Received: 24 January 2024 | Accepted: 4 August 2024

DOI: 10.1002/poi3.420

RESEARCH ARTICLE

Structuring different manifestations of misinformation for better policy development using a decision tree-based approach

Olivia Hägle[1,2] | **Stephan Escher**[3] | **Reinhard Heil**[4] | **Jutta Jahnel**[4]

[1]Institute for Information and Economic Law, Center for Applied Legal Studies, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany

[2]University of Freiburg, Freiburg im Breisgau, Germany

[3]Institute of Information Security and Dependability (KASTEL), Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany

[4]Institute for Technology Assessment and Systems Analysis, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany

Correspondence

Jutta Jahnel, Institute for Technology Assessment and Systems Analysis, Karlsruhe Institute of Technology (KIT); P.O. Box 3640, 76021 Karlsruhe, Germany.
Email: jutta.jahnel@kit.edu

Funding information

Helmholtz Association, Grant/Award Number: 46.24.01

Abstract

The spread of false and misleading information in digital communication spaces has enormous potential for harm in democratic societies, but has so far been insufficiently addressed by policy makers. This problem has been exacerbated by recent technological developments such as deepfakes. But deepfakes, fake news, and disinformation are only manifestations of misinformation. It is therefore essential to establish a clear definition to develop an appropriate policy. Since misinformation is a hydra whose countless heads cannot be cut off by a single heroine, we conclude that an interdisciplinary approach is necessary to appropriately "regulate" the different dimensions of misinformation. Therefore, we develop a decision tree that allows us to structure the problem of misinformation. Since deepfakes are particularly well suited to characterize the various manifestations of misinformation, we use them as an application example to illustrate our decision-tree-based approach. Based on this systematization, it will be possible to identify the capabilities and limitations of strategies from different disciplines and to develop a bundle of measures to address the various manifestations of misinformation. The basic principles of this bundle of measures will already be outlined here.

KEYWORDS

decision tree, deepfakes, digital communication, misinformation, policy development

INTRODUCTION

False and misleading information is not a phenomenon unique to the digital age; however, it is increasingly being used not only to discredit individuals but also to manipulate opinions. To date, there have been concerns that new technologies based on artificial intelligence will contribute to the qualitative and quantitative intensification of the problem (Chesney & Citron, 2019, p. 1776 et seq.), as these technologies enable the simple and cost-efficient creation and distribution of false and misleading information. Recently, in the wake of Russia's war of aggression against Ukraine, a deepfake of the Ukrainian president, Volodymyr Zelenskyy, was disseminated, showing him allegedly calling on the Ukrainian army to surrender (see, e.g., Sardarizadeh, 2022). This fake was still quite primitive and, therefore, easy to debunk, but it illustrates the enormous potential for harm from the dissemination of false and misleading information, especially as the technology for creating and disseminating such information continues to evolve (Wakefield, 2022). Surprisingly, these dangers of opinion manipulation by single abusive actors arise for the discourse in democratic societies in particular due to the openness of their systems.[1]

However, the problem is not limited to the deliberate dissemination of false information. Time and again, information that is initially disseminated with no malicious intent can nevertheless cause immense harm and may subsequently be exploited by malicious actors for their own purposes. Recently, for example, images purporting to show Donald Trump being arrested (Higgins, 2023) and Vladimir Putin kneeling in front of Xi Jinping (Smart, 2023) were disseminated on social media, originally—at least in the case of the Donald Trump fakes—accompanied by a note clarifying that the content was synthetically generated and therefore not intended to be misleading. Nevertheless, they had the potential to confuse recipients when the images were shared again shortly thereafter without clarifying notes (Buttcrack Sports, 2023).

Furthermore, misinformation is a major problem not only due to its potential to mislead, but has a negative impact on fundamental trust in the media and institutions, as well (van Duyn & Collier, 2019). Technological advances make it increasingly difficult for recipients to recognize misinformation. This “information uncertainty” (Stubenvoll et al., 2021) makes it easier to discredit information from reliable sources, for example by labeling it as fake news. Paradoxically, measures that enable recipients to better assess information, such as fact-checkings, may improve their ability to distinguish between right and wrong in specific cases, but at the same time make recipients even more insecure about their own ability to distinguish between right and wrong (York et al., 2020).

Although the regulation (in a broader sense)[2] of the individual phenomena such as deepfakes, fake news, and disinformation has been discussed in the literature of the various disciplines, including law, computer science, technology assessment, communication science, psychology, and philosophy (Aïmeur et al., 2023; Bayer et al., 2019; Bennett & Livingston, 2018; Chesney & Citron, 2019; Davis & Fors, 2020; Fallis, 2015; Guo et al., 2019; Pennycook & Rand, 2021; Pielemeier, 2020; van Boheemen et al., 2020; van Huijstee et al., 2021; Yamaoka-Enkerlin, 2019), a broader connection between these different manifestations of misinformation is still rarely established for the purposes of policymaking (this connection regarding the regulation of deepfakes and misinformation is only suggested by Chesney & Citron, 2019, p. 1776 et seq.). In the previous literature, a variety of terms is used in connection with the dissemination of false and misleading information, as well as forged and manipulated images, such as deepfakes, fake news, disinformation, and misinformation. Too often, however, a sufficient distinction between the different terms is missing. They are sometimes used interchangeably to describe the general problem of misleading information. At the same time, the regulation of misinformation manifestations is often discussed in the context of other phenomena such as hate speech, conspiracy theories, and propaganda, but also, for example, satire and hoaxes. For example, the German Network Enforcement Act follows the approach of a joint regulation of fake news and hate speech (Entwurf eines Gesetzes zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken [Netzwerkdurchsetzungsgesetz], 2017, p. 1). However, the regulatory assessment of misinformation on the one hand and, for example, hate speech on the other is not entirely congruent but only partially convergent (cf. Buchheim, 2020, p. 161 et seq.). These other phenomena do not qualify as misleading in the strict sense of the term, yet they may ultimately lead to misinformation.

Therefore, for effective policy development, it is essential to find a clear definition of misinformation and to distinguish misleading information from other information-related phenomena.

Several attempts have been made to find a solution to this "confusion of terms" (Bayer et al., 2019, p. 22 et seq.; Jack, 2017, p. 1 et seq.; Kapantai et al., 2021; Tandoc et al., 2018;). Particularly noteworthy is a proposal by Wardle (Wardle & Derakhshan, 2017, p. 4 et seq., esp. 15 et seq.; Wardle, 2019) in which she systematizes this complex phenomenon using the term "information disorder" by characterizing both the content and the context of information. The authors also stressed the need to identify different categories of problematic information based on clear definitions to understand the specific challenges that should be considered for policy development. Nevertheless, there is still uncertainty in the handling of the different terms. Since misinformation is a complex issue that can only be resolved through a combination of several measures, various scientific disciplines are developing approaches to the broad topic of misinformation. However, different scientific disciplines naturally have very different purposes when approaching a definition. So far, a comprehensive discussion beyond the boundaries of the individual disciplines has only taken place with restraint. It is precisely this kind of comprehensive discourse that is needed to improve policy development in this field.

Therefore, this article aims to support policymaking by providing an interdisciplinary framing and systematization of *misinformation*, which we consider to be the general term covering all kinds of phenomena associated with problematic misleading information, and simultaneously differentiating phenomena that are to be distinguished from misinformation and that include information that is misleading in the relevant context (See section "Systemizing misinformation"). Based on this systematization, we want to enable the development of a regulatory system (in a broader sense) that can deal with the different manifestations of misinformation. The basic elements of this system are outlined in the final section "Developing a comprehensive misinformation policy: A bundle of different approaches from various disciplines needed."

SYSTEMIZING MISINFORMATION

Scrutinizing the common approach to distinguishing misinformation from disinformation from an interdisciplinary perspective

In the legal literature (Ferreau, 2021, p. 204 et seq.; Mafi-Gudarzi, 2019, p. 65 et seq.; Pielemeier, 2020, p. 917 et seq.; Yamaoka-Enkerlin, 2019, p. 728) and the (legal) policy discourse (Alaphilippe et al., 2019, p. 5 et seq.; the strengthened EU Code of Practice on Disinformation, 2022, p. 1; German Government, 2023), the discussion on dealing with misleading information is based on the distinction between the terms dis- and misinformation. Disinformation is mostly defined, with reference to the definition of the High-Level Expert Group on Fake News and Online Disinformation of the European Commission, as verifiably false or misleading information that is created, presented, and disseminated for economic gain or to intentionally deceive the public and may cause public harm (European Commission, 2018, p. 3 et seq.). In contrast to disinformation, misinformation describes false or misleading content that is passed without the deliberate intention of causing harm; however, its effects may nevertheless be harmful (European Commission, 2020, p. 18; Zimmermann & Kohring, 2020, p. 23).

However, from a regulatory perspective (Dreyer et al., 2021, p. 11 et seq.) and for the purposes of this study, in particular, to enable an interdisciplinary discussion of the issue and to assist policy makers, this distinction between mis- and disinformation and their respective definitions fall short for several reasons.

First, the general distinction between dis- and misinformation does not seem to be purposeful. This is the case, at least to the extent that the differentiation is used not only to distinguish between different regulatory measures, but already primarily (German Government, 2023) determines whether a piece of information is to be considered problematic or not. Both forms of misleading information have the inherent potential to be disruptive (cf. European Commission, 2020, p. 18; Fallis, 2015, p. 402; Jack, 2017, p. 2). Therefore, the need for policy exploration should not be determined by the distinction between mis- and disinformation. However, the form and intensity of regulation must differ between mis- and disinformation. It should also be emphasized that the strict differentiation between mis- and disinformation and especially the further classification into the categories of fake news, hate speech, etc. is often not suitable as a basis for further interdisciplinary discussion and for the development of suitable solutions due to the difficulties of interpretation associated with the definitions. This is particularly true from the perspective of the law, as the categories formed are frequently not legal concepts, yet their consequences do affect the law.

In addition to the criticism of the starting point for regulatory measures, difficulties arise in relation to some of the specific characteristics of the definitions of dis- and misinformation.

The discussion is still dominated by the idea of disinformation in the form of deliberate lies. This is reflected in the fact that the classical definition of disinformation explicitly mentions false information. However, it is already ambiguous what constitutes truth (Glanzberg, 2006; Gloy, 2004) and, accordingly, what constitutes untruth, which makes the criterion of untruth a difficult basis for dealing with mis- and disinformation (Dreyer et al., 2021, p. 7 et seq.). Moreover, the problem of misleading information goes far beyond deliberate lies that can be clearly identified as false. Rather, in many cases, the harmful potential of misleading information lies precisely in the combination of *false* and *true* information, or the information itself is *true at all* and is misleading only because of the combination or the context in which it is presented. Thus, depending on how the point of reference for the assessment of truth or falsity is chosen, some of the misleading information in question may not be considered untrue at all. Consequently, regardless of its controversy, false information is at most a special case of misleading information. In addition, not only is it difficult to determine retrospectively whether information is false from the legal perspective, but it also poses difficulties in other disciplines that deal with this issue, such as for policy makers trying to address the problem and computer scientists developing technical solutions to detect misleading information. Thus, the core aspect is not falsity but rather the misleading effect of the information in question, whether true or false.

Because of the requirement of fraudulent intent, the common definition limits the problem of misleading information to one specific manifestation of misinformation: disinformation. First, there is a wide variety of information that is originally disseminated without the intention to harm and that can cause immense damage, as in the case of hydroxychloroquine (McCarthy, 2022); second, the subsequent determination of this subjective attribute is extremely difficult (Dreyer et al., 2021, p. 11; Pielemeier, 2020, p. 922 et seq.). Despite these difficulties, reference to the subjective intention of the person making the statement is justified at the secondary level in the context of determining the specific type of regulation. Individual accountability for problematic information can be justified by subjective elements. However, subjective intentions should not be relevant in assessing the need for regulation. Rather, the need for regulation should be determined solely on the basis of the potential impact of misleading information.

Finally, the common definition is restricted to information that may cause harm to the public. However, the conditions under which this public harm occurs remain unclear. Under certain circumstances, even misleading information that may initially cause harm only to a definable group of persons or individuals may later become problematic beyond that group of persons and eventually reach the public. For instance, primarily individual-related misinformation can lead to silencing effects (Gelber & McNamara, 2016), and due to the exclusion of certain (groups of) people from the shaping of a public opinion, the process of democratic opinion formation as a whole can subsequently be affected (Markard & Bredler, 2021, p. 867 et seq.; Hong, 2022, p. 140 et seq.; MacKinnon, 2020, p. 1243 et seq.; Völzmann, 2021, p. 620 et seq.). This is especially true for the special case of deepfakes. The vast majority of victims of deepfakes are female, especially when the fakes have pornographic content (Ajder et al., 2019, p. 2). And if (these) women withdraw from the democratic opinion-forming process as a result of the defamation through deepfakes, a significant portion of the democratic public is excluded from this process, demonstrating the gendered dimension of silencing effects in the context of these forms of online misinformation (Nadim & Fladmoe, 2021).

As the common approach to defining misinformation and disinformation poses several challenges from an interdisciplinary point of view, a modified approach is proposed here.

Our approach: Systemizing misinformation using a decision tree

Critical perspective: The recipients

We consider *misinformation* as the starting point for our considerations, under which we subsume all misleading and deceptive information that inherently has a certain potential to cause harm and therefore gives rise to a need for "regulation" in the broader sense, which includes not only legal regulation but also technical measures and preventive educational measures. To systematize misinformation and enable the development of an appropriate and comprehensive policy through suitable interdisciplinary measures, we introduce a decision tree that enables the qualified allocation of certain information as misinformation or not. Consequently, the decisive elements for this assessment are the effects and impacts of the information in question. Since the deceptive effect of misinformation largely depends on its impact on the recipients, their position becomes the key perspective for our interdisciplinary approach.

Therefore, to qualify content as misinformation, it is not crucial to subsume a piece of information under a rigid definition which is difficult to formulate unambiguously and can therefore easily be politicized. Instead, we have focused on specific misinformation-related effects as a first step. Depending on the presence or absence of additional attributes, misinformation can be further differentiated. Although this differentiation is not crucial for assessing the need for regulation, it is relevant for developing specific and concrete regulatory measures.

Distinguishing the different levels

To distinguish misinformation from other communication phenomena and systematize different manifestations of misinformation, we have identified four levels that can be distinguished: (a) misleading information, (b) potential to cause damage, (c) infringements of rights or the imminent danger of this infringement, and (d) intent to deceive (see Figure 1).

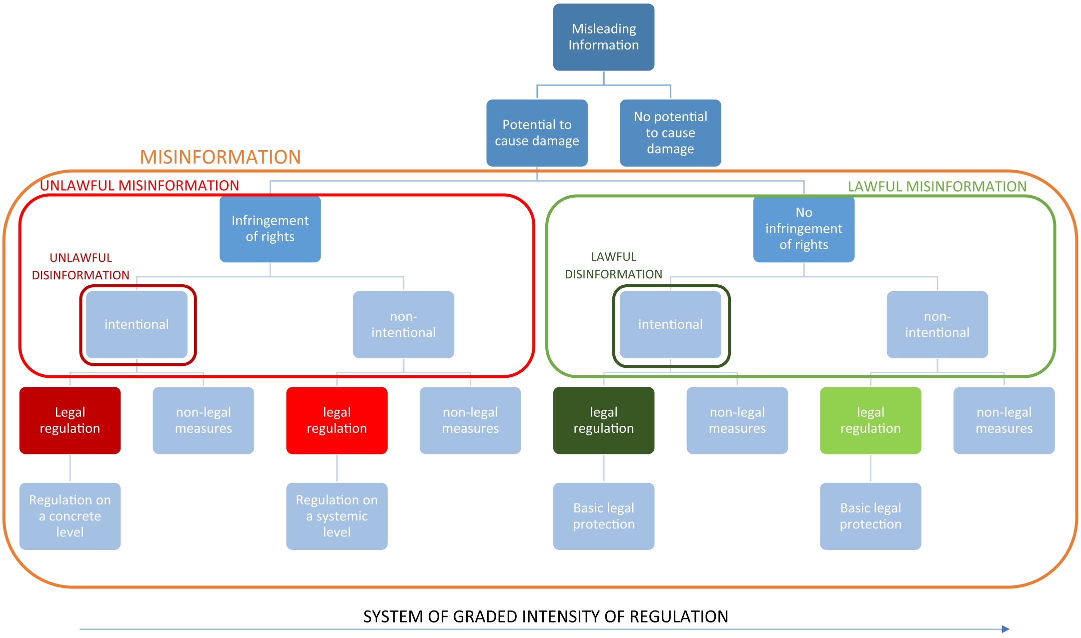

FIGURE 1 Decision tree for the systematization of misinformation.

Depending on the level in the decision tree, at which a specific piece of information is to be classified, various measures from different disciplines are required, which differ in terms of type and intensity (see the distinction between the "regulation on a concrete level," the "regulation on a systemic level" and the "basic legal protection" in Figure 1. See further on the different measures below in the section "Developing a comprehensive misinformation policy: A bundle of different approaches from various disciplines needed").

Misleading information

The basic prerequisite for qualifying pieces of information as misinformation is that they must be misleading,[3] or, in other words, that they must have an inherent potential to deceive.[4] According to our approach, it is irrelevant whether the information in question can be qualified as true or false according to objectified standards or whether it is not open to such an evaluation. The decisive factor is whether the information is misleading in a particular case.

Information is considered misleading if the recipients' perception legitimately differs from the actual circumstances.[5] This can be the case, for example, if the piece of information in question is false, but also if information is presented incompletely or if a piece of information is placed in a different context and thus has a different meaning. However, actual deceptive success is not required; it is sufficient if the information has the potential to deceive (Fallis, 2015, p. 406; Zimmermann & Kohring, 2020, p. 25). To assess the potential to deceive, the objectified point of view of an average recipient, to whom the information is addressed or whom it can be expected to reach, must be taken into account.[6] It must be considered, for example, whether the information in question is predominantly directed at an audience of adults or experts in the relevant field, on the one hand, or primarily at credulous minors or other persons who, for objective reasons, lack the ability to determine the credibility of the information and can be misled, on the other hand. Purely subjective misconceptions without a legitimate basis, by contrast, are irrelevant for assessing the misleading nature of the information. Additionally, the particularities of communication in digital spaces (e.g., the speed of digital communication, the digital characteristics of echo chambers and filter bubbles as well as the general problem of polarization in digital environments) must be considered, which leads to different due diligence requirements compared to the analogue space.

Potential to cause damage

Because our objective is to enable the development of a comprehensive misinformation policy, supplemented by specific measures to address problematic misleading information, only information that has an inherent potential to cause damage is discussed here. For misleading information to qualify as misinformation, the information must have some potential to cause damage. However, this harm must not be considered legal damage. Initially, it is sufficient if such potential can be identified at the societal level.[7] This characteristic serves to distinguish harmful misinformation from harmless, misleading information that may be capable of deceiving the recipient, but has no inherent potential to cause damage. Such cases of harmless, misleading information may include, for example, newspaper hoaxes ("Zeitungsenten") or satire. This does not mean, however, that these categories of information (hoaxes, satire, etc.) (Kapantai et al., 2021; Pielemeier, 2020, p. 918 et seq.) should always be classified as harmless, the assessment must be made on a case-by-case basis. This evaluation is made from an objective perspective, so it is irrelevant whether the actor also had a corresponding intention (to the contrary, however, German Government, 2023; Wardle & Derakhshan, 2017, p. 5). Here, too, the assessment may depend on who the information is aimed at or who it reaches (see already in the section "Misleading information").

As soon as this second characteristic is fulfilled, we consider the misleading information to be misinformation, according to our approach. Already at this level, the issue needs to be appropriately addressed by policymakers and certain measures from different disciplines can be applied (e.g. a "basic legal protection of information", see further on these measures below in the section "Developing a comprehensive misinformation policy: A bundle of different approaches from various disciplines needed").

Infringement of rights

Misinformation is generally not considered problematic from a legal perspective; specific legal regulation is only necessary and may only intervene if the misinformation in question is accompanied by some kind of violation of rights and is therefore problematic from a regulatory point of view.

The problem of misinformation is characterized by a conflict of fundamental rights in the digital space. Thus, the rights and interests of the actors involved need to be balanced with the conflicting fundamental rights of others, which also affect the relations between private actors.[8] As a rule, creators and disseminators of misinformation, as well as intermediaries, can also claim legally protected rights and interests—especially communicative freedoms—as well as professional freedoms and potentially other fundamental rights.

From a legal perspective, pieces of misinformation are (untrue) factual claims (Steinebach et al., 2020, p. 149) that, even if untrue, may be protected by freedom of expression under Art. 10 ECHR and Art. 11 (1) CFR (ECHR Salov v. Ukraine, 2005, para. 113; Calliess, 2022, para. 10).[9] However, the untruthfulness of factual claims may be of essential importance for the balancing of conflicting fundamental rights, so that untrue factual claims must regularly be subordinated to conflicting rights (ECHR Sorguc v. Turkey, 2009, para. 29; Cornils et al., 2021, para. 15; in this sense also argued at the national German level, e.g. Jestaedt, 2011, para. 38). Freedom of communication is of outstanding importance in democratic societies. In the words of the ECHR: "Democracy thrives on freedom of expression" (ECHR *United Communist Party of Turkey and others* v. Turkey, 1998, para. 57; ECHR Herri Batasuna and Batasuna v. Spain, 2009, para. 76). Nevertheless, communication freedom does not enjoy absolute priority but must be reconciled with conflicting rights through balancing. Therefore, regulatory measures on misinformation restricting communicative freedoms must be justified by serving a legitimate purpose and also being otherwise proportionate (Dreyer et al., 2021, p. 14). In principle, this legitimate purpose can only be to protect the rights and legal interests of third parties or the public.

While legal interests worthy of protection are easily identifiable when false information is related to individuals and measures are taken to protect the individual rights, especially the personality rights, of directly affected third parties, the regulation of misinformation is more difficult when problematic information is not related to individuals or when the problem should be addressed directly at the systemic level (Buchheim, 2020, p. 166). At the systemic level, especially communication freedoms may be identified as legal interests to be weighed.[10] Since communicative freedoms are not only important for the formation of individual opinions, but also constitutive of the democratic opinion-forming process (Demokratischer Willensbildungsprozess), it is crucial for democracy itself (ECHR Grigoriades v. Greece, 1997, para. 44; ECHR Hertel v. Switzerland, 1998, para. 46; ECHR Stoll v. Switzerland, 2007, para. 101; ECHR Mouvement Raelien Suisse v. Switzerland, 2012, para. 48; BVerfG *5.* Rundfunkentscheidung, 1987, p. 323). In the context of misinformation, freedoms of communication can be impaired if the autonomy of single/multiple individuals is affected in such a way that they form their will on an uncertain factual basis (Buchheim, 2020, p. 163 et seq.). Only in the rarest of cases, however, will one single misleading piece of information be capable to sufficiently influence or jeopardize the opinion-forming process.

Since an infringement of rights can only rarely be identified, regulatory measures are not intended to be repressive at the specific and individual level, but are primarily aimed at the systemic level. Therefore, they do not refer to specific misleading information, but follow a preventive and general approach of securing the basic prerequisites for an orderly formation of individual and collective opinions. If these basic requirements are met and a free battle of opinions is ensured, the public discourse itself should be able to prevent the spread of misinformation by uncovering and correcting misinformation in discussions, so that the truth prevails (cf. BVerfG Parteienfinanzierung I, 1966, p. 99; Milker 2017, p. 219 et seq.; Steinebach et al., 2020, p. 164).

Intent to deceive

In contrast to previous approaches, any form of subjective attribute is no longer relevant for assessing whether problematic information exists at all, but only unfolds its effect at the level of legal regulation, namely in two respects:

On the one hand, the awareness of the untruthfulness of the information to be expressed may have an impact on its protectability,[11] on the other hand, subjective characteristics also serve to link an event with problematic consequences for third parties or the general public to certain persons to establish legal responsibility for these consequences (see further below in the section “Developing a comprehensive misinformation policy: A bundle of different approaches from various disciplines needed”).

According to our approach, any intentional form of misinformation, whether or not accompanied by a violation of legally protected rights, is considered disinformation. Therefore, disinformation is not the primary connecting factor of regulation, but rather a subcategory (Fallis, 2015, p. 402; different, however, Hernon, 1995, p. 134) of the regulatory object misinformation.

Application example: Deepfakes

Using a decision tree, it is possible to systematize the different forms of misleading information and, based on this, to develop a differentiated policy approach consisting of a bundle of different measures to deal with various misinformation phenomena. In the following, our decision tree is applied to several examples of expressions in the form of deepfakes. Like misinformation, deepfakes are not a distinct category of misleading information that can always be assessed in the same way, but can be encountered in a variety of different forms for which appropriate solutions need to be developed. Therefore, these solutions should not be developed depending on the classification (e.g., deepfakes, fake news, hoaxes etc.), but based on the characteristics of a particular piece of information and its impact on the recipients, as described in the decision tree presented here.

The initial filtering of all forms of expressions regarding the qualification as misinformation and its subtypes is carried out via the "misleading" criterion. There are various forms of expressions that can ultimately be misleading, and deepfakes are a great example of how the same type of expression (according to the conventional distinction between different phenomena) can lead to very different effects, and must therefore be treated differently. If recipients can recognize and are not deceived by information such as deepfakes, then this kind of information should not be subject to misinformation-related measures. This does not mean, however, that such deepfakes are not amenable to regulation at all, they are simply not subject to specific misinformation-related measures. They may be problematic for reasons other than deceiving characters; for example, if images of third parties are used to create them, personality rights may be affected. In principle, all types of deepfakes can be misleading. However, this does not apply to appropriately labeled applications of deepfake technologies in art (The Next Rembrandt, 2016) or appropriately labeled content of on-demand services for customers, where text is converted into high-quality videos with AI avatars (Synthesia—#1 AI Video Generator, 2023).

However, even if a statement is misleading, it is only amenable to specific misinformation-related measures if it has at least some potential to cause harm. Certainly, there are some examples of deepfakes that lack the potential to cause damage. For example, the Tom Cruise impersonation on social media may be misleading to a large number of social media users, but does not have the potential to cause damage. However, the EU AI Act (Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonized rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act)[12]) is so broad that it may cover these cases of deepfakes under the definition in Art. 3 para. 60. On the basis of this broad definition, Art. 52 para. 3 subpara. 1 establishes a disclosure obligation for deepfakes because they are generally considered problematic from a regulatory perspective due to their authenticity. However, according to Art. 52 para. 3 subpara. 2, an exception to the transparency obligation should apply if the use of AI systems is "authorised by law to detect, prevent, investigate and prosecute criminal offences or it is necessary for the exercise of the right to freedom of expression and the right to freedom of the arts and sciences guaranteed in the Charter of Fundamental Rights of the EU, and subject to appropriate safeguards for the rights and freedoms of third parties."

Problematic forms of misleading statements are qualified as misinformation according to our decision tree-based approach, and are amenable to certain forms of intervention from different disciplines. However, the need for legal regulation is only triggered by the infringement of rights or the imminent danger to this infringement. Such a potential legal threat may also exist in the case of unintentional misinformation and is not necessarily related to a malicious intent. For example, even deepfakes with a satirical background that usually lack an intention to deceive can be problematic in individual cases. In the case of a deepfake of the German Chancellor Olaf Scholz by the collective "Zentrum für politische Schönheit" (Center for political beauty) in which he is supposedly announcing the application to ban a far right wing political party in Germany, the court of first instance ruled in the proceeding for interim relief that the deepfake video may not be published due to its misleading potential. (Reporting e.g. LTO, 2024). In the case of satirical deepfakes, the necessary balancing of rights may often be in favor of the freedom of the arts and freedom of expression, however, if the satirical background is not evident as such, as assumed by the court in this case, the assessment may be different, so that there might be an infringement of rights.

If the statement is intentional, individual responsibility can be established so that regulation can also take place at the individual level. Such forms of intentional misinformation qualify as disinformation. An example of such unlawful disinformation is the deepfake of Zelenskyy (see, e.g., Sardarizadeh, 2022), which spread shortly after the beginning of the Russian war of aggression against Ukraine. Regulation here applies not only at the systemic level (e.g., the Digital Services Act and nonlegal measures such as social measures of empowerment of identification and appropriate action), but also at the individual level (e.g., copyright, personality rights).

Another example of unlawful misinformation with the intent to deceive (unlawful disinformation) are cyberattacks such as the case in which an audio deepfake was used to deceive the CEO of a British energy company during a phone call. The manager was asked to transfer money to a supplier using an impersonated voice of the CEO (Stupp, 2019). These cases require regulation at the individual level, where a variety of regulations already exist for certain forms of disinformation, in this specific case, for example, the criminal law provisions on fraud.

In addition to these "traditional" forms of misinformation, there are other forms of harmful expressions that focus primarily on harassing or defaming other persons but also contain misleading elements and may, therefore, be subject to misinformation regulation. For example, in the case of pornographic deepfake content of Taylor Swift which spread on X (See e.g. Saner, 2024), misleading elements can be observed, too, although the harassment and abuse may prevail. Hence, these deepfakes may be considered not only as some form of online sexual harassment or hate speech but might also qualify as misinformation.

DEVELOPING A COMPREHENSIVE MISINFORMATION POLICY: A BUNDLE OF DIFFERENT APPROACHES FROM VARIOUS DISCIPLINES NEEDED

Appropriate and comprehensive management of misinformation is a transdisciplinary task. Even without a clear definition, misinformation can serve as a boundary object; that is, an object to which different groups refer together and coordinate their activities, even if they do not mean the same thing by it. However, the decision tree has introduced structures into the field, making it possible to develop policy options, identify necessary alliances, and show the limits of the possibilities of the individual disciplines and actors. The law, in particular, is constantly challenged to solve problems that are beyond its domain and possibilities. Moreover, law enforcement now depends on technical assistance because of the quality and quantity of misleading information. The same applies to computer science. Technical (detection) solutions depend on legal science to determine what information is problematic from a legal perspective. However, law and computer science are not the only players.

Despite its inherent limitations, the law is an important element in combating misinformation. Three levels of legal regulation can be distinguished in the context of misinformation regulation: legal regulation at the individual level, legal regulation at the systemic level, and basic legal protection (see the decision tree framework above, Figure 1). The strict measures of legal intervention (e.g., threat of punishment, compensation, omission, stipulation of certain obligations, and responsibilities) are generally only permissible for the protection of legally protected goods and interests and, thus, in connection with unlawful misinformation and, in particular, disinformation. At the individual level, regulation can only occur in cases where individual responsibility can be established, particularly in cases of unlawful disinformation. However, legal regulation can also be applied at the systemic level, where, liability is established not by the direct creation or dissemination of misinformation, but by providing a platform for such misinformation. Systemic legal regulation is complemented by measures of self-regulation by platforms. Beyond cases of unlawful mis- and disinformation, the law can also play an essential role for the "regulation" of lawful mis- and disinformation that does not involve an infringement of rights. We refer to this form of regulation as "basic legal protection." While democratic legal systems assume that opinions are self-regulating without strict legal intervention through the free battle of opinions (cf. Fehling & Leymann, 2020, p. 110 et seq.; BVerfG Parteienfinanzierung I, 1966, p. 98 et seq.; Steinebach et al., 2020, p. 164), this will only work if this process can take place unhindered. However, to enable a proper battle of opinions, some basic conditions must be met, which the phenomenon of misinformation is about to challenge. These basic requirements for a system of free, democratic opinion-forming include, inter alia, freedom, openness, and plurality of public communication (see extensively Heldt et al., 2021, p. 6 et seq.). The law can help ensure these basic conditions; however, it must be enacted by society.

Preventive approaches with an active role for the recipients of misinformation should include safety- and security-based legal provisions. At the level of general education, the awareness, reflection, and empowerment of media users can be enhanced to promote a critical reception of media content. This so-called media literacy approach (Aufderheide, 1997) includes educational methods focussing on the occurrence and recognition of misinformation or propaganda techniques, but also on the consequences of possible actions that enable an active change in one's own behavior (Fazio, 2020; Kahne & Bowyer, 2017; Lutzke et al., 2019). Media literacy is often characterized by the concept of media competence, which includes media criticism, media knowledge, media use, and media design. Through these learned skills, children as well as adults should be encouraged to identify and combat problematic content, which is also useful in countering the systemic effects of misinformation. However, there are also limitations of this knowledge-based empowerment due to subjective and emotional factors. These include possible misperceptions, preferences and cognitive effects, such as the confirmation bias, and the possibility of influencing people's attitudes. It is also highly relevant to bear in mind that measures intended to contribute to knowledge-based empowerment, such as fact checking, can themselves have negative effects (York et al. 2020). This highlights the important role of behavioral science, psychology and communication science for the development of effective interventions such as accuracy nudges (Pennycook & Rand, 2021 and 2022). In combination with technical measures of computational science for the detection of misleading content these disciplines are crucial for effective labeling or warning-based approaches (Kirchner & Reuter, 2020).

Besides these interventions on the individual level, specific measures on a systemic macro level, too, were proposed to strengthen democratic societies and their resilience to misinformation (Humprecht et al., 2020). The starting point for these solutions is the assumption that some countries are more resilient to misinformation than others due to different institutional and structural factors and specific contextual conditions.

Other preventive measures focus on due diligence obligations in the production and distribution of information, ethical principles, and quality guidelines for the responsible actors involved, such as professional journalists, technology developers, and platform operators. These tools are mainly based on self-regulation and voluntary agreements of organizations, as opposed to mandatory legal provisions (e.g. the strengthened EU Code of Practice on Disinformation, 2022). Finally, there is a need for independent interdisciplinary research and knowledge transfer between actors to strengthen research and fact-checking communities.

In the technical domain, identification and authentication approaches help to verify users distributing information within digital services (Roy & Karforma, 2012; Shah & Kanhere, 2019). However, the use of verification also leads to a loss of privacy, as all information becomes unambiguously traceable (Camenisch & Lysyanskaya, 2001; Camenisch & Van Herreweghen, 2002). This is particularly evident in the inverse approach, which attempts to disprove malicious information through full-scale tracking of one's existence rather than verifying legitimate information (Chesney & Citron, 2019). Other approaches attempt to detect malicious information using manual (expert/crowd-based), automated, or hybrid detection methods. The drawbacks of manual detection approaches include the need for experts, associated costs, low coverage rates, and slow response times. Automated detection solutions, on the other hand, can directly cover a wide range of topics, but usually cannot explain their decisions. However, the discrepancy between security and privacy is exacerbated by automated analysis of information (Sharma et al., 2019; Zhou & Zafarani, 2020). Once malicious content is detected, there are several ways to respond. Detected misinformation can be reduced in its exposure, either through deleting or blocking it, or by reducing its visibility or demonetizing it. However, blocking and filtering information can also lead to trade-offs between freedom of speech, utility, and security. Detected misinformation can also be labeled or rated to reduce the potential for deception and improve the ability to assess the credibility of the information (Kaiser et al., 2021; Kirchner & Reuter, 2020; Moravec et al., 2020). Some approaches attempt to contextualize misinformation to provide users with a more complete picture of an issue. However, user studies have also demonstrated unintended or even opposing effects of labeling and contextualization (Wittenberg & Berinsky, 2020).

In summary, a bundle of activities is needed to combat misinformation, consisting of a large number of individual measures from a wide range of disciplines. Although there are already various individual measures from different disciplines that attempt to address this problem, these forces need to be more effectively combined to comprehensively tackle misinformation; however, there are also areas, particularly in relation to nonindividual-related misinformation, that have so far only been addressed in a limited way by regulatory means. The need for further regulation to address these gaps is a matter for policy makers to negotiate. Our decision tree-based approach for systemizing misinformation could be an essential element for the comprehensive discussion of misinformation beyond the boundaries of different disciplines to enable the development of a package of measures addressing the various challenges posed by misinformation. Ultimately, this requires a broad societal discourse. While this view, that dealing with misinformation is a challenge for society as a whole, is unsatisfactory—and could also mislead people into thinking that no one in particular feels responsible—it is correct in principle. Therefore, it is important to clarify which disciplines and actors can contribute to solving the problem of misinformation and how they need to cooperate to develop a comprehensive, adequate, and effective misinformation policy. These questions should therefore be the subject of future studies. We therefore encourage representatives of other disciplines (e.g. psychology, behavioral science, communication studies) to test our decision tree-based approach for its functionality and to expand and modify it with regard to the respective needs of the various disciplines, especially at the lower levels, so that a comprehensive misinformation policy can ultimately be developed on this basis.

ACKNOWLEDGMENTS

The research was financially supported by the Helmholtz Association, Research Field Information, Program Engineering Digital Futures (46.24.01). Open Access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

ORCID

Olivia Hägle http://orcid.org/0009-0007-9805-9391
Stephan Escher http://orcid.org/0009-0005-7628-2352
Reinhard Heil http://orcid.org/0000-0001-7265-7376
Jutta Jahnel http://orcid.org/0000-0001-7104-1908

ENDNOTES

[1] According to a recent study by the Pew Research Center, 70% of respondents in 19 countries considered "the spread of false information online" to be a major threat to their country. This was the most common response in Germany and Canada. "The spread of false information online" was mentioned more often than "global climate change," "cyberattacks from other countries," "the condition of the global economy," and "the spread of infectious diseases" (Poushter et al., 2022, p. 3, 6).

[2] Here, the term "regulation" should not be limited to the legal regulatory field but refers to policy-making in general and also includes "regulatory" measures in a broader sense from other disciplines, including measures of self-regulation, technical measures, and such measures of a social nature.

[3] If one searches for familiar forms of "misinformation" in law, one will be successful, for example, in unfair competition law. Here, too, the term "misleading" is of crucial importance. See, e.g., Art. 6 Directive 2005/29/EC of the European Parliament and of the Council of May 2005 concerning unfair business-to-consumer commercial practices in the internal market (Unfair Commercial Practices Directive), in the following: UCP-Directive and Art. 2 lit. b) Directive 2006/114/EC of the European Parliament and of the Council of 12 December 2006 concerning misleading and comparative advertising (Advertising Directive), in the following: Ad-Directive.

[4] The terms "misleading" and "likeliness to deceive" are also used synonymously in European Unfair Competition Law. See, for example, the definitions in Art. 6 (1) UCP-Directive and Art. 2 lit. b) Ad-Directive.

[5] Art. 2 lit. b) Ad-Directive, for example, defines "misleading advertising means advertising which in any way, including its presentation, deceives or is likely to deceive the persons to whom it is addressed or whom it reaches […]".

[6] Such objectified assessment standards in connection with the evaluation of the misleading character of certain information are also known, for example, from the European Unfair Competition Law. Here, the reference is the "average consumer" (see, e.g., Art. 6 (1) UGP-Directive).

[7] For this legal damage, however, see the "Infringement of rights" subsection.

[8] In German law, this is achieved through the figure of the 'mittelbare Drittwirkung'. See fundamentally German federal constitutional court (BVerfG) Lüth, 1958, p. 198 et seq.; For transferability of the evolved principles on information intermediaries such as online platforms see the development of the jurisdiction of the BVerfG FRAPORT, 2011, p. 226 et seq.; Stadionverbot, 2018, p. 267 et seq.; Dritter Weg, 2019, p. 1935 et seq.; While there is no equivalent to the figure of the 'mittelbare Drittwirkung' in European law, similar approaches can be found which enable the applicability of fundamental rights in relationships between private individuals. For example, the ECHR also refers to state actions in constellations under private law and thus enables the indirect application of the ECHR in relationships between private individuals. See e.g. ECHR Young, James and Webster v. The United Kingdom, 1981, para. 49; ECHR Costello-Roberts v. The United Kingdom, 1993, para. 26; The European Court of Justice has also already conceded a direct application of certain fundamental rights for example in the context of Art. 21 of the Charter of Fundamental Rights, see for example, CJEU, Egenberger, 2018, para. 77; For further discussion of a possible third party effect in European law see Kingreen, 2022, para. 24 et seq.

[9] Also in German law, misinformation may potentially be protected by freedom of expression. The German Federal Constitutional Court excludes at most factual claims if their untruth is undoubted or the propagator even knows about the falseness: See BVerfG Auschwitzlüge, 1994, p. 247 et seq.; Jarass, 2020, para. 7. In any case,

the boundaries of protected expression of opinion are blurred and the scope of protection of freedom of expression is to be interpreted extensively, so that even unobjective expressions may eventually be protected. Instead of many, see Epping, 2021, p. 120, BVerfG Postmortale Schmähkritik, 1990, p. 283 et seq. Detailed on the protection of deliberately false information under German constitutional law Buchheim, 2020 p. 159 et seq.; Critical on the traditional differentiation in protection under freedom of speech according to the German Constitution Steinbach, 2017, p. 653 et seq.

[10] For an overview of the potentially affected legal interests see Dreyer et al., 2021, p. 30 et seq.

[11] Refer above in connection with the illustration of the decision tree at the “infringement of rights” section.

[12] At the end of 2023, the Parliament and the Council of the European Union reached a political agreement on the draft AI-regulation, which has now also been formally approved by the member states of the European Union. The Artificial Intelligence Act (Regulation (EU) 2024/1689) was promulgated in the Official Journal of the European Union on 12 July 2024 and shall enter into force on the 20th day following that of its publication. It shall apply from 2 August 2026, although other dates apply to individual provisions, cf. Art. 113 AI-Act.

REFERENCES

Aïmeur, E., Amri, S., & Brassard, G. (2023). Fake news, disinformation and misinformation in social media: A review. *Social Network Analysis and Mining*, *13*(1), 30. https://doi.org/10.1007/s13278-023-01028-5

Ajder, H., Patrini, G., Cavalli, F., & Cullen, L. (2019). *The state of Deepfakes: Landscape, threats, and impact*. https://regmedia.co.uk/2019/10/08/deepfake_report.pdf

Alaphilippe, A., Gizikis, A., Hanot, C., & Bontcheva, K. (2019). Automated tackling of disinformation. European Parliament STOA. https://data.europa.eu/doi/10.2861/368879

Aufderheide, P. (1997). Media literacy: From a report of the national leadership conference on media literacyIn: R. Kubey (Ed.), *Media literacy around the world* (1st ed.). Routledge.

Bayer, J., Bitiukova, N., Bárd, P., Szakács, J., Alemanno, A., & Uszkiewicz, E. (2019). *Disinformation and propaganda—Impact on the functioning of the rule of law in the EU and its Member States*, 202. https://www.europarl.europa.eu/thinktank/en/document/IPOL_STU(2019)608864

Bennett, W. L., & Livingston, S. (2018). The disinformation order: disruptive communication and the decline of democratic institutions. *European Journal of Communication*, *33*(2), 122–139. https://doi.org/10.1177/0267323118760317

van Boheemen, P., Munnichs, G., & Dujso, E. (2020). *Digital threats to democracy*. Rathenau Instituut. https://www.rathenau.nl/en/digital-society/digital-threats-democracy

Buchheim, J. (2020). Der staat. *Der Staat*, *59*(2), 159–194.

Buttcrack Sports. (2023, March 21). *Breaking: Donald Trump was just arrested by New York law enforcement*. Twitter. https://twitter.com/ButtCrackSports/status/1638191345234857989; [https://archive.ph/IJLET].

Calliess, C. (2022). EU-GRCharta Art. 11 Freiheit der Meinungsäußerung und Informationsfreiheit. In C. Calliess, & M. Ruffert (Eds.), *EUV/AEUV—Das Verfassungsrecht der Europäischen Union mit Europäischer Grundrechtecharta Kommentar* (6th ed.). EUV/AEUV.

Camenisch, J., & Van Herreweghen, E. (2002). Design and implementation of the idemix anonymous credential system. *Proceedings of the 9th ACM Conference on Computer and Communications Security* (pp. 21–30). https://doi.org/10.1145/586110.586114

Camenisch, J., & Lysyanskaya, A. (2001). An efficient system for non-transferable anonymous credentials with optional anonymity revocation. In B. Pfitzmann (Ed.), *Advances in cryptology—EUROCRYPT 2001* (Vol. 2045, pp. 93–118). Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-44987-6_7

Chesney, R., & Citron, D. K. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. *California Law Review*, *107*, 1753–1819. https://doi.org/10.2139/ssrn.3213954

Cornils, M., Gersdorf, H., & Paal, B. (2021). EMRK Art. 10 Freiheit der Meinungsäußerung, *BeckOK Informations—und Medienrecht* (35th ed.). BeckOK.

Court of Justice of the European Union (CJEU 2018). Egenberger, ECLI: EU:C:2018:257 https://curia.europa.eu/juris/document/document.jsf;jsessionid=98BDBFF6E4538271063C2F9B47AFEBA3?text=&docid=201148&pageIndex=0&doclang=EN&mode=req&dir=&occ=first&part=1&cid=721851

Davis, M. J., & Fors, P. (2020). Towards a typology of intentionally inaccurate representations of reality in media content. In D. Kreps, T. Komukai, T. V. Gopal, & K. Ishii (Eds.), *Human-Centric Computing in a Data-Driven Society* (pp. 291–304). Springer International Publishing. https://doi.org/10.1007/978-3-030-62803-1_23

Dreyer, S., Stanciu, E., Potthast, K. C., & Schulz, W. (2021). *Desinformation: Risiken, Regulierungslücken und adäquate Gegenmaßnahmen—Wissenschaftliches Gutachten im Auftrag der Landesanstalt für Medien NRW*. https://www.medienanstalt-nrw.de/fileadmin/user_upload/NeueWebsite_0120/Themen/Desinformation/Leibnitz-Institut_LFMNRW_GutachtenDesinformation.pdf van Duyn, E., & Collier, J. (2019). Priming and fake news: the effects of elite discourse on evaluations of news media. *Mass Communication and Society*, *22*(1), 29–48. https://doi.org/10.1080/15205436.2018.1511807

Entwurf eines Gesetzes zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken (Netzwerkdurchsetzungsgesetz). (BT-Drs. 18/12356 2017). https://dserver.bundestag.de/btd/18/123/1812356.pdf

Epping, V. (2021). *Grundrechte* (9th ed.).

EU Code of Practice on Disinformation. (2022). https://digital-strategy.ec.europa.eu/en/policies/code-practice-disinformation

European Commission. (2018). *Tackling online disinformation: A European Approach* (COM(2018) 236 final). https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52018DC0236&from=EN

European Commission. (2020). European Democracy Action Plan (COM(2020)790 final).

European Court of Human Rights. (1981, August 13). ECHR 13 The United Kingdom: Young, James and Webster v. 7601/76; 7806/77. ECHR. https://hudoc.echr.coe.int/eng?i=001-57608

European Court of Human Rights. (1993, March 25). The United Kingdom: Costello-Roberts v. 13134/87. ECHR. https://hudoc.echr.coe.int/eng?i=001-57804

European Court of Human Rights. (1997, November 25). 9121/1996/740/93. Grigoriades v. Greece. ECHR.

European Court of Human Rights. (1998, August 25). Hertel v. Switzerland. 59/1997/843 /104 9. ECHR.

European Court of Human Rights. (1998, January 30). United Communist Party of Turkey and others v. Turkey. ECHR. https://hudoc.echr.coe.int/eng%23%7B%22itemid%22:[%22001-58128%22]%7D133/1996/752/951

European Court of Human Rights. (2005, September 6). Ukraine: Salov v. ECHR. https://hudoc.echr.coe.int/eng%23%7B%22itemid%22:[%22001-70096%22]%7D65518/01

European Court of Human Rights. (2007, December 10). Switzerland: Stoll v. 69698/01. ECHR. https://hudoc.echr.coe.int/eng?i=001-83870

European Court of Human Rights. (2009, June 30). Spain: Herri Batasuna and Batasuna v. ECHR. https://hudoc.echr.coe.int/eng%23%7B%22itemid%22:%5B%22001-93475%22%5D%7D25803/04and25817/04

European Court of Human Rights. (2009, June 23). Turkey: Sorguc v. ECHR. https://hudoc.echr.coe.int/eng%23%7B%22fulltext%22:[%22Sorguc/Turkey%22],%22documentcollectionid2%22:[%22GRANDCHAMBER%22,%22CHAMBER%22],%22itemid%22:[%22001-93161%22]%7D17089/03

European Court of Human Rights. (2012, July 13). Switzerland: Mouvement Raelien Suisse v. ECHR. https://hudoc.echr.coe.int/eng%23%7B%22itemid%22:[%22001-112165%22]%7D16354/06

Fallis, D. (2015). What is disinformation. *Library Trends*, *63*(3), 401–426. https://doi.org/10.1353/lib.2015.0014

Fazio, L. (2020). Pausing to consider why a headline is true or false can help reduce the sharing of false news. *Harvard Kennedy School Misinformation Review*, *1*(2), 1. https://doi.org/10.37016/mr-2020-009

Fehling, M., & Leymann, M. (2020). Der neue strukturwandel der öffentlichkeit: wie lassen sich die sozialen medien regulieren. *AfP*, *51*, 110–119.

Ferreau, F. (2021). Desinformation als herausforderung für die medienregulierung. *AfP*, *52*(3), 204–210.

Gelber, K., & McNamara, L. (2016). Evidencing the harms of hate speech. *Social Identities*, *22*(3), 324–341. https://doi.org/10.1080/13504630.2015.1128810

German Federal Constitutional Court (1958). Lüth, 7 BVerfGE 198. BVerfG. https://www.bundesverfassungsgericht.de/SharedDocs/Downloads/DE/1958/01/rs19580115_1bvr040051.pdf;jsessionid=460C602C6EDF3A65F57B3E9F2F1935B3.1_cid329?__blob=publicationFile&v=1

German Federal Constitutional Court (1987). 5. Rundfunkentscheidung, 74 BVerfGE 297. BVerfG. https://www.servat.unibe.ch/dfr/bv074297.html

German Federal Constitutional Court (1990). Postmortale Schmähkritik, 82 BVerfGE 272. BVerfG. https://www.servat.unibe.ch/dfr/bv082272.html

German Federal Constitutional Court (1994). Auschwitzlüge, 90 BVerfGE 241. BVerfG. https://www.servat.unibe.ch/dfr/bv090241.html

German Federal Constitutional Court (2011) FRAPORT, 128 BVerfGE 226. BVerfG. https://www.bundesverfassungsgericht.de/SharedDocs/Downloads/DE/2011/02/rs20110222_1bvr069906.pdf?__blob=publicationFile&v=3

German Federal Constitutional Court (2018). Stadionverbot, 148 BVerfGE 267. BVerfG. https://www.bundesverfassungsgericht.de/SharedDocs/Downloads/DE/2018/04/rs20180411_1bvr308009.pdf?__blob=publicationFile&v=1

German Federal Constitutional Court (2019). Dritter Weg, 2019 NJW 1935. BVerfG. https://www.bundesverfassungsgericht.de/SharedDocs/Downloads/DE/2019/05/qk20190522_1bvq004219.pdf?__blob=publicationFile&v=3

German Federal Constitutional Court. (1966). Parteienfinanzierung I, 20 BVerfGE 56. BVerfG. https://www.servat.unibe.ch/dfr/bv020056.html

German Government. (2023, September 4). *Was ist Desinformation*. https://www.bundesregierung.de/breg-de/themen/umgang-mit-desinformation/was-ist-desinformation-1875148

Glanzberg, M. (2006). Truth. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy (Fall 2018 Edition)*. https://plato.stanford.edu/archives/fall2018/entries/truth/

Gloy, K. (2004). *Wahrheitstheorien eine Einführung*. Francke.

Guo, B., Ding, Y., Yao, L., Liang, Y., & Yu, Z. (2019). *The Future of Misinformation Detection: New Perspectives and Trends*. arXiv:1909.0365 4. arXiv. https://doi.org/10.48550/arXiv.1909.03654

Heldt, A. P., Dreyer, S., Schulz, W., & Seipp, T. J. (2021). *Normative Leitbilder der europäischen Medienordnung: Leitvorstellungen und rechtliche Anforderungen an die Governance für eine demokratische Öffentlichkeit* (Projektergebnisse Nr. 54; Arbeitspapiere des Hans-Bredow-Instituts). Hans-Bredow-Institut. https://www.hans-bredow-institut.de/uploads/media/default/cms/media/agnsi1n_AP54_BKM_Leitbilder-Gutachten_HBI.pdf

Hernon, P. (1995). Disinformation and misinformation through the Internet: findings of an exploratory study. *Government Information Quarterly*, *12*(2), 133–139.

Higgins, E. (2023, March 20). *Making Pictures of Trump getting arrested while waiting for Trump's arrest*. Twitter; https://twitter.com/EliotHiggins/status/1637931151410216960; [https://web.archive.org/web/20230329190613/].

Hong, M. (2022). Hassrede und desinformation als gefahr für die demokratie—und die meinungsfreiheit als gleiche und positive freiheit im zeitalter der digitalisierung. *Rechtswissenschaft*, *13*(1), 126–174.

van Huijstee, M., van Boheemen, P., Das, D., Nierling, L., Jahnel, J., Karaboga, M., & Fatun, M. (2021). European Parliament, European Parliamentary Research Service, & Scientific Foresight Unit, *Tackling deepfakes in European policy*. http://www.europarl.europa.eu/RegData/etudes/STUD/2021/690039/EPRS_STU(2021)690039_EN.pdf

Humprecht, E., Esser, F., & van Aelst, P. (2020). Resilience to online disinformation: A framework for cross-national comparative research. *The International Journal of Press/Politics*, *25*(3), 493–516. https://doi.org/10.1177/1940161219900126

Jack, C. (2017). *Lexicon of Lies: Terms for problematic information*. Data & Society Research Institute. https://datasociety.net/pubs/oh/DataAndSociety_LexiconofLies.pdf

Jarass, H. D. (2020). GG Art. 5 [Kommunikationsfreiheiten sowie Kunst- und Wissenschaftsfreiheit. In H. D. Jarass, & M. K. Pieroth (Eds.), *Grundgesetz für die Bundesrepublik Deutschland Kommentar* (16th ed.).

Jestaedt, M. (2011). § 102 Meinungsfreiheit. In D. Merten, & H.-J. Papier (Eds.), *Handbuch der Grundrechte in Deutschland und Europa* (1st ed.). Otto Schmidt.

Kahne, J., & Bowyer, B. (2017). Educating for democracy in a partisan age: confronting the challenges of motivated reasoning and misinformation. *American Educational Research Journal*, *54*(1), 3–34. https://doi.org/10.3102/0002831216679817

Kaiser, B., Wei, J., Lucherini, E., Lee, K., Matias, J. N., & Mayer, J. (2021). *Adapting security warnings to counter online disinformation* (pp. 1163–1180). ResearchGate. https://www.usenix.org/conference/usenixsecurity21/presentation/kaiser

Kapantai, E., Christopoulou, A., Berberidis, C., & Peristeras, V. (2021). A systematic literature review on disinformation: toward a unified taxonomical framework. *New Media & Society*, *23*(5), 1301–1326. https://doi.org/10.1177/1461444820959296

Kingreen, T. (2022). EU-GRCharta Art. 51 Anwendungsbereich. In C. Calliess, & M. Ruffert (Eds.), *EUV/AEUV—Das Verfassungsrecht der Europäischen Union mit Europäischer Grundrechtecharta Kommentar* (6th ed.). EUV/AEUV.

Kirchner, J., & Reuter, C. (2020). Countering fake news: A comparison of possible solutions regarding user acceptance and effectiveness. *Proceedings of the ACM on Human-Computer Interaction*, *4*(CSCW2), 1–27. https://doi.org/10.1145/3415211

LTO. (2024, February 25). Gericht verbietet Scholz-Fake-Video zu AfD-Verbot. Legal Tribune Online. https://www.lto.de/recht/nachrichten/n/lg-berlin-ii-15o579-23-olaf-scholz-bundeskanzler-deep-fake-afd-verbot-zentrum-politische-schoenheit/

Lutzke, L., Drummond, C., Slovic, P., & Árvai, J. (2019). Priming critical thinking: Simple interventions limit the influence of fake news about climate change on Facebook. *Global Environmental Change*, *58*, 101964. https://doi.org/10.1016/j.gloenvcha.2019.101964

MacKinnon, C. A. (2020). Weaponizing the first amendment: an equality Reading. *Virginia Law Review*, *106*(6), 1223–1283.

Mafi-Gudarzi, N. (2019). Desinformation: Herausforderung für die wehrhafte Demokratie. *Zeitschrift für Rechtspolitik*, 3.

Markard, N., & Bredler, E. M. (2021). Grundrechtsdogmatik der beleidigungsdelikte im digitalen raum. *JuristenZeitung*, *76*(18), 864–872.

McCarthy, J. (2022, August 26). *Hydroxychloroquine in Australia: A cautionary tale for journalists and scientists*. Reutersinstitute.Politics.Ox.Ac.Uk. https://reutersinstitute.politics.ox.ac.uk/hydroxychloroquine-australia-cautionary-tale-journalists-and-scientists

Milker, J. (2017). »Social-Bots« im meinungskampf wie maschinen die öffentliche meinung beeinflussen und was wir dagegen unternehmen können. *Zeitschrift für. Urheber- und Medienrecht*, *3*, 216–222.

Moravec, P. L., Kim, A., & Dennis, A. R. (2020). Appealing to sense and sensibility: system 1 and system 2 interventions for fake news on social media. *Information Systems Research*, *31*(3), 987–1006. https://doi.org/10.1287/isre.2020.0927

Nadim, M., & Fladmoe, A. (2021). Silencing women? gender and online harassment. *Social Science Computer Review*, *39*(2), 245–258.

Pennycook, G., & Rand, D. G. (2021). The psychology of fake news. *Trends in Cognitive Sciences*, *25*(5), 388–402. https://doi.org/10.1016/j.tics.2021.02.007

Pennycook, G., & Rand, D. G. (2022). Nudging social media toward accuracy. *The Annals of the American Academy of Political and Social Science*, *700*(1), 152–164. https://doi.org/10.1177/00027162221092342

Pielemeier, J. (2020). Disentangling disinformation: what makes regulating disinformation so difficult. *Utah Law Review*, *2020*(4), 917.

Poushter, J., Fagan, M., & Klein, H. (2022). *Climate Change Remains Top Global Threat Across 19-Country Survey*. Pew Research Center. https://www.pewresearch.org/global/2022/08/31/climate-change-remains-top-global-threat-across-19-country-survey/

Roy, A., & Karforma, S. (2012). A survey on digital signatures and its applications. *Journal of Computer and Information Technology*, *3*, 45–69.

Saner, E. (2024, January 31). *Inside the Taylor Swift deepfake scandal: 'It's men telling a powerful woman to get back in her box'*. The Guardian. https://www.theguardian.com/technology/2024/jan/31/inside-the-taylor-swift-deepfake-scandal-its-men-telling-a-powerful-woman-to-get-back-in-her-box

Sardarizadeh, S. (2022, March 16). *Zelensky Deepfake*. Twitter. https://twitter.com/Shayan86/status/1504131692411432966

Shah, S. W., & Kanhere, S. S. (2019). Recent trends in user authentication – A survey. *IEEE Access*, *7*, 112505–112519. https://doi.org/10.1109/ACCESS.2019.2932400

Sharma, K., Qian, F., Jiang, H., Ruchansky, N., Zhang, M., & Liu, Y. (2019). Combating fake news: A survey on identification and mitigation techniques. *ACM Transactions on Intelligent Systems and Technology*, *10*(3), 1–42. https://doi.org/10.1145/3305260

Smart, J. J. (2023, March 20). *Putin attempting to persuade Xi*. Twitter. https://twitter.com/officejjsmart/status/1637836851619807232; https://archive.ph/uKrSR.

Steinbach, A. (2017). Meinungsfreiheit im postfaktischen umfeld. *JuristenZeitung*, *72*(13), 653–661.

Steinebach, M., Bader, K., Rinsdorf, L., Krämer, N. C., & Roßnagel, A. (2020). *Desinformation aufdecken und bekämpfen: Interdisziplinäre Ansätze gegen Desinformationskampagnen und für Meinungspluralität*. Nomos.

Stubenvoll, M., Heiss, R., & Matthes, J. (2021). Media trust under threat: antecedents and consequences of misinformation perceptions on social media. *International Journal of Communication*, *15*, 2765–2786. https://ijoc.org/index.php/ijoc/article/view/15410

Stupp, C. (2019, August 30). Fraudsters used AI to mimic CEO's voice in unusual cybercrime case. *Wall Street Journal*. https://www.wsj.com/articles/fraudsters-use-ai-to-mimic-ceos-voice-in-unusual-cybercrime-case-11567157402

Synthesia—#1 AI Video Generator. (2023). https://www.synthesia.io/

Tandoc, E. C., Lim, Z. W., & Ling, R. (2018). Defining "fake news". *Digital Journalism*, *6*(2), 137–153. https://doi.org/10.1080/21670811.2017.1360143

The Next Rembrandt. (2016). The Next Rembrandt. https://news.microsoft.com/europe/features/next-rembrandt/

Völzmann, B. (2021). Freiheit und grenzen digitaler Kommunikation—Digitale gewalt als herausforderung der bisherigen meinungsfreiheitsdogmatik. *Multimedia Und Recht*, *8*, 619–624.

Wakefield, J. (2022, March 18). Deepfake presidents used in Russia-Ukraine war. *BBC News*. https://www.bbc.com/news/technology-60780142

Wardle, C. (2019). *First draft's essential guide to understanding information disorder*. https://firstdraftnews.org/wp-content/uploads/2019/10/Information_Disorder_Digital_AW.pdf?x76701

Wardle, C., & Derakhshan, H. (2017). *Information disorder: Toward an interdisciplinary framework for research and policy making* (DGI(2017)09; Council of Europe Report). Council of Europe. https://rm.coe.int/information-disorder-toward-an-interdisciplinary-framework-for-researc/168076277c

Wittenberg, C., & Berinsky, A. J. (2020). Misinformation and Its Correction. In N. Persily, & J. A. Tucker (Eds.), *Social media and democracy* (pp. 163–198). Cambridge University Press. https://www.cambridge.org/core/books/social-media-and-democracy/misinformation-and-its-correction/61FA7FD743784A723BA234533012E810?utm_campaign=shareaholic&utm_medium=copy_link&utm_source=bookmark

Yamaoka-Enkerlin, A. (2019). Disrupting disinformation: Deepfakes and the Law. *NYU*. *Journal of Legislation & Public Policy*, *22*, 725.

York, C., Ponder, J. D., Humphries, Z., Goodall, C., Beam, M., & Winters, C. (2020). Effects of fact-checking political misinformation on perceptual accuracy and epistemic political efficacy. *Journalism & Mass Communication Quarterly*, *97*(4), 958–980. https://doi.org/10.1177/1077699019890119

Zhou, X., & Zafarani, R. (2020). A survey of fake news: fundamental theories, detection methods, and opportunities. *ACM Computing Surveys*, *53*(5), 109:1–109:40. https://doi.org/10.1145/3395046

Zimmermann, F., & Kohring, M. (2020). Aktuelle Desinformation—Definition und Einordnung einer gesellschaftlichen Herausforderung. In R. Hohlfeld, M. Harnischmacher, E. Heinke, L. Lehner, & M. Sengl (Eds.), *Fake News und Desinformation*. Nomos Verlagsgesellschaft mbH & Co.

How to cite this article: Hägle, O., Escher, S., Heil, R., & Jahnel, J. (2024). Structuring different manifestations of misinformation for better policy development using a decision tree-based approach. *Policy & Internet*, 1–18. https://doi.org/10.1002/poi3.420