Such Thing as a Computer-Authored Work And It is a Good Thing, Too", *Columbia Journal of Law and the Arts* 39, pp. 403-416.

The United Kingdom was the first jurisdiction to provide specific protection to computer-generated creativity.[522] Section 9(3) of the Copyright Designs and Patents Act 1988 (CDPA)[523] clarified that for computer-generated works, the author is the person who undertakes the arrangements necessary for the creation of the work. In addition, Section 178 provides that "computer-generated, in relation to a work, means that the work is generated by computer in circumstances such that there is no human author of the work". Under this regime, the term of protection for computer-generated works would be 50 years from when the work was made. Shortly thereafter, other common law countries, including Hong Kong, India, Ireland, Singapore, and New Zealand, enacted similar legal arrangements.[524]

There are, however, two issues that challenge this arrangement. The first is fundamental and systemic. Would this approach be sustainable under a legal framework that builds upon the notion of originality as an expression of the author's personality as adopted by EU law as well as most international jurisdictions? Of course, programming, training and imparting instructions would be unlikely to fulfil the requirement of an original contribution from the human counterparts, as ultimately any 'expression' would be the result of the AI creative process. As long as the present subjective standard for originality is in place, any fictional human author theory would crumble in the face of the lack of originality of AI-generated creativity. The work itself, whose fictional authorship is attributed to a human agent, would actually remain unoriginal, thus unprotectable. It is worth noting that "fictional human author theory" and 'necessary arrangements' approaches have been enacted in the UK and other common law countries when 'sweat of the brow' or 'skill and labour' originality standards were still dominant in those jurisdictions. Since then, as previously mentioned, changes have been occurring, with the personality standard for originality fully or partially replacing any alternative approach,[525] challenging the systemic compliance policy approach of Section 9(3) CDPA.

The second issue is of more practical nature. This approach makes it tricky to allocate who is the person in charge of the necessary arrangements.[526] Does the AI-generated work belong to the person who built the system, such as the software developer, the manufacturer, the person who trained it, or the person who fed it specific

---

[522] Guadamuz A., "Do Androids Dream of Electric Copyright? Comparative Analysis of Originality in Artificial Intelligence Generated Works", *Intellectual Property Quarterly*, pp. 169-186.

[523] Copyright, Designs and Patents Act 1988. Available at https://www.legislation.gov.uk/ukpga/1988/48/contents (UK)

[524] Copyright Ordinance cap 528, Section 11(3), https://www.elegislation.gov.hk/hk/cap528 (Hong Kong); Copyright Act 1957, Section 2(d)(vi), https://www.wipo.int/edocs/lexdocs/laws/en/in/in122en.pdf (India); Copyright and Related Rights Act 2000, Section 21(f), http://www.irishstatutebook.ie/eli/2000/act/28/enacted/en/html (Ireland); Copyright Act 1987 chapter 63, Section 7A, https://sso.agc.gov.sg/Act/CA1987#pr27- (Singapore); Copyright Act 1994, Section 5(2), http://www.legislation.govt.nz/act/public/1994/0143/latest/DLM345634.html (New Zealand).

[525] *Temple Island Collections v New English Teas* (No. 2) [2012] EWPCC 1. Guadamuz A., *op.cit.*, p. 178-180; Rahmatian A., *op.cit.*, pp. 4-34.

[526] Dorotheu E., "Reap the benefits and avoid the legal uncertainty: who owns the creations of artificial intelligence?", *Computer and Telecommunications Law Review* 21, p. 85.

inputs like a user?[527] In Guadamuz's view, however, the system's ambiguity should actually be seen as a positive feature that deflects the user/programmer dichotomy question and renders a case-by-case analysis basis.[528] In any event, the question "who is in charge of the necessary arrangements?" has been answered in multiple ways both by case law and scholarship.

5.4.2.1. Should the programmer be the author?

A first possible answer to the question "should the programmer be the author?" was provided in *Nova Productions v Mazooma Games*, a case in which Section 9(3) of the CDPA was applied. The case concerned copyright for frame images generated by a computer program using bitmap files and displayed on the screen when the users played a snooker video-game. The court refused to grant authorship to the user as their input was not artistic in nature:

> *The appearance of any particular screen depends to some extent on the way the game is being played. For example, when the rotary knob is turned the cue rotates around the cue ball. Similarly, the power of the shot is affected by the precise moment the player chooses to press the play button. The player is not, however, an author of any of the artistic works created in the successive frame images. His input is not artistic in nature and he has contributed no skill or labour of an artistic kind. Nor has he undertaken any of the arrangements necessary for the creation of the frame images. All he has done is to play the game.*[529]

Instead, the court found the programmer to be the sole author and the person who made the necessary arrangements, noting that "[t]he arrangements necessary for the creation of the work were undertaken by [the plaintiff] because he devised the appearance of the various elements of the game and the rules and logic by which each frame is generated and he wrote the relevant computer program.[530]" In truth, the Nova Productions outcome may have been a direct consequence of the rudimental technology at stake, and, as Guadamuz argued, a different allocation of authorship might result depending on the specifics of the case and technology under review.[531] Nonetheless, the approach has been proposed in other jurisdictions, such as the United States, requiring the courts to bend the language of the Copyright Act, so that, where neither the programmer nor the user meet the requirements of authorship of a copyrightable work, the court should assign the copyright to whoever owns the copyright to the computer program.[532]

---

[527] Bonadio E. and McDonagh L., *op.cit.*, pp. 117-119. Kasap A., *op.cit.*, pp. 364-376.
[528] Guadamuz A., op.cit. p. 177.
[529] *Nova Productions Ltd v. Mazooma Games Ltd & Ors Rev 1* [2006] EWHC 24 (Ch) (20 January 2006) (UK). See also Farr E.H., *op.cit.*, 75-78.
[530] *Nova Productions Ltd v. Mazooma Games Ltd & Ors Rev 1* [2006] EWHC 24 (Ch) (20 January 2006) (UK). See also Farr E.H., *op.cit.*, 73-74.
[531] Guadamuz A., *op.cit.*, p. 177.
[532] Wu A.J., *op.cit.*, pp. 173-174.

Vesting authorship in the programmer of AI-generating content prompts a fundamental critique. The allocation of authorship to the software developer — or to owners of AI technologies such as companies and investors —may constitute a blatant misperception[533]. In fact, at least in state-of-the-art neural network (GAN and CAN) based creativity, there appears to be no direct causal connection between the software developers and the final AI-generated output, as the expression embedded in that output is the result of the training of the machine and the instructions given to create that specific output. In light of this, first, from a systemic perspective, it could be argued that this arrangement opens a fundamental inconsistency. Actually — as also the Beijing Internet Court highlighted in a case mentioned earlier — the software developer has been already rewarded with exclusive rights over the software that generates works.[534] In addition, given that precisely this legal fiction is meant to provide incentives to create AI-generated works, where its public domain status would presumptively fail to do so, a sound economic analysis would probably discourage a policy option that rewards the same market player twice. Finally, from a more practical perspective, this policy solution would potentially entitle coders to aggressive copyright protection for innumerable pieces of creativity,[535] which would also lower any incentive for the original programmer to create more software.[536]

US case law appears also to endorse the conclusion that software and output are two very distinct entities, and ownership over the former does not imply rights over the latter. In two cases, the courts have ruled that the output created using an infringing copy of software or of a programme is not considered an infringing derivative work. In *Design Data v. Unigate Enterprises,* affirming the district court decision, the court of appeals ruled that the plaintiffs' copyright over the computer programme did not extend to the output created by the programme (drawings and data for steel buildings).[537] In *Rearden v. Walt Disney Co*., movies created using an (infringing) copy of the plaintiff's software were not considered derivative works of the software because although the software did a significant amount of work, the lion's share of creative expression in the movie was attributable to the defendants.[538]

---

[533] Abbott R., "Artificial Intelligence, Big Data and Intellectual Property: protecting computer generated works in the United Kingdom" in Aplin T. (ed.), *Research Handbook on intellectual property and digital technologies*, Edward Elgar, England, pp. 323-324. Svedman M., *op.cit*., 10-11. Bridy A. (2016), "The Evolution of Authorship: Work made by Code", Columbia Journal of Law and the Arts 39, pp. 400-401. Bridy A. (2012), *op.cit*., 24-25, para 62. Samuelson P. (1986), *op.cit*., pp. 1207-1212.

[534] *Beijing Feilin Law Firm v Baidu Corporation* (26 April 2019) Beijing Internet Court, (2018) Beijing 0491 Minchu No. 239.He K. (2020b), *op.cit.* See also Bonadio E. and McDonagh L., *op.cit*., p. 117. Chen M., *op.cit*., pp. 14-18. Samuelson P. (1986), *op.cit*., pp. 1207-1212.

[535] Svedman M., op.cit., p. 14.

[536] Huson G., *op.cit*., p. 74.

[537] *Design Data Corp. v. Unigate Enterprise*, No. 14-16701 (9th Cir. 2017) (USA).

[538] *Rearden v. Walt Disney Co*, No. 17-cv-04006-JST (N.D. Cal. 2018).

5.4.2.2. Should the user be the author?

Allocating rights in AI-generated output to the user of the generator programme might be a more sound solution.[539] The recent Beijing decision earlier described stressed such a conclusion.[540] Samuelson has argued that the user is the reason the AI-generated work comes into being, thus "[i]t is not unfair in these circumstances to give some rights to a person who uses the work for its intended purpose of creating additional works".[541] This solution would not be novel under copyright standards. For example, in the United States, copyright —and authorship — are given to users for being the instrument of fixation[542] as in the case of a person who tape-records a jazz performance.[543] In this scenario, the user would be the author of the sound recording, rather than the jazz performers. Similarly, the user could be construed as the author of the fixation of the AI-generated work. Of course, a specific provision, such as 9(3) CDPA, should be introduced to that end. Most likely, in some exceptional cases, such as when the user does not have any control over the software other than running it, awarding copyright to the user would be a sub-optimal policy choice at odds with copyright incentive theory.[544] In this case, joint authorship between users and programmers could be a possible solution,[545] depending on the legal scheme for joint authorship made available by different jurisdictions.

In a "Draft Report on intellectual property rights for the development of artificial intelligence technologies", the European Parliament seemingly endorsed the same view and proposed to entrust AI users with copyright over AI-generated works. The draft report "[t]akes the view that consideration must be given to protecting technical and artistic creations generated by AI, in order to encourage this form of creation; considers that certain works generated by AI can be regarded as equivalent to intellectual works and could therefore be protected by copyright".[546] Therefore, the draft report "recommends that ownership of rights be assigned to the person who prepares and publishes a work lawfully, provided that the technology designer has not expressly reserved the right to use the work in that way".[547] Apparently, this proposal implies ownership of rights to be also assigned to the users, as the draft report makes reference to copyright protection, rather than a sui generis protection, stating that "it is proposed that an assessment should

[539] CONTU, *op.cit.*, p. 45. See also Ralston W.T., *op.cit.*, pp. 303-304.
[540] *Beijing Feilin Law Firm v Baidu Corporation* (26 April 2019) Beijing Internet Court, (2018) Beijing 0491 Minchu No. 239. He K. (2020b), *op.cit.*
[541] Samuelson P. (2020), "AI Authorship?" *Communications of the ACM* 63(7), pp. 20-22. Samuelson P. (1986), "Allocating Ownership Rights in Computer-Generated Works", *University of Pittsburg Law Review* 47(4), pp. 1200-1204.
[542] 17 U.S.C. § 114.
[543] Samuelson P. (1986), *op.cit.*, pp. 1200-1204. However, most countries favor neighbouring rights protection for sound recordings, rather than copyright as in the Unites States.
[544] Ralston W.T., *op.cit.*, pp. 304-305.
[545] E.g. Bonadio E. and McDonagh L., *op.cit.*, 117-118.
[546] European Parliament (2020), *op.cit.*
[547] European Parliament (2020), *op.cit.*

be undertaken of the advisability of granting copyright to such a 'creative work' to the natural person who prepares and publishes it lawfully [...]".[548]

5.4.2.3. Should the employer be the author?

Scholars have proposed the work-made-for-hire (WMFH) doctrine[549] as a legal framework to ensure the protectability, ownership and accountability of works generated by AI systems.[550] This model would be based on the fiction that the AI system is a creative employee or independent contractor of the users — humans or legal entities — that use AI systems and enjoy its benefits.[551] As Samuelson argues, "one who buys or licenses a generator program has in some sense 'employed' the computer and its programs for his creative endeavours, similar considerations to those that underlie the work made for hire rule support allocation of rights in computer-generated works to users".[552] Thus, ownership of the copyright and liability for any infringements arising from the work would be imposed on the human or legal entity considered the employer or commissioner of the AI system that creates the work.

Adoption of this regulatory framework would trigger a few critiques. First, an argument is often made that employers are treated as authors of work-for-hire works despite having no role in the output, thus similar arrangements could be devised for AI-generated creativity.[553] This position, however, seems to miss the fact that as part of the WMFH legal fiction, the underlying work has been created by a human author and fulfils the originality standard under the present legal framework. This would not be the case with AI-generated creativity. Second, this arrangement would face challenges on the grounds that it would be a misapplication of the WMFH doctrine, as it is difficult to define a legal, contractual employment or agency relationship between a human and a machine.[554] The application of the WMFH doctrine would be especially problematic in jurisdictions such as France where transfers of ownership to employers or commissioning parties must be explicitly provided for in the employment or commissioning agreement

[548] European Parliament (2020), *op.cit.*

[549] Under US copyright law, work made for hire constitutes an exception to the rule that only the author (or those deriving rights from the author) can claim copyright over a protectable work. Under this doctrine, the employer is considered the author of a work even if an employee created the work. It is defined under Section 101 of the Copyright Act. A work created by an employee during their course of employment or specially ordered or commissioned for use by the employer is a work made for hire if the parties so agree in a signed agreement (United States Copyright Office, "Works Made for Hire". Available at copyright.gov/circs/circ09.pdf).

[550] Bridy A. (2016), *op.cit.*, pp. 400-401. Bridy A. (2012), *op.cit.*, 26, para 66. Hristov K., *op.cit.*, 431-454. Kaminski M., *op.cit.* See also Pearlman R., "Recognizing Artificial Intelligence (AI) as Authors and Inventors under U.S. Intellectual Property Law", *Richmond Journal of Law & Technology* 24, pp. 1-38. Yanisky-Ravid S. (2017), "Generating Rembrandt: Artificial Intelligence, Copyright, and Accountability in the 3A Era, The Human-Like Authors are Already Here: A New Model", Michigan State Law Review, pp. 659-726.

[551] Yanisky-Ravid S., *op.cit.*, pp. 659-726

[552] Samuelson P. (2020), *op.cit.*, pp. 20-22; Samuelson P. (1986), *op.cit.*, pp. 1200-1204.

[553] Brown N., *op.cit.*, p.39. Kaminski M., *op.cit.*, 602.

[554] Bonadio E. and McDonagh L., *op.cit.*, pp. 114-115. Bridy A. (2012), *op.cit.*, p. 27, para 68. Butler T., *op.cit.*, pp. 739-742. Huson G., *op.cit.*, pp. 73-75. Ramalho A., *op.cit.*, pp. 18-19.

and will not be implied. It seems obvious that in order for the WMFH doctrine to apply to AI-generated works some substantial statutory and jurisprudential reconstruction of the notion of 'employer' and 'employee' must in any event first occur.[555]

5.4.3. Should a robot be the author?

One policy option — although quite residual according to the scholarship[556] — would be to construe the AI as the author. A fiction would have to be established in the law to provide AI with legal personality, so that it can author a work and own a copyright,[557] or at least the law would have to be amended to reflect the fact that a computer can be an author in a joint work with a person[558]. According to Pearlman, the law should recognise sufficiently creative AIs as authors when the AI creation is original and developed independently from human instructions, so that the AI is the cause of creativity, not a mere machine working under the instructions of a human author.[559] Once the AI is declared the author, rights would be immediately assigned to a natural or legal person, such as the creator/programmer of the AI, the user of the AI, or as a joint work. The law should identify the person entitled to receive the transfer and exercise the rights.

Still, also in this scenario, meeting the requirement of originality could be an insurmountable burden for a machine. The notion of originality would most likely have to be tweaked to include works originating from a machine, as mentioned before.[560] However, if legal personhood is granted to a machine, an argument could also be made that, once recognised as a (legal) person, the machine would be capable of original creativity according to the personality approach that governs copyright law and construes originality as an expression of the personality of the author. In any event, allowing AI as author would require substantial amendments to the legal framework. As noted, given the early state of technological development, amending the law before truly intelligent machines have even materialised — and whose materialisation and evolution remains as of today just hypothetical speculation — would be a sub-optimal policy option.[561]

---

[555] Hristov K., *op.cit.*, pp. 445-447.
[556] E.g. Bonadio E. and McDonagh L., *op.cit.*, p. 116. Farr E.H., *op.cit.*, p. 79. Ralston W.T., *op.cit.*, pp. 302-303. Samuelson P. (1986), *op.cit.*, pp. 1199-1200.
[557] Ihalainen J., "Computer creativity: artificial intelligence and copyright", *Journal of Intellectual Property Law & Practice* 13(9), pp. 724–728.
[558] Abbott R., *op.cit.*
[559] Pearlman R., *op.cit.*, pp. 1-38).
[560] Yanisky-Ravid S. and Velez-Hernandez L.A., *op.cit.*, pp. 40-48.
[561] Huson G., *op.cit.*, pp. 77-78. Liebesman Y., "The Wisdom of Legislating for Anticipated Technological Advancements", J. Marshall Rev. Intell. Prop. L. 10, p. 172.

5.4.4. Sui generis protection for AI-generated creativity

Proposals have also suggested the creation of a related sui generis right (where no authorship or originality would be a necessary requirement) which might protect the investment made in developing and training AI-generating creativity.[562] For example, McCutcheon has suggested a sui generis rights regime for AI-generated creativity similar to database rights, thereby protecting the investment in the creation but not requiring an author, nor authorship, nor originality.[563] While denying protectability under the traditional copyright scheme, the Australian Copyright Law Review Committee noted that, if computer-generated creativity needs protection, this should be "more akin to that extended to neighbouring rights [...] the protection extended to performers, producers of phonograms and broadcasting organizations".[564]

Japan has been considering a novel sui generis regime for non-human-created intellectual property based on a trademark-like approach with an emphasis on protection against unfair competition.[565] This approach seeks to limit the protection of AI works by allowing flexibility in levels of protection based on the popularity of the AI-generated works as a proxy for goodwill.[566] This would leave out obscure works created for the sole aim of copyright protection. The proposal would allocate ownership of the work to the individual or company that had created the AI.[567] In the same vein, with the goal of limiting overbroad protection of algorithmic creativity, some authors propose a thin scope of the sui generis right, coupled with strong fair use safeguards, with a short duration of around three years or so.[568]

5.4.5. Providing rights to publishers and disseminators

Finally, further proposals suggest providing rights to publishers and disseminators of AI-generated works. On the one side, the regime for anonymous/pseudonymous works could be applied to AI-generated works. According to several national regimes, such as Spain, France, Italy and Sweden[569], the person who publishes the work will exercise the rights.

---

[562] Ciani J., "Learning from Monkeys: Authorship Issues Arising From AI Technology" in Moura Oliveira P., Novais P., Reis L. (eds), *Progress in Artificial Intelligence EPIA 2019 Lecture Notes in Computer Science* vol. 11804, Springer, Cham.
[563] McCutcheon J., *op.cit.*, pp. 965-966.
[564] Ricketson S. (2012), "The Need for Human Authorship - Australian Developments: Telstra Corp Ltd v Phone Directories Co Pty Ltd (Case Comme nt)", *E.I.P.R.* 34(1), 54.
[565] Intellectual Property Strategic Program 2016, pp. 10-11, http://www.kantei.go.jp/jp/singi/titeki2/kettei/chizaikeikaku20160509_e.pdf.
[566] Intellectual Property Strategic Program 2016, *op.cit.*, p. 11.
[567] Intellectual Property Strategic Program 2016, *op.cit.*
[568] Bonadio E. and McDonagh L., *op.cit.*, 136-137.
[569] The Intellectual Property Act 1996, Article 6. Available at https://www.wipo.int/edocs/lexdocs/laws/en/es/es177en.pdf (Spain); Code de la propriété intellectuelle [Intellectual Property Code] 1912, L113-6. Available at

On the other side, an additional policy option could provide to the disseminator of AI-generated creativity a right similar to the EU's publisher rights relating to previously unpublished works, as in Art. 4 of Directive 2006/116/EC.[570] Under this scheme, the protection covers the first lawful publication/communication of previously unpublished public domain works. Similarly, AI-generated works would be in the public domain, therefore the 'disseminator' scheme would only reward someone for the dissemination of AI-generated creation. The duration of the right could be limited to 25 years for example, as in the case of Art. 4 of Directive 2006/116/EC.[571]

5.5. Conclusions

Anthropocentrism and the personality right approach strongly influence the present copyright legal framework, in particular the notion of originality, from the Berne Convention to major jurisdictions. Thus, AI-generated creativity falls short of all fundamental requirements for granting copyright protection, including legal personality, authorship and originality.

Utilitarian/incentive approaches push for the adoption of legal fictions that do not satisfactorily address and overcome internal systemic inconsistencies. Still, even if the law fictionally claims that the work is human-made rather than AI-made, the work itself remains unoriginal as machines will be inherently incapable of originality under a personality standard. Only a fundamental overhaul of the copyright system, sidelining the current anthropocentric approach, can provide full copyright protection to AI-generated creativity proper, when no human intervention can be construed as an original expression. This would be ill-considered, especially given the primitive stage of technological development in the field. Residual sui generis approaches are also available and, most likely, a preferable option, if policy-makers choose to provide monopolistic incentive to AI-generated creativity. In that case, the incentive should fall upon the users if they contributed any meaningful labor and effort to the AI-generated output, because otherwise programmers, marketers and investors would be double-dipping into earlier rewards linked to the AI-generating content software.

But are incentives really needed? The need for such incentives should be empirically proved together with the positive externalities that they would bring about for the creative ecosystem as a whole. In fact, there is well-established historical evidence that property rights are not the only incentive to creativity.[572] Miscellaneous research and market evidence show that open and free access to creative works or alternative business

---

https://www.wipo.int/edocs/lexdocs/laws/en/fr/fr467en.pdf (France); Law for the Protection of Copyright and Neighbouring Rights 1941, Article 9. Available at https://www.wipo.int/edocs/lexdocs/laws/en/it/it211en.pdf (Italy); Act on Copyright in Literary and Artistic Works 1960, Article 7. Available at wipo.int/edocs/lexdocs/laws/en/se/se124en.pdf (Sweden).

[570] Ramalho A., *op.cit.*, pp. 22-24.

[571] Ramalho A., *op.cit.*

[572] E.g. Frosio G., *op.cit.*

models may provide stronger incentive to AI-generated creativity than IP-based protection models,[573] without creating the negative externalities of propertisation and exclusive rights. AI creative capacities might scale up at singularity pace, swamping the cultural marketplace with an unmanageable mesh of rights to clear and copyright trolling escalating towards ultimate computational doom. As noted, the copyright soup is already too thick. Infinite AI monkeys[574] may eat it all up, and then there will be nothing left to use

[573] E.g. Bonadio E. and McDonagh L., *op.cit.*, 122-123. Svedman M., *op.cit*, pp. 13-14.
[574] Borel É., "Mécanique Statistique et Irréversibilité", J. Phys. (Paris) 5(3), pp. 189–196.

Advertising

The worries raised by AI are, to a certain extent, nurtured by science fiction literature and films. And yet, science fiction tends to foresee the future. Think, for example, of Steven Spielberg's "Minority Report"[575]*. There is a scene in which a man walks down an alley inside a shopping centre. His eyes are flashed by a multitude of cameras equipped with eye-recognition software. Immediately, the shop windows start to show on flashy screens advertising specially tailored to him. Science fiction, really? It is actually not so far removed from what we already experience in real life today. In this age of the Internet, connected TV sets and 'second screens', the possibilities of obtaining the personal data of media users in both legal and illegal ways have multiplied exponentially. Such data are a very important commodity for advertisers, which can be used to provide individually targeted ads on online services and on all sorts of connected devices. Furthermore, personal data obtained via search engines, social media and connected devices can be used as a means to provide a better experience for the user of an online service. In her contribution to this publication,* ***Justina Raižytė*** *reminds that "building a sustainable AI framework for advertising and the use of data and technology for good is in the interest of the advertising community itself", since it is the only way to earn consumer trust, which "is, and always will be, a true gold standard in advertising".*

[575] https://www.imdb.com/title/tt0181689/.

6. AI in advertising: entering Deadwood or using data for good?

Justina Raižytė, European Advertising Standards Alliance[576]

6.1. Introduction

Artificial intelligence (AI) jumped from futuristic movie scripts and fictional stories straight into our living rooms, cars, pockets and shopping bags a while ago. It didn't take long before advertising, always fuelled with creativity and hungry for innovation, saw the potential of automated technologies and adopted AI as its right-hand man, who can lead consumers through their journeys to becoming customers and loyal clients.

Machine learning technologies allow advertisers to more quickly analyse data on people's interests, preferences, locations and demographics, and even to predict their desires, create tailored targeting plans and deliver more relevant ads. In the process, the cost of advertising is reduced, while responsiveness and effectiveness rates increase. Consumers are also supposed to become more satisfied and even be entertained with personalised ads, which provide fewer annoying or irritating commercial messages.

However, massive data flows that allow marketers to segment audiences and target consumers leave many people uneasy. Anxiety and the feeling of being constantly tracked while browsing, listened to by electronic devices in pockets, and targeted by unfairly priced offers, raise fears of loss of privacy and autonomous decision-making to algorithms. That's why today's technology-fuelled advertising ecosystem today is frequently painted in colour tones of the classic westerns: with dark shades of lawlessness and tinted principles of ethics.

Which brings us to the gates of Deadwood: one of the hometowns of settlers, who were caught by the "Gold Fever" and rushed across the oceans in pursuit of wealth. The legends of the town have recently been brought back to life on our screens with movie magic and impeccable storytelling that depict the essence of the chaotic spirit of the wild west, thereby offering an interesting allegory for the debates on technology and market developments that we are having today.

[576] Disclaimer: the views expressed in this paper are those of the author and do not necessarily reflect the official policy or position of any other organisation, company, or individual mentioned in the text.

Where do we find AI in advertising nowadays? What are the ethical principles behind harnessing people's personal data and adopting automated technologies for marketing purposes? What are the risks and benefits of using the power of algorithms in creating, delivering and monitoring the ads? These and other questions will be investigated in this paper through the prism of existing market practices and policy frameworks, as well as explored through the prism of targeted qualitative interviews conducted with experts working in the AI and advertising fields (including practitioners from the advertising agencies and tech companies; independent analysts; and experts in data privacy law and advertising self-regulation).

The first part will look at the key applications of AI in advertising and will discuss how various machine learning tools help to better target, optimise and even design commercial communications. The second chapter will present the most prominent concerns and issues, which are linked to machine learning applications in marketing, and will investigate the safeguards which are (or should be) in place to mitigate them. The third part will explore how technologies are being used to harness the power of data for good causes, such as protecting consumers from bad ads, and how AI is potentially going to shape the advertising regulatory field in the future.

The paper will conclude on the relationship between AI and advertising through the lens of the earlier introduced wild west metaphor. The last part will summarise the arguments by discussing similarities and differences between the historic gold rush era and the data rush phenomenon put forward by the author, and will ask a final question: are we about to enter a virtual Deadwood?

6.2. AI in advertising: From tracing online footprints to writing ad scripts

"There is no data like more data" is a well-known quote attributed to Robert Mercer,[577] frequently used to emphasise the essential role of data in developing and adopting AI algorithms in different industries and human activities. AI use in advertising ecosystems is not an exception. From crunching personal data in tailoring the delivery of ads based on user preferences and browsing history in a split second, to training algorithms to recognise patterns in vast volumes of historic and contextual data (and ultimately create new content), AI and machine learning techniques in advertising are applied broadly and for a variety of purposes. This first chapter will investigate the main use of AI in the advertising ecosystem today, and will discuss how such advanced algorithms have been transforming the ad industry.

[577] Lee, K.F., *AI Superpowers: China, Silicon Valley, and the New World Order*, Houghton Mifflin Harcourt, Boston.

6.2.1. Programmatic advertising: The stock market of ads and data

According to the Interactive Advertising Bureau (IAB), most of the growth of AI in marketing today is attribute to programmatic advertising, which is the use of AI and other machine learning technologies to buy and optimise advertising in real-time, with the goal of increasing efficiency and transparency for both the advertiser and the publisher.[578] In programmatic advertising, AI is used in ad planning, placement, automatic adjustment, optimisation and campaign metrics.[579] The key elements of programmatic advertising are outlined in the "programmatic advertising glossary" in the table below.

Table 1. Programmatic advertising glossary

Real-time bidding (RTB)
■ Real-time bidding is a way of buying and selling online ads on a case-by-case basis through real time auctions that occur in the time it takes a webpage to load on a user's browser (i.e. around 100 milliseconds). During this time, the information about the page where an ad will be placed and the user loading it passes through an ad exchange, which auctions it off to the advertiser willing to pay the highest price for it, and the winning bidder's ad is then loaded into the webpage.[580] RTB allows advertisers to target better and quicker, enabling ads which, for example, only show X brand's ads to users who previously visited X brand's website but didn't make a purchase. Advertisers bid based on their interest and how the data about the user measures up against their targeting parameters: the higher the 'match', the higher the price.[581]
Ad exchange
■ An ad exchange is essentially a stock market for programmatic advertising, and it is where ad inventory (i.e. ad space) gets auctioned. It's a virtual platform where publishers meet advertisers, are matched and agree on a price to display their ads, which all happens in milliseconds, due to automated technology. Unlike ad networks, which used to focus on pre-packaged ad inventory, an ad exchange focuses on audience metrics.[582] It sits in the middle of the programmatic ecosystem, plugged into respective platforms on both the advertiser's and the publisher's side.
Demand-side platform (DSP)
■ A demand-side platform is a tool or software programme that allows advertisers to

[578] Rask, O., *What is Programmatic Advertising? The Ultimate 2020 Guide*, Match2One, https://www.match2one.com/blog/what-is-programmatic-advertising/.
[579] IAB (2019), *Artificial Intelligence in Marketing Report*, IAB Data Center of Excellence, https://www.iab.com/insights/iab-artificial-intelligence-in-marketing/.
[580] Marshall, J., *WTF is real-time bidding?*, Digiday, https://digiday.com/media/what-is-real-time-bidding/.
[581] SMAATO, *Real-Time Bidding (RTB): The Complete Guide*, SMAATO, https://www.smaato.com/real-time-bidding/.
[582] IAB Europe (2017), *The advent of RTB*, IAB Europe, https://iabeurope.eu/blog/laypersons-programmatic/.

	buy ad placements automatically. Advertisers sign up with a DSP, which is in turn connected to an ad exchange.[583] When a user triggers a website that is connected to the ad exchange, an auction signal is sent to the exchange. The exchange then asks the DSP if the advertiser has any ads that might fit the placement and, if so, the bid for an ad space in sent back to the auction in real time. A winning bidder gets to show their ad to the website visitor.[584]
	Supply-side platform (SSP)
■	On the other side, publishers use a supply-side platform to manage their ad space. The SSP connects to an ad exchange and makes clear what kind of inventory is available. Through RTB this inventory is automatically auctioned off to the highest bidder. While a DSP's purpose is to buy programmatic ad space as cheaply as possible from publishers with desirable inventory, the SSP has the opposite function: selling ad space for the highest possible price, connecting to several different ad exchanges in order to maximise the publisher's exposure to potential buyers.[585]

Although programmatic advertising is mostly referred to in the context of online digital channels, including display, mobile, video and social media, programmatic ad purchasing has also found its way into traditional media, including TV[586], and even to digital out-of-home billboards and signage,[587] where powerful algorithms paired with mobile location data and visual sensors enable ad placement on digital screens at bus stops, which can now focus on specific targets - young commuters in the morning rush hour, for example.

6.2.1.1. The AI role in programmatic advertising

All key elements involved in the programmatic advertising value chain highlight the crucial role of machine learning applications in this process, which allow ad selection and placement decisions to happen in the blink of an eye, rendering AI inseparable from the notion of programmatic advertising itself.

These sophisticated, automated technologies permit all market actors to enjoy numerous speed-, efficiency- and predictive analysis-based advantages. Overall, these

[583] Rask, O., *op.cit.*

[584] Wang, J., Zhang W. & Yuan, S., *Display Advertising with Real-Time Bidding (RTB) and Behavioural Targeting*, ArXiv, https://arxiv.org/abs/1610.03013.

[585] Rask, O., *op.cit.*

[586] Although advanced, beyond broad demographic-based audience measurement issues are still named among the key challenges in transforming traditional linear TV advertising buying into programmatic, some station groups are nevertheless beginning to embrace impression based TV ad buying, with some companies working to standardize advanced TV audience segments and opening up more addressable broadcast TV inventory. See Blustein, A., *The programmatic TV dream is edging closer to reality*, The Drum, Carnyx Group Ltd, https://www.thedrum.com/news/2020/02/19/the-programmatic-tv-dream-edging-closer-reality.

[587] Côté, R., *What is programmatic DOOH?*, Broadsign, 2020. https://broadsign.com/blog/what-is-programmatic-digital-out-of-home/

broad benefits are summed up in the following key objectives of AI in programmatic advertising today:[588]

- Personalisation of ads, where AI and machine learning can take real-time behavioural data from consumers and serve highly tailored and relevant ads, based on attributes such as age, gender, location and millions of other data points. The process is powered by strong predictive algorithms helping to determine which consumers are likely to engage with advertisers' commercials.
- Audience targeting, where leveraging AI, marketers can scan through content on the Internet and determine which ads are best suited for particular audiences or channels. Through the use of image recognition, these systems also help place ads correlated with images that can be found on the page of an article or website.[589] Moreover, when set up correctly, AI is able to continuously evolve the audience based on actual performance, and expand it to other segments that may share the same purchasing behaviour.[590]
- Performance optimisation, where machine learning algorithms can automatically analyse how ads are performing across specific platforms and offer recommendations. These AI systems are also able to track the metrics not only of the advertiser's campaigns but also those of the advertiser's competitors. They have built-in 'situational awareness', which allows machine learning algorithms to adjust quickly, shifting ad spend to alternate channels and changing advertising messages to reflect market patterns and consumer behaviours.[591]
- Media mix modelling, where AI is used to identify consumers most receptive to their campaigns on different media channels, thus increasing digital advertising return on investment (ROI). AI can continuously issue recommendations on how to refine the media mix; brands and agencies can therefore completely automate their marketing mix allocation — saving valuable time and money.

To summarise, AI and machine learning allow advertising ecosystem players to analyse huge volumes of data in real-time, tailoring messages through AI-enabled hyper-personalisation, and to find the best times and channels to communicate. Therefore, it comes with no surprise that over 80% of surveyed advertising executives and early adopters of AI, reported positive ROI for their AI initiatives and strong intentions to increase such investment in the future.[592]

---

[588] IAB (2019), *op.cit.*
[589] Schmelzer, R., *AI Makes A Splash In Advertising*, Forbes, https://www.forbes.com/sites/cognitiveworld/2020/06/18/ai-makes-a-splash-in-advertising/#24c0287c7682.
[590] Rowan, M., *The Impact of Artificial Intelligence in Advertising*, AW360, https://www.advertisingweek360.com/the-impact-of-artificial-intelligence-in-advertising/.
[591] Schmelzer, R., *op.cit.*
[592] Deloitte, *State of AI in the Enterprise*, 2nd edition, the Deloitte AI Institute, Deloitte, https://www2.deloitte.com/us/en/insights/focus/cognitive-technologies/state-of-ai-and-intelligent-automation-in-business-survey.html.

6.2.2. Algorithmic creativity: AI dipped in the ink of imagination

"Being creative can be quite similar to having the sword of Damocles hanging above your head" - many ad script creators struggling with approaching deadlines, too many ideas and too little time, would probably agree with these words of apparent hyperbolised wisdom. Except, the phrase was not uttered by an ad executive, but was in fact produced by InspiroBot – an AI software programme "dedicated to generating unlimited amounts of unique inspirational quotes", which has been trained to create random phrases based on countless inspiring lines found on the Internet.

Although the 'sword of Damocles' example is a fun one created by AI in the preparation of this paper, the actual application of AI in the creation of ads is much more advanced. AI-powered systems can already partially or fully design ads, leveraging natural language processing (NLP) and natural language generation (NLG), two AI-powered technologies used to write ad copy.[593] Moreover, computer vision, paired with image recognition technologies, promises a future where users can 'shazam'[594] any video content or picture and instantly match it with existing promotions.

Some tech-powered companies already provide AI-based solutions to generate images for marketing purposes. Usually supported by machine learning technologies falling into the category of Generative Adversarial Networks (GANs), such companies are able to generate 'artificial models' increasingly difficult to distinguish from real people,[595] and can create faces for a marketer's advertising campaigns, easily adaptable to fit different audiences and demographics.

Moreover, although mouse clicks are the essential driving force for digital marketing, AI is possibly bringing in a new advertising trend: the era of Voice.[596] According to 2019 data, 89% of surveyed marketing professionals believed that voice assistants would be a significant marketing channel over the coming three to five years, with over one third of respondents calling them an "extremely important channel".[597] Alexa, Siri, Google Home and other devices and voice assistants popularised interactive voice interfaces and paved the way for conversational voice advertising. Powered by AI and

---

[593] Kaput, M., *AI for Advertising: Everything You Need to Know*, Marketing AI Institute, https://www.marketingaiinstitute.com/blog/ai-in-advertising.

[594] Based on Shazam Entertainment Ltd - a mobile application which can identify music, movies, advertising, and television shows, based on a short audio sample played; other types of image recognition-based apps have been prototyped in the industry, such as Smartify (available at https://smartify.org/about-us) – Shazam for art and museums.

[595] Gonfalonieri, A., *Integration of Generative Adversarial Networks in Business Models,* A Medium, Towards Data Science, https://towardsdatascience.com/integration-of-generative-adversarial-networks-in-business-models-47e60263aec4.

[596] Tushinskiy, S., *Voice is the New Click: Why Voice Commands Will Replace the Click as the Standard Measure for Brands*, Medium, Instreamatic, https://medium.com/@instreamatic/voice-is-the-new-click-why-voice-commands-will-replace-the-click-as-the-standard-measure-for-61225e4e7caa.

[597] Kinsella, B. and Mutchler, A. (2019), *The State of Voice Assistants as a Marketing Channel Report*, Voicebot.ai, https://voicebot.ai/the-state-of-voice-assistants-as-a-marketing-channel-report/.

harnessing speech recognition and natural language understanding, they offer new ad formats for marketers, publishers, and consumers.

According to Charlie Cadbury, the CEO of Say It Now[598], the five most exciting trends in voice include: various service providers empowering voice assistants (e.g. managing food delivery); improved discovery of third-party skills and name-free skill invocation in voice assistants; voice commerce; digital interactive audio engagement; and location aware voice services, delivered in cars or ear buds. Asked for a "moon-shot" idea for voice advertising application, Cadbury noted that the "clue is in the 'voice assistant'", where he wants to see consumers treating their digital assistants like a real aid in their life, trusting their advice and meaningfully delegating tasks.

6.2.3. From creative games to gains

Some of the world's top advertisers have already invested in AI-enhanced creative processes, delivering noteworthy results and ad campaigns which have captured consumers' attention, and made media headlines. The campaigns and brand-agency partnerships presented in table 2**Error! Reference source not found.** provide a selection of cases where AI tools were used for creative marketing purposes and resulted in the production of advertising content at least partially designed by automated technologies. Kerry Richardson, one of the co-founders of Tiny Giant[599], believes that consumers will soon see more creative applications for AI, and more cases of humans and creatives working together to create campaigns, where agency creatives will think of an idea or concept and then use AI tech to help execute it.

Table 2. Advertising and marketing campaigns enabled by creative AI technologies

■ Lexus in partnership with IBM's Watson, The&Partnership London and Visual Voice [600]
■ This particular Lexus commercial is the first TV ad entirely scripted by AI. Using a number of data points as input, including 15 years' worth of Cannes Lion award-winning ads and 10 years of the best ads in the 'luxury' sector, as well as points relating to 'intuition' and how people make decisions, IBM Watson technology identified elements common to award-worthy commercials.[601] An AI engine then formed the script flow and outline. The creators themselves expressed surprise over

[598] Charlie Cadbury from Say It Now (available at https://www.sayitnow.ai/) was interviewed on 19 June, 2020 for the purposes of this paper.
[599] Kerry Richardson from Tiny Giant (available at https://www.tinygiant.io/) was interviewed on 22 June, 2020 for the purposes of this paper.
[600] The ad campaign can be viewed at Variety.com. Available at https://variety.com/2018/digital/news/lexus-ai-scripted-ad-ibm-watson-kevin-macdonald-1203030693/.
[601] Spangler, T., F*irst AI-Scripted Commercial Debuts*, Directed by Kevin Macdonald for Lexus, Variety, https://variety.com/2018/digital/news/lexus-ai-scripted-ad-ibm-watson- kevin-macdonald-1203030693/.

	the fact that instead of the expected "mad and weird" output, they received the narrative with a footnote to every single line with a data point, explaining why the decision had been made.[602]The result was a rather perplexing story of a Lexus engineer putting the finishing touches on a new model, a self-aware car and a televised crash test, but it proved that machine-written creative copy is more than a distant possibility.
■	Malaria No More in partnership with RG/A, Ridley Scott Associates and Synthesia[603]
■	In the "Malaria Must Die" campaign, David Beckham lends his voice to the fight against malaria for non-profit organisation Malaria No More, in its battle to eradicate the mosquito-borne disease. The novel campaign depicts Beckham apparently fluently speaking in nine languages as he invites listeners to get involved.[604] As he shifts between different languages, Beckham's various voices are actually those of malaria survivors whose features have been digitally mapped onto those of the famous sportsman with the help of AI-powered video synthesis technology.[605]
■	Deutsche Bahn in partnership with Ogilvy Germany, Frankfurt, Getty Images and Spirable[606]
■	Deutsche Bahn, the German rail company, launched a campaign encouraging domestic travel using photos of German locations that mirror famous world tourist destinations. The "No Need To Fly" campaign invites Germans to enjoy the benefits of cheaper train travel. It used AI to identify German locations resembling iconic world landmarks. Then, using Facebook data, it targeted travel enthusiasts and local influencers with dynamic video ads including the real-time juxtaposition of travel costs associated with international landmarks and their German counterparts.[607]
■	Cheltenham Science Festival in partnership with Tiny Giant[608]
■	"AIDA: AI Science Festival Curator", an AI-generated festival curator for the Cheltenham Science Festival, took 10 years of festival talks as a dataset, trained it on a recurrent neural network, and generated around 800 new potential talks for the festival.[609] AIDA's suggestions were submitted to a Twitter poll for the world to select a winner. The talk on "Introvert Narwhals" was delivered during the event by

[602] Faull, J., *Lexus reveals ad 'created by AI'. Is it a gimmick? No. Will it win any awards? Probably not,* The Drum, Carnyx Group Ltd, https://www.thedrum.com/news/2018/11/16/lexus-reveals-ad-created-ai-it-gimmick-no-will-it-win-any-awards-probably-not.

[603] The ad campaign can be viewed at TheDrum.com. Available at https://www.thedrum.com/news/2019/04/09/david-beckham-lends-his-voice-malaria-ai-petition..

[604] Butcher, M., The startup behind that deep-fake David Beckham video just raised $3M, TechCrunch, https://techcrunch.com/2019/04/25/.

[605] Glenday, J., *David Beckham lends his voice to Malaria AI petition*, The Drum, Carnyx Group Ltd., https://www.thedrum.com/news/2019/04/09/david-beckham-lends-his-voice-malaria-ai-petition.

[606] The ad campaign can be viewed at wersm.com. Available at https://wersm.com/how-deutsche-bahn-increased-sales-by-24-thanks-to-instagram/.

[607] Desreumaux, G., *How Deutsche Bahn Increased Sales By 24% Thanks To Instagram*, Wersm, https://wersm.com/how-deutsche-bahn-increased-sales-by-24-thanks-to-instagram/.

[608] The ad campaign can be viewed at TinyGiant.io. Available at https://www.tinygiant.io/case-study-one-aida.

[609] Tiny Giant, *AI Festival curator Cheltenham Science Festival*, Tiny Giant, https://www.tinygiant.io/case-study-one-aida.

	AIDA, whose actual audio voice was created using deep learning to turn input text into a nuanced, human-sounding voice. AIDA proved a great success, and later appeared on BBC Radio, and was also honoured with the Data & Marketing Association gold award for best use of AI in 2019.[610]
■	JPMorgan Chase; ongoing partnership with Persado
■	In 2019, JPMorgan Chase announced a five-year, enterprise-wide deal with Persado, one of the leading agencies in the use of AI, to generate high-performing marketing creatives. Their successful pilot proved that that AI-enabled marketing copy is highly effective, delivering a lift of up to 450% in click-through rates on ads rendered by Persado, compared with others in the 50-200% range.[611] Persado's proprietary technologies were used to rewrite copy and headlines based on an advanced marketing language knowledge base of more than one million tagged and scored words and phrases. Through the tool, JPMorgan Chase redrafted marketing messages in its card and mortgage businesses, to create the most compelling message possible for individuals and targeted groups of customers.[612]

6.2.3.1. "Virtuality" of influencer marketing: A case study

Lil Miquela is a musician streaming music on Spotify, a designer who owns her clothing brand, a model working with luxury fashion brands and a social media star with over 2.4 million followers on Instagram.[613] She presents herself as "Musician, change-seeker and robot" – a computer-generated (CGI) character and a first world-famous virtual influencer.

Although Lil Miquela isn't truly an AI creation, she has inspired companies to invest heavily in virtual humans and envision future digital beings completely powered by AI and existing autonomously on social media platforms.[614] With production of high-quality 3D models becoming more affordable, some creators are already envisioning virtual humans 'living' their own lives without any human involvement – from posting

[610] Data & Marketing Association, *Gold Best Use of AI*, Data & Marketing Association Awards, https://dma.org.uk/awards/winner/2019-gold-best-use-of-ai.
[611] Persado, *JPMorgan Chase Announces Five-Year Deal with Persado For AI-Powered Marketing Capabilities*, Persado, https://www.persado.com/press-releases/jpmorgan-chase-announces-five-year-deal-with-persado-for-ai-powered-marketing-capabilities/.
[612] Business Wire, *JPMorgan Chase Announcement Concerning Preferred Stock*, Business Wire, https://www.businesswire.com/news/home/20191031005537/en/JPMorgan-Chase-Announcement-Preferred-Stockl.
[613] As of 1 July, 2020.
[614] Alexander, J., *Virtual creators aren't AI — but AI is coming for them*, The Verge, Vox Media, https://www.theverge.com/2019/1/30/18200509/ai-virtual-creators-lil-miquela-instagram-artificial-intelligence.

pictures or videos, and the captions that go with them, to interacting with their followers.[615]

While the idea of virtual humans, empowered by machines, is not new, the "proliferation of smartphones and the popularity of image sharing sites such as Instagram have accelerated our awareness of these virtual humans and elevated them to a position of influence," says Scott Guthrie[616], an independent influencer marketing consultant and analyst. The fashion industry was first to embrace the potential benefits of virtual models: synthetic humans do not need beauty regimes, adjusted clothing sizes or multiple takes at a photoshoot; they always show off the sponsoring-brand garments in the best possible way, always turn up on time and always deliver content adhering to the brief.[617]

6.2.3.2. Real dangers in artificial reality

Despite their virtual nature, digital beings are raising real concerns, foremost among them: transparency. Rupa Shah, founder of Hashtag Ad Consulting[618], notes that deployment of advanced technologies means it is becoming increasingly difficult to distinguish virtual influencers by sight alone. Attention to detail in the rendering of every image allows them to appear in any context or scene, at any destination, to achieve the brand's desired narrative, and NLP makes their communication feel natural and responsive.

Advertising self-regulatory codes, enforced by ad standards bodies across Europe, already require that all commercial communications must be immediately and unambiguously identifiable, using appropriate disclosure.[619] However, these rules will arguably need to be extended to virtual influencers, additionally requiring their owners and creators to inform consumers about their virtual nature – which is under the complete control of a brand.

Dudley Neville-Spencer, director and head of data analysis at the Virtual Influencer Agency[620], agrees that virtual influencers can exacerbate already-existing issues

---

[615] Bradley, S., *Even better than the real thing? Meet the virtual influencers taking over your feeds*, The Drum, Carnyx Group Ltd, https://www.thedrum.com/news/2020/03/20/even-better-the-real-thing-meet-the-virtual-influencers-taking-over-your-feeds.

[616] Scott Guthrie[616], an influencer marketing management consultant, event speaker and blogger. Aavailable at https://sabguthrie.info/. Interviewed on 22 June, 2020 for the purposes of this paper.

[617] Guthrie, S., "Virtual Influencers: More Human Than Human", Ch. 15 in Yesiloglu, S. and Costello, J. (ed.), *Influencer Marketing: Building Brand Communities and Engagement*, Routledge, London.

[618] Rupa Shah from Hashtag Ad Consulting. Available at: https://www.hashtagad.co.uk/). Interviewed on 22 June, 2020 for the purposes of this paper.

[619] European Advertising Standards Alliance, *Best Practice Recommendation on Influencer Marketing*, EASA, https://www.easa-alliance.org/products-services/publications/easa-best-practice-guidance.

[620] Comments from Dudley Neville-Spencer, director and head of data analysis at the Virtual Influencer Agency and strategy & insights director at Live & Breath, are taken from a live online webinar hosted by Persollo on 30 June, 2020, see Persollo, *Ethics, Influencers and Growth*, Persollo Webinar 3, 30 June 2020, https://www.blog.persollo.com/post/persollo-webinar-3-ethics-influencers-and-growth.

in influencer marketing and advocates for a special watermark for virtual beings.[621] He is working alongside Shah to develop a virtual influencer code of ethics, with a focus not just on transparency, but also on other areas of social responsibility, such as body image and diversity. Many prominent voices in the industry have expressed concern that virtual influencers may lead to problems of self-esteem and mental health, related to harmful human portrayal and idealised, unrealistic beauty standards.[622] Shah adds that regulators should be ready for this, and prepared to tackle any concerns consumers may have.

6.2.3.3. The future of virtual influencers

Globally, the influencer market is growing fast, and is predicted to expand to USD 9.7 billion in 2020.[623] AI is not only used to create virtual influencers, it is also adopted in the selection and evaluation phases of influencer marketing, helping brands identify the most appropriate social media influencers for potential campaigns and vet them (e.g. checking for potentially fraudulent follower numbers).

Guthrie believes further technology advances will allow virtual influencers to become "unshackled" from pre-scripted animation paths, and freely interact and learn from each human conversation. He also predicts that in the future the subgenre of virtual influencers will splinter into at least three smaller categories: virtual brand assistants (virtual humans designed and operated by a brand, driven entirely by a brand's purpose); customer service representatives (virtual beings functioning as chatbots, but with human-like emotions and body form); and virtual influencers (virtual human influencers either owned by a sponsoring brand or operating in their own right).

Finally, as virtual influencers continue to evolve, the audiences will decide how quickly they are ready to embrace digital avatars in their social media feed and engage with them. Perhaps some experts' predictions that there won't be a brand without some sort of virtual representative will become true in the near future.[624]

6.2.4. Conclusion: AI enabled intelligent advertising

Intelligent advertising is said to be a third phase and a new frontier of digital advertising, building on interactive and programmatic marketing.[625] It builds on interactivity and automation from the previous phases, but adds new attributes, such as personalisation that goes beyond a user's interests, and shifts towards predicting their needs in various

[621] Persollo, *op.cit.*
[622] Bradley, S., *op.cit.*
[623] Influencer Marketing Hub, *The State of Influencer Marketing 2020: Benchmark Report*, Influencer Marketing Hub, https://influencermarketinghub.com/influencer-marketing-benchmark-report-2020/.
[624] Persollo, *op.cit.*
[625] Li, H., "Special Section Introduction: Artificial Intelligence and Advertising", *Journal of Advertising*, Aug/Sep2019, Vol. 48 Issue 4, p. 333-337.

contexts and time-frames, issuing highly individualised commercial content in real-time and at scale.[626] It merges programmatic buying and programmatic creativity techniques, and opens the door to uniquely tailored advertising experiences that promise to be even more relevant and useful content.

However, for intelligent advertising to be effective and earn consumer trust, it cannot be sealed in a non-transparent black box which generates promotional messages based on illegitimately collected user data and unfair algorithms. The next part of this chapter will therefore investigate challenges related to AI in advertising and discuss necessary safeguards.

6.3. Concerns regarding Big Data and AI

Most uses of AI in advertising today rely on algorithms and large datasets containing information about users' characteristics and personal preferences, and serving up tailored, 'hand-picked' commercial content.[627] Naturally, such an optimisation process raises ethical and legal challenges, particularly those regarding privacy and biased algorithms. This chapter will therefore look into existing legal safeguards protecting citizens and consumers from potential harms related to issues of data protection and automated decision-making. Although this part will focus on the European context, it should be noted that policies addressing privacy and data protection issues exist also elsewhere in the world and could be the subject of a similar analysis.[628]

Scholars have argued that good privacy legislation in the age of AI should protect consumers from potential AI-based discrimination, lack of consent, and data abuse, and should include several key components: a requirement for transparency, so that AI has a deeply rooted right to the information it is collecting; an opt-out for consumers; data collected and the purpose of the AI limitation by design; and the option for data deletion by consumer request.[629] Civil society groups have also been calling for "clear red-lines for impermissible use, democratic oversight, and a truly fundamental rights-based approach

---

[626] Chen, G, et. al, "Understanding Programmatic Creative: The Role of AI", *Journal of Advertising*, Aug/Sep2019, Vol. 48 Issue 4, 347-355.

[627] Lee, K. F., *op.cit.*

[628] For example, the California Consumer Privacy Act (CCPA), which introduces new consumer rights around businesses' use, deletion, withdrawal and access to personal information (see Paka, A., *How Does The CCPA Impact Your AI?*, Forbes Technology Council, Forbes, https://www.forbes.com/sites/forbestechcouncil/2020/02/20/how-does-the-ccpa-impact-your-ai/#7d27ce7c43c7), or the Act on Protection of Personal Information (APPI) in Japan, which is expected to be further amended in 2020 with introduction and promotion of pseudonymised data in the context of feeding the AI, see GLI, *AI, Machine Learning & Big Data Laws and Regulations Japan*, Global Legal Insights, https://www.globallegalinsights.com/practice-areas/ai-machine-learning-and-big-data-laws-and-regulations/japan. See also Chapter 2 of this publication.

[629] Intel AI, *Rethinking Privacy For The AI Era*, Forbes, Insight team, https://www.forbes.com/sites/insights-intelai/2019/03/27/rethinking-privacy-for-the-ai-era/#693cda737f0a.

to AI regulation", to create a trustworthy AI system.[630] Although the scope of this paper is limited to automated technology use in marketing, the human rights approach, particularly in relation to the right to privacy and data protection (Art. 8, ECHR) and the prohibition on discrimination (Art. 14, ECHR) are also worth taking into account, especially concerning the ways in which collected personal data can be ultimately repurposed and how algorithms built with intentional and non-intentional bias might lead to segmentation and differentiated treatment of targeted social (or consumer) groups.[631] Finally, scholars have also warned of "power asymmetry between those who develop and employ AI technologies, and those who interact with and are subject to them".[632] Taking these issues into account, the following section will look into existing accountability and responsibility mechanisms to address and mitigate these risks.

6.3.1. Existing legal framework in Europe

In February 2020, the European Commission published a white paper proposing a strategy to ensure the successful uptake of AI in the European Union via an appropriate policy and regulatory framework and the creation of an "Ecosystem of Excellence and Trust".[633] The white paper pointed to a strict legal framework in the EU, which already ensures inter alia consumer protection, addresses unfair commercial practices and protects personal data and privacy, notably the "General Data Protection Regulation and other sectorial legislation covering personal data protection, such as the Data Protection Law Enforcement Directive".[634] Although the GDPR does not refer to AI specifically, it is set up in a technologically-neutral manner, to face any technological change or evolution, and therefore fully captures the processing of personal data through an algorithm.[635] Furthermore, even AI systems that rely on anonymised data may still be subject to GDPR regulation, since some anonymisation techniques are not necessarily able to annul the risk of re-identification.[636] Furthermore, according to an assessment by the European Data

[630] EDRi, *Can the EU make AI "trustworthy"? No – but they can make it just*, EDRi, https://edri.org/can-the-eu-make-ai-trustworthy-no-but-they-can-make-it-just/.
[631] Wagner, B., *Study On The Human Rights Dimensions of Automated Data Processing Techniques (In Particular Algorithms) And Possible Regulatory Implications*, Council of Europe, Committee of Experts on internet intermediaries (MSI-NET), https://rm.coe.int/study-hr-dimension-of-automated-data-processing-incl-algorithms/168075b94a.
[632] Yeung, K., *Responsibility and AI,* Council of Europe study DGI(2019)05, Council of Europe, Expert Committee on human rights dimensions of automated data processing and different forms of artificial intelligence, (MSI-AUT), https://rm.coe.int/responsability-and-ai-en/168097d9c5.
[633] European Commission, *White Paper On Artificial Intelligence - A European approach to excellence and trust*, Brussels, 19.2.2020, COM(2020) 65, https://ec.europa.eu/info/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en.
[634] European Commission, *op.cit.*
[635] European Data Protection Board, *EDPB Response to the MEP Sophie in't Veld's letter on unfair algorithms*, EDPB, https://edpb.europa.eu/our-work-tools/our-documents/letters/edpb-response-mep-sophie-int-velds-letter-unfair-algorithms_en
[636] Zaccaria, E., *Artificial Intelligence: addressing the risks to data privacy and beyond*, PrivSec Report, https://gdpr.report/news/2020/06/01/artificial-intelligence-addressing-the-risks-to-data-privacy-and-beyond/.

Protection Board (EDPB), a "risk based approach, the data minimisation principle and the requirement of data protection by design and by default", as well as provisions regarding accountability, transparency and the prohibition of any decision-making based solely on automatic processing, make the current legal framework suitable to address many of the potential risks and challenges associated with the processing of personal data through algorithms.[637] Finally, a report on the impact of GDPR on artificial intelligence, published in June 2020 by the European Parliamentary Research Service (EPRS), concludes that GDPR "provides meaningful indications for data protection in the context of AI applications", and that data "controllers engaging in AI-based processing should endorse the values of the GDPR and adopt a responsible and risk-oriented approach".[638]

Geraldine Proust, director for policy at the Federation of European Direct and Interactive Marketing (FEDMA)[639], who shared her expertise on data protection laws in the EU, concurs, saying that the GDPR does indeed provide principles that protect personal data processed by technologies using AI, and that it requires that organisations be accountable, and that they not only respect these principles, but also be able to demonstrate that they are respecting them. Proust also points out that "organisations must process only necessary, adequate and accurate personal data". She adds: "Moreover, the processing must be fair, transparent, lawful and for a legitimate purpose. If the purpose may be achieved without personal data, an alternative approach must be taken, for example the data can be anonymised". This calls for IT experts, managers and data protection officers to work together as early as possible in the marketing creative process, to achieve the right balance between innovation and ethics. The EDPB also stated that considering the already extensive existing legal framework, the focus should be on the development of existing norms, accountability and Data Protection Impact Assessments (DPIAs) in the context of machine learning algorithms. Proust supports such an approach and thinks that current legislation should be properly implemented and assessed first, and that any new laws "must be balanced to avoid hindering the development and use of this technology and aligned with data protection legislation to avoid contradictions".

Guidelines issued by relevant authorities could be a good way to provide additional clarity and advice for the application of specific legal requirements, where necessary. A recent EPRS report emphasises the need for such guidance, stating that controllers and data subjects "should not be left alone" and should be "provided with guidance on how AI can be applied to personal data consistently with the GDPR".[640] The authors of the EPRS report call for a multilevel approach and for institutions to actively engage in broad societal debates with all stakeholders, including controllers, processors, and civil society, in order to develop appropriate responses and high-level indications

[637] European Data Protection Board, *op.cit.*

[638] European Parliamentary Research Service (2020), The impact of the General Data Protection Regulation (GDPR) on artificial intelligence, Study, European Parliamentary Research Service, Scientific Foresight Unit (STOA), PE 641.530, https://www.europarl.europa.eu/RegData/etudes/STUD/2020/641530/EPRS_STU(2020)641530_EN.pdf.

[639] Geraldine Proust from FEDMA. Available at: https://www.fedma.org/. Interviewed on 7 July, 2020 for the purposes of this paper.

[640] European Parliamentary Research Service, *op.cit.*

"based on shared values and effective technologies".[641] To this end, several useful guidance documents have already been published by the European institutions, including the European Commission, Council of Europe and the EDPB,[642] as well as by industry stakeholders.[643] Many more are likely to be introduced in the coming years and together with the regulatory and self-regulatory instruments they will hopefully be useful tools contributing to the overall success of AI adoption, consumer protection and trust in technologies.

6.3.2. Conclusion: (Mostly) the Good, the Bad and the Ugly

The response to existing key concerns over the use of AI in marketing fit well into the borrowed frame of the famous Sergio Leone western entitled: "The Good, the Bad and the Ugly".

Overall, there is mostly good news, relating to the fact that the legislation and policy guidelines existing today provide a solid framework to build sustainable AI systems for advertising which safeguard human rights and protect consumers' interests without hampering further innovation. Moreover, businesses, from global tech corporations to start-ups, appear to have reevaluated the long-term benefits of consumer trust, and are calling for a human-centric approach in designing and applying AI, from its functioning, sensing, cognitive and learning abilities,[644] to considerations on how it affects end-users (micro approach), as well as society and the ethical environment in general (macro approach). Dutch company DEUS calls it a "humanity-cantered" approach in marketing, where it is particularly relevant, because advertising addresses people's needs, desires and, frequently, pressure points leading through their purchasing journeys. As consumers become more aware of the use of their data in advertising and eager to know even more,[645] it appears inevitable that the topic of the ethical use of AI in marketing will

---

[641] European Parliamentary Research Service, *op.cit.*

[642] For example *Guidelines on Artificial Intelligence and Data Protection* (2019) by the Consultative Committee of the Convention for the Protection of Individuals with regard to the Processing of Personal Data (Convention 108). Available at https://rm.coe.int/guidelines-on-artificial-intelligence-and-data-protection/168091f9d8. Council of Europe; *Ethics Guidelines for Trustworthy Artificial Intelligence (*2019*)* by the High-Level Group on Artificial Intelligence, European Commission. Available at: https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines#Top); *Guidelines 4/2019 on Article 25 Data Protection by Design and by Default* (2019) by the European Data Protection Board. Available at https://edpb.europa.eu/our-work-tools/public-consultations-art-704/2019/guidelines-42019-article-25-data-protection-design_en).

[643] For example, High Level Principles on Artificial Intelligence (2020) by the European Tech Alliance (EUTA). Available at http://eutechalliance.eu/wp-content/uploads/2020/02/EUTA-High-Level-Principles-on-AI.pdf); The Seven-Step Ad Tech Guide (2020) by the Data & Marketing Association (DMA) and the Incorporated Society of British Advertisers (ISBA) addressing privacy challenges of RTB in programmatic advertising. Available at https://dma.org.uk/uploads/misc/seven-step-ad-tech-guide-v9.pdf).

[644] Yamakage, Y. and Okamoto, S., *Toward AI for human beings: Human centric AI*, Zinrai. Fujitsu scientific & technical journal. January 2017, Vol.53. no. 1, pp. 38-44.

[645] According to an EDAA survey on how EU citizens perceive digital advertising since the adoption of the GDPR: 97% of consumers are aware that data are used for online advertising, 62% have some understanding

continue to be prioritised by civil society, policy-makers and responsible industry players, who all appear to agree on the need for a trustworthy AI framework.

The bad news comes with fact that all the above-mentioned policies and strategies will only be useful as long as sufficient enforcement mechanisms are in place and there is active demand for accountability and transparency. In short, there must be resistance to desensitisation regarding all the data that consumers are providing in exchange for more relevant information and services. This transparent exchange needs to remain a focus point at the individual and society levels.

Finally, the ugly truth is that the speed of technological development is always likely to surpass the efforts of lawmakers regarding the application of new technologies, for example facial recognition, thus creating new challenges. That is why it is essential to insert ethical standards into the design of AI and other technologies used in the advertising ecosystem and beyond. As noted by all interviewed experts, technology is important, but people and organisational culture are more important, and they can steer tech adoption towards any chosen ethical principles.

6.4. Using AI for intelligent ad regulation

Advertising self-regulatory organisations (SROs) are independent ad standards bodies ensuring responsible commercial communications through self-regulation (SR) and enforcement of advertising codes. Based on a set of Best Practice Recommendations on Digital Marketing Communications, issued by the European Advertising Standards Alliance (EASA),[646] all digital online advertising fall under SROs' remit, meaning all new or emerging digital advertising formats and practices, such as product promotion delivered by a voice home device or through a post by a virtual influencer sponsored by a brand, are subject to the same ethical standards requiring that they be legal, decent, honest and truthful.[647]

Although the primary responsibility of SROs is to ensure the compliance of ads, by handling consumer and competitor complaints and providing voluntary ex ante copy advice, ad standards bodies are increasingly dedicating their resources to providing tech-assisted ad monitoring, particularly in the online environment, which is seeing a yearly double-digit increase in ad spend.[648] In fact, in 2019 two SROs received EASA Best Practice

---

of how it works, and 72% would like know more about how information about them is used online. See European Interactive Digital Advertising Alliance, *Consumer Research – How EU citizens perceive digital advertising since GDPR*, EDAA, https://www.edaa.eu/consumer-research-how-eu-citizens-perceive-digital-advertising-since-gdpr/.

[646] The EASA membership comprises 28 such SROs across Europe and 13 industry associations, representing the entire advertising value chain (i.e. brands, agencies, media, publishers and digital platforms).

[647] European Advertising Standards Alliance, *Best Practice Recommendation on Digital Marketing Communications*, EASA, https://www.easa-alliance.org/products-services/publications/easa-best-practice-guidance.

[648] IAB Europe (2020), *AB Europe AdEx Benchmark 2019 Study,* IAB Europe,

Awards for their use of technology in online monitoring: the ad standards body in the UK - the Advertising Standards Authority (ASA) - won a Platinum Award for its Avatar Monitoring project, while the French SRO - *Autorité de Régulation Professionnelle de la Publicité* (ARPP) - received a Gold Award for its AI programme delivering "Compliance as a Service".[649]

In order to foster the exchange of best practices and help scale SR tech-empowered innovations and knowledge on AI and machine learning applications in the ad regulatory environment at the European level, in March 2020 the EASA established a working group on "Data Driven SR". The following sections will look into ongoing SR AI projects and discuss how these, and other related tech initiatives, can help promote trustworthy advertising.

6.4.1. Avatars gathering data for good

In 2019 the UK ad standards body, the ASA, used avatars – automated data capture programmes, which replicated the online profiles of specific age groups – to monitor online display advertising of restricted products, such as alcohol, gambling, HFSS (high fat, salt, or sugar) foods and soft drinks, directed at British consumers.

The avatar monitoring involved avatars representing seven different age profiles, which visited 250 websites and YouTube channels, and related sites, chalking up a total of 196,000 page visits, and capturing information on over 95 000 ads served in a two-week period – a volume impossible to attain using traditional investigative approaches.[650] The exercise was the first time avatar technology had been used for regulatory monitoring in this way, and allowed the ASA to gather information on whether ads for restricted products were being inappropriately targeted at specific audiences or served alongside inappropriate media.[651]

Although the automated technology used in this successful pilot for automated monitoring approaches wasn't employing AI systems, the ASA continues to focus on bringing machine learning algorithms into SROs' everyday activities as part of their long-term *Tech4Good* programme, currently comprising four projects. The ASA has also employed Brandwatch technology, a machine learning-based social intelligence software programme, to discover illegal or non-compliant ads on social media. In the first

https://iabeurope.eu/knowledge-hub/iab-europe-adex-benchmark-2019-study/.

[649] European Advertising Standards Alliance, *Press Release: Easa's Best Practice Awards 2019*, EASA, https://easa-alliance.org/news/easa/press-release-best-practice-awards-2019.

[650] European Advertising Standards Alliance, *2018 European Trends in Advertising Complaints, Copy Advice and Pre-clearance*, EASA, https://www.easa-alliance.org/products-services/publications/easa-statistics.

[651] The Advertising Standards Authority (2019), *ASA monitoring report on HFSS ads appearing around children's media,* The Advertising Standards Authority Ltd, https://www.asa.org.uk/resource/asa-monitoring-report-on-hfss-ads-appearing-around-children-s-media.html

monitoring exercise, 12 000 problem posts featuring Botox ads[652] were removed over a period of three months.[653]

Guy Parker, the chief executive of the ASA[654], thinks AI technologies have the potential to deliver efficiencies in SROs' operations and improve the effectiveness of ad regulation. "We believe that regulators that don't embrace it will be left behind", Parker adds. He thinks AI applications in the ad compliance field "offer the potential to respond to the scale challenge of regulating online advertising", including "numerous small and medium sized businesses who are not always aware of the rules and their importance."

6.4.2. AI advancements for advertising compliance in France

The French ad standards body, the ARPP, started using AI in its operations in 2019, as part of its "Compliance as a Service (CaaS)" programme. The first three projects focused on detecting industry's priorities for AI use via dedicated R&D workshops, development of chatbot assistant "Jo" (an interactive chatbot guide for requesting advertising pre-clearance and copy advice services from the ARPP), and a deep-learning application for alcohol, super-imposed text, and gender representation detection and analysis in TV advertising.

In early 2020, the ARPP extended the scope of its supervised learning system for gender depiction from face detection to automatic analysis of voices and age groups, and ran an exhaustive analysis of television advertising received for ARPP pre-clearance approval. Nearly 131 hours of video material were analysed by three supervised learning models using image, voice recognition technologies and tracking actors in a scene, allowing the ARPP to assess the evolution of the representation of men and women.[655]

Mohamed Mansouri, deputy director of the ARPP[656], says that the main reasons the SRO is pursuing AI technologies are: time efficiency thanks to the machine learning automation of low-added-value tasks; and a desire to create modern, agile advertising self-regulation taking advantage of the latest technologies, for everyone - industry professionals, public authorities and consumers. The consultation with the industry during the workshops allowed further definition of the priority areas for the ARPP's monitoring services and further evolution of its AI programme.

---

[652] Botox products and services cannot be advertised to the public in the UK.

[653] The Advertising Standards Authority (2020*), Annual Report 2019,* The Advertising Standards Authority (ASA) and Committee of Advertising Practice (CAP), https://www.asa.org.uk/news/using-technology-for-good-our-annual-report.html.

[654] Guy Parker from the ASA. Available at http://https://www.asa.org.uk. Interviewed on 23 June, 2020 for the purposes of this paper.

[655] Mansouri, M., *Intelligence artificielle et représentation féminine/masculine dans la publicité audiovisuelle. Après les visages et le genre : la voix et l'âge !*, ARPP blog, ARPP, https://blog.arpp.org/2020/03/05/intelligence-artificielle-et-representation-feminine-masculine-dans-publicite-audiovisuelle-visages-genre-voix-age/.

[656] Mohamed Mansouri from the ARPP. Available at https://www.arpp.org/ Interviewed on 24 June, 2020 for the purposes of this paper.

6.4.2.1. Invenio and beyond: A case study

In 2020, Mansouri together with ARPP's tech-partner company Sicara[657] and other experts from the French SRO launched the Invenio project, comprised of four modules:

- a web-crawler: an automatic ad collector on a series of sites, currently used in display advertising;
- detection of potential breaches using AI based on computer vision and text analysis;
- reporting of breaches and validation by the ARPP's lawyers and experts, looped back into the model to improve its accuracy (currently based on misleading therapy claims and prohibited financial advertising);
- an automatically triggered alert system when an ad pointing to the specified prohibited sites is found.

Invenio means 'I found' in Latin, and it's a suitable name for the project, which helps to significantly save time in identifying non-compliance cases and tackle more breaches, according to Mansouri. However, its value is also seen on an educative level, and enabling various players in the ad value chain to audit their program flow.

Mansouri explains that the next phases for Invenio expansion involve: detection of other types of common breaches in areas like advertising of cosmetics and electronic cigarettes; lack of disclosure of commercial collaborations and other forms of adverting with regard to influence marketing and video ads. He also adds that the future vision for Invenio includes the European dimension, "which is more than ever essential to demonstrate the validity of self-regulation at the European level".

Asked whether he sees a danger that AI tools may be used too widely in the future and may thus misjudge and over-police ads, Mansouri offers reassurance: "The human element is fundamental in the system" and nothing, he adds, is decided without prior legal analysis and validation by experts. "AI allows to process loads of data very efficiently, but on very basic things. Even if the models improve its accuracy with use, the legal syllogism to date cannot be automated".

6.4.3. Harnessing technology to bring more trust to the Dutch ad market

Another frontrunner when it comes to exploring AI and its application to ad self-regulation is the Dutch SRO Stichting Reclame Code (SRC), which is currently exploring different possible applications of automated technologies. SRC's core objective is "building trust in the ad industry". Building on its dominant position in the field of offline ad self-regulation, SRC is seeking to become a force for good in the online environment

[657] More information about Sicara available at https://www.sicara.fr/.

as well. Otto van der Harst, the director of SRC,[658] is convinced that advertising self-regulatory organisations "need to take action and get a deep understanding of the advertising tech world" while maintaining their position as independent regulators. He believes that, together, the Dutch SRO and the ad industry are "standing at the dawn of a new decade in online commerce and advertising".

SRC's technology partner, DEUS, concurs, saying that the "use of data science and new technologies offers an opportunity to start using the enormous amounts of data in the online advertising ecosystem for good". Nathalie Post from DEUS[659] adds that one of the key focus areas for SRC in the Netherlands is the protection of vulnerable people (e.g. minors, people with addiction problems, the elderly) from potentially harmful commercial content (e.g. targeted alcohol, gambling, slimming, plastic surgery and Botox product ads, as well as excessive volumes of HFSS ads, and misleading medical claims). The DEUS team sees five areas where automated technologies can be particularly helpful in supporting SRC in these areas of activity: using avatars to simulate profiles of specific (vulnerable) target groups; using social listening tools to monitor social media influencers; using NLP and computer vision to automate the registration, categorisation and reporting of ads that breach the advertising code; using computer vision to identify ads that contain certain biases, such as a lack of diversity; and gaining a better understanding of data flows in the RTB ecosystem, to address potential breaches of advertising codes.

Otto van der Harst is confident that the dream of having AI-enabled SR services is already on the horizon, and that the coming years will be about experiments, discussions with advertising self-regulatory bodies, platforms and agencies, where he expects to align all serious parties involved: "AI and machine learning are just instruments to see that the online world remains a safe advertising space".

6.4.4. Tech solutions from the ad industry powerhouse

SROs are not the only ones tracking irresponsible ads and trying to make the online space safer by employing AI technologies. Some of the online ad industry giants, standing at the very forefront of tech development are also actively applying AI tools to ensure a safe and sustainable ad environment online.

In April 2020, Google published its annual "bad ads" report, claiming to have blocked and removed 2.7 billion bad ads in 2019 (0.4 billion more than in 2018), in other words 10 million ads per day, 5,000 ads per minute and more than 100 per second.[660]

---

[658] Otto van der Harst from SRC. Available at https://www.reclamecode.nl/. Interviewed on 20 June, 2020 for the purposes of this paper.
[659] Nathalie Post from DEUS. Available at https://deus.ai/. Interviewed on 22 June, 2020 for the purposes of this paper.
[660] Spencer, S., *Stopping bad ads to protect users*, Google Ads, Google, https://blog.google/products/ads/stopping-bad-ads-to-protect-users.

"We take great pride in being a resource for people around the world searching for important information," says the Google team[661] interviewed for this paper "Along with trusting the information surfaced, we know users should also be able to trust the ads they are seeing."

As the digital advertising ecosystem grows, new threats are rising, requiring continuous adaptation of companies' policies, and the improvement of technology. "As we get better at detecting trends and patterns with scammers and fraudsters, we are responding in kind with new technology to stop these emerging threats and take more account-level action," the Google experts continue. In 2019, they doubled down on the tech to address long standing abuse of phishing and trick-to-click ads. A dedicated task force at Google tracked and analysed the tactics bad actors were using to circumvent Google's systems and collect personal information from users. "While phishing ads range from those that exploit people who are interested in cryptocurrency to others looking for information on passport renewals to ads leading users to fake bank websites, we found common denominators in how they evade our systems that allowed us to improve our tech to stop them," the Google team says.

Although technology helps Google spot potential violations, it is the combination of tech and talent that allows effective enforcement. "We have thousands of people at Google dedicated to helping us fight bad ads, bad sites and scammers online," the team notes. "We're constantly reviewing ads, sites and accounts to ensure that they comply with our policies.". Implementing AI and other machine learning technology allows Google to charge their enforcement efforts 24/7 and enables them to assess multiple variables when taking decisions about taking down an ad.

Asked about the fact that more bad ads are being removed every year, Google's team replied that the increase in numbers is indicative of a couple of key factors, including the dynamic nature and evolving size of the digital advertising ecosystem, as well as ongoing improvements in adapting to the different ways in which bad actors try to game the system. According to Google's experts: "No system will be 100% perfect, but we're vigilant and always working to improve our tools."

6.4.5. Future frontier for advertising self-regulation

Intelligent advertising was mentioned earlier as a potential aim of those creating and adopting AI applications for advertising. It's therefore also worth asking what the future of AI-enhanced advertising regulation should be and how it can be made intelligent, that is to say not merely reactive and punitive, but anticipatory of emerging issues and supportive of industry actors in proactive ways.

[661] A team of experts from Google (available at: https://about.google/) was interviewed on 8 July, 2020 for the purposes of this paper, and provided the response from the company on the topics discussed in this chapter.

The interviewed SROs all agree that ad standards bodies have to respond to ad innovations with tech innovations of their own. Guy Parker from the ASA thinks that AI and other forms of automated tech will provide SROs with better intelligence on where and how they should be intervening or providing services. He adds that although the ASA "will never achieve perfection", the technology will be at the heart of reducing irresponsible and harmful online ads in the future. Mohamed Mansouri from the ARPP adds that AI will essentially help SROs increase efficiency, go deeper and wider, and act quickly in the digital ad realm, where channels, formats and content grow exponentially. He adds that collaboration in creating and employing AI technologies is very important as it allows the development of collective intelligence and responds to the challenges associated with the inherently cross-border digital environment. Otto van der Harst from SRC agrees, saying SROs need to get a grip on the ad ecosystem if they "want to be relevant in the future" and use tech assistance to have some form of independent oversight and be able to deliver a credible response to the current market developments.

Finally, European cooperation will likely play an important role in the scaling of existing initiatives, the sharing of expertise and the mutual learning of the various actors. The SRO network in Europe, coordinated by the EASA, has seen new formats of commercial communications disrupting the ad industry many times in the past and has proved agile and flexible in adapting to changes in technology and societies, while respecting European cultural diversity. Through the sharing of best practices and collective discussion of future challenges, the SROs have continued to inspire and support the ad industry in the creation of responsible advertising and the preservation of consumer trust. The ongoing cooperation on the use of AI and other automated technologies is one the most exciting European-wide initiatives, one that will likely be an incubator for the promising future of intelligent advertising self-regulation.

6.5. Conclusion: 'The great data rush'

Many practitioners and scholars writing about AI describe its increasing and expanding applications in different industry sectors, including advertising, using a big data wave analogy - with reference to a tsunami-like power forcefully transforming and reshaping the environment, sometimes in an unrecognisable way. However, unlike tsunamis, the operationalisation of data does not occur as natural phenomenon. It needs to be engineered

That's why I believe that a more useful analogy to describe AI applications in marketing and advertising is the gold fever of the 19th century, which led to massive changes in the world economy and trade patterns, and spurred migration and rapid social mobility. Not unlike those new settlers who sailed across oceans in pursuit of the American dream, the advertising industry today is experiencing a great data rush, which promises even greater benefits to those who harness and properly channel data, turning consumers into profit-bringing customers.

Much like the innovations introduced during the gold rush era, AI applications in advertising have already generated numerous benefits for both industry and consumers. AI has made advertising easier to deliver and cheaper to place in front of the relevant consumers. Machine learning technologies have also made it possible to seamlessly integrate new formats of advertising in our everyday lives, from voice assistants offering us tailored products and services, based on our past purchasing history and other contributing factors, to virtual influencers naturally interacting with us on digital platforms, exchanging holiday tips with audiences and promoting aspirational lifestyles.

On the other hand, gold fever was also frequently associated with the lawless pursuit of wealth, encapsulated by the term "wild west". It alludes to a disruptive nature of the new AI-enhanced forms and formats of advertising. Although today's market race to develop and implement the most effective AI tools for advertising should not be compared to the dramatic face-offs depicted in westerns, there are still challenges and concerns regarding a potential lack of oversight of AI technologies.

Civil society and policy-makers are troubled by increasing personal data collection, automation of decisions and market asymmetry. Accountability and transparency are paramount to ensure that people don't feel alienated by AI and can instead embrace and fully benefit from the technologies that help them find relevant information faster and get their queries answered quicker - and may even help them with their creative projects.

In conclusion, is AI a gateway to a Deadwood-like advertising ecosystem, functioning without adherence to existing rules, exploiting consumers and designed to profit only those who develop the technological expertise and grab the golden power source of data? No, and in reality, such a gloomy future, and indeed the wild west scenario, are very unlikely. As discussed in this paper, the constant pressure from civil society, as well as the detailed requirements of the regulatory frameworks and self-regulatory initiatives, put the ethical use of AI at the centre of continuing public debates. Recent policy developments and the views of the interviewed experts show a strong commitment to human-centric AI and its application in advertising. However, strategies and guidelines have to be continuously implemented in practice and safeguards only work if there is a will to trigger and enforce them - which is why consumers also have the responsibility to know their rights, be active and remain vigilant about demanding proper protection and accountability from market actors.

Finally, and perhaps most importantly, building a sustainable AI framework for advertising and the use of data and technology for good is in the interest of the advertising community itself. Only by embedding ethical principles in the machine learning algorithms, collecting and using personal data in a respectful way that reflects high standards of transparency and responsibility, and fostering AI applications in marketing that help to identify bad actors and hold them to account, can the industry expect to earn lasting consumer trust. After all, trust is, and always will be, a true gold standard in advertising.

6.6. Acknowledgements

Much like AI algorithms, which need good data input to learn, this article could not have been developed without the insights of the experts interviewed, who kindly agreed to share their expertise, knowledge and future predictions related to the use of AI in advertising. I would like to thank Kerry Richardson from Tiny Giants, Charlie Cadbury from Say It Now and Nathalie Post from DEUS for sharing their expertise on the intersections between AI and creativity, technology-enabled new advertising formats and the development of a human-centric AI approach. A big thanks to independent influencer marketing consultant Scott Guthrie and Rupa Shah from Hashtag Ad Consulting for sharing their thoughts on the development of virtual influencers, ethics, and the possible future of virtual humans as marketing channels. I also want to thank Geraldine Proust from FEDMA who kindly guided me through the legal frameworks concerning AI use in advertising and shared her insights on the key challenges for machine learning applications in advertising going forward. Many thanks to the Google team that provided insights and further explanations about the company's efforts to use advanced technology and human expertise to remove bad ads and keep their platforms safe. Last but not least, I am very grateful to my colleagues from the advertising self-regulatory organisations - Guy Parker from the ASA, Otto van der Harst from the SRC and Mohamed Mansouri from the ARPP - for taking challenging steps to adopt AI and other machine learning technologies to continue to ensure high ethical standards in advertising and for allowing me to join them on this exciting journey, to learn along the way and to discuss the future developments together, searching for innovative solutions with an emphasis on a strong European collaboration.

AI in advertising is in some ways a labyrinth in which it is impossible to visit every corner. However, with the guidance of all the interviewed experts I hope I have found several possible paths through it. Those paths, analysed in this chapter, represent my own take on AI in advertising and should not be attributed to any particular organisation or individual mentioned in the text. Finally, the European Audiovisual Observatory should not be held responsible for any of the quotes and opinions provided in this paper, including those of all the interviewed experts and the author herself.

6.7. List of interviews

- Charlie Cadbury from Say It Now, https://www.sayitnow.ai/. Interviewed on 19 June 2020.
- Google expert team, https://about.google/. Interviewed on 8 July 2020 for the purposes of this paper, and provided the response from the company.
- Scott Guthrie, an independent influencer marketing management consultant, event speaker and blogger, https://sabguthrie.info/. Interviewed on 22 June 2020.
- Otto van der Harst from SRC, https://www.reclamecode.nl/. Interviewed on 20 June 2020.
- Mohamed Mansouri from the ARPP, https://www.arpp.org/. Interviewed on 24 June 2020.
- Guy Parker from the ASA, https://www.asa.org.uk. Interviewed on 23 June 2020.
- Geraldine Proust from FEDMA, https://www.fedma.org/. interviewed on 7 July 2020.
- Nathalie Post from DEUS, https://deus.ai/. Interviewed on 22 June 2020.
- Kerry Richardson from Tiny Giant, https://www.tinygiant.io/. Interviewed on 22 June 2020.
- Rupa Shah from Hashtag Ad Consulting, https://www.hashtagad.co.uk/. Interviewed on 22 June 2020.

Personality rights

AI can be so creative it can go way beyond helping in the scriptwriting process. After all, a script is only the beginning of the creative process. The story and the ideas in a script have to be translated into images. In most cases, these stories talk about people. People played by actors. AI can not only write the script and play the music but it can also provide the actors. Or at least turn any actor into the actor you always wished to have in your film. Making him or her younger or older, for instance. There are very recent examples of this: in "Gemini Man", the character played by Will Smith has to fight against a younger clone of himself. A similar de-aging procedure was applied to the main characters in Martin Scorsese's "The Irishman". In "Star Wars: Rogue One", Carrie Fisher looks younger than ever, and Peter Cushing, who died in 1994, also has his moment of post-mortem glory. But normally, after hype comes hysteria. If you read the newspapers these days, you may come across headlines such as: "In the age of deepfakes, could virtual actors put humans out of business?"[662] *Indeed, imagine for example a gangster film with a digital Marlon Brando but without his notorious backstage behaviour. Which director would not want that? For this, you just need the relevant hardware and software … and a ghost actor. That is, an actor whose face is replaced by that of the more famous one. Cheaper and, in the case of Brando, probably better behaved. On top of that, AI makes it substantially easier to create digital extras. As you can imagine, these developments, both technological and artistic, raise personality rights issues. These legal issues are regulated by law and then settled by contract. But a contract can be unfair to the party with less bargaining power. Like an unknown actor. Or a dead person. There is also a darker side of this issue. Deepfakes. They can be used in different, harmful ways. First of all, commercial misappropriation. Deepfakes can be used for fake endorsements of products. There is another issue: identity abuse (mostly in porn films). In her contribution to this publication,* ***Kelsey Farish*** *remarks that "especially in the case of novel technologies such as deepfakes and ghost acting […] [P]ersonality rights require a careful consideration of situational context".*

[662] Kemp L., "In the age of deepfakes, could virtual actors put humans out of business?", *The Guardian*, https://www.theguardian.com/film/2019/jul/03/in-the-age-of-deepfakes-could-virtual-actors-put-humans-out-of-business.

7. Personality rights: From Hollywood to deepfakes

Kelsey Farish, DAC Beachcroft LLP

"The body is not a thing, it is a situation: it is the instrument of our grasp upon the world"

Simone de Beauvoir

7.1. Introduction

Lawyers do not often turn to Hollywood actors for professional insight, but when it comes to new technologies that may harm one's public image, Scarlett Johansson is a worthy exception. During an interview with *The Washington Post,* Johansson, one of the most recognisable and highly-paid film stars in the world, explained "nothing can stop someone from cutting and pasting my image or anyone else's onto a different body and making it look as eerily realistic as desired".[663] Although computer-generated special effects are nothing new, Johansson was referring to a then relatively unknown practice of superimposing celebrities' faces into pornographic videos, using a sophisticated artificial intelligence system. Today, we refer to this sort of face-swapping video, whether pornographic or otherwise, as a deepfake.

When asked about initiating court proceedings, Johansson admitted she felt it was a "useless pursuit, legally" because "every country has their own legalese regarding the right to your own image". She added: "So while you may be able to take down sites in the U.S. that are using your face, the same rules might not apply in Germany."[664] These rules, which seek to protect how a person's likeness appears in publications and videos, are

[663] Harwell, D., *Scarlett Johansson on fake AI-generated sex videos: 'Nothing can stop someone from cutting and pasting my image'*, The Washington Post, www.washingtonpost.com/technology/2018/12/31/scarlett-johansson-fake-ai-generated-sex-videos-nothing-can-stop-someone-cutting-pasting-my-image/.
[664] Harwell D., *op.cit.*

commonly referred to collectively as personality rights. However, personality rights are actually comprised of many different laws which have complicated exceptions and nuances. Generalisations are difficult to make, the outcomes of lawsuits are often unpredictable, and as correctly noted by Johansson, the laws differ drastically from country to country.

This chapter seeks to clarify some of the confusion surrounding personality rights, and aims to offer some practical commentary on the risks and advantages deepfakes and ghost acting present to the film, television, broadcasting, and non-news media industries. The section "AI sets the scene" introduces the technology, and the following one on "Personality rights and implications" covers the legal framework from four different angles with real-world examples. The section "Laws in selected jurisdictions" builds upon the legal framework and summarises how personality rights are recognised in Germany, France, Sweden, Guernsey, the United Kingdom[665] and California. Finally, the last section "What next for Europe's audiovisual sector?" discusses certain notable shortfalls in the laws as well as potential trends.

7.2. AI sets the scene: Deepfakes and ghost acting

Research suggests that the incredible diversity of human faces is the result of an evolutionary pressure to make each of us easily recognisable due to the highly visual nature of our personal interactions.[666] Because we are particularly adept at distinguishing faces from each other, we are likewise quick to sense when faces look eerie or unnatural,[667] and scientists and visual special effects (VFX) specialists have long struggled to accurately recreate animations of facial expressions. This changed in 2014 thanks to profound advancements in the subset of artificial intelligence known as deep machine learning.[668]

[665] The laws of the United Kingdom are referred to throughout this chapter, but the country has distinct legal systems in Scotland, England and Wales, and Northern Ireland. Unless otherwise noted, references to any of the United Kingdom's laws are as typified by the courts of England and Wales, or "English law".

[666] Sanders R., *Human faces are so variable because we evolved to look unique*, Berkeley News, University of California Berkeley, https://news.berkeley.edu/2014/09/16/human-faces-are-so-variable-because-we-evolved-to-look-unique/.

[667] Mori M., *The Uncanny Valley: The Original Essay by Masahiro Mori*, IEEE Spectrum, translated by K. F. MacDorman and N. Kageki [Japanese orig. 不気味の谷 (Bukimi No Tani) 1970]. English translation available at: www.spectrum.ieee.org/automaton/robotics/humanoids/the-uncanny-valley.

[668] Han J., *Ian Goodfellow: Invented a Way for Neural Networks to Get Better by Working Together*, MIT Technology Review, www.technologyreview.com/lists/innovators-under-35/2017/inventor/ian-goodfellow.

7.2.1. Deepfakes

Deepfakes first gained notoriety when users on the website Reddit shared hyper-realistic pornographic videos depicting Scarlett Johannson, Gal Gadot, and Emma Watson, amongst others.[669] Today, deepfakes thrive beyond that world of image-based sexual abuse, and are available as a form of entertainment for anyone to make or enjoy. The deepfake creation software is free to download on file distribution sites such as BitTorrent and GitHub, YouTube tutorials provide step-by-step instructions, and some freelance creators even sell bespoke deepfakes for as little as EUR 5 per video on marketplaces such as Fivver. Mobile apps like ZAO, Doublicat, and AvengeThem can generate face-swapped videos using just one selfie as their source,[670] and even the mainstream apps Instagram and Snapchat have 'filters' which can easily do the same. Although the technology will undoubtedly continue to improve, most deepfakes made casually and quickly are often easily detectible as fakes upon closer inspection. Nevertheless, they remain a popular phenomenon because no specialised technical knowledge is required. Besides, minor inconsistencies or glitches are no deterrent for those who make them simply out of curiosity, or for a laugh.

7.2.2. Ghost Acting

Making someone look older, younger, or otherwise different for theatrical purposes is as old as the art of performative storytelling itself, and computers have been used to animate the human face since the 1970s.[671] But it was not until the 2000s that film-worthy facial alterations were possible, evidenced particularly by *The Lord of the Rings* (2001) and *The Matrix Reloaded* (2003). Many VFX techniques include the mapping and scanning of an actor's face and body, to generate a virtual model known as a digital double. The digital double is then modified and superimposed on the body of a stand-in performance double.[672] The performance double may be the very person whose face was originally scanned, and often, the final cut depicts an actor as either a younger or an older version of themselves. More controversially, the digital double can depict an actor whose death has made their physical presence in the production impossible.

These practices are known as ghost acting or hologram acting, and may be integral to narrative continuity and the cohesive vision of the production. Rather than rely on facial mapping captured during new performances, VFX specialists use existing film

---

[669] Cole S., *AI-Assisted Fake Porn Is Here and We're All Fucked*, Motherboard Tech by VICE, www.vice.com/en_us/article/gydydm/gal-gadot-fake-ai-porn.

[670] Beebom staff, *8 Best Deepfake Apps and Websites You Can Try for Fun*, Beebom, www.beebom.com/best-deepfake-apps-websites.

[671] Parke F. I., "Computer generated animation of faces", Association for Computing Machinery '72 Proceedings of the ACM Annual Conference, 1: 451–457.

[672] Cosker D., Eisert P, and Helzle V., *Facial Capture and Animation in Visual Effects, pgs. 311-321 in Digital Representations of the Real World: How to Capture, Model, and Render Visual Reality*, University of Bath, Department of Computer Science, http://cs.bath.ac.uk/~dpc/papers/VFX_2015.pdf.

footage of the deceased actor to create their digital double, which is then again modified and superimposed on the body of a stand-in performance double. By way of recent example, after Carrie Fischer's untimely death in 2016, Lucasfilms used archived footage of the late actress to recast her as Princess Leia in the 2019 film, *Star Wars: The Rise of Skywalker*. As a notable aside, this was despite assurances made by a studio executive that they had "no intention of beginning a trend of re-creating actors who are gone".[673]

7.3. Personality rights and implications

Personality rights are aimed at providing an individual the ability to control the publication of his or her appearance, or certain characteristics of it. Protected characteristics typically include photographs and pictures of the face and body, but can also include names, distinctive styles, signatures, voices, or even mannerisms. Accessories deeply associated with the individual may also be protected, as seen in Italy when the use of a famous singer's hat and glasses for an advertisement without his consent was considered a publicity violation.[674] When taken together, these attributes may be called one's "persona" or "image", with the latter referring not to a picture as such but to one's public image. For this reason, some jurisdictions including the United Kingdom and France refer to personality rights as "image rights". But regardless of what they are called, they can present some curious legal complications for the audiovisual sector.

Because technology and society have changed in ways that have outpaced the law, lawyers must turn to older, more established doctrines to resolve disputes concerning the use of someone's image. Personality rights are therefore often said to have two distinct aspects: publicity and privacy. The right of publicity concerns use of one's image for commercial purposes, and the right of privacy seeks to prohibit unwarranted intrusion into one's intimate or family life.

It is worth pausing here to remember that the law is a nuanced language, whose obscure rules and ancient precedents have certain mythological qualities which reflect the social norms from which they develop. The proper use of personality rights depends on context as well as culture, and even within the European Union there is no unified approach. In a sense, what we observe from a legal perspective is analogous to a dimly-lit nightclub, in which the laws concerning intellectual property rights, branding and advertising, reputation management, freedom of speech, emotional distress and consumer protection all dance together, but not necessarily in harmony. The remainder of this section attempts to organise the dancefloor somewhat, by considering personality rights from four different angles.

---

[673] Robinson J., *Star Wars: The Last Jedi—What Happened to Leia?*, Vanity Fair, www.vanityfair.com/hollywood/2017/12/star-wars-the-last-jedi-does-leia-die-carrie-fisher-in-episode-ix.
[674] *Dalla v Autovox SpA Pret. Di Roma*, 18 apr. 1984, Foro It. 1984, I, 2030, Giur. It. 1985, I, 2, 453.

7.3.1. Angle 1: Publicity as (intellectual) property

Movie stars, professional athletes and other celebrities can earn substantial amounts of money beyond their day job at the studio or stadium. Lucrative uses of their *persona* may include paid advertisements, official merchandising, or collaborations with fashion houses. It is not only the individual whose financial interests are at play: his or her management team, corporate sponsors, film studios, record labels, and so on will care about the monetary value of his or her *persona*, too. With that in mind, this first angle looks at the publicity aspect of personality rights, as a form of economic protection.

Under the labour theory of property, an individual is entitled to the fruits of his or her own labour.[675] Intellectual property is, by extension, a right acquired through mental and creative labour. Copyright laws are a form of intellectual property which recognises the effort and investment involved in producing creative works, ranging from feature films to Tiktok clips, and everything in between. Copyright laws prohibit others from copying, modifying or using such creations without the creator's permission and, in the online ecosystem, copyright infringement claims have become a pervasive way for creators and studios to assert control over their content.

Unfortunately, there are several practical challenges to this copyright infringement approach to personality rights. Firstly, a somewhat paradoxical complication can arise when attempting to assert copyright protections on behalf of the person depicted. Recall that it is the labour of the photographer or videographer which is protected by copyright, and those rights are often passed directly from the creator to the studio or company producing the film in question. For this reason, a person cannot bring a copyright lawsuit just because he or she is featured in the audiovisual content. Although some rightsholders may be willing to initiate lawsuits on behalf of the person depicted, this will not always be the case, and especially not when the dispute is between the rightsholder and the performer.

Hundreds or even thousands of images may have been used to generate a digital double. Once those images have been blended, ascertaining the relevant copyright owners of those original images may be impracticable, if not impossible. Additionally, there are certain uses of copyrighted material which are lawful despite a lack of authorisation. These exceptions notably include satire and parody, and in the United States, transforming a creative work to create new expressions or meaning may likewise be permissible.

[675] Hughes J., *The Philosophy of Intellectual Property*, Georgetown University Law Center and Georgetown Law Journal, 77 Geo. L.J. 287.

7.3.2. Angle 2: Publicity and brand recognition

As with angle 1, angle 2 concerns commercial use of one's publicity, but shifts focus away from viewing one's image in audiovisual content as property. Instead, under the principle of unjust enrichment (also known as unjustified enrichment), it is unlawful to unfairly benefit financially from the goodwill or reputation of another. Just as specific symbols and names are protected through trademarks to distinguish a brand's services and goods from its competitors, it is possible to protect the image of a person when that image is used to signify a certain quality, provenance or affiliation. But unlike the trademarking of a discrete symbol or a phrase, it is impossible to trademark every aspect of a celebrity's *persona*. Firstly, video footage is not capable of trademark protection. Secondly, even if a particular shot or picture has been trademarked, this will not prevent someone from using different photographs to create the deepfake or ghost acting performance. To bridge this gap, several jurisdictions are seeking to redress the financial harm arising from a misappropriated image.

This is especially relevant in cases of false or misleading endorsements, as when British fashion retailer Topshop sold a T-shirt which prominently displayed a photograph of the Grammy Award-winning singer Rihanna. Topshop had a proper copyright licence to use the image, but because Rihanna demonstrated that the shirt damaged the attractive magnetism of her personal brand, she successfully sued Topshop for misuse of her publicity.[676] In another lawsuit coincidentally involving a photograph of a pop star on T-shirts, the German Federal Court awarded the *99 Luftballons* singer Nena a fee on the basis that she had been deprived of payment for use of that particular image.[677] Worth noting here is that in some jurisdictions, this brand recognition angle may only be successful for the person who is able to prove the financial damage to his or her publicity. Understandably, this leaves much to be desired for lesser-known figures, or those who cannot prove the monetary value of their *persona*.

7.3.3. Angle 3: Privacy protections

The two angles explored thus far have considered how publicity laws can protect the use of one's image for advertising and other commercial purposes. Angle 3 considers the ways in which privacy laws operate to prohibit unwarranted intrusion into one's intimate or family life, and how those laws might apply to deepfakes and ghost acting. Whereas property rights are inherent to ownership of some tangible or intellectual asset, the right to privacy arises automatically by virtue of being human.

---

[676] *Robyn Rihanna Fenty v Arcadia Group Brands Ltd* (t/a Topshop), Court of Appeal (Civil Division) - [2015] EWCA Civ 38.

[677] Nena, BGH, Urt. V. 14 October 1986 VI ZR 10/86 Oberlandesgerichtsbezirk Celle, Landgericht Lüneburg.

The European Convention on Human Rights (hereafter, "the Convention")[678] is the key legislation to protect human rights and political freedoms in Europe, and applies to all member states of the European Union as well as 20 other countries. Article 8 of the Convention grants everyone a fundamental right to privacy, which includes protection against unwanted intrusion into one's personal space. As anyone familiar with documentaries or reality television will know, visual images enable the viewer to act as a spectator or voyeur with regard to the subject's life.[679] Photographs and film footage are accordingly treated with caution at law, and are regarded by the courts with heightened scrutiny.

Even so, privacy laws by no means provide absolute protection for one's image. Most societies recognise that privacy must be balanced against – and in some cases yield to - other competing rights. For instance, the news media will have a right, and in some situations a duty, to share otherwise 'private' photographs of notable figures, if doing so is in the public interest. Although privacy is a fundamental human right under the Convention, so too is the right to freedom of expression under Article 10. Most countries beyond the remit of the Convention have a similar provision: in the United States, this is chiefly the First Amendment to that country's Constitution.[680]

In most privacy cases therefore, analysis will turn on whether the image was captured in public, or whether the subject had a reasonable expectation of privacy at the time. Using a telescopic lens to take paparazzi shots of a celeb lounging topless by the pool of her gated villa will, naturally, be construed differently than a photograph which captures her fully dressed on the red carpet at a film festival. A problem we encounter from a privacy law angle, however, is that in reality, the millions of photographs taken and shared each day will mostly fall somewhere between these two extremes. As such, it is not easy to predict which images will be deemed an intrusion into an individual's private life, or misuse of his or her confidential information. Additionally, as a matter of popular opinion, it is understood that a celebrity who benefits financially from his or her fame must necessarily give up some aspect of his or her privacy. Whether or not this is fair is another matter.

Where audiovisual content discloses confidential details, for example if a person speaks to reveal private facts, an invasion of privacy may be at issue. However, beyond this scenario, privacy laws are likely not suitable tools for countering unwanted deepfakes or ghost acting performances. Firstly, multiple images and videos are often blended together to make a digital double, and determining which if any of those used was initially private could be next to impossible. Secondly, and perhaps more importantly, deepfakes by their very definition depict something that never happened, and fantasy scenarios cannot constitute an invasion of privacy. As for using footage depicting a deceased actor, the heirs or estate of a performer would struggle to assert post-mortem

[678] European Convention on Human Rights (formally the Convention for the Protection of Human Rights and Fundamental Freedoms), www.echr.coe.int/Documents/Convention_ENG.pdf.
[679] *Michael Douglas, Catherine Zeta-Jones & Ors v Hello! Ltd. & Ors* [2005] EWCA Civ 595.
[680] First Amendment to the Constitution of the United States, www.constitution.congress.gov/constitution/amendment-1/

privacy rights. The European Court of Human Rights has consistently hesitated to recognise privacy rights for the deceased, unless their privacy is connected to those who are living.[681] In all but a few circumstances, a deceased individual has no privacy to speak of which may be infringed and, because privacy rights are inalienable to the person concerned, they are neither inheritable nor transferable.

7.3.4. Angle 4: Dignity and the neighbouring rights

The perspectives discussed above have aimed to provide some useful commentary on the publicity and privacy aspects of personality rights. Although the laws mentioned offer adequate protections in some instances, they can also leave open the possibility for misappropriation. To resolve this, personality rights could be expanded to cover emotional and mental well-being, irrespective of financial or privacy implications. This fourth and final angle, therefore, centres on the notion that every person has a fundamental right of dignity and personal integrity. It is also arguable that laws that currently exist to protect the integrity of recorded performances may be available to cover virtual performances, too.

The Geneva Convention recognises that "respect for the personality and dignity of human beings constitutes a universal principle which is binding even in the absence of any contractual undertaking".[682] This explicit recognition of dignity is enshrined in the post-war constitution of Germany, and elsewhere throughout the case law and legislation of most European countries. It is reasonable to conclude that these laws were inspired by a desire to protect a person's integrity, individuality, and self-determination.[683] By way of example, the widely-cited privacy case of *von Hannover v Germany No. 2*[684] concerned Princess Caroline of Hanover, the eldest daughter of Prince Rainier III of Monaco. Princess Caroline had long attempted to keep photographs of herself out of the German press. When photographs of her and her family were published without her consent, the matter ultimately landed before the European Court of Human Rights. In its judgment, the court declared that "a person's image constitutes one of the chief attributes of his or her personality, as it reveals the person's unique characteristics and distinguishes the person from his or her peers. The right to the protection of one's image is thus one of the essential components of personal development."

---

[681] Buitelaar, J.C., *Post-mortem privacy and informational self-determination*, Ethics and Information Technology, 19(2), pp.129–142, https://link.springer.com/article/10.1007/s10676-017-9421-9.

[682] Geneva Convention preamble, Convention (IV) relative to the Protection of Civilian Persons in Time of War. Geneva, 12 August 1949, https://ihl-databases.icrc.org/ihl/385ec082b509e76c41256739003e636d/6756482d86146898c125641e004aa3c5#:~:text=(1)%20Persons%20taking%20no%20active,on%20race%2C%20colour%2C%20religion%20or.

[683] Abraham, K. and White E., "The Puzzle of the Dignitary Torts", *Cornell Law Review* 104 (2), pp.317–380, www.core.ac.uk/download/pdf/228302795.pdf.

[684] Von Hannover v. Germany (no. 2) 40660/08 [2012] European Court of Human Rights 228.

Obviously, the Internet makes it remarkably simple to publish audiovisual content that has the potential to embarrass and humiliate. The discursive nature of social media also makes such content subject to exponential exposure and further ridicule. In addition to damaging the depicted individual's self-esteem and mental health, such videos could have considerable knock-on effects for his or her co-stars and corporate partners. Although still relatively nascent in legal scholarship on the subject, arguments for legal solutions centred on dignity embrace a more holistic understanding of identity. In due course, legal recognition of this facet of personhood may suitably modernise personality rights, and thereby protect people from the harms of misused technologies.[685]

Somewhat frustratingly, however, during court battles where a person's honour and peace of mind is at issue, the concept of dignity is often used as a mere placeholder to express an abstract ideal. Whilst the text of the law may emphasise the importance of a person's dignity, we frequently see academics and practising lawyers struggle to define cohesive rules for how such a right should operate in reality. The resolution of many cases, including *von Hannover*, ultimately defers to privacy or defamation laws. As explained above, these will often require the harmed individual to overcome the publisher's rights of expression.

There is, however, a faint hope for those in the performing arts. Whereas copyright seeks to protect the interests of a work's author or creator, neighbouring rights respect the necessary input made by a musician, dancer or actor in a recorded performance. These neighbouring rights are so named as they "neighbour" the concept of copyright (they may also be called "performer's rights" to avoid confusion with "authors' rights"). Neighbouring rights generally seek to secure credit for the performer, as well as to prohibit modifications to a recording which may damage the performer's intellectual or personal interests. In essence, neighbouring rights address problems arising from pirated copies or broadcasts of shows which do not adequately compensate the performers involved.

Despite their obvious importance with respect to performances shown through an audiovisual medium, neighbouring rights are very rarely discussed in the context of personality rights. Usually, personality rights disputes concern the misuse of static photographs, which can be easily altered to realistic effect, rather than a recorded performance. But as we have seen, advances in artificial intelligence and VFX have made manipulated videos more likely to be confused or implied as genuine. This new era of creating audiovisual works that imitate true performances may warrant an update to how neighbouring rights legislation has historically defined 'performance'.

Recognition of non-consensual virtual performances has precedence in California, where courts have compensated individuals for their digital 'enslavements' in video games. Musicians in the band No Doubt, for example, successfully sued the makers of Band Hero, a game which allowed players to select highly realistic digital avatars to

[685] Dunn S., *Identity Manipulation: Responding to advances in artificial intelligence and robotics*, presented as a working paper at the WeRobot Conference April 2020, and provided to Kelsey Farish through private correspondence in July 2020 with Ms Dunn's kind permission to cite.

simulate performing in a rock band.[686] If deepfakes or ghost acting performances are recognised as 'performances' under the law, this may protect an individual's image in cases where he or she has been forced to become a digitised performer against his or her will. In the interests of completeness, therefore, neighbouring rights should be considered together with any other rights of publicity or privacy that may comprise the personality rights framework.

7.4. Laws in selected jurisdictions

By way of summary, as explained in section 2, personality rights are often split into two broad concepts. The first concerns the exploitation of one's fame and reputation for financial gain, and is regarded as the right of publicity. We can view the right of publicity from several related but distinct angles, namely, intellectual property rights (angle 1, above), and brand protection laws (angle 2, above). Separate from commercial publicity, the second concept seeks to protect one's personal life by prohibiting the publication of certain images by way of privacy laws (angle 3, above). Dignity as an aspect of personality has received less attention in personality rights scholarship, and neighbouring rights even less so (angle 4, above). Whether or not any of these rights extend beyond death varies, but post-mortem rights will typically be viewed as a form of property which an individual may pass down to his or her heirs.

It remains to be seen just how future litigation concerning deepfakes and ghost acting will play out. Arguably, for the selected jurisdictions in this section 3 at least, laws concerning brand recognition and reputation will likely be most relevant. Rights to dignity and neighbouring rights may also be applicable, and could potentially offer a more modernised approach to protecting one's *persona*. On the other hand, intellectual property and privacy rights are perhaps less important when considering manipulated images of individuals appearing in audiovisual content. In any event, the publicity, privacy and dignity of an individual must be analysed on a case-by-case basis, and will always involve a balancing exercise against the competing rights of others, freedom of expression.

7.4.1. Germany

Home to Europe's largest audiovisual market, Germany's media sector was worth some EUR 23 billion in 2018[687] and employed more than 520 000 media workers as of 2017.[688] Recognition for personality rights is broad, and rooted in the German Constitution. As a

[686] *No Doubt v. Activision Publishing, Inc*., 122 Cal.Rptr.3d 397 (Cal. Ct. App. 2001).
[687] European Audiovisual Observatory, *Yearbook 2019/2020 Key Trends – Cinema, television, video, and audiovisual services on demand – The pan-European landscape. European Audiovisual Observatory*, http://yearbook.obs.coe.int/s/document/key-trends/2019.
[688] Weidenbach B., *Beschäftigte in der Medienbranche in Deutschland 2017*, Statista.

near-perfect example of the dignity protections discussed under angle 4 above, the German Constitution provides that "human dignity shall be inviolable" and "every person shall have the right to free development of his personality" (Grundgesetz, Articles 1 and 2).[689] These fundamental protections are complemented by the civil code, under which one who commits unlawful injury to another's right of personality is liable for compensation (BGB, §823).[690] In Germany, it is possible for one's dignity to survive beyond their natural life, so close relatives may protect a deceased person against disrepute of his or her life image.[691]

All aspects of a person's identity fall within Germany's unified right of personality, but images and photographs of an individual are subject to heightened protections. The German Copyright Act provides that where images depict an identifiable individual, they "may only be distributed or publicly displayed with the consent of the person depicted" (UrhG, §19a and §22).[692] Recognisability includes not only depictions of the face, but other distinguishing characteristics, as well as cartoons and even *doppelgängers* or lookalikes. Importantly, the concept of distribution is interpreted widely, and thus even the casual photographer would technically require consent before posting an image of another person to social media, or sharing it with friends through text messaging apps. Consent is likewise required from a performer before broadcasting or otherwise communicating their recorded performance (Section III, UrhG).

Consent is however not required when depicting persons of contemporary history, nor for pictures in which the people appear only incidentally (as "accessories"), nor where the publication serves a greater interest in art and culture (KunstUrhG, §23).[693] However, this balancing exercise has shifted towards protecting the image rightsholder in recent years.[694] As for particularly intimate images, for example those depicting someone in their home, the Criminal Code criminalises the taking, transmission, and use of photographs

[689] Grundgesetz, formally Grundgesetz für die Bundesrepublik Deutschland [Basic Law for the Federal Republic of Germany]. English translation used www.gesetze-im-internet.de/englisch_gg/englisch_gg.html.

[690] BGB, formally Bürgerliches Gesetzbuch [German Civil Code] in the version promulgated on 2 January 2002 (Federal Law Gazette page 42, 2909; 2003 page 738) last amended by Article 4 para. 5 of the Act of 1 October 2013. English translation used www.gesetze-im-internet.de/englisch_bgb/englisch_bgb.html.

[691] Seyfert C., *Regional Court Frankfurt am Main: Postmortem personality right of a Holocaust survivor prevails over publications by a British history professor (Case 2-03 O 306/19)*, Zeller & Seyfert Rechtsanwältern, www.zellerseyfert.com/en/litigationblog-detail/items/regional-court-frankfurt-am-main-postmortem-personality-right-of-a-holocaust-survivor-prevails-over-publications-by-a-british-hi.html.

[692] UrhG, formally Urheberrechtsgesetz [Act on Copyright and Related Rights] Copyright Act of 9 September 1965 (Federal Law Gazette I, p. 1273), as last amended by Article 1 of the Act of 28 November 2018 (Federal Law Gazette I, p. 2014). English translation used www.gesetze-im-internet.de/englisch_urhg/englisch_urhg.html.

[693] KunstUrhG, formally Gesetz betreffend das Urheberrecht an Werken der bildenden Künste und der Photographie [Law on the copyright in works of the fine arts and photography] Law of January 9th, 1907 (RGBl. I p. 7) last amended by the Act of February 16, 2001 (BGBl. I, p. 266), 1 August 2001, www.gesetze-im-internet.de/kunsturhg/__23.html.

[694] Coors C., "Image Rights of Celebrities vs. Public Interest – Striking the Right Balance Under German Law", *Journal of Intellectual Property Law & Practice* (2014) 9 (10): 835-840, www.ssrn.com/abstract=2738514.

which violate another's intimate privacy, punishable by up to two years' imprisonment or a fine (StGB, §201a).[695]

Regarding the publicity aspect of personality rights, German courts are increasingly willing to defend individuals against unwanted commercial exploitation of their image.[696] Unlike the approach seen in the United Kingdom, prior commercialisation of one's *persona* is not expressly a precondition for having a protectable right of publicity. In cases where a deceased person is shown, consent from the subject's relatives is required for a period of 10 years following their death. In a judgment concerning photographs of screen star Marlene Dietrich (who had died several years prior) the Federal Court of Justice held that in any unauthorised exploitation of a picture, the owner of the personality right is entitled to compensation irrespective of the gravity of the infringement.[697]

7.4.2. France

Widely regarded as the birthplace of cinema, France has by many metrics the most productive film industry in Europe: nearly 300 films or more have been made in France each year since 2015.[698] French personality rights, which may literally be translated as the rights of (or to) one's image, include privacy laws which aim to protect a person from unwanted exposure, as well as commercial rights to allow such images to be exploited as a marketable asset.[699] As a general rule, before the image of any individual is communicated to the public, consent must be obtained from the person shown. As elsewhere, France defines image widely to cover an individual's likeness, voice, photograph, portrait, or video reproduction. French courts have confirmed blurring the face of a model may not in itself resolve an image rights violation, where other parts of the body are still visible.[700]

This philosophy is largely rooted in France's strong protections for one's privacy or intimate family life. The Civil Code states everyone has the right to respect for their private life, and importantly, empowers French courts to utilise any measures that are appropriate to prevent or end an invasion of the intimacy of private life (Code civil,

[695] StGB, formally Strafgesetzbuch [Criminal Code] in the version promulgated on 13 November 1998 (Federal Law Gazette I p. 3322) last amended by Article 3 of the Act of 2 October 2009 (Federal Law Gazette I p. 3214). English translation used www.gesetze-im-internet.de/englisch_stgb/.
[696] Peters M., *The Media and Entertainment Law Review – Germany*, The Law Reviews, www.thelawreviews.co.uk/edition/the-media-and-entertainment-law-review-edition-1/1211744/germany.
[697] Marlene Dietrich, BGH 1 December 1999 - 1 ZR 49/97 - Kammergericht LG Berlin.
[698] Lemercier F., *301 feature films produced by France in 2019*, Cineuropa - the best of European cinema, https://cineuropa.org/en/newsdetail/387425/.
[699] Logeais E. and Schroeder J-B., "The French Right of Image: An Ambiguous Concept Protecting the Human Persona", *18 Loyola University Entertainment Law Review* 511, https://digitalcommons.lmu.edu/cgi/viewcontent.cgi?article=1366&context=elr&httpsredir=1&referer=.
[700] Mr X v Umanlife, TGI de Paris, judgment of 16 November 2018.

Article 9).[701] Accordingly, French judges have over time progressively developed stronger personality rights for claimants on a case-by-case basis.[702] The quality of consent will be a dispositive issue, and is not always clear-cut. Recently, the Paris Tribunal found that a professional model participating in a photo shoot for advertising purposes did not imply her consent to all forms of commercial exploitation.[703]

Exceptions to the consent requirement include where footage is captured in a public place, or where images depict well-known figures partaking in official duties or activities otherwise connected with their notoriety, subject to the person's right to dignity. Parody and strictly private use are also exceptions. French courts may also award enhanced damages in cases where a celebrity has avoided endorsements or other commercial exploitations of his or her image in the past. The Penal Code also establishes a criminal offence for privacy violations, which carries a sanction of one year's imprisonment and a fine of EUR 45 000 (Code penal, Article 226-1).[704] Beyond the purview of privacy law, the French Intellectual Property Code provides that a performing artist shall enjoy the right to respect of his or her name, quality and performance (IPC, Article 212-2).[705] As this is a perpetual right attached to the individual, it may be passed down to his or her heirs so that they may protect the artist's performance and memory.

7.4.3. Sweden

For a country with a relatively small population, Sweden has produced a remarkable number of international film and television stars, notably including Greta Garbo, Ingrid Bergman, Max von Sydow, Stellan Skarsgård, and Alicia Vikander. In addition to being a hotspot for the Scandi Noir genre, the country has created several programmes that have gone on to achieve wide acclaim in other markets, such as the *Wallander, The Bridge,* and the *Girl With The Dragon Tattoo* franchises. That the country has no separate personality right as such may therefore come as a surprise.

In stark contrast to Germany and France, personality rights are essentially omitted from Swedish law.[706] In cases where the news media has misappropriated one's image, the Freedom of Press Act is applicable, but Sweden tends to emphasise freedom of expression over an individual's right to privacy. This approach is somewhat unusual when compared

[701] Code civil [French Civil Code], www.legifrance.gouv.fr/
[702] Sullivan C. L. and Stalla-Bourdillon S., *Digital Identity and French Personality Rights – A Way Forward in Recognizing and Protecting an Individual's Rights in His/Her Digital Identity*, Computer Law & Security Review (2015), www.ssrn.com/abstract=2584427.
[703] *Mrs X v SARL Denim*, TGI de Paris, 17th chamber, judgment of 21 November 2018.
[704] Code penal [French Criminal Code], www.legifrance.gouv.fr/
[705] IPC, formally Code de la propriété intellectuelle [French Intellectual Property Code], www.legifrance.gouv.fr/.
[706] Ondreasova E., "Personality Rights in Different European Legal Systems: Privacy, Dignity, Honour and Reputation", The Legal Protection of Personality Rights, pp.24–70. Brill | Nijhoff, https://brill.com/view/book/edcoll/9789004351714/B9789004351714_004.xml.

to many other European countries, and is perhaps attributable to the fact that Swedish law has deep reverence for transparency and openness, and therefore also for public access to information. This is evidenced by the controversial *Utgivningsbevis*, a sort of publishing licence which permits companies to openly publish personal details about individuals.[707] Similarly, a person's taxable income is also a matter of public record.[708] Somewhat confusingly, the "principle of publicity" in Sweden refers not to the publicity aspects of personality rights, but rather the principle that government documents are largely unclassified and subject to public scrutiny.

Notwithstanding the above, Sweden's Act on Names and Pictures in Advertising mandates that consent must be obtained before using someone's name or picture, or a representation which clearly indicates that specific person, for commercial purposes (*Lag om namn och bild i reklam* (1978:800)).[709] The Swedish Tort Liability Act also states that non-financial harms may be compensated, but only if the violation constitutes a criminal act or otherwise endangers health and life [*Skadeståndslag* (1972:207), Chapter 2].[710] While defamation is a criminal offence in Sweden under the Criminal Code (*Brottsbalken*)[711] it occurs only where someone falsely or without "reasonable grounds" accuses another of "being a criminal or of having a reprehensible way of living, or otherwise furnishes information intended to cause exposure to the disrespect of others" (*Brottsbalken*, Chapter 5 §1). The Criminal Code also permits family members or the public prosecutor to initiate prosecutions for "disturbing the peace" of the deceased, if doing so is in the public interest (*Brottsbalken*, Chapter 5 §4). These narrow protections aside, the gaps in Sweden's personality rights legislation are palpable, and judges in Sweden have been largely unwilling to fill them through case law.[712]

7.4.4. Guernsey

The Bailiwick of Guernsey is an island off the coast of northern France and, while a British Crown dependency, has a legal system distinct from that of the United Kingdom. Although home to only 63 000 inhabitants, its favourable corporate tax treatments, scenic environs and English-speaking population have contributed to its burgeoning film production and

[707] Herlin-Karnell E., *Corona and the Absence of a Real Constitutional Debate in Sweden*, Verfassungsblog on Matters Constitutional, www.verfassungsblog.de/corona-and-the-absence-of-a-real-constitutional-debate-in-sweden/

[708] Marçal, K., *Sweden shows that pay transparency works*, The Financial Times, www.ft.com/content/2a9274be-72aa-11e7-93ff-99f383b09ff9.

[709] Lag om namn och bild i reklam (1978:800) [The Act On Names And Pictures In Advertising]. English translation used www.kb.se/Dokument/Bibliotek/biblioteksjuridik/names_pictures.pdf.

[710] Skadeståndslag (1972:207) [Swedish Tort Liability Act]. Available at www.riksdagen.se/sv/dokument-lagar/dokument/svensk-forfattningssamling/skadestandslag-1972207_sfs-1972-207.

[711] Brottsbalken, SFS 1962:700 [Swedish Criminal Code] English translation used www.legislationline.org/download/id/1700/file/4c405aed10fb48cc256dd3732d76.pdf.

[712] Ondreasova E., *op.cit.*

film finance sector.[713] Interestingly for our purposes, Guernsey established the world's first statutory registration regime for personality under its Image Rights Ordinance 2012.[714] Broadly speaking, this legislation functions similarly to other laws concerning trademarks.

Protected images may include photos and pictures of the individual, but also film footage, as well as his or her name, voice, signature, likeness, mannerisms and personal attributes, such as a sports jersey number. Living persons, or persons who have died within 100 years of the application, as well as groups and teams or even fictional characters, may be registered. Applicants must be proprietors of the personality in question, and may therefore be the individual, or his or her authorised representatives or heirs.

An infringement occurs if the registered protected image (or one similar to it) is used for a commercial or financial benefit without the proprietor's consent, and such use either confuses the public or damages the reputation of the person depicted. When assessing damages, the court will consider all relevant factors, to include any economic consequences and lost profits, as well as any moral prejudice suffered by the victim. Registrations are valid for an initial 10-year period, and may be renewed for another 10. Although there is no requirement that the applicant be established or resident in Guernsey, the enforcement will only concern infringements or misappropriation in Guernsey. However, the law has been designed with modern media and cross-border digital services in mind[715] and so clearance searches prior to broadcasts made available on the island, regardless of where based, would be prudent.

7.4.5. United Kingdom

At the time of writing, the current government of the United Kingdom has emphasised the importance of separating the country from the political and economic ecosystem of the European Union. Despite Brexit however, UK television shows and films remain an inexorable aspect of the audiovisual market in wider Europe, and are enjoyed by audiences throughout the continent. This notwithstanding, the United Kingdom, like Sweden, does not formally recognise personality rights in its legislation. And unlike many of its counterparts in Europe, English courts have also resisted any temptation to recognise such rights through case law.

To quote the judgment from Rihanna's case mentioned above: "There is today in England no such thing as a free-standing general right by a famous person (or anyone else) to control the reproduction of their image." Rather than attempt to fashion some discrete personality right, judges in the United Kingdom normally prefer to rely on the

[713] Tustin B., *Guernsey Film: More Than You Might Think*, Mondaq, www.mondaq.com/guernsey/film-television/171220/guernsey-film-more-than-you-might-think.
[714] Image Rights (Bailiwick of Guernsey) Ordinance, 2012.
[715] Shires S (2015), *Guernsey image rights exposed*, Lexology, www.lexology.com/library/detail.aspx?g=9829178f-e93d-4ead-94d0-bfbfaa81f8b7.

more traditional legal tools at their disposal. In this sense, there are several ways in which an individual can protect his or her image which typically focus on disclosure of private facts, and harm to one's business interests or reputation.

Due to the nature of how the common law in the United Kingdom has developed, and the fact that the United Kingdom has no codified written Constitution, privacy is not regarded as a distinct right as such. Instead, judges respect the fact that privacy is an important social value, which in turn underpins specific laws related to the breach of confidence or the misuse of particularly sensitive information. From the late 1990s, courts have regarded the privacy protections available under the UK's Human Rights Act 1998, which mirrors the European Convention on Human Rights. That said, on the relatively rare occasions when privacy matters pertaining to images have come before English courts, judges have tended to favour the protection of free speech and other press freedoms. There are notable exceptions for images which constitute confidential information, or depict children or particularly intimate scenes.

When protecting one's publicity as a commercial asset, the most relevant option is often the tort of 'passing off', but success will depend on whether the celebrity has "a significant reputation or goodwill" in the first instance.[716] From the dignity and reputational perspective, unauthorised use of a person's image may give rise to the common law offence of malicious falsehood, but only insofar as it contains false words which result in quantifiable monetary loss.[717] Falsehoods that damage an identifiable individual's reputation may constitute defamation, but following reforms to the Defamation Act 2013,[718] only where this causes "serious harm" to the individual depicted.

The above should be read also in the context of the Copyright, Designs and Patents Act (CDPA)[719] which allows a performer in some circumstances "to object to derogatory treatment of performance, with any distortion, mutilation or other modification that is prejudicial to the reputation of the performer" (CDPA, s. 205). Before using a recorded performance, consent must be obtained from the actor, musician, dancer, or other performer in question (CDPA, s. 182). This statute was inspired by a case concerning clips featuring the actor Peter Sellers which were used after his death to make a new Pink Panther film. His personal representatives then successfully argued for Sellers' post-mortem right to prevent the unauthorised use of his performances.[720]

7.4.6. California

There is no federal right to privacy in the United States of America: privacy protections are regulated by reference to specific sectors or topics, such as financial or healthcare

[716] Edmund 'Eddie' Irvine v Talksport [2002] EWHC 367 (Ch).
[717] Marathon Mutual Ltd v Waters [2009] EWHC 1931 (QB)).
[718] Defamation Act 2013, www.legislation.gov.uk/ukpga/2013/26/contents/enacted.
[719] Copyright, Designs and Patents Act 1988, www.legislation.gov.uk/ukpga/1988/48/contents.
[720] *Rickless v. United Artists Corporation* [1988] QB 40.

information. Although there are federal intellectual copyright statutes and codes concerning unfair competition and advertising, it is largely the laws of a particular state which will be of key importance to protect one's publicity and privacy. As home to both Hollywood and Silicon Valley, perhaps no place better reflects the *zeitgeist* of celebrity and technology than California.

California protects publicity both in its civil codes as well as at common law, thanks in part to its world-renowned entertainment industry and the Californian propensity to embrace and respect emotions. The California Civil Code (CIV)[721] prohibits the unauthorised usage of another's name, voice, signature, photograph or likeness for advertising purposes without their consent (CIV §3344). A guilty defendant will be ordered to pay the injured party the greater of either USD 750 (approx. EUR 635 as of August 2020) or the actual damages suffered, which will include a disgorgement of profits. The Fred Astaire Celebrity Image Protection Act later amended §3344 to protect the commercial use of a deceased individual's name, image or voice for 70 years post-mortem.[722] However, this section applies to merchandise, advertisements, and endorsements only, and exempts fictional and nonfictional entertainment, dramatic, literary, and musical works from liability. The common law publicity right covers any misuse of an individual's identity, which is broader than the specific characteristics listed in the statute, for the "defendant's advantage, commercially or otherwise".[723]

California courts recognise that the right of a person to be free from unauthorised and unwarranted publicity is an aspect of privacy.[724] Several privacy laws are applicable to the misuse of one's image, and each is distinguished based on whether the harm is economic or dignitary in nature. False content that injures a person's reputation may fall under the tort of defamation, and content that is not technically false but nevertheless harms the victim's mental or emotional well-being may constitute the tort of false light.[725] Due to having residents who are frequently under the spotlight (both metaphorically, and literally), the state is regarded as one of the most claimant-friendly jurisdictions in which to bring a personality rights infringement case.

It would however be remiss to ignore the fact that the competing free speech protections available through the First Amendment are strong, even for images and

---

[721] CIV (California Civil Code), www.codes.findlaw.com/ca/civil-code/.

[722] Fred Astaire Celebrity Image Protection Act, see §3344.1 of the California Civil Code.

[723] Clint Eastwood v. National Enquirer, Superior Court, 149 Cal.App.3d 409 (Cal. Ct. App. 1983).

[724] Fairfield v. American Photocopy Equipment Co., (1955) 138 Cal. App. 2d 82, 291 P.2d 194

[725] In common law jurisdictions, if a publication is false and harms a person's reputation, defamation might have occurred. In some U.S. states including California, however, there is a separate tort (legal harm) which covers communications which are not technically false, but are nevertheless misleading. To this point, courts in California recognise that false *implications* can lead to false and ultimately harmful impressions about an individual. This distinction is subtle and nuanced, but as a generalisation, defamation seeks to remedy damage to a person's reputation. Truth is a defence to defamation; in other words, where a statement is true, it is not defamation. Conversely, a factually accurate statement or photograph can "place someone under a false light". California's tort of false light therefore seeks to address damage to a person's feelings – irrespective of its veracity. CACI (Judicial Council of California Civil Jury Instructions) 2017 edition. No. 1802. False Light. https://www.justia.com/trials-litigation/docs/caci/1800/1802/.

videos published by corporations. As a notable example, the actor Dustin Hoffman sued a magazine over a digitally-manipulated image used in a fashion story, which purported to show him dressed in designer clothes. The courts found that as the magazine had no intent to commit harm and the image itself was not purely for commercial reasons, the publisher was entitled to free speech.[726]

Persons of notoriety, including individuals who become involved in events worthy of public interest, will on balance have fewer privacy protections than an ordinary member of the public. Importantly for those seeking to alter the appearance of an actor through ghost acting or similar techniques, California uses a transformative work test to determine whether a use of a person's image is protected by the First Amendment. Under this test, the more a new work 'transforms' original footage to provide a different meaning or message, the more likely it is that it will be exempt from copyright protections (Copyright Code, §107).[727] With respect to images of individuals in particular, California courts will consider whether the "celebrity's likeness is so transformed that it has become primarily the defendant's own expression rather than the celebrity's likeness."[728]

7.5. What next for Europe's audiovisual sector?

This chapter has sought to establish the multifaceted nature of personality rights, which most typically comprises various rights of publicity and privacy. There is also scope to incorporate rights to dignity and integrity, together with a performer's neighbouring rights, into this framework. But regardless of how such laws are theoretically classified, there are some practical tips and observations which may assist in mitigating risks associated with deepfakes and ghost acting performances.

Some jurisdictions have passed or proposed deepfake-specific laws to address image-based sexual abuse or electoral interference. For example, the U.S. state of Virginia became the first to update its so-called 'revenge porn' laws, making it a misdemeanour to publish manipulated photos or videos which depict someone nude or their genitalia without consent.[729] Texas was the first state to criminalise the sharing of deepfake videos made with intent to injure a political candidate in the days leading up to a vote.[730] Of course, deepfakes can be deeply damaging without being sexual or political in nature.

---

[726] Dustin Hoffman v. Capital Cities/ABC, Inc., 255 F. 3d 1180 (9th Cir. 2001).
[727] Copyright Code, formally U.S. Code, Title 17 Copyrights, §107 Limitations on exclusive rights: Fair use, www.law.cornell.edu/uscode/text/17/107.
[728] Comedy III Prods., Inc. v. Gary Saderup, Inc. - 25 Cal. 4th 387, 106 Cal. Rptr. 2d 126, 21 P.3d 797 (2001).
[729] Virginia House Bill 2678, formally a Bill to amend and reenact § 18.2-386.2 of the Code of Virginia, relating to unlawful dissemination or sale of images of another; falsely created videographic or still image, https://law.lis.virginia.gov/vacode/title18.2/chapter8/section18.2-386.2/#:~:text=Any%20person%20who%2C%20with%20the,or%20female%20breast%2C%20where%20such.
[730] Texas Senate Bill 751, formally a bill Relating to the creation of a criminal offense for fabricating a deceptive video with intent to influence the outcome of an election, www.legiscan.com/TX/text/SB751/2019.

In lieu of statutory instruments, big tech companies have attempted to regulate deepfakes appearing within their remit. Facebook and its sister company Instagram have officially 'banned' deepfakes, as have Twitter, Reddit and Pornhub. Despite having once embraced the deepfake trend, Tiktok has also recently 'banned' deepfakes following takeover discussions with U.S. corporations.[731] The use of inverted commas around the word banned is intentional: regardless of any official statements or policies, removal or moderation of deepfakes is exceptionally difficult, meaning the prohibition on deepfakes is often in name only.

Firstly, distinguishing a deepfake from an authentic video is a considerable challenge in and of itself, as evidenced by the Deepfake Detection Challenge led by Facebook, Microsoft, and a small army of artificial intelligence experts.[732] Some efforts include meta-tagging source images and watermarking, but no widely available solution exists as of yet. Secondly, even if a deepfake were itself detected, the ease and desirability of remaining anonymous online make the discovery of a deepfake's creator a potentially impossible feat. In addition to these clear technical issues, we must also consider the complications surrounding context and intention. The term 'deepfake' speaks to the method of production or its face-swapping characteristics, and not its substance: the medical and educational fields are two areas in which the technology has been used admirably for social benefit.[733] In short, there is not always a bright line separating an unwanted deepfake from one which is acceptable.

Leaving the specifics of how 'bad' deepfakes could be separated from the 'good', novel and entertaining content has an incredible propensity to go viral, even if fake or misleading. Though a somewhat cynical view to take, platforms which depend on advertising clicks and page views may find it suits their business interests to keep exciting content online. Furthermore, even if a deepfake is deemed suitable for removal, doing so may have a chilling effect on freedom of expression. Finally, even where a deepfake has been removed from a platform, this remedy may be superficial, as the video could still very easily appear elsewhere. What's more, the so-called Streisand effect[734] has shown that attempting to suppress information may unintentionally make it more widespread.

---

[731] Statt N., *TikTok is banning deepfakes to better protect against misinformation*, The Verge, www.theverge.com/2020/8/5/21354829/tiktok-deepfakes-ban-misinformation-us-2020-election-interference.
[732] Wiggers K., *Facebook, Microsoft, and others launch Deepfake Detection Challenge*. VentureBeat.
[733] Kalmykov M., *Positive Applications for Deepfake Technology*, Hackernoon.com, www.hackernoon.com/the-light-side-of-deepfakes-how-the-technology-can-be-used-for-good-4hr32pp.
[734] The Streisand effect is named for the 2003 lawsuit launched by American actress and singer Barbra Streisand against a photographer, Kenneth Adelman. Streisand sought to suppress the publication of the photographs Adelman had taken of her house, located on the Californian beaches of Malibu. However, Adelman's photography was for the California Coastal Records Project, a scientific coastal erosion research project which provides pictures of the California coast for study. Her lawsuit was ultimately dismissed and, as a result of the publicity garnered over the lawsuit, led to even more interest in the photographs of her home. See Cacciottolo M., *The Streisand Effect: When censorship backfires*, BBC News, https://www.bbc.co.uk/news/uk-18458567.

When audiovisual content is published online, it may be difficult to protect it from being used in ways not originally intended, or otherwise misappropriated by third parties. Nevertheless, it is still possible for private parties to contractually agree as to how personality rights are to be exploited or protected. Performers often enter into new contracts for each separate project undertaken, and setting clear parameters from the outset can help avoid disputes later. Legal jargon need not be used, but it is prudent to specify in writing the ways in which a person's image may be used, shared, and transferred to others.

Both parties would be well-served to carefully consider how the content should be used beyond the scope of the original production, as well as any provisions for additional remuneration. Where a studio uses footage from an earlier performance and then repurposes it, additional payments to the performer for re-use of footage are unlikely unless specifically stipulated in a contract. From a reputational perspective, it may also be appropriate to include anti-disparagement clauses or so-called morality clauses, both of which seek to prevent one party from injuring the reputation of the other. For the actor, this will be especially important in jurisdictions such as the United Kingdom, Sweden, and the United States, as in those places defamation and harm to dignity are notoriously difficult to establish.

From a purely financial perspective, unwanted modification of appearances may harm or even destroy the working relationship between an actor and those working behind the camera. Creative differences and demands, whether reasonable or not, have been known to derail or even sink productions. The loss or recasting of a star midway through shooting can cost a studio thousands, or even millions, of euros. Substantial modifications beyond artistic necessity should therefore be discussed in advance, especially where such alterations have the potential to injure the feelings or interests of the actor.

As a final point, as exciting and innovative as it may be to take someone's image to create a digital double or ghost acting performance, doing so carries unavoidable risk. When we sit down to enjoy a film or television show we are, to paraphrase Ingmar Bergman, consciously priming ourselves for illusion. We put aside will and intellect to make way for a fictional narrative to unfold in our imagination.[735] But the recent verisimilitude of digitally created faces is something altogether different, because it has the potential to remove the autonomy and self-determination of the actors concerned. It also has the tendency to manipulate the evaluative or decision-making processes of the audience. When individuals are falsely depicted in non-fictional videos, including documentaries, interviews and advertisements, the risk of financial, reputational and societal harm is even more palpable. It stands to reason that as artificial intelligence becomes more sophisticated, the distinction between authentic performances and those digitally generated will blur. If the human eye is unable to discern the difference, perhaps

[735] Bergman I., *Four Screenplays*, Secker & Warburg, London. Translated from the Swedish by L. Malmstrom and D. Kushner.

the law will likewise cease to distinguish the two, and thus extend rights of personality to cover one's virtual self.

If only one thing is remembered from this chapter, let it be that personality rights require a careful consideration of situational context. Where such rights are litigated, courts are entrusted to interpret rather than to create the law, and thus merely affirm the applicability of established rules which have evolved from that society's customs. Regardless of one's legal training or authority, it is no straightforward task to deliberate upon questions of art, truth, expression and identity – each of which speaks to the core of what it means to be human. When asked about personality rights exploitation and protection, especially in the case of novel technologies such as deepfakes and ghost acting, it appears that the typical lawyer's answer must suffice, at least for now: it depends.

Regulating AI

The "consideration of situational context" mentioned by Kelsey Farish in her contribution to this publication should probably be extended to all the legal issues dealt with in these pages. And there is an important fact (already mentioned in the foreword) to be recalled: computers will be computers, stupid machines that only know the difference between a one and a zero, and as such, the results of their soulless calculation efforts will depend on and/or require human intervention and oversight. And very often, human intervention means regulation. ***Atte Jääskeläinen*** *outlines in his contribution to this publication some principles that, in his view, should be applied to the regulation of AI. As observed before, transparency is the most fundamental principle here, since it "serves human needs to make sense of how the systems work and address responsibilities to the right persons". Jääskeläinen suggests, however, that we need to accept that "unknown risks may be impossible to regulate, at least if the regulation is based on the technology, not on goals". Moreover, regulation of AI should "reduce public risk without destroying creativity and innovation", and "unnecessary obstacles to using data to create well-being and doing good" should be removed.*

8. Approaches for a sustainable regulatory framework for audiovisual industries in Europe

Atte Jääskeläinen, LUT University, Finland, and LSE, London[736]

8.1. Introduction

A software programme detects the ball and the players in the picture of a football match and zooms in to where the action is. And since the cameras are extremely sharp, the 'virtual director', a computer relying on machine-learning algorithms, is able to produce a full match broadcast featuring multiple cameras angles – without a human touch. The aim of Dutch football association KNVB and Dutch media company Talpa is to broadcast 80 000 amateur football matches live yearly, with no people involved.

A software programme analyses your behaviour in social media and can detect your personality in the widely-used psychological model Big 5, attaining the same accuracy as the filling in of a questionnaire with information about some 200 Facebook likes. The result: a system that could be used to influence elections by targeting political messages for those most likely to be affected by a certain style of personalised advertising.

A software programme developed by regional Swedish publisher Mittmedia uses personalisation combined with journalistic target-setting with a clear approach to contextualising data and using machine learning. I think we will see total personalisation and automation of publishing within the near future says the company's chief technology officer.

A software programme developed by Swiss company Largo predicts from an early script what an audience's response will be, on a country by country basis. Based on this analysis,

[736] Atte Jääskeläinen is professor of practice at LUT University, Finland, and visiting senior fellow at LSE, London. He was director of news and current affairs of The Finnish Broadcasting Company 2006-2017 and CEO of The Finnish News Agency 2004-2006. He co-authored with Maike Olij the EBU News Report 2019 "The Next Newsroom. Unlocking the Power of AI for Public Service Journalism".

data scientists advise which parts of the script should be changed to increase the market appeal of a film and maximise revenues generated.

All these real-life examples illustrate how artificial intelligence is already affecting the audiovisual industry or its environment in a strategic way: audiovisual production may be largely automated; audience behaviour may be analysed like never before; this information can be used to both optimise the value customers get from the market, but also to guide creative and journalistic work and use it in democratic processes - sometimes in a way in which democracy was not supposed to work.

Some have predicted that this evolution leads to a doomsday for humankind, if not properly regulated. As the Future Today Institute's Amu Webb wrote in 2019: "The lack of nuance is one part of AI's genesis problem: some dramatically overestimate the applicability of AI in their workplaces, while others argue it will become an unstoppable weapon."[737]

Regulation of AI, especially regulation of AI in areas that are relevant to the audiovisual industry, is a complex question with no clear answers. That complexity affects decision-making in the field. For both authorities and politicians, the field of AI is hard to get a grip on, as the whole concept of AI is unclear and the target is moving fast: just when you think you 'get' it, new forms have already emerged. Regulation is a reactive process and in a fast-changing area it often becomes obsolete before it can be applied.[738]

Some of the examples of AI generate the feeling that computers are creating something almost magical. Typically, this illusion is based simply on the use of smart mathematics with the help of huge computational resources available through cloud computing and fast telecommunications. What appears as creativity is just freedom of biases that limit human thinking combined with the ability to check and produce all alternatives, including those that humans would not consider even when called on to think out of the box.

The authors of a recent book about the "simple economics" of predictive AI explain: "All human activities can be described by five high-level components: data; prediction; judgement; action; and outcomes. As machine intelligence improves, the value of human prediction skills will decrease because machine prediction will provide a cheaper and better substitute. However, the value of human judgement skills will increase."[739]

The OECD has estimated that almost half of all professions will either disappear or fundamentally change within 15-20 years because of automation and new self-learning

[737] Webb A., "The big nine before it's too late", WMG Weekly, https://www.wmgweekly.com/post/2019/06/08/the-big-ninebefore-it-s-too-late.
[738] Petit N., "Law and regulation of artificial intelligence and robots - Conceptual framework and normative implications", https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2931339.
[739] Agrawal A., Gans J. and Goldfarb A., "Prediction machines: The simple economics of artificial intelligence", Harvard Business Press.

technologies.[740] From a practical point of view, it appears that the challenge is aligning processes in such a way that the strength of both machines and humans can be used in the best possible manner.

Technologies don't exist in isolation from culture and values. Some academics see increased automation as leading to a major redefinition of values in closer connection with computing.[741] Others argue that communication can no longer be defined as only a human-to-human phenomenon.[742] One fundamental observation is that we are shifting into a many-to-many communication era, in which billions and billions of connections exist between individuals, powered by automation, defining our understanding of the networked world we live in.[743]

This chapter seeks to clarify some of the ethical and moral issues that should be taken into account when designing effective and ethically sound regulation for this technology area, which is at the same time fascinating and, for some, frightening.

8.1.1. The basics of AI, simplified

The term "artificial intelligence" is not clear. One common definition is that AI describes machine processes that would require intelligence if performed by humans. The term was originally coined by John McCarthy, who began research into AI in the 1950s. He assumed that human learning and intelligence could be simulated by a machine. But in its most basic form, artificial intelligence is "a system that makes autonomous decisions, a branch of computer science in which computers are programmed to do things that normally require human intelligence".[744] One of the pioneers of artificial intelligence, Marvin Minsky, described artificial intelligence as a "suitcase term": there are many concepts packed inside.

Most applications of artificial intelligence today use technologies that fall into the domain of **machine learning**. With machine learning, computers learn from data without being explicitly programmed.

---

[740] OECD, The Future of Work. OECD Employment Outlook 2019. Highlights, https://www.oecd-ilibrary.org/employment/oecd-employment-outlook-2019_9ee00155-en.

[741] Coddington M. "Clarifying journalism's quantitative turn", Digital Journalism, Routledge, 3(3), pp. 331–348. Milosavljević M. and Vobič I., "'Our task is to demystify fears': Analysing newsroom management of automation in journalism", Journalism, SAGE Publications, https://journals.sagepub.com/doi/abs/10.1177/1464884919861598?journalCode=joua.

[742] Lewis S. C., Guzman A. L. and Schmidt T. R., "Automation, journalism, and human–machine communication: Rethinking roles and relationships of humans and machines in news" Digital Journalism, Routledge, 7(4), pp. 409–427.

[743] Jääskeläinen A. and Olij M., "The next newsroom: Unlocking the power of AI for public service journalism", European Broadcasting Union, https://www.ebu.ch/publications/news-report-2019.

[744] https://futuretodayinstitute.com/trend/artificial-intelligence/..

- In traditional programming, data are run on the computer, which produces the output. Programmers therefore have to know the rules if the desired output is to be achieved.
- In machine learning, data and examples of the desired output are first run on the computer, which then learns from them to create its own rules. These are then used to produce the output.

A very powerful and resource-consuming kind of machine learning, **deep learning**, uses algorithms called 'artificial neural networks', which are modelled on the way neural networks operate in the human brain. Advances in deep learning have made language technologies and image recognition much more sophisticated and therefore have opened up substantial opportunities in the audiovisual sector.

The key benefit of deep learning is that it is able to absorb huge amounts of data. This has allowed machine learning to accomplish tasks it never could have managed before. On the other hand, it also requires huge amounts of training data and expensive computational resources, so using it on a large scale has only recently become possible - and is still limited.

The different types of machine learning are based on how the machines use data to learn rules. The oldest kind of machine learning is called **supervised learning**. It uses a set of desired outcomes to train the computer. The algorithm then comes up with rules that will allow the computer to produce results similar to those of the training dataset. In **unsupervised learning**, the computer is used to group a huge dataset in a meaningful way. This approach does not require a set of training data, as the data set is typically clustered based on its internal logic.

One of the more recent models is **reinforcement learning**, in which the system learns on the fly from feedback it receives from its environment. This field is developing fast and is quite useful as these systems based on reinforcement learning can quickly adapt to new situations and learn from user behaviour – which can suddenly change, for any reason. For example, the most advanced recommendation systems use reinforcement learning as a way to monitor and adapt to the behaviour of users.

All three learning models can be combined in a complex system. An advanced system typically has multiple algorithms and can conjoin different approaches to achieve the desired outcome for a specific situation. For this kind of system, it is crucial to understand the problem, the context and the information contained in the data.

What is confusing about the different areas of artificial intelligence is that people often mix up the dimensions of AI. For example, machine learning often appears alongside image recognition and natural language processing in lists of key AI application areas. But machine learning is present in almost all modern applications of artificial intelligence, including medical diagnosis, self-driving cars, prediction systems and automatic classification.

Concerning the need to regulate AI, a very important categorisation is narrow or weak AI, in which the system performs a single task related to a specific problem. A much more difficult area of regulation is so-called artificial general intelligence (AGI), which is a

type of AI that can solve complex problems in any context and define its goals autonomously. While there are bold initiatives to reach this kind of autonomy of computers, like Google's DeepMind and OpenAI, this kind of AI has not yet been achieved. Most researchers believe that we are still decades away, at least, from reaching this level of AI.[745]

8.2. How is AI used in audiovisual industries?

By audiovisual industries, we traditionally mean the production and marketing of movies, television and the Internet's audiovisual content. However, as it is difficult to define clear boundaries for AI, it is not wise to limit our thinking only to the specifics of this particular sector. Delivering messages nowadays often happens with mixed methods - and these methods may even be autonomously selected, based on individual preferences, by data-driven AI systems. Therefore, the challenges of regulating AI in the audiovisual sector are strongly related to other sectors close to it. The key question is: what are we trying to achieve with regulation and how should it be applied in areas relevant to this sector. The range of potential applications is already diverse, and developing fast into areas that we can't even imagine yet. However, it would be neither wise nor effective to apply the same rules to automatic translations, self-driving cars, sensitive personal data and advanced camera technology, to mention just a few of the application areas.

How to get a grip on this issue then? First, we should try to increase our understanding of the consequences of different usages of AI in audiovisual industries. As they vary substantially, even a rough categorisation helps to understand the values that should be protected with regulation.

In the absence of a better and more sophisticated categorisation, one can use here a categorisation employed in the report for the European Broadcasting Union on how to use AI in public service journalism: "The Next Newsroom"[746], which could work as a basic framework to understand the strategic relevance of different types of AI technologies for the audiovisual industry.

First, artificial intelligence can be considered as a growing set of practical tools. This is AI at a purely operational level, aiming primarily to automate repetitive tasks and reduce costs.

For example, the solutions including AI systems for editing and media management tasks are numerous, and their adoption is increasing with substantial speed. Tools used for transcribing and translating languages, and for detecting specific material

[745] See e.g. Joshi N., et al. "How far are we from achieving artificial general intelligence?", https://www.forbes.com/sites/cognitiveworld/2019/06/10/how-far-are-we-from-achieving-artificial-general-intelligence/#5ebe24f06dc4 and
Fjelland R., "Why general artificial intelligence will not be realized", https://www.nature.com/articles/s41599-020-0494-4.

[746] Jääskeläinen A. and Olij M., *op.cit.*

in archives, make reusing material easier and faster, and change the value-creation logic in production. For example, Deutsche Welle uses AI-based language processing just to keep the newsroom on track with what is happening real-time in its news operations in multiple languages. Swiss Broadcasting is one of the companies that has developed advanced systems to detect persons and places in archived video footage[747].

Irish broadcaster RTE and Al Jazeera even collaborated to create a system that measures the airtime of politicians during election campaigns and flags content that might carry a regulatory problem. The system is based on an advanced method of detecting not only items and persons in the pictures, but also their context.[748]

One could reasonably ask: Is there anything special in these tools that requires specific regulation? Or is it simply enough that these tools be used in a responsible way. After all, these are just tools and technologies to help get the work done.

Second, AI allows the creation of a data-savvy culture that rests on defining and knowing your objectives and learning ways to measure and optimise, based on them. This also allows for a very strategic use of AI: targeting messages and optimising value for individual customers based on information about their preferences and behaviour.

The same type of optimisation and personalisation AI is also used to optimise the financial results of audiovisual operations, for example by creating more efficiency in marketing campaigns by testing the effectiveness of different messages or identifying interesting market clusters and opportunities.

One of the fundamental challenges in the era of abundance that digital technologies have created is finding good content amidst too much clutter. So there's a need to connect content with the audience that is interested in it. This kind of optimisation is not limited to online offerings. For example, Spain's RTVE has a promising research project on designing television scheduling using AI algorithms. They ask: which TV programs fit the taste for the audience at a specific time of the day or on a specific day of the week?[749]

It is technically already possible to connect all a user's devices submitting information about their musical preferences, movies watched, television routines and even to detect their mood from their personal health devices, and to then direct them to the best content. Add information about weather, work calendar and local traffic, and you have quite a powerful personal assistant serving you what may be interesting right now, right here. All these services already exist, and most of them are based on audiovisual content. In reality, some of the global tech giants are already offering something like this. Just think of what Google Assistant or Apple's portfolio of services are able to do, and all

[747] Rezzonico P., "Artificial intelligence at the service of the RTS audiovisual archives", FIAT/IFTA, http://fiatifta.org/index.php/artificial-intelligence-at-the-service-of-the-rts-audiovisual-archives/
[748] TM Forum, "AI indexing for regulatory practise", https://www.tmforum.org/ai-indexing-regulatory-practise.
[749] Cibrián E. et al. "Artificial intelligence and machine learning for commercial analysis in the audiovisual sector: A case study of designing TV schedules", http://www.kr.inf.uc3m.es/artificial-intelligence-and-machine-learning-for-commercial-analysis-in-the-audiovisual-sector-a-case-study-of-designing-tv-schedules/.

those functionalities are combined on our mobile phone's operating system – controlled by those same companies.

Third, AI can be used in unique processes aiming to create better and more distinctive content. This involves not only optimising and repeating what has already been done, but creating completely new approaches without the limitations and biases of the human brain. This area poses interesting challenges as sometimes a good outcome created by a machine does not actually feel better than one created by a human, or may not be acceptable according to our ethical code. Another interesting angle is the question of how to maintain creativity and artistic motivation, and identity when computers take part in the process and work in a way that could be defined as creative - at least according to some definitions.

8.3. Is AI somewhat different than previous technologies?

How to prevent AI from causing harm? How to mitigate risks involved with creating such AI systems? Is the world safe when cars drive autonomously on the streets? Or, is the world fair, if crime is detected based on where you live or how you look into the camera?

Before jumping into regulating AI we should ask: how is it different? What makes AI so special that it would need special attention from regulators? Are the problems really new ones, looked at from a human's and society's perspective? Or are they just new versions of the same fundamental problems that the law may already address?

8.3.1. Who is responsible when AI causes harm?

The key concept resulting from the question of who is responsible when AI causes harm is AI's assumed ability to act human-like without human-like responsibility and control. In other words: If AI has potential autonomy, should it be controlled and regulated?

Another special feature is the fact that AI's actions are difficult to foresee when systems are designed. This is especially true if systems are supposed to be 'creative', as you would expect in the audiovisual sector. What if the systems create results so unexpected that it is impossible to say that the designer of the system should have foreseen them and is therefore responsible for the outcomes?

Typically and traditionally, we have not considered machines responsible for their actions. The responsible party is the one who uses the machine or the one who built the machine - without enough care[750].

In theory, it is reasonable to imagine a world where machines would have responsibilities and would have to compensate for the harm they cause. We do have

[750] Petit N., *op.cit.*

constructs like companies with a legal personality which may carry duties and claim invoices.

However, in most cases we have considered humans the ones best suited to carry moral responsibilities and do ethical value-weighing. In all high-end technology, it is however difficult to locate the relevant actor responsible for outcomes that may occur only later and in contexts the original builder never thought about.

Autonomy is, as the European Group on Ethics in Science and New Technologies (EGE) states in their report "Artificial Intelligence, Robotics and 'Autonomous' Systems", an important aspect of human dignity which should not be relativised and shifted to technology. Machines can't take the moral standings of humans nor inherit human dignity. All our moral and legal institutions are based on the idea of humans taking on moral responsibilities, like being accountable or liable, or carrying duties.[751]

8.3.2. It's not just the economy

It is indisputable that personalisation is valuable for individuals, but especially in an industry with substantial cultural and societal relevance it is crucial that these systems serve interests beyond those of the individual, in other words the interests of the audience or society as a whole. It is essential to find an acceptable balance between these potentially competing interests. In recent times, some of the most fundamental challenges of democracies have been created by 'targeting machines' in the wrong hands: "fake news" on a massive scale, deep fakes, or the ability to fashion a filter bubble in which to live.

Audiovisual industries have special relevance in the democratic processes of our societies, both at the national and the European levels. And, we live in a global arena, especially when AI is considered. Data and software services are increasingly offered through global systems and often in a market controlled by US or Chinese companies. How a sustainable creative and audiovisual sector can be fostered when strategic use of AI in the field is becoming more and more important, is also a question of European competitiveness and will define what kind of soft power role Europe has in the future.

The audiovisual sector as a whole is tightly connected with the cultural and political needs of societies. However, while culture has higher societal values and relevance, it also has economic value in the marketplace and economic reality. In their understanding of the economic impact of culture, some divide it into 'institutional value' connected to the macroeconomic effects on the national or international level for economies, and 'micro-economic' value, which is the basis of single transactions in which

[751] European Group on Ethics in Science and New Technologies, "Statement on artificial intelligence, robotics and 'autonomous' systems", European Commission, Directorate-General for Research and Innovation, http://ec.europa.eu/research/ege/pdf/ege_ai_statement_2018.pdf.

people pay for cultural experiences. Institutional value is the theoretical basis on which economic subsidies have been granted to cultural industries[752].

The concept of AI for social good (AI4SG) provides some inspiration for how to design systems that promote societal value. These include trustworthiness as a basis of society and using all technologies in the aim to positively impact life.[753]

Other recommendations include developing systems in steps and testing them in the lab before launching them in production, and preventing manipulation through outside intervention. Promoting social good is often successful when users are involved in designing the systems, the autonomy of the user is respected and informs the system design, and the user is given the ability to provide the systems and their outputs with sense and meaning.

8.4. We have a moral obligation to do good with AI

When we consider that best way of regulating an industry, there's always a trade-off of costs of regulation against benefits. When we see only risks, and try to avoid harm and establish accountability with regulation, we may lose some of the benefits, as regulation typically disincentivises innovation and leads to costs.[754]

The problem with tight regulation is that it typically restricts risk-taking, creativity and value creation. Therefore, we have to ask what level of risk is acceptable and who should carry the risk if something goes wrong.

So, in designing regulation it is essential that we look at both sides of the phenomenon. On the philosophical side, one has to address the question of how to inspire innovation and creativity, and how not to establish a system that discourages risk-taking. Risk is fuel to the economy and well-being. Risk-taking is part of creativity and value creation in audiovisual industries as well.

Therefore, both in the field of AI in general, as also in the field of audiovisual industries, before regulating a specific - or in this case unspecific set of technologies - one has to ask whether there is a problem that has to be solved. Going even further: is there some regulation that should be removed to enable doing good with AI? Or should regulation be put into place to enable easier access to data or rights needed for the shift towards a new industrial era, also in audiovisual industries?

The discussion about regulation of AI has its roots in the field of basic human instincts and feelings which make us fear that machines will take ultimate control over us.

---

[752] La Torre M., "Defining the audiovisual industry", in La Torre M. (ed.) The economics of the audiovisual industry: Financing TV, film and web. Palgrave Macmillan UK, pp. 16–34.
[753] Taddeo M., "Trusting digital technologies correctly", *Minds and Machines*, 27(4), 565–568. Taddeo M., Floridi, L., "The case for e-trust", *Ethics and Information Technology*, 13(1), 1-3.
[754] Gurkaynak G., Yilmaz I. and Haksever G., "Stifling artificial intelligence: Human perils", Computer Law & Security Review, 32(5), pp. 749–758.

This is especially the case with the imagined Artificial General Intelligence with a will of its own. This ability to act independently of humans and the partly fictional construct of human-like intelligence is the basis of doomsday predictions and underpins calls for specific regulation of AI.

While often regulation is considered necessary to manage risks and prevent harm, in the case of audiovisual industries, regulation could also be seen from another ethical angle: we have a moral and ethical obligation to do good; how can we encourage the good usage of AI in a responsible way, and could regulation create an environment that fosters creativity and innovation? In other words: richer European life and economic prosperity.

Regulation is typically established when there is a problem that needs solving. But, we can only lose something if we have first gained it.

8.5. Regulation should be human-centric and goal-based

Let's remind ourselves that in a world full of risks, some of the risks are desirable, because taking them is the basis of creativity and well-being. Fundamentally, the question of regulating AI in the audiovisual industries is a question about which risks are so undesirable that they deserve to be regulated.[755] If we want an AI system to be creative, should we allow it to create unforeseen results? What are the real consequences of these risks if taken?[756]

Applying ethical rules mechanically can sometimes be tricky. For example, sometimes, demanding transparency may result in a ridiculous situation. With creative work in the audiovisual sector, who has ever known how an artist came to a particular artistic conclusion? Do we even want to know? Isn't the mystique part of the fascination? Why should we demand transparency from the machines that are used in the creative part of the industry if we don't demand the same from humans? In this case, there is no public or private interest in protecting someone from creativity with the help of technology.

On the other hand, the benefits of using AI in audiovisual industries can be especially promoted through effective regulation, because it minimises risksrelated to a lack of clarity. Are there obstacles in the present regulatory system that should be removed? Should we foster the creation of regulated data-sharing arrangements to make new businesses easier to establish and more value for customers and citizens easier to create?

---

[755] Buiten M. C., "Towards Intelligent Regulation of Artificial Intelligence". *European Journal of Risk Regulation*. 10 (1): 41–59, https://www.cambridge.org/core/journals/european-journal-of-risk-regulation/article/towards-intelligent-regulation-of-artificial-intelligence/AF1AD1940B70DB88D2B24202EE933F1B.

[756] Scherer M. U., "Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies". *Harvard Journal of Law & Technology*, Vol. 29, No. 2, Spring 2016, p. 364, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2609777.

In my view, Europe suffers from an overly risk-centred approach to regulation. While we evaluate risks and consider managing them, US and Chinese players are already on the global playing field, and establishing defences for their already-dominant roles in the industry.

8.5.1. Major risks should be addressed

Our technology already allows us, together, to destroy nature and even the whole of humankind. According to Huber, these are "public risks" that are so broadly distributed and outside the individual risk bearer's understanding and control that they pose threats to human health and safety[757]. In the audiovisual sector, risks that may cause most harm are actually of this kind, and they are those that result in the unintentional destruction of the foundations of our societies because citizens are manipulated and nudged for commercial and political purposes. In a way, it's one example of the "tragedy of the commons",[758] a still-discussed concept in which individuals acting according to their self-interest behave contrary to the common good.

One approach to this set of risks is to use human rights as a lens and a policy tool. Our legal system places the obligation to avoid the infringement of human rights. However, if democracy is in danger because of increased usage of targeting, who is responsible? And even if the concept of responsibility could be created, are judges and national legal systems capable of identifying the responsible parties at a speed that makes the regulation effective?

Machine learning systems have repeatedly been both accused of making – and, it has been revealed, actually make - of biased decisions or predictions.[759] It is actually possible that the system may be biased because it was designed to be so. More often, though, the bias comes from the data used in training the system. And, finally, these data often just reflect and reveal the bias of current reality, and the decisions made by humans. They therefore mirror our present societal values and choices, and human and cultural biases. This, again, reflects the societal nature of AI systems especially in the audiovisual industries. We have to be careful not to blame the AI systems for something that actually might have societal value: revealing how biased the decisions humans have made until now really are. Instead, we should welcome the systems, as they make us more conscious of our values – taking, of course, care that unethical decisions are not applied in practice before testing and the analysis of results.

---

[757] Huber P., "Safety and the second best: The hazards of public risk management in the courts", Columbia Law Review, 85(2), pp. 277–337.

[758] Feeny D. et al., "The tragedy of the commons: Twenty-two years later", Human ecology, 18(1), pp. 1–19.
Hardin G., "The tragedy of the commons", Journal of Natural Resources Policy Research. Taylor & Francis, 1(3), pp. 243–253.
Stavins R. N., "The problem of the commons: Still unsettled after 100 years", The American economic review, 101(1), pp. 81–108.

[759] See chapters 1,2 and 3 of this publication.

8.5.2. Humans are the responsible ones

We can't avoid concluding that a well-functioning regulatory system in a fast-changing world should be flexible and based on fundamental principles rather than constituting an attempt to regulate the technology. There are already approaches that may function well: focus on the goals and be human-centric, not technology-centric.

One of the fundamental principles to build on is the question of allocating legal responsibility. In all societies, there is a system in which responsibility, liability, and in some cases even accountability, is allocated in cases where damage to others is caused. In our present systems, this responsibility can't be allocated to technologies or machines, but can be assigned to humans, companies and other legal entities. In some industries, the risks have been considered so huge that organisations carry a heavier responsibility for results even in the absence of negligence. AI is created by humans who are well-educated and often aware of the possible risks their technology may create for users or objects. It is not unbearable for them - or in practice for the institutions they represent - to carry the responsibility of their actions, and in some fields even to carry strict liability without fault.

So far, whenever regulation appears to be about technology, it is in reality about the persons who created or used that technology, and organisations in which they are employed. And, interestingly, while AI-based technology may have its own faults, often humans are even more faulty. The issue is that we have become used to human faults but consider the same faults caused by technology to be less acceptable.[760] One core question is whether there really is a new risk that AI creates, perhaps even as yet unknown.[761]

8.5.3. Transparency as an interim solution?

Without transparency, citizens and consumers face decisions they do not understand and have no control over. To assess whether there should be liability for an AI-based decision, the courts, too, need to understand how the AI made its decision. So, requiring transparency and explainability might be a suitable solution to some of the ethical and legal problems, and has already been recommended as a tool for regulation.[762] Actually, in the discussion about sustainable AI regulation, transparency is a central concept, at least as an interim solution.

Transparency in relation to data refers to the obligation to keep users, customers or clients informed about how their data are being used. In the world of algorithms, transparency means the need to explain the way they work to the extent that they are understandable for the users. Explainability - a concept close to transparency - means

[760] Petit N., *op.cit.*
[761] Scherer M. U., *op.cit.*, p. 364.
[762] See chapter 1 of this publication.

ultimately, even, that the values on which the systems base their behaviour should be traceable.

However, a balance must be struck between the benefits and costs of this transparency. Sometimes, requiring transparency may in practice mean implementation of the system is rendered technically unfeasible.[763] Requiring more transparency may even oblige us to accept that systems are less accurate than they could technically be. An interesting choice, actually: if a diagnostic system is more accurate as an unexplainable black box, would you still save a life with it - without knowing how? So, black and white regulation requiring transparency in all cases in which AI is applied is not feasible either.

There are other problems with transparency, as well. There may be trade secrets that can't be revealed or the costs of maintaining transparency may result in a concentration of industries that is undesirable.[764] Requiring costly transparency may have negative effects on innovation.[765] Sometimes the logic of machines simply can't be expressed in a language understandable by humans. And how can accountability be traced in a technology field largely based on sharing resources like pieces of code and AI algorithm models globally?[766]

8.6. Human-centricity, not technology-centricity

The discussion about regulation of AI has followed two dominant routes: one based on the point of view of the legal system; the other aligned with the notion of starting from the technologies, and then building regulatory needs bottom-up from specific AI applications.[767] However, the basis of a good and functioning regulatory system rests not on these concepts but on the needs of humans and societies.

While all this discussion about the special distinctive features of AI technology is valuable, fundamentally AI is just a technology - a piece of computer software whose core domain is mathematical calculations - and it is fair to question whether it is so fundamentally novel as often presented.[768]

In conclusion, human-centricity in regulating AI may mean:

- transparency serves human needs to make sense of how the systems work and address responsibilities to the right persons;

[763] Buiten M. C., *op.cit.*
[764] Scherer M. U., *op.cit.*
[765] Buiten M. C., *op.cit.*
[766] Leonelli S., "Locating ethics in data science: Responsibility and accountability in global and distributed knowledge production systems", Philosophical transactions. Series A, Mathematical, physical, and engineering sciences, 374(2083), https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5124067/.
[767] Petit N., *op.cit.*
[768] Edelman R. D., "Here's how to regulate artificial intelligence properly", https://www.post-gazette.com/opinion/Op-Ed/2020/01/14/R-David-Edelman-Here-s-how-to-regulate-artificial-intelligence-properly/stories/202001140013.

- accepting that unknown risks may be impossible to regulate, at least if the regulation is based on the technology, not on goals;
- we should try to reduce public risk without destroying creativity and innovation;
- we should look at the present legal environment, especially in Europe, and remove unnecessary obstacles to using data to create well-being and doing good;
- humans should have more real control over how their data is used; however the present GPDR framework does not function well towards that goal,[769] as the consents are in reality given without real understanding and a large part of personal data usage happens through third parties; the European Union should critically review its policies in the field of AI and data, and centre more on enabling the doing of good, without forgetting that major risks, especially to democracies, should be addressed.

[769] See chapter 2 of this publication.

Concluding remarks

The common sense expression "with great power comes great responsibility" (made famous by Spider-Man's comic books but dating back at least to the French Revolution)[770] *fits AI like a glove. AI has enormous potential for doing both good and evil. That is why we find it fascinating and scary at the same time, and while some will tend to worry, others will set the accent on the marvellous things that can be achieved with this ground-breaking technological development. Indeed, a reading of the different contributions published in this report shows there is no single vision of how AI should be regulated, and yet there are certain principles that appear to be (in one way or another) in the minds of all of the authors: explainability, trust, privacy, pluralism, but also freedom of expression, creativity and innovation. If we manage to combine all those goals, AI can be a blessing for humanity in many ways.*

Unless, of course, one day Elon Musk's worst nightmares become reality and machines take over the world. But such dystopian future is not on the horizon.

Not yet, at least.

[770] https://quoteinvestigator.com/2015/07/23/great-power/.

A publication
of the European Audiovisual Observatory