

STUDY DISINFORMATION
DISINFORMATION STUDY
EXCERPT DISINFORMATION
DISINFORMATION STUDY
STUDY DISINFORMATION
DISINFORMATION STUDY
STUDY DISINFORMATION
DISINFORMATION STUDY

DISINFORMATION



DISINFORMATION

RISKS, REGULATORY
GAPS AND ADEQUATE
COUNTERMEASURES

Dreyer | Stanciu | Potthast | Schulz

CONTENT

3 SUMMARY

8 KEY FINDINGS

9 Starting points for countermeasures

10 Options for action, their requirements and relevant
implementing actors:

10 – Area of measures A

11 – Area of measures B

12 – Area of measures C

14 – Area of measures D

15 – Area of measures E

17 IMPRINT

KEY FINDINGS

Disclaimer: The following findings and options for actions are excerpts from a larger legal study that will be published shortly. The full report systematically examines the possibilities and limits of legal intervention regarding disinformation, based on potential risks of disinformation for legally protected rights and goods. Here, the core findings of the study as well as an overview of possible options for action are documented.

1)

Scientific definitions of disinformation focus on the untruthfulness of a statement and the speaker's intention to mislead. From a legal perspective both criteria are difficult to determine and therefore are **rarely suitable for regulatory debates**.

For the study, we therefore use the following working definition: "Disinformation describes utterances,

- the truth of which can be doubted with good reason,
- which can easily be disseminated and shared,
- which due to the person making the statement or due to their design claim to be truthful from an objective recipient's perspective, and
- which impair legally protected rights and goods."

2)

Such statements can be further analysed by applying a number of **dimensions**, enabling a more **differentiated assessment** and policy discourse. These dimensions are

- the type of statement,
- the context of the statement,
- the structure of the actor making the statement,
- the motive for making the statement and underlying incentives,
- the degree of its potential public visibility,
- a recognisable intention to mislead or deceive in a given case, and
- pivotally, its potential risk for legally protected rights and goods (with the dimensions of the immediacy of danger, the probability of a violation, the intensity of the violation, and the importance of the legally protected rights concerned).

3)

In order to answer the question to what extent there is a need for action against the spread of disinformation and what options are available, we first need to identify **legally protected rights** that might be affected by disinformation. We identify the following rights (in some cases additional rights may be affected):

On the individual and group level

- autonomy,
- the freedom of political decision-making and opinion-forming,
- right to free election,
- freedom of expression,
- freedom of information,
- the general right of personality,
- the right to unimpaired personal development, and
- the rights to life and health.

At the societal level

- the freedom of public opinion formation,
- equal opportunities to communicate,
- diversity of opinion,
- the democratic formation of will and the integrity of elections,
- trust in democratic institutions,
- the construction of societal reality and social cohesion, as well as
- public safety and order and public health.

4)

Contrasting the affected interests and legally protected rights with the current legal framework – using Germany as an example – shows that **individual and group-related legal rights are essentially protected against dangers** emanating from (online) communication. The effectiveness of this protection may be subject to criticism, for example with regard to the possibilities of the legal system to deal with the quantity and rate of dissemination of potentially harmful communications on platforms. However, this problem exists in relation to all content that endangers legally protected rights and is not limited to content that involves dangers due to their untruthfulness.

5)

The latter represent a problem category when it comes to the possibilities of platform-specific legal reactions in order to make protection more effective. But if the violation of a legal right specifically rests upon the untruthfulness of, for example, an allegation about a person, an **adequate solution to the legal dispute presupposes that – depending on the burden of proof – it must be possible to prove the truth or untruthfulness**. Procedures such as the NetzDG require platforms to take this decision, although they do not have court-like proceedings to find the truth. An adequate conflict resolution is not only about determining the facts of the case, but also about the question of whether there is a factual claim at all and, if so, what the exact content of the statement encompasses. The incentives for platforms to delete content due to the obligation to check statements lead to the question whether such regulatory concepts can be designed in ways that would be in accordance with fundamental human rights.

6)

With regard to societal interests and rights, specific regulations in Germany only exist for few legally protected interests such as public peace and the counterfactual stabilisation of trust in journalistic content through provisions regarding journalistic duties of care (§ 19 MStV). Any regulatory options in this area face **the problem that in order to assess true/untrue statements, certain bodies would have to be able to determine the truth**. Apart from adversarial proceedings in independent courts, however, declaring the truth is **not a task of the state, but a social process of communicative construction in which substantiation and doubt play an important role**. At least temporarily, shared understandings of “what is the case” emerge in such processes.

7)

The **functioning of such discursive processes can only be guaranteed to a very limited extent by state measures**. If trust in actors or individuals with a particular role in the construction of reality erodes, this has an impact on the democratic self-understanding of a society. This is the case, for example, when political-strategic claims are currently being made in the USA that the presidential election was manipulated, although courts have rejected this. Political actors should be aware of the fundamental danger of changing the practices of reality construction. The preservation of political culture appears to be the central factor in this case. Self-commitments, especially regarding campaign communication, might be an option. However, areas remain where the state cannot enable this social process, but might at least be able to support it through legal frameworks.

8)

Across all countermeasures examined, it appears that **traditional forms of regulation are either not permissible for such discourse-supporting approaches, or are not helpful, or do not appear to be feasible in practice**. One reason for this is that private actors would regularly be responsible for their implementation, who themselves are entitled to basic freedoms when it comes to shaping their offers and contractual conditions. Here, **in addition to classic forms of self-regulation, new forms of “hybrid governance” are needed**, where state regulation and platforms’ own areas of governance are intertwined. State-based control of communication can rarely contribute to solving the problems associated with disinformation; or only at the price of endangering the very freedom that it aims to protect.

9)

Against this background, only **a few paths appear to be expedient when it comes to containing problems of disinformation by regulation**, and only a few of them use the criterion of (un)truth. Apart from the reasons mentioned, the difference between true/untrue is related to the type of statement. However, this does not always correspond to the actual use of language: currently, language seems to be used in a way that underlines a political statement by offensive denial of facts (e.g. “I don’t care that it is disproved, the election was stolen”). Here, the underlying problem is not solved by checking the truth of the statement. Based on the findings of the study the following paths could be pursued further:

- Legal measures based on untruthfulness can be considered (only) if there is **a high probability of immediate danger to the highest individual legal rights such as life and physical integrity**. This includes statements that may mobi-

use a lynch mob or factual allegations that might directly cause self-harm. In these cases, after balancing the legally protected interests, statements might exceptionally first be deleted and then checked for their untruthfulness in legal proceedings. In these cases, public discourse cannot fulfil its function because it might be too late by then.

- Another area of legal provisions can target untrue statements made in **direct temporal proximity to a democratic election**. In these cases, too, society is deprived of the possibility to negotiate the truthfulness of the statement. Where voters can be manipulated in such a way, society's interest in the freedom to vote outweighs the rights of the person making that statement.
- Measures that **make doubts about a statement visible and check the statement** (tagging or labelling, fact-checking procedures, notices and warnings as well as combinations of these) cannot be introduced on a mandatory basis by law with regard to legal utterances, but are subject to the **voluntary measures of the platforms**. However, expectations and ideas can be exchanged through cooperation between the state and platforms as well as among the platforms themselves. If platform providers introduce such measures and procedures, the (EU or national) legislator should provide a legal framework that safeguards fundamental rights regarding these procedures. This might include the obligation to transparently provide information in the terms of use regarding the existence of such procedures and their decision-making processes, possible sanctions and the rights of those affected by these decisions.
- The current legal framework for the counterfactual stabilisation of trust in journalistic content requires compliance with journalistic duties of care. This is one way of **linking statements claiming truthfulness with their increased obligations to seek truth**. If a sliding scale of duties of care is applied, **non-journalistic actors with high relevance for opinion formation** (such as influencers or activists) might also be covered by such obligations. Such measures have the additional advantage that they contribute to stabilising expectations of certain types of offerings. In the light of the variety of content provided on platforms, this is a significant aspect for both users and providers. Due to the potential for abuse of such a regulatory approach the aspects of selection, proof of compliance and sanctions must be designed in a particularly careful, transparent and comprehensible manner to prevent any suspicion of forms of supervision that target specific opinions. With regard to necessary sanctions in the case of violations of duties of care, forms of labelling can be seen as milder means compared to injunctions or deletion orders.
- Access to or duties to disclose platform data, usually a rather generic demand in policy discussions, makes specific sense in the area of disinformation: only then can society learn which discourse-oriented measures to make doubt visible (e.g. by labelling or counter-speech) show impact. Regulation could work towards this, ideally in such a way that an established procedure or data broker does not have to negotiate the conditions of data access for each individual case.

10)

Governing **inauthentic behaviour on platforms** can also be a way to slow down the viral spread of disinformation. However, this approach points to a general problem of platform regulation that can only be touched upon here: **a platform has a legitimate interest in defining what it considers to be authentic communication** by

its users. In this area of “hybrid” public/private governance, new forms of interaction between state and private norm-setting bodies seem effective. In this case, this could consist of government-appointed regulators formulating requirements for authenticity from the perspective of societal interests. However, it would then be up to the platforms to implement this in a more detailed way in their terms of service, and then control their implementation. This instrument is only apparently content-neutral, as certain actors can be recognised through specific patterns of sharing or liking, for example.

11)

A key future challenge in dealing with disinformation is the search for **possibilities of cross-platform measures that aim at making doubt** and fact checking results visible. If the validity of a statement is disputed based on valid evidence and arguments on one platform, interoperable forms of making this doubt visible can help prevent the same statement from remaining unquestioned on other platforms.

12)

Measures that are not specifically related to disinformation but nevertheless compensate for it can be seen in systematically improving the information literacy of children, adolescents and adults, in positively labelling providers who are committed to journalistic duties of care as well as in discussing forms of (more) disinformation-sensitive reporting by traditional media and journalistic outlets.

13)

Governance measures designed to reduce the potential risks of disinformation should not distract from the fact that the increased occurrence of disinformation may have deeper societal causes. If disinformation is a symptom, states and societies can only solve the issue in the long run if they also address such underlying causes in parallel.

OPTIONS FOR ACTION FOR THE RELEVANT ACTORS

The study reveals a number of possible approaches to countering the risk dimensions of disinformation. Four major areas of action that can be identified are: (a) measures to improve regulatory knowledge, (b) measures in cases of objectively falsifiable content, (c) measures in cases of doubts about non-falsifiable statements, (d) measures targeting technological- and distribution-related aspects independently of the content, and (e) Official statements. The following overview (Table 1) summarizes possible countermeasures, the requirements regarding their design and the actors called upon to implement them.

AREA OF MEASURES A

Measures to improve regulatory knowledge

AREA OF MEASURES B

Measures in cases of objectively falsifiable statements

AREA OF MEASURES C

Measures in cases of doubts about not or not completely falsifiable statements

AREA OF MEASURES D

Measures targeting technological- and distribution-related aspects

AREA OF MEASURES E

Official statements

STARTING POINTS OF COUNTERMEASURES

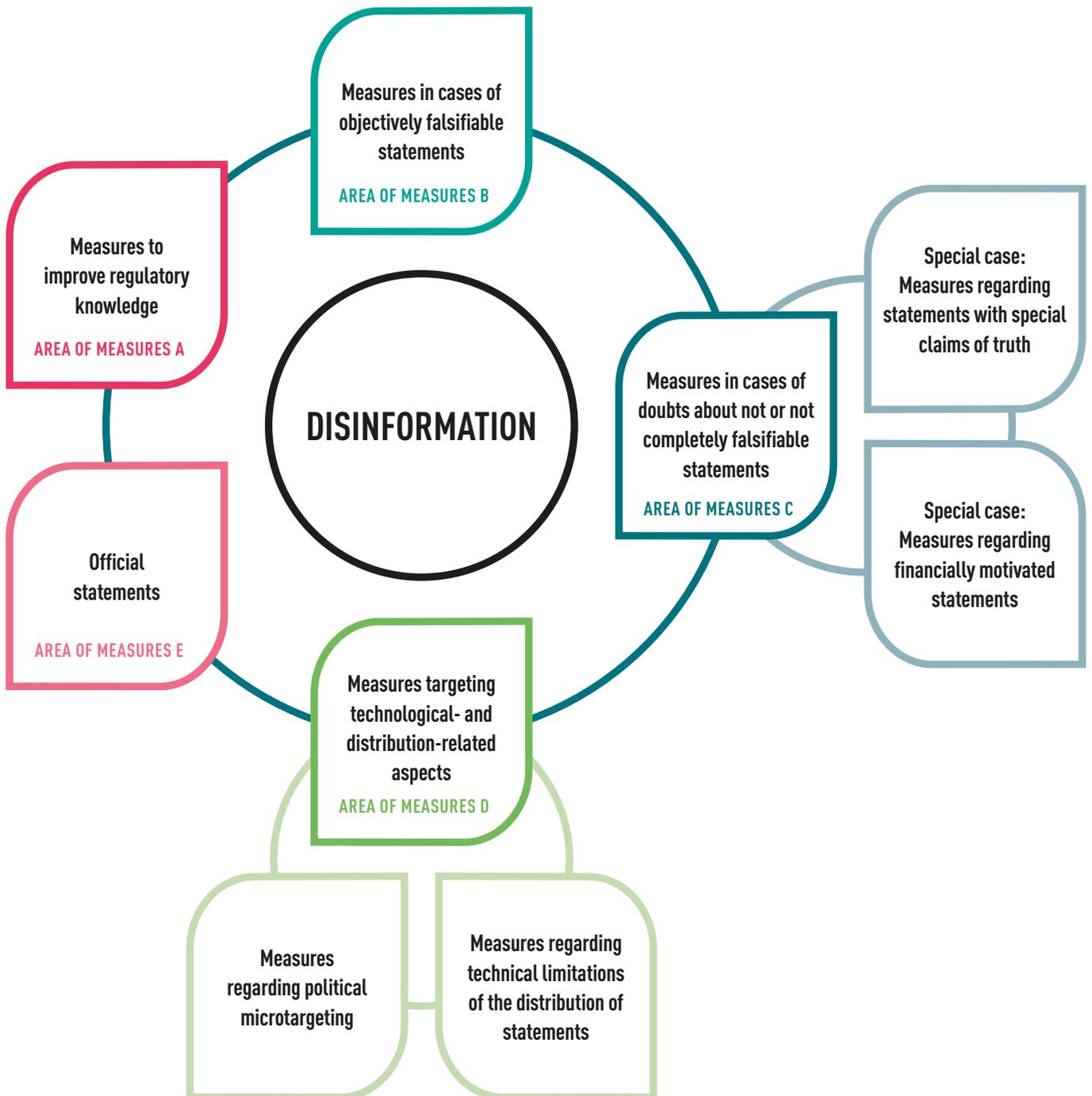


Table 1: Options for action, their requirements and relevant implementing actors

AREA OF MEASURES A

Measure	Requirements and risks	Design	Actor
Information/disclosure obligations; access rights	<ul style="list-style-type: none"> Accuracy of the data is not fully verifiable Access rights can enable validation 	<ul style="list-style-type: none"> Congruent and comparable report structure and data structures Provision of country-specific data 	
MEASURES TO IMPROVE REGULATORY KNOWLEDGE			



* in the case of threats to legal interests with legal (minimum) requirements for the procedure

AREA OF MEASURES B

Measure	Requirements and risks	Design	Actor
Legal prohibitions	<ul style="list-style-type: none"> Legal case-by-case decision regarding the trueness of a claim required Great and imminent danger for legally protected rights necessary (life and health; public safety; free elections) 	<ul style="list-style-type: none"> Court-like, uniform proceedings for each individual case 	
Reservation of rights to delete or block content	<ul style="list-style-type: none"> Accessory to legal prohibitions, or voluntary reservations in terms of use (in case of the latter a threat to legally protected rights is not necessary) 	<ul style="list-style-type: none"> Uniform proceedings; duty to give reasons Option for persons affected to object Transparent procedures 	
Downranking/ Deprioritization	<ul style="list-style-type: none"> Legal provisions only in case of a threat to legally protected rights Reservation of rights to apply such measures on voluntary basis in terms of use (in this case a threat to legally protected rights is not necessary) Potential threat to freedom of expression due to excessive deprioritization Recognizability of deprioritization for affected persons is limited 	<ul style="list-style-type: none"> Uniform proceedings; duty to give reasons Notification of affected persons; option for these persons to object Transparent procedures 	
Tagging of falsified content	<ul style="list-style-type: none"> Content remains available for discourse; no "censorship" Risk of misuse in community-based approaches 	<ul style="list-style-type: none"> Uniform proceedings; duty to give reasons Option for persons affected to object Transparent procedures 	

MEASURES IN CASES OF OBJECTIVELY FALSIFIABLE STATEMENTS



* with legal (minimum) requirements regarding the procedure

AREA OF MEASURES C

Measure	Requirements and risks	Design	Actor
Reporting/ flagging mechanisms	<ul style="list-style-type: none"> ▪ Possible requirement for subsequent fact checking procedures ▪ User-based flagging has high requirements and is prone to misuse 	<ul style="list-style-type: none"> ▪ Uniform processes ▪ Trusted flaggers as an option for more reliable reports and starting point for prioritized processing 	
Fact checking procedures	<ul style="list-style-type: none"> ▪ Dominant position of fact checking bodies ▪ High costs with low scalability ▪ Requirement for follow-up measures (esp. tagging, counter-speech based measures) 	<ul style="list-style-type: none"> ▪ Easily accessible and understandable information regarding the criteria, procedure and follow-up measures ▪ Implementation by independent institutions/bodies ▪ Uniform processes; cross-platform bodies as an option ▪ Option for persons affected to object with external bodies ▪ Transparent and auditable processes ▪ Exceptions for small and non-profit platforms 	
Labelling/tagging	<ul style="list-style-type: none"> ▪ Risk of misuse in context of community-based approaches ▪ Design options: notice or warning, with or without reference to refuting information ▪ also possible at the level of whole accounts or profiles (strong intervention) ▪ Knowledge of effects of labelling/tagging is still limited 	<ul style="list-style-type: none"> ▪ Uniform processes ▪ Option for persons affected to object ▪ Transparent and auditable processes ▪ Accompanying research on effects necessary 	
Addition of contradicting information/debunking	<ul style="list-style-type: none"> ▪ Risk of misuse in context of community-based approaches 	<ul style="list-style-type: none"> ▪ Uniform processes ▪ Option for persons affected to object ▪ Transparent and auditable processes ▪ Accompanying research on effects necessary 	

MEASURES IN CASES OF DOUBTS ABOUT NOT OR NOT COMPLETELY FALSIFIABLE STATEMENTS



* with legal (minimum) requirements regarding the procedure

AREA OF MEASURES C

SPECIAL CASE: MEASURES REGARDING STATEMENTS WITH SPECIAL CLAIMS OF TRUTH

Measure	Requirements and risks	Design	Actor
Obligation to exercise journalistic duties of care in cases of journalistic appearance	<ul style="list-style-type: none"> High conformity with expectations of journalistic functions Danger of misuse a duty of care obligations by the state Potential interpretative power of platforms regarding what is considered journalistic appearance Legal provisions that oblige platforms to monitor/control are not possible 	<ul style="list-style-type: none"> Exclusion of state intervention, e.g. through arbitrary or politically motivated selection of targets Procedures must be independent from the state Guidelines for the distinction of journalistic appearance necessary Development of criteria for providing evidence required 	
Obligation to exercise journalistic duties of care in cases of accounts with high relevance or wide reach	<ul style="list-style-type: none"> Difficulty in determining the threshold for duties of care to be applicable Necessity of a case-by-case assessment 	<ul style="list-style-type: none"> Development of flexible standards of duties of care varying from case to case 	

SPECIAL CASE: MEASURES REGARDING FINANCIALLY MOTIVATED STATEMENTS

Capping of economic incentives/ demonetisation	<ul style="list-style-type: none"> Infringement with the freedom to choose and carry out one's career as well as the right to an established and operating business Demonetisation of entire outlets or profiles based on few false statements is disproportionate 	<ul style="list-style-type: none"> Potential anti-trust issues regarding cross-provider agreements Options for advertisers to select advertising contexts and to subsequent review 	
--	--	--	---

MEASURES IN CASES OF DOUBTS ABOUT NOT OR NOT COMPLETELY FALSIFIABLE STATEMENTS



* with legal (minimum) requirements regarding the procedure

AREA OF MEASURES D

Measure	Requirements and risks	Design	Actor
Prohibition of social bots	<ul style="list-style-type: none"> General legal prohibition disproportionate; restriction to blatant cases of interference with the right to equal chances to communicate Challenging legal criteria since evidence of automation is difficult to supply; general issue of the burden of proof Permission and limitation by platforms possible 	<ul style="list-style-type: none"> Uniform processes Option for persons affected to object Transparent and auditable processes 	
Labelling of social bots	<ul style="list-style-type: none"> Obligations to disclose automated communication is proportionate 	<p>When labelling is carried out by platforms:</p> <ul style="list-style-type: none"> Uniform processes Option for persons affected to object Transparent and auditable processes 	
Real name policies	<ul style="list-style-type: none"> Legal obligation to provide real name seems questionable with regard to fundamental rights 	<ul style="list-style-type: none"> Obligation to provide real name on basis of platform terms possible 	
Positive labelling of persons/institutions with valid special claims to truth	<ul style="list-style-type: none"> Legal obligation only possible in cases where state independent processes have identified actors who can be granted a positive label Otherwise: Strong position of the state or the platform regarding the decision-making power over criteria and verification processes Indirect potentials for misuse (e.g., through positive labelling of only certain outlets) 	<ul style="list-style-type: none"> Cooperation with external, independent verification bodies (Non-binding) state guidelines regarding possible criteria feasible External review of the concretisation by platforms 	

MEASURES TARGETING TECHNOLOGICAL- AND DISTRIBUTION-RELATED ASPECTS



* with legal (minimum) requirements regarding the procedure

AREA OF MEASURES D

Measure	Requirements and risks	Design	Actor
Obligations regarding transparency and mandatory identification for political advertisers	<ul style="list-style-type: none"> Legal requirements possible, but interference with key functions of political parties Challenge: definition of "political advertising" 	<ul style="list-style-type: none"> Focus on visibility and public discourse regarding booked political ads Alternative approach: Limitation of selectable segments in political advertising Self-commitments by parties as a less restrictive measure 	
MEASURES REGARDING POLITICAL MICROTARGETING			

Restrictions of sharing functionalities	<ul style="list-style-type: none"> Strong interference in freedom of expression; legal provisions questionable with regard to human rights 	<ul style="list-style-type: none"> Implementation and limitation by platform possible Counterproductive with regard to social discourse 	
Prohibition or labelling of cases of buying artificial reach	<ul style="list-style-type: none"> Legal provisions seem problematic; Tagging by platforms is legitimate 	<ul style="list-style-type: none"> Uniform processes Option for persons affected to object 	
MEASURES REGARDING TECHNICAL LIMITATIONS OF THE DISTRIBUTION OF STATEMENTS			

AREA OF MEASURES E

Measure	Requirements and risks	Design	Actor
Official Statements	<ul style="list-style-type: none"> Potential for misuse: Possibility of discrediting specific persons/groups/opinions by propagating a certain governmental viewpoint 	<ul style="list-style-type: none"> Restriction to exceptional cases of high and imminent danger 	
OFFICIAL STATEMENTS			



* with legal (minimum) requirements regarding the procedure

The overview (Table 1) shows that legislative options for regulation of disinformation are limited to a few and severe types of disinformation. Where classic legal instruments such as legal prohibitions reach their limits in cases of possible disinformation, further possibilities exist for platforms to shape the rules of communication within their offerings, especially based on their contractual autonomy. However, these margins to shape their platform governance do not apply indefinitely. The (new) power of platforms in shaping public and private communication is bound to take human rights into account when implementing internal processes. However, this is not an insight that applies to disinformation specifically, but a part of the general legal debate in the field of regulating platforms with user generated content (“platform governance” or “governance of platform governance”).

In light of the countermeasures that platforms can introduce to counter disinformation the importance of minimum legal requirements regarding such procedures has become clear. In this field, there is an opportunity for EU and national legislators to develop principles, guidelines, and benchmarks that apply in cases where platforms decide to provide certain countermeasures. By doing so, legislators are able to safeguard the respect for fundamental rights once platforms implement content-related processes. Article 12 of the draft Digital Services Act, for example, seems to be a first step in such direction with regard to the formulation of terms of service. In the medium term the DSA might become a regulatory platform for formulations of legal guidelines which are then given concrete forms and are implemented by private parties, open for their subsequent monitoring and review by socially accountable institutions and bodies. The evaluation exercise of the Code of Practice on Disinformation shows hints that the EU is increasingly thinking in terms of forms of co-regulation in this area; this would be consistent with the line of thinking in this study. The form of societal self-efficacy by using new actors could be guaranteed by different organizational or procedural provisions. One possible form for such bodies can be seen in media regulators that are independent from the state and consist of pluralistic decision-making bodies, such as e.g. the German state media authorities.

IMPRINT

Publisher:

Landesanstalt für Medien NRW
Zollhof 2
40221 Düsseldorf
T +49 211 77007-0
F +49 211 727170

info@medienanstalt-nrw.de
www.medienanstalt-nrw.de

Responsible:

Sabrina Nennstiel (Leitung Kommunikation)
Dr. Meike Isenberg (Leitung Forschung)

Editorial:

Dr. Meike Isenberg, Thomas Wierny

Authors:

Dr. Stephan Dreyer, Elena Stanciu, Keno Potthast, Prof. Dr. Wolfgang Schulz

Design:

Nadine Hawle (NH CORPORATE – Designstudio)

This publication is released under the Creative Commons license
(CC BY-SA 4.0).

→ <https://creativecommons.org/licenses/by-sa/4.0/legalcode.de>