



D5.3

Assessing Externalities: Toxic Debate and Pluralistic Values



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101004488

The information and views in this report are those of the author(s) and do not necessarily reflect the official opinion of the European Union. Neither the European Union institutions and bodies nor any person acting on their behalf may be held responsible for the use which may be made of the information contained therein.

Document information

Grant Agreement #:	101004488
Project Title:	EUROPEAN MEDIA PLATFORMS: ASSESSING POSITIVE AND NEGATIVE EXTERNALITIES FOR EUROPEAN CULTURE
Project Acronym:	EUMEPLAT
Project Start Date:	01/03/2021
Related work package:	WP5 Power: People and Platforms
Related task(s):	5.3 Assessing Externalities: Toxic Debate and Pluralistic Values
Lead Organisation:	P9: ISCTE-IUL – University Institute of Lisbon
Author(s):	Mehmet Ali Üzelgün Ioanna Archontaki Iliana Giannouli Klára Odstrčilová Barbara Thomass Dessislava Dankova Cláudia Álvares
Contributor(s):	Nico Carpentier Miloš Hroch
Status	Final
Submission date:	07/11/2023
Dissemination Level:	Public

Table of Contents

Introduction	4
I. Theoretical reflections	8
1.1 Online toxicity	8
1.1.1 Incivility	9
1.2 Public debate and the sources of normativity	12
1.2.1 Dialectical reasonableness	12
1.2.2 Civility and civil society	14
1.2.3 Pluralism and ethics of communication	16
1.3 Toxic debates	17
1.3.1 Who?.....	18
1.3.2 What?.....	18
1.3.3 How?.....	20
II. EUMEPLAT data and findings	24
2.1 Pluralism in the EU legislation and directions	25
2.2 A slice of Europe on social media.....	28
2.3 Anti-EU fake news and the quest for reliable information: Fact-checking policies	32
2.4 Civil Society on European platforms.....	34
2.5 Summary	35
III. Future scenarios	37
3.1 Method and material	37
3.2 An overview of the future scenarios: “Educate people, not machines!”	40
3.3. Discussion of the future scenarios.....	45
Conclusion	49
References	52

Introduction

The gap between the limited attention of the social media user and the practically infinite number of available messages has become a prime issue of mediated communication. The structure of social media platforms, where the user moves quickly from one message to the next, poses a major challenge for anyone trying to attract the interest of an audience, a phenomenon coined as the “attention economy” (Webster, 2014; Williams, 2018). Perhaps the first thing social media algorithms have learnt, in this economy, is that the more provocative a message to a user, the greater the chances of attracting their attention. Research has shown that news stories conveying emotions of anger and surprise are shared among social media users with greater frequency and speed (Fan & Gordon, 2014; Ferrara & Yang, 2015). The same goes for populist messages that provoke anger (Hameleers et al., 2017; Bobba et al., 2017), as well as emotional posts in general (Stieglitz & Dang-Xuan, 2013). That being so, findings also suggest that platform algorithms enhance emotional, political partisan and polarizing content, particularly tweets expressing anger and animosity towards out-groups (Milli et al., 2023).

In platformed communication designed for exploring emotions and extracting reactions, phenomena grouped under the label “online toxicity” have become a core concern (e.g. Pascual-Ferrá et al., 2021; Rossini, 2019). Hate speech and toxic exchanges have a serious impact on willingness to participate in public debate, the formation of personal and public opinion, and people’s interpretation of polarization around issues of common concern (Anderson et al., 2014; 2018). Besides the huge implications of media ranking algorithms for shaping public discussions and participation, there are also other elements involved in online toxicity. The nature and characteristics of online communication such as lack-of-face interaction, anonymity, and virtually instant access to an unprecedented audience make platforms an ideal environment for toxic encounters. The new speech context social media platforms provide for people to express themselves more freely than in offline settings, a phenomenon dubbed as the “online disinhibition effect” (Suler, 2004), increased the exposure of the public to the toxic effect of online hate speech. In this context, the rise of political polarization and hate-based violence should not be perceived as an abnormality, but as the “logical result of individuals who have spent years inhabiting hate-filled spaces where racist, sexist and anti-semitic views were normalized” (Munn, 2020, p. 3, also Anderson et al., 2014; 2018).

In their analysis of toxicity in online news comments, Gautam and Taboada (2019) suggest constructiveness and toxicity as two principal axes along which we can evaluate the contribution of each public comment to the elaboration of a given discussion. Constructive comments are considered to provide well-founded and reasonable arguments, while toxic comments are characterised by profanity, insults or attacks towards the people involved in

the debate (p. 1). Toxic speech and narratives challenge the premises of democratic deliberation since they “undermine the traditional understanding of reason, violate the traditional understanding of mutual respect, and threaten the traditional notion of inclusive engagement” (Boswell, 2015, p. 315).

As we will develop in more detail in the following sections, online toxicity is not a monolithic concept and can be manifested in various forms from name-calling to insults and threats. This fluidity renders the analysis of online toxicity a rather challenging task. In addition to the problem of definition consensus among different research plans, most models for online toxicity detection are applied to a single platform, leaving limited space for cross-platform evaluation of online toxicity classifiers (Salminen et al., 2020b).

Toxicity is not an on/off phenomenon; it refers to a spectrum of communicative infelicities that unfold at various levels of public debate (e.g., micro, meso, macro). Thus, instead of “toxic” versus “healthy” communication, research may better focus on a gradient composed of notions such as negative sentiment, incivility, clickbait, populism, hate speech, toxic outrage and threats. We will also use the notion of potentially toxic to designate those acts of speech – and repercussions – that are not detectable by the linguistically grounded perspectives. Thus, instead of focusing solely on computational techniques to detect online toxicity, researchers and all interested parties should put more effort into understanding the social phenomena described under the broader scope of toxicity (Mondal, Silva & Benevenuto, 2017, p. 86). The relationship between news topics and online toxicity is a case in point. Research on online toxicity shows that a significant part of the troubled comments is directed to the topic rather than individual users or groups, and that levels of toxicity vary significantly between topics (Salminen et al., 2020). Accordingly, this “topic-driven toxicity” suggests the potential impact the topic selection and the framing of news stories have on the shape and quality of social media discussions.

Even if the best practices of the previous eras of journalism and broadcasting can be applied to the newly emerging contexts, comment threads on platforms are not going to be deprived of acrimony anytime soon. It is therefore also relevant to explore how normative concepts of toxicity relate to the notion of harm. After all, how are we to impose standards on the harmful? For example, a linguistic expression could be considered harmful in itself, due to its illocutionary force or performativity. Alternatively, only clearcut incitements to violence could be regarded as harmful, i.e., with registered consequences. Such concerns are encapsulated in Article 19 of the code drawn up by the European Commission against Racism and Intolerance in 2015, which delivers recommendations to EU member-states on issues regarding discrimination. This take on the limits of free speech voices a tension, according to which speech is not seen as action and, as such, any action taken, based on speech, does not fall under the responsibility of those who deliver or disseminate toxic

speech, but rather of those who act on the basis of such speech (cf. Titley, Keen & Földi, 2014, p. 16).

Finally, as we will address centrally in this report, there is a tension between incivility underpinning toxic debates and incivility integral to political expression and participation. This tension suggests that we can neither get rid of regulation and content moderation, nor apply standard – and automated – restrictions on public debate. There are thus no easy solutions to the predicaments facing online discussions and technologically mediated multi-party debates; users and institutions need to muddle through reflexively and iteratively. For an iterative and innovative process of overcoming some of the worst communicative predicaments impacting democracies today, the aim of our study can be described as twofold:

(1) We explore online incivility and toxicity with the aim of understanding better the dynamics that trigger and bring about toxic debates. This involves the conceptualisation of toxicity as a public-political issue rather than an interpersonal issue of psychological harm. In addition, by proposing that toxicity is an umbrella term that could be better understood as a gradient, we try to investigate which degrees or which aspects of incivility may permit deliberation, which, despite being emotionally laden, still leaves room for rational understanding. Indeed, the presupposition that an ideal speech situation would be rational rather than emotionally charged can be contested on the grounds that fallacies are frequently camouflaged through the rational construction of arguments. The pathos/logos dichotomy thus ceases to hold ground in much the same manner as any distinction between private/public domains becomes increasingly hazy in the present conjuncture of social media use, where the performative effects of language in a private context easily seep into public space.

(2) We also aim to explore the normative concepts that can be useful to examine and tackle the phenomena grouped under the label of toxicity. In this regard, the sources or grounds of normativity are distinguished into interpersonal or micro, intersubjective or meso, and institutional or macro levels. Beyond these levels, we establish a connection between pluralistic values and toxic debates, taking into account that pluralism is intertwined with free speech and inclusive open debate, which may entail the freedom to use language in a way that collides with the sensibilities of those with whom we share public space.

The paper is structured into three main parts. After defining toxicity and incivility, Part I offers some basic theoretical reflections, namely on the roots or sources of the normative concepts of toxicity, and the *who*, *what* and *how* of toxic debates. Part II reviews the relevant research findings conducted in the EUMEPLAT Project to connect these reflections to the present-day experience of platformed communication. Part III presents an analysis of the Future Scenario Essays produced within the framework of the Project's Work Package 5: it offers

a view of the futures of platformed and algorithmically mediated communication and some prescriptions by the experts taking part in the Delphi+ workshops. We conclude by bringing together and discussing the diverse issues and interests that centrally relate to the theme of toxic debates and pluralistic values.

I. Theoretical reflections

1.1 Online toxicity

Captain Vimes believed in logic, in much the same way as a man in a desert believed in ice – i.e., it was something he really needed, but this just wasn't the world for it. (Theatre of Cruelty, Pratchett, 1993)

Online toxicity and hateful commentary are prevalent in all forms of online communication, from news websites and discussion forums (Coe et al., 2014) to social media platforms (Pascual-Ferrá et al., 2021; Konikoff, 2021). Some consensus is required as to what exactly is defined under the term toxic communication, in order to tackle its effects and causes, and to draw substantial guidelines for its moderation.

Due to the bulk of work on the detection of online hate speech (e.g. Burnap et al., 2015; Davidson et al., 2017; Mondal, Silva & Benevenuto, 2017; Waseem & Howy, 2016), toxicity is mostly connected to this notion. Indeed, according to Petlyuchenko and colleagues (2021), toxicity as a media phenomenon is synonymous with hate speech and involves “intentional statements or messages with discriminatory content” (p. 114). In a similar vein, online toxicity is seen to manifest itself in the forms of hate speech, bullying, trolling, harassment, physical threats, and online stalking (Patel et al., 2021). The list can be continued with “obscenity, insults, and identity-based hate” (Adams et al., 2017, p. 1) and “rude language, harsh criticisms, anger, hatred, even threats” (Suler, 2004, p. 321). From these designations, toxicity emerges as a term that designates a variety of acts of speech that have the common feature of harm and denigration, namely of some addressee: “a rude, disrespectful, or unreasonable comment that is likely to make you leave a discussion” (Jigsaw LLC in Pascual-Ferrá et al., 2021, p. 3). In this regard, the literature seems largely focused on the ‘who’, namely parties subject to harm. On this question of ‘who’, a distinction must be drawn between types of platform users, such as individuals and institutional actors. This is particularly important, especially since research has revealed the role of organised cyber attacks and smear campaigns against individuals based on their physical appearance, race or ethnicity, age, gender, sexuality, profession, and political views (Duggan, 2017; Jubson et al., 2020; Posetti et al., 2020; Bradshaw & Henle, 2021; Dunn et al., 2023).

Another concept sometimes used interchangeably with toxic speech is “abusive language” (Nobata et al., 2016; Waseem et al., 2017). This use highlights the verbal violence aspect of online harassments (Guberman et al., 2016). According to Waseem et al. (2017) abusive language over the web can be categorised on the basis of (a) the target of the abuse: a specific or generalised Other, and (b) the nature of the language used: implicit or explicit.

Similarly, comparing hate speech to concepts such as flaming, toxic comments, and abusive language, Fortuna and Nunes (2018) designate specific preconditions. Accordingly, hate speech has (a) specific targets, (b) incites violence or hate, (c) uses language that denigrates, and (d) can be subtle or explicit (Fortuna & Nunes, 2018, 85:5).

Offering an apparently similar but partly different two-fold typology of verbal aggression, Kumar and colleagues (2018) propose looking into (a) 'how' it is expressed: overt or covert aggression, and (b) 'what' exactly constitutes the aggression: physical threat, sexual threat, identity threat with its subcategories (racial, political, geographical, etc.), as well as non-threatening aggression. Thus, while the question of what seems to pertain to the content and context of a particular act of speech, the question of how seems to designate its level of explicitness (Waseem et al., 2017).

In sum, online toxicity can be broadly viewed as an interpersonal communication phenomenon reflecting the violation of personal boundaries and psychological harm (Petlyuchenko et al., 2021, p. 106). The detection and moderation of toxic speech in this framework of violation is a daunting task due to the subjective factors involved in interpretation and the literature is bent on linguistically grounded approaches. One type of the proposed solution employs content analysis, more specifically a specific lexicon of "coarse language", more sophisticated types also use words connected to intelligence, appearance, race, and sexual preference (Rezvan et al. in Sheth et al., 2022). This approach is keyword-based and provides limited options when it comes to automated toxicity detection. Problems proliferate, as in different cultures common words are in certain contexts used as insults: e.g., basic, cancel, Karen, shade, snowflake, and thirsty (Sheth & Kapanipathi in Sheth et al., 2022). The use of slurs also poses a challenge for this approach as they can be commonly used in non-toxic conversations (Wang et al. in Sheth et al., 2022).

While linguistically grounded approaches that operate at the micro-level of speech components are certainly useful in detecting instances of toxic speech, they only scratch the surface of a more complex and multi-dimensional problem. The who, what, and how questions briefly mentioned in this section pose serious challenges to the identification and moderation of toxicity, and will be revisited. We propose that these questions are best answered at the level of the whole debate, therefore we now incrementally expand the scope of toxicity from hate and toxic speech to potentially-toxic debates.

1.1.1 Incivility

Incivility, together with toxicity, finds a place and becomes salient a term in research on social media. In communication studies, the term is broadly used in connection with online

debates and digital discourse, whereas political communication researchers use the notion of political incivility (Brooks & Geer, 2007). Incivility is a concept closely interrelated with toxic debate and online toxicity because of the nature and characteristics of the online environment as well as factors like anonymity, lack-of-face interaction, and the ability to communicate with a large audience quickly and easily. According to Rajadesingan, Resnick and Budak (2020, p. 559) incivility is closest to toxicity. In their comparison of the two terms Hede and colleagues (2021) suggest that incivility is “more subtle and nuanced than toxicity, which includes identity slurs, profanity and threats of violence along other unacceptable incivility” (p. 2620). In the same vein, Hiaseshutter-Rice and Hawkins (2022, p. 2) argue that uncivil language could be conceived as rude or impolite, whereas toxic language seems to be poisonous – used with the intent to harm.

While conceptualization and operationalizations of incivility vary significantly, a common theme is the expression of disrespect towards others (Chen, 2017; Chen, Muddiman, Wilner, Pariser, & Stroud, 2019; Muddiman & Stroud, 2017; Mutz, 2016; Papacharissi, 2004; Rossini, 2019; Shmargard, Coe, Kenski, & Rains, 2021; Sobieraj & Berry, 2011). For example, Sydnor (2019) defines incivility as “any statement that is not respectful of individuals’ desire to maintain their self-image”, and Coe et al. (2014) associate it with an “unnecessarily disrespectful tone toward the discussion forum, its participants, or its topics” (p. 660).

Given this brief description of incivility, one might consider that the remedy for this form of antisocial behaviour in the online public sphere would be to deprive the abusers of their participation rights in online discussion communities. Nonetheless, this form of punishment might generate greater dangers for the quality of online deliberation. Disrespectful language may serve the minority or the discriminated groups who are otherwise not heard at all in public debate (Jamieson et al., 2017), and thus is integral to the formation and transformation of public opinion. Incivility may also serve social purposes among like-minded people and be conducive to reasoned arguments (Chen, 2017; Chen et al., 2019; Rossini, 2019). Rossini’s (2019) contribution to this discussion conceptualizes incivility as a “communicative practice rather than a normative violation” (p. 4). This implies that different forms of incivility exist and that under specific forms of interactions, incivility might be used as a reasonable rhetorical tool. In such a conceptualization, incivility presupposes a distinction based on “the locus of the attack”, e.g., the arguments of other participants in a discussion or particular groups and identities.

Yet, while an uncivil comment may be considered relevant and just, its implications for a particular debate may be detrimental, if not toxic. Gervais (2015) and Kim et al. (2021) offer evidence that exposure to incivility in an online discussion environment induces more incivility in subsequent posts, with the precondition that the original message was aligned with participants’ political views. On the other hand, Han and Brazeal (2015) point out that

this “modelling effect” is distinguishable also when people read civil comments on online platforms, thus, resulting in more civil comments. Research on the impact of online incivility has also revealed that the “modelling effects” of uncivil comments might be less severe than what previous studies suggested, even though there seems to be a notable correlation between uncivil comments and aggressive cognitions (Rösner, Winter & Kramer, 2016). Han, Brazeal, and Pennington (2018) highlight the role of metacommunication in online discussions as a possible form of moderating the tone of the discussion towards acceptable discursive standards, by stressing the discussants’ power to improve the foundations of the online communicative environments. Molina and Jennings (2017) also report evidence of the use of metacommunication as an instrument to encourage online participants to be highly engaged in the discussion while trying to influence the conversation towards a civil tone. More precisely, according to the research findings, even though metacommunication doesn't automatically generate more civil comments on the comment thread, it does have a modelling effect, as more people become engaged in the public debate with the aim of restoring the civility of the conversation. As the authors state, “cues that scold incivility can encourage individuals to engage in a Facebook conversation by fostering more elaboration when processing arguments” (Molina & Jennings, 2017, p. 16).

On the whole, incivility is a relative concept and, thus, one should be critical about the actual benefits of a general ban on uncivil comments on platforms, since, as Chen and colleagues (2019) suggest, the norms and values of civility are decided by people who have the power to define the rules of public discourse. Labelling as “uncivil” every comment that might hold a critical stance against the status quo might drive the voices of already marginalised groups into obscurity. In this context, Mouffe’s recommendation that “a democracy cannot treat those who put its basic institutions into question as legitimate adversaries” (2005, p. 120) can act as a normative guideline as to the counter-publics that should be incorporated into the agonistic public sphere, maintaining the vibrancy of productive conflict alive, in opposition to those groups that, rather than counter-publics, would be more akin to anti-publics (Cammaerts, 2009), the objective of which would be to antagonize and polarize the public sphere, using freedom of expression to incite hatred. This agonistic view on the public sphere stands in contrast to that of Habermas, due to the fact that instead of arguing for rationality that subsumes contraries, Mouffe claims that “public spaces should be places for the expression of dissensus, for bringing to the floor what forces attempted to keep concealed” (Mouffe, quoted in Carpentier & Cammaerts, 2006, p. 973).

1.2 Public debate and the sources of normativity

Designating some communicative acts or encounters as “toxic” and all sorts of moderation of such acts, entails a normative perspective. In order to distinguish the acceptable or reasonable forms of public discourse and criticism from toxic ones – with some degree of legitimacy – we would argue for the clarification of the norms and principles, as well as the source(s) of normativity underpinning the decision of moderation.

Norms pertinent to public debate can usefully be divided into three scalar contexts of action, the micro, meso, and macro levels (Zenker et al., 2023). The micro-level norms apply to linguistic-communicative units, such as speech acts, and these norms aim to promote the smoothness and effectiveness of the interaction between two or more actors (e.g., Grice’s maxims). The meso-level norms apply to the whole debate rather than the molecular units in it, and promote a certain internal goal or output (e.g., to resolve a difference of opinion, van Eemeren & Grootendorst, 2004; to reach a consensus, Habermas, 2003). Complexity is increased, and norms such as open-mindedness, consistency, and fairness are added at this level to soundness, justifiability, dialectical acceptability, and charity. The macro-level of public debate acquires further complexity with issues of participation, institution, and representation: a parliament or courtroom decision binds those beyond its local boundaries, and precisely for this reason, norms that apply to the macro-level of policy are characterised by a procedural component. As it applies to platformed communication, inclusion, participation and (dis-)engagement are the most relevant normative concepts at this macro level.

In this section, we identify three sources of normativity that may apply to platformed debates: *Dialectical reasonableness* pertains mostly to the *micro-level* of toxic speech and commentary. *Civility* pertains primarily to the *meso-level* of a particular issue and the debate around it. Finally, *pluralism and ethics* of public communication, although crosscutting at all levels, pertain primarily to the *macro-level* of institutional norms that lay the foundations of the modern public sphere.

1.2.1 Dialectical reasonableness

The classical concept of dialectic captured the norms and principles of reasonable discussion between two interlocutors. As a method to achieve true knowledge, dialectic developed in intense opposition to Sophistry from Socrates onwards, and the critical rationality underpinning the method had truth as the ultimate value or source of normativity. Dialectics is today still a very much respected discipline among scholars, and home to

contemporary proposals (see e.g., van Eemeren & Grootendorst, 2004), which include attempts to make some basic proposals more relevant to present-day contexts characterised by the multiplicity of players and positions (Lewiński & Aakhus, 2014). Yet, when the discussions on both the broadcast media and the social media platforms are considered, the critical rationality of dialectics appears in a state of crisis. According to the proposal of *post-dialectics* (Paliewicz & McHendry, 2020; Paliewicz, 2022), many norms of critical discussion are eroded to an extent that they no longer apply to a significant part of public discourse, perhaps most evidently on social media platforms. Instead, without any commitment to dialogue, and through “the successful use of verbal aggression, hostility, and control in argumentative situations”, social actors effectively create communicative contexts in which strong evidence and reasoning bite the dust in propelling public policy (Paliewicz & McHendry, 2020, p. 139). Accordingly, present-day discussions, especially those fueled by populism and authoritarianism, “flow through networks, forces, and affects”, rendering the cartography of those networks and forces the critical task ahead (Paliewicz & McHendry, 2017; Paliewicz & McHendry, 2020, p. 139).

Does the crisis of dialectics, or rationality at large, suggest that norms of reasonable argument are outdated and useless on the whole? We would argue not. With new technologies and new affordances come new possibilities of action and interaction, and new predicaments attached to them. Our task is to collectively draw the boundaries of the democratically acceptable, for instance, with regard to “click speech” (Sklan, 2013). Recognising that we may be witnessing a “paradigm shift” of deliberative norms and discursive ethos, Sklan suggests that “if in this new Internet era civic social networks are the new public square and social media sites are the new water cooler, then a click of a button surely constitutes speech” (2013, p. 389). To provide further detail, let us distinguish what may be called toxicity as a verbal manifestation of deep-rooted bias aiming to diminish or harm the self-image of the opponent (Sobieraj & Berry, 2011; Sheth et al., 2022) from the strategic use of incivility as rhetorical practice (Rossini, 2019). The notion of informal fallacies captures such biases as failures and with its Latin roots in the notion of deception, has both unintentional and intentional aspects.

To update the norms of geometrical rationality (originating in logic) and offer a new perspective to public argument as a process, pragma-dialectics resorts to discussion rules that are grounded in critical rationality (van Eemeren & Grootendorst, 2004). In its focus on the shape and procedure of argumentative discussions, pragma-dialectics offers a procedural – instead of logical – account of fallacies. Due to their direct relation to political debate and conversations, let us only mention two well-known fallacies as examples: ad hominem and strawman.

Personal attack or ad hominem fallacy, known as a fallacy of relevance, is regarded as a

violation of the first principle, called the freedom rule (van Eemeren & Grootendorst, 2004, p. 190): “Discussants may not prevent each other from advancing standpoints or from calling standpoints into question”. Attacks targeting persons instead of their positions, whether it is their gender, age, or other personal qualities that are not directly relevant to the disagreement, are thus viewed as prevention from freedom all discussants are endowed with unconditionally (see also Walton, 2006). The second exemplary fallacy, well known as strawman, concerns a misrepresentation of someone's position in order to easily refute that position (Oswald & Lewinski, 2014). Whether strawmanning is done unwillingly or strategically, it essentially aims to downplay and delegitimize an opponent's position, for instance making it look extreme and ridiculous. Such a move is regarded as the violation of pragma-dialectics' third principle, known as the standpoint rule: “Attacks on standpoints may not bear on a standpoint that has not actually been put forward by the other party.” Principles of the kind exemplified can inform both the analyses and the design of debates, as whether unwitting or intentional, violation of such norms can be regarded as failures in reasonable communication at the micro-level.

1.2.2 Civility and civil society

Civility as an expression of mutual respect has long played a central role in online public sphere research and offers a central normative dimension to assess the democratic quality of online discussions (Friess & Eilders, 2015). The normativity of the notion is rooted both in deliberative theory, which postulates that mutual respect increases the openness for opposing arguments (Kies, 2010), and in liberal notions of communicative restraint, which aim to prevent escalation of conflicts (Ackerman, 1989; Cheng et al., 2017; Shmargad et al., 2021).

The notion of civility cannot be fully grasped without placing it in the context of the modern public sphere and civil society. In return, the notions of the public sphere and civil society cannot be fully grasped without the civil rights movements – e.g. feminist, environmentalist, antifascist – of the end of the 20th Century. The place they claimed outside the structures of the political parties was framed as civil society – understood as the “enlargement of democracy” (Arenhövel, 2000, p. 55, transl. B. Thomass).

This enlargement may be regarded from two perspectives of civil society, emphatic and moderate (Arenhövel, 2000, p. 55). The emphatic version of civil society is guided by the normative idea of enlargement and enrichment of democracy and of democratization in its manifold possibilities. Dubiel identifies the civil society approach as a highly normative concept directed towards a “utopian program” (Dubiel, 1994) that aims at the deepening of

democracy and the transformation of the societal status quo. On the other hand, there is a more moderate perspective of the notion, used to describe given activities of citizens beyond the classical democratic obligations to elect representatives (Kocka, 2002; Pankoke, 2004). The moderate version regards civil society as a link in complex societies that facilitates cohesion in times of globalization and uncertainty and reaffirms some feeling of tradition and national belonging; it is especially important in societies that are not democratic. In contrast, the emphatic version underlines new possibilities of citizens' participation and taking over responsibilities. This emphatic version is also connected to the idea of compensatory activation of citizens, where provisions of the welfare state are reduced because of globalization and to social movements which criticize globalization as aggressive neoliberalism. The emphatic interpretation of civil society is supported by the examples of best practices in citizen journalism (Crespo et al., 2022).

This fluid sphere beyond state and market, and the private circle of friends and family, constantly changes form, expands and retreats in response to pressure, and has at the same time persistent manifestations. From trade unions to self-help groups, "civil society is not a homogenous and united entity, but rather a complex arena where diverse and often competing values, ideologies and interests interact and power struggles occur" (Mati et al., 2010, p. 20–21). As civil society actors tend to bring into the political debates those claims and interests which are missing within the formalised institutional positions, they expand the scope of arguments and enrich the debate, and bring in citizens and perspectives not represented by political or economic elites. Thus, while civil society may include both 'civil' and 'uncivil' actions, it deepens the democratic processes of a society (Putnam et al., 1994).

Examining the differences between civility and politeness, Papacharissi (2004) notes that civility depends on the existence of repercussions on the common good, instead of private individuals. Indeed, bad manners – or impoliteness – may be unpleasant but hardly represent a threat to democracy, due to the lack of impact on the collective. Nevertheless, should offence be directed to a social group to which one of the members of the discussion group belongs, then indeed this attempt at exclusion can be seen as toxic due to potentially endangering ideals of participatory and democratic debate.

Since civility as a normative concept seems to apply to the public sphere variably, how can it be used to inform the research and policies on toxicity? One way to appreciate the relevance of civility in this context goes through the notion of meta-level criticisms (Krabbe & van Laar, 2011). Accordingly, unlike ground-level reactions in a debate that takes aim at the issue of the debate, meta-level criticisms focus on the argumentative situation, the discussion itself, the persona, or the strategies of the arguers. By this definition, meta-level criticisms are reactions to comments or posts, they cannot start a conversation. Arguably, the normative force of civility applies differentially to ground-level and meta-level criticisms,

as it is regarded as a prerequisite for preventing cross-cutting dialogue from spiralling into an exchange of insults (Cheng et al., 2017; Gervais, 2015; Shmargad et al., 2021; Ziegele et al., 2018). In other words, critical posts of civil society actors that take issue with societal problems and issues of common concern may be regarded as legitimate even when involving uncivil elements, whereas comments and criticisms with similar linguistic features that take issue with preceding posts may be regarded as having a place in the gradient of toxicity.

This discussion suggests that civility can be seen as an actor's consideration of adversarial positions with a view of the common good. The salience of civility as a concept in addressing and evaluating the problem of toxic speech can thus be regarded as having partially ethical grounds. This focus in literature can be expanded by considering other sources of normativity, namely institutional and societal ones.

1.2.3 Pluralism and ethics of communication

Applied ethics always become necessary when new possibilities for action arise as a result of scientific and technical developments, and with them new problems of evaluation (Funiok, 2002, p. 41).

To address the question, "What should we do?", ethics refers both to procedural regulations and moral concepts, and seeks to justify the meaning and the binding nature of directives for action. While ethics examines the structure of morally correct action and claims to provide insights for guiding action (Pieper, 1991), norms are the outcomes of ethical reflection. As norms stipulate what to wish or to avoid and have the function of regulating the co-existence of individuals, norms of public communication can also benefit from ethical reflection.

A historically new condition that has given communication ethics social importance is online communication, which seems to sweep away all previous ethical postulates and break all regulations. While the applied ethics of public communication in journalism can rely on the fact that professional communicators are integrated into professional structures, have established bodies and codes of self-regulation and that ethical content is – more or less – present in their respective training, the ethics of online communication is free of such professional-structural prerequisites.

However, the institutional norms that had been developed in the context of legacy media still form the basis for evaluation standards for public communication – only that their enforcement now follows completely different rules and prerogatives. The conflictual nature

of these norms has increased in view of the globalised dimension of public communication. These conflicts unfold against the backdrop of conflicting goals of communication: freedom of opinion, democracy, pluralism, and protection of privacy are not necessarily in line or conflict with universal access, cultural autonomy, priority of religious values, and national ethos, just to name a few (Thomaß, 2013).

Pluralistic values are grounded in the self-understanding of modern democratic societies, which are composed of a multiplicity of actors and their interactions, and whose continued existence depends on the quality of communication in their debates on relevant issues. Pluralism thus emerges as a fundamental value underpinning norms of discussion and debate on all three levels discussed in this section. We would like to argue, however, that rather than leaving it to the private sphere of individual self-regulation and ethics, pluralism should be regarded as a design principle and a community ethos in tackling toxicity. With a broader understanding of communication ethics, which also deals with questions of implementation and enforcement of norms, ways can be shown how public communications could be organised according to pluralistic values.

Platformed online communication has the capacity to offer what appears to be a limitless public sphere of deliberative democracy (Hummel, 2016). Yet, studies grounded in different disciplines depict a reality far from the ideals of deliberation and democracy (e.g. Mondal et al., 2017; Paliewicz & McHendry, 2020). This section identified a series of norms that apply to online discussions and debates and built a case for their implementation in the design of social media platforms. An ongoing debate on these and other relevant norms is required to make it possible to tackle toxic debates, namely with regard to the underlying values and the violation of recognised and accounted-for standards.

1.3 Toxic Debates

While toxic speech, as uncivil comments of a speaker towards a particular addressee, appears primarily as an issue of the micro-level, the notion of toxic debate, as unfolding across various participants and positions around a particular topic, points to issues and norms at the meso-level. There is also space to argue that conflicts on social media platforms sometimes attain a significant part in macro-level socio-political debates, thanks both to the quality of their participants (e.g., a political figure such as V. Zelensky taking part in it), and to their quantity or mobilization (e.g., a hashtag such as “EUvaccineXX” becoming a top tweet). There is then space to apply all the norms discussed in the previous section to the notion of toxic debate. In the three subsections of this part of the report, we propose to examine the toxicity of debates encompassing an exploration of three key aspects: the

individuals involved (who), the topics discussed (what), and the methods employed (how).

1.3.1 Who?

If Section 1.1 on online toxicity depicted the phenomenon as primarily *inter-personal*, resulting from the addresser-addressee relations in public communication (Petlyuchenko et al., 2021; Muddiman & Stroud, 2017), this was in order to distinguish toxic debates as a primarily socio-political activity, characterised by unequal relations of power, discrimination, and political exclusion. In other words, the relevance of the 'who' question is surely not circumscribed by the micro-level norms. Especially in cases where the actors hold institutional powers, such as a president posting a negative comment concerning a particular person due to her gender, the act of speech bears on norms that pertain to macro-levels.

While most studies pivot on the "locus of the attack" (Rossini, 2019), namely the target or the victim, it is also important to look at "who speaks". If the concepts of the public sphere and civil society were initially euphemistically associated with "democratically valuable" actors, it turned out in increasingly fragmented and polarised societies that actors hostile to democracy also raise their voices. Toxic debates are thus closely related to the question of "who speaks". Whereas platforms, especially the so-called social media, were once seen as offering hope that civil society actors would have more forums at their disposal and thus be democratically valuable, it is becoming clear with precisely the more complex understanding of civil society that these forums are accessible to very different actors and thus provide gateways for toxic debates.

On this account, Section 1.2.2 has already suggested that a distinction must be drawn between two types of platform users: private persons and institutional actors. This is simply on account of their differentiated representativeness, visibility (thus impact), and accountability, broadly. But also, the "position" or "character" of these actors in the particular socio-political conflict interwoven around a topic is relevant in the detection and interpretation of toxicity. This is crucial in accounting for the authority and power relationships surrounding toxic speech, as legal entities such as corporations and governments, as well as their particular representatives, have a different footing than ordinary users in platformed debates and conflicts.

1.3.2 What?

We have so far explored the 'who' of toxicity as produced by someone (attacker) and as acting upon someone (a target identity or persona). While such a view of toxicity is certainly helpful in detecting and understanding the speech act of toxic comment, further context is a requirement to make a fuller sense of it. To complement the contextualization of the toxic encounter, the 'what' of toxic commentary, namely the topic and the very content of the debate is a necessity, namely to stipulate what goes between the attacker and the target.

In a study on the connection between online news topics and toxicity of user comments, Salminen and colleagues (2020a) propose the notion of "topic-driven toxicity", suggesting a view of toxicity as directed to the topic rather than its participants. Accordingly, both the topic selection and the framing of news stories have an impact on users' responses on social media platforms, suggesting that media organisations as well as other content creators can unintentionally or deliberately provoke hostile reactions. Journalistic practices enabled by technology such as clickbait journalism and manipulation of search rankings are also considered among the factors that promote toxicity online. Journalists are thus held accountable for the news framing they provide, since they should "be aware of the content topic's inflammatory nature and possibly use that information to report in ways that mitigate negative responses rather than encourage them" (2020a, p. 17). This research also reveals that topics with political connotations are more divisive for the online community: key topics such as the environment, health, politics, race, religion, and sexism, among others, generate more negative user comments and hostility.

It is possible to characterise toxicity associated with such topics as situated between the uses of incivility as a rhetorical strategy (Rossini, 2019) and deep-rooted toxicity connected to identity-based hate. In regard to the first, Hopp and colleagues (2018) offer a view of confirming online political incivility as "a conscious act to disrespect those perceived as oppositional others when conducting political discussion" (p. 12). The act of incivility involved has as its target a position in around a topic; this presupposes an understanding of the position and the politics interwoven around such an understanding. In regard to the second, where the opposition is mostly sedimented and crystallised into identities, the topic's political relevance is somewhat suppressed: it is the lack of a common ground that characterizes such toxic acts, but as mediated by the subjective histories projected on the topic. In other words, the level of polarisation is a determining factor in understanding and moderating such toxic exchanges. Indeed, in a study conducted by Kim et al. (2021) analysing Facebook comments, the authors found that individuals who frequently comment on Facebook tend to exhibit a higher level of political interest, possess more polarised viewpoints, and are more prone to employing toxic language in their comments during an elicitation task.

How can the notion of topic-driven toxicity be better understood beyond some polarised issues such as abortion or the Israeli-Palestinian conflict producing more toxic encounters?

Goodwin's (2020) "system-level" investigation of communicative activity in climate change controversies offers a view of the complex argumentative relations unfolding in societal debates. Accordingly, at the system level, arguments are "abstract entities which cannot be equated with the specific makings, 'products' or 'speech acts' that instantiate them" (Goodwin, 2020, p. 169). The "hypocrisy argument" directed at climate scientists, for example, consists of several instantiations, but denotes a well-known position established by the sceptic/contrarian camp in the debate. Goodwin's analysis shows that, within a given societal debate, particular arguments take place in relation to such established positions, which constitute the "argumentative content knowledge" (Goodwin, 2019) of that debate. This topical knowledge of a debate or controversy is not about the knowledge in a particular field such as atmospheric science, it is simply about what participants do – or have done – in a particular debate, i.e., how that content knowledge is deployed in a controversy (See e.g., Üzelgün et al., 2016).

The topical knowledge of a debate may involve inflammatory content that can be exploited easily. A "standing standpoint" (Mohammed, 2019), for instance, refers to well-established positions in a conflict such that, even when not explicitly expressed as a standpoint, may be called out in association due to belonging to this background knowledge. An example can be drawn from the case of Turkey's Kurdish question involving violent conflict: a televised witness testimony that has the features of an argument from pity, and that draws on human rights, can without reservations be portrayed by the politically parallel media as an act of terrorism, on the account that the testimony aligns with the terrorists (Üzelgün, 2023).

The question of what, pertinent to the content and context of the act of speech in public-political debates, can be useful in examining particular topics that grow into toxic debates online. It suggests that topic-driven toxicity or toxicity towards adversarial positions in a debate can be better understood in relation to communication ethics and norms. Those actors with the power to frame the issue and offer starting points have particular responsibilities, namely in offering common grounds and using inclusive categories and language.

1.3.3 How?

As mentioned in the section on online toxicity, the question of 'how' pertains to the degree of explicitness of a particular act of speech (Kumar et al., 2018; Waseem et al., 2017). However, as it applies to toxic debates, the degree of explicitness may work in manifold ways in its contribution to the exacerbation of a conflict, intersecting with expertise, topic, and cultural norms. The use of incivility as a rhetorical strategy (Rossini, 2019; Zarefsky,

2016; also Paliewicz & McHendry, 2020; Paliewicz, 2022) as opposed to a manifestation of deep-rooted racist or sexist bias, for instance, suggests that the level of implicitness or explicitness of an act of speech depends on the whole background of the controversy as well as the rhetorical situation. This means that the degree of explicitness is a complex phenomenon that can be better addressed at the level of the whole controversy, or the issue, and even as an inter-issue phenomenon (Mohammed, 2019).

As it applies to the notion of toxic debate as a whole, the 'how' question involves not only the levels of explicitness/implicitness of the message, but also the contexts and the processes through which people become toxic commenters. This involves not just a momentary failure in an otherwise well-functioning interaction, but the involvement of the whole person in a debate and the material conditions interwoven around it. As Rajadesingan, Resnick and Budak (2020, p. 559) argue, toxic behaviour is not "an isolated phenomenon but a consequence of more structural factors" that have to do with each platform's design and specific traits, their policies to moderate users' content and community culture.

Users don't introduce incivility blindly into public discussions, but they are "sensitive" to other people's actions. In this regard, Shmargad and colleagues (2022) suggest the relevance of injunctive norms: when incivility is rewarded with Up Votes, implying that a commenter's viewpoint is endorsed, this may influence the perception of what is acceptable in the general discussion context. Similarly, drawing on a Social Learning Theory perspective, Hmielowski, Hutchens and Cicchirillo (2014) also argue in favour of a socialization process in which the engagement in online political discussions teaches people to see, for example, flaming as an acceptable form of behaviour.

Norms and exemplars of online conduct can be connected to emotional contagion, particularly the effects of emotional charge in the expression of specific opinions and stances on various issues (Kim et al., 2021; Kramer et al., 2014; Larsen et al., 2015; Li et al., 2016; Sailunaz et al., 2019). A user's emotional state can elicit similar effects in another user without the latter being aware of it. A noteworthy example is the study by Kramer et al. (2014), which demonstrates how emotional states can be transmitted to others through emotional contagion. In an experiment involving Facebook users, researchers tested whether emotional contagion could occur even without direct user interaction by manipulating the amount of emotionally charged content on the users' homepage. When positive expressions were reduced, people generated fewer positive posts and more negative posts, and vice versa. Han and Brazeal (2015) also provide evidence that when people are engaged in discussions in line with the norms of democratic deliberation, they are more inclined to produce civil comments themselves. Similarly, Kim and colleagues (2021) found that exposure to toxic language in comments increases the toxicity of subsequent comments as well. These findings indicate that the emotions expressed on

Facebook impact others' emotions, providing experimental evidence of large-scale contagion through social networks.

The utter differences online users have in their situated contexts and topical knowledge (Goodwin, 2019) in participating in the very same discussion may warrant a pessimistic view, as their understanding of the same discourse may differ in unexpected ways. The interaction of contagious emotions with complex content only complicates the issue. The interdiscursivity built into the very notion of the hypertext renders the view even more complicated. To illustrate this, let us imagine a social media post addressed to a friend with the speech act of promise to join together in a certain protest. The felicity conditions of the speech act may be well met between the two friends, but for the third parties witnessing such an online communication act, the implications of such a post may depart significantly from the context of the act, interpreted for instance as show-off, threat, or a commitment to another established position. Thus, while such a message may be regarded as non-problematic at all by its producers, it may acquire quite different implications as it travels across contexts and interacts with different posts.

This means that the acts of speech that are classified as non-toxic due to not having any hate or harm aspect may still stir strong emotions and take a role in toxic debates. To highlight once again, toxicity at this level may be “nobody’s fault” in the sense that it cannot be easily traced to specific toxic comments. Instead, it slowly builds up around sensitive topics that may have different connotations in different communities, geographies, and cultures. A useful concept in this regard is deep disagreements (Fogelin, 1985) that ensue from deep-seated beliefs and preferences that one does not consider to be questionable (Godden & Brener, 2010). In other words, deep disagreements arise from the lack of sufficient common ground which would allow for a productive disagreement and perhaps even its resolution. The problem with standoffs and persistent disagreements is that the interaction cannot just stop: Since the parties cannot extend the arguments together, they can only provide meta-level criticisms and evaluations of the available positions, reframing the terms of the debate and what is at stake at each turn. If persistence is a relevant condition for a debate to be recognised as toxic (Cinelli et al., 2021), a better understanding of persistent disagreements may help in examining toxic speech and encounters.

The EUMEPLAT research focuses on the socio-political toxicity of political figures and other actors on social media platforms. As already discussed, online platforms are structured to prioritize engagement and interaction, often at the expense of reasonable and nuanced discussion. The vast majority of comments on social media or news websites do not reach the public at all, except the few that “go viral” due to the organic spread, algorithmic amplification, or the interaction between the two. It is therefore important to consider whether “popular” comments are particularly toxic compared to unpopular ones. Likewise, following

the connection identified between online news topics with political connotations and increased levels of toxicity of user comments (Salminen et al., 2020), it is important to pay attention also to the salient topics around which toxic speech concentrates.

II. EUMEPLAT data and findings

EUMEPLAT Project's Work Package 5 aims to highlight the main issues that are expected to affect European landscape, culture and legislation in the years to come. To this end, it draws on the research findings of previous work packages, which are re-evaluated under the theoretical scope of the theme of Toxic Debate and Pluralistic Values. A note of import before moving on with EUMEPLAT research findings that are connected to the theme of Toxic Debate and Pluralistic Values: the research to be reviewed has not been designed to detect or study toxic speech, as identified at the outset of this document. In other words, this part of the report is focused on findings that are not directly derived from research built on key concepts related to the theoretical study of toxicity. Yet, they partly are still relevant for the purposes of our study, since they disclose issues and predicaments that might jeopardize the future of reasonable public communication. In view of the discussion carried out in this section on toxic debates, we look at toxicity as a spectrum of communicative troubles that unfold at various levels of public debate, and in contrast to pluralistic values embedded in European culture and institutions. Thus, this part mainly focuses on the pluralistic background of the European media and public sphere, as well as "who speaks" in the platformed debates with toxic elements.

In section 2.1 entitled *Pluralism in the EU Legislation and Directions*, we start by providing a brief description of the European legislation that aims at tackling media concentration and safeguarding media pluralism, derived from the deliverables from WP1. This provides the background of pluralistic values in European media culture and institutions, helping to better understand the institutional development and innovations at the face of new technological innovations and new social problems.

In the next section 2.2 entitled *A Slice of Europe on Social Media*, we selectively present data from WP2, which highlight some trends and insights on the social media content regarding Europe and European issues of public concern. As the relevant findings reveal, polarizing and emotional content, mostly exploited by populist actors, holds the lion's share of the online public sphere. Polarizing issues such as COVID-19 vaccination can trigger spirals of toxicity, augmented by governmental measures against the anti-vaccination camp. Understanding the type of content that platforms foster, the structures of conversation they promote, and the actors or personas more likely to initiate inflammatory discussions, is conducive to revealing the mechanisms which provide a context for uncivil and toxic comments to proliferate.

In sections 2.3 *Anti-EU fake news* and 2.4 *Fact-checking policies and strategies*, we briefly discuss data on misinformation and strategies to combat the "pandemic" of fake news,

respectively. In polarizing issues, social media platforms seem to provide an ideal space for misinformation or disinformation to spread, and there are instances where bits of unverified information or unreliable sources have been used from one or the other side of a given debate to make a solid argument. It is this kind of information that users should be prepared to evaluate and debunk if they aspire to be a part of a public sphere that promotes the ideals and the norms of a truly participatory and well-informed deliberative space.

In the last section of this part, *2.5 Civil Society on European platforms*, we explore the place of democracy-promoting civic actors in the most popular social media content. This draws on the idea that, concerning pluralism and toxicity, it is crucial to know the extent to which civil society actors are faced with toxicity and how they deal with toxic debates. At least within the temporal and topical slice our WP2 had, it appears that quite different actors than the democracy-promoting ones are the main orchestrators of the online discussions.

2.1 Pluralism in the EU legislation and directions

Pluralism, as a value and a source of normativity as outlined in 1.2, faces a dilemma that lies in the Janus-faced nature of media content as being a public good, serving the democratic, social, and cultural needs of society and being a private market good serving the profit interest of its producers. Looking at the same object from two different perspectives shows very different logics with conflicting demands. How this has been dealt with from a legal perspective and for constituting a common base for all EU member states is explained in this section. As reviewed in EUMEPLAT project's deliverable [D1.4 of WP1](#) entitled "European Media Legislation: Overview", in EU media policies this translates into a conflict of objectives of, on the one hand, strengthening media pluralism through strict antitrust measures versus, on the other, trying to nurture European media heavyweights that can compete in the global market. One can observe this throughout the development of the discourse on European media legislation over decades. The market-focussed Europeanisation of media entailed rules on pluralism and a level playing field – but not on content-related diversity.

With US-American companies dominating the global platform market, emerges a rerun of the "invasion" rhetoric of the 1980's commercial broadcasting in Europe. The negative effects of platformisation are attributed to foreign technological hegemony which is not informed by European values: surveillance, misinformation, polarization of debates, toxic online discussions impacting a pluralism of opinions, exclusion of independent voices from web monopolies, spread of anti-EU disinformation, intransparent algorithmic control

optimised to serve commercial monopolies, electronic pollution. This is countered by emphasizing positive effects like European co-productions and cross-European success stories, the Creative Media Programme, virtuous grassroots phenomena and media narratives and practices able to bring people out of the information bubble (see Vaccari & Valeriani, 2021), but most of all with the narrative of regaining European sovereignty.

A centrepiece of the EU media law framework aiming at the safeguard of pluralism is the first Television without Frontiers Directive (1989), followed by the Audiovisual Media Services Directive (AVMSD, 2007). Public Service Media (PSM) is an important element in safeguarding pluralism in media, therefore the EU permitted the member states to fund their national PSM under an exception to the state aid rules – introduced in the Amsterdam Protocol (1997). Being preoccupied with the Common Market, the EU developed its own body of competition law. Yet, even though a 1992 Green Paper analysed the need for special competition rules in order to safeguard the pluralism of media and thus diversity of information and opinion in society, this issue has been too controversial to produce any regulation to date.

The Television without Frontiers Directive, adopted in 1989, renewed in 1997 and transported to national law by the member states, had as aims to guarantee the free broadcasting of TV programmes in the EU and safeguard cultural diversity, the right of reply, consumer protection, and the protection of minors. This legal framework of broadcasting which was provided by the TWFD and by competition rules is aimed at preventing the development of oligopolistic and monopolistic market structures. A fundamentally revised directive came into force in 2007, covering under the title Audiovisual Media Services Directive all audiovisual media but did not change its content in substance. The importance of pluralism is addressed in Recital 25, which explains that “content of general interest” is “under defined general interest objectives such as media pluralism, freedom of speech and cultural diversity.”

In preparation for the revision of the 2007 AVMSD, an evaluation of its current version was conducted in 2016 to evaluate its performance. As to media freedom and pluralism, the evaluation concluded that the Directive's rules had been only partly effective, due to the differences in independence and effectiveness of national regulators.

Another important decision to safeguard pluralism is the Amsterdam Treaty, which resolved the uncertainty of the status of the public broadcast fee, not as a proper exception to the prohibition of state aid, but in the form of a Protocol on the System of Public Broadcasting in the Member States, which considered “that the system of public broadcasting in the Member States is directly related to the democratic, social and cultural needs of each society and to the need to preserve media pluralism” (Protocol No 29 TFEU). In further documents and

resolutions, the importance of PSB for pluralism, as well as creating a dynamic, digital environment, was underlined several times.

As media concentration has since long been an ongoing threat to the high-standing value of pluralism, the Commission issued its 1992 “Green Paper on Pluralism and Media Concentration in the Internal Market”, pointing out that “protection of pluralism as such is primarily a matter for the Member States” (p. 7). However, there is no common understanding but a variety of uses of the word ‘pluralism’ (p. 14). From the rulings, the Commission did identify two common features: (i) The concept of pluralism serves to limit the scope of the principle of freedom of expression. It allows “to refuse a broadcasting licence or permission for the takeover of a newspaper, a monolithic corporate structure, a holding in a media company, etc.” (p. 15). (ii) The purpose of such limitation is to guarantee diversity of information for the public. If the application of the principle of freedom of expression would result in preventing another beneficiary of that freedom from using it, thus diminishing diversity, it may be necessary to limit that application.

The 1992 Green Paper on Media Pluralism produced no consensus on EU media pluralism legislation and a lot of controversy. It seems to have fallen in between two contrary trends. On the one hand, virtually all EU Member States had restrictions on media ownership, implying agreement on the need for special rules beyond general competition law in order to safeguard the pluralism of media and thus diversity of information and opinion in society. On the other hand, the deregulation of broadcasting since the early 1990s had led to increasing concentrations across media and territories. There had been European conglomerates (e.g. News International, Bertelsmann, Hachette and Fininvest) before. In the new dual system of public and commercial broadcasting and the growing influx of US-American content, MS became less willing to restrict domestic media ownership, hoping to promote economies of scale. At the same time, pressure against limiting concentrations grew, particularly by press publishers.

Safeguarding pluralism and promoting competition therefore appeared as conflicting aims. The European Parliament sided with pluralism and diversity and called for action to address the growing number of media concentrations. The Council of Europe also consistently argued that pluralism needs to be protected as essential to the principle of freedom of speech. But despite all these fundamental statements, a regulation addressing pluralism in the sense of combating media concentration failed. While there are no specific EU rules on concentrations in the media sector, only the general EU competition law has regularly been applied to media.

However, there is progress in monitoring media concentration. The CoE's European Audiovisual Observatory (EAO), the Centre for Media Pluralism and Media Freedom (CMPF)

at the European University Institute, along with the Media Pluralism Monitor and the Euromedia Ownership Monitor (EurOMo) are some notable entities that offer various reports in this context. The letter is part of a broader effort in the field of media freedom and pluralism, as outlined in the European Democracy Action Plan.

To sum up, we can state a fundamental dilemma in the field of media concentration. On the one hand, concentration processes are fundamentally inherent in a profit-oriented media market due to the reigning economies of scale. On the other hand, media concentration with its consequences of restricting competition is in obvious contradiction to the most noble aim of the EU to safeguard competition. Thus, media concentration has been a long-standing issue of EU media policy debates, but never came to a coherent regulation due to the divergent positions of the Member States and strong lobbying. Nevertheless, recent activities of the Commission show the problem has become so pressing that the Commission is beginning to take the first steps.

2.2 A slice of Europe on social media

EUMEPLAT project's deliverable [D2.2 of WP2](#) entitled Platformisation of News in Ten Countries aimed to assess how news about Europe and the issues of common concern for European citizens are produced and circulated on social media platforms. More specifically, the aim was to identify what types of content are most relevant on those platforms, which actors are more prominent, which subjects are predominantly addressed, and how Europeanisation is viewed in those publications with higher interactions. The methodological framework was to collect, categorize and analyse the most relevant social media posts in all 10 countries from September to November 2021 (Cardoso et al., 2021). In examining the discussion about Europe on social media, the focus was both (i) on the actors driving the discussion and (ii) on the main topics of discussion. The research selected three issues of common concern drawing on the Eurobarometer; accordingly, to be taken into the corpus, a post should concern economy, health, or climate change, besides centrally concerning Europe.

The study showed that, despite politicians' notable presence on both Facebook and Twitter, political actors and the topics promoted tend to be different on each platform. In the Greek case, the most popular posts came from a variety of institutional actors, the most popular post on Facebook was a news piece from the Orthodox Church, The second post with the most interactions was on the same news piece from a Greek radio station. The third and fourth most popular posts were from the opposition leader and ex-Prime Minister, Alexis Tsipras. Similarly in Twitter, the most popular tweets in the "all users" category came from

Greek politicians. This confirms the case that Twitter, in Greece, is more like a niche media for politicians, journalists, and political movements.

As in most countries of our sample, both social networks (Facebook and Twitter) in Belgium have different target audiences. Twitter is a platform mainly used by professionals looking for news (Roginsky, 2020), whereas Facebook is used by a less specifically defined audience. It is striking that half of the posts over the three months on Facebook were made by politicians. In Flanders, the posts were mainly posted by members of the N-VA, Theo Francken and Zuhair Demir. In Wallonia, the posts were mainly posted by Manon Aubry and Philippe Lamberts. Manon Aubry's posts are mostly related to things she is doing or about to do as a politician (Cf. re-elected as co-chair of The Left in the European Parliament or the climate march in Glasgow), while Philippe Lamberts' posts are only about decisions made by the European Parliament on climate. Regarding health, however, the posts on both Facebook and Twitter were mostly made by media actors and to a lesser extent by virologists and politicians.

Professionally or institutionally created content, whether from the news media or political actors, is the most prevalent kind of content in the Czech case. This can be explained by the role that news media still play in the Czech Republic in mediating the public sphere, especially when it comes to European matters, and the pre-election circumstances, which increased politicians' online prominence as they fought for voters' support.

In Italy, both professional and non-professional content is popular and shared on social media platforms. Regardless of whether it comes from the news media or political actors, professionally generated content is most prominent in the Italian case as well. Interestingly, non-institutional organisations (various citizen groups) publish mainly in Facebook groups which are the non-professional channel *par excellence* (100% of contents are not institutional). Political actors, on the other hand, prefer to share content on Facebook pages (57%) and Twitter accounts (40%) that are not managed by the media. Political actors focused mainly on issues relating to health, in particular with regard to the issue of the green pass, especially during September 2021. Non-professional social media content and media actors instead focus their attention mostly on climate change, while economic issues of Europe are more present in the content published by media actors.

The Spanish case highlights another interesting point. Researchers note that comparing Facebook and Twitter in terms of posted content and the emotional charge they offer is a useful strategy. Media professionals use both social platforms to convey users to their official websites. Their posts therefore try to attract users' attention with clickbait titles and captivating images. In contrast, politicians and influencers tend to engage with their followers by boldly expressing their opinions, often challenging their followers' viewpoints. The

common thread running through this form of social media content is its emotional intensity, as it involves taking a firm stance in favour of or against specific issues. This style of online communication captures the interest of other users and is employed in both platforms to increase user engagement.

Bulgaria communicates predominantly on Facebook with more than 98,52% of all online users active on that platform and only 1,48% on other social media. Twitter is not popular in Bulgaria and is mostly used by people active in politics, often businessmen and active citizens. Facebook posts in Bulgaria had mostly a national scope, with only 1% European scope. Twitter has a strongly prevailing European scope (98%) more than twice the national (44). Bulgarian social media users seem to show a constant and deep interest in politics and the economy. The two topics are the main focus of all social media, with COVID-19 being also a strong interest. Domestic and international politics dominate as the most viewed topics in all media and by all users. This is possibly connected to the fact that many of the posts come from politicians, people in political groups, and the mainstream media that are traditionally very much oriented towards politics in their content.

Above all, our analysis of the platformisation of news indicates the success of far-right politicians: The content with which the users of the three platforms interacted the most was that which was published by far-right nationalist politicians (23,5% of the sample, D2.2, p. 17). The difference was also notable between the platforms themselves: 31.4% of the posts on Facebook were posted by far-right nationalist political actors, whereas on Twitter it made up 8.7% of the most popular content, and on Youtube 14.9%. Alexis Tsipras, who led the left-wing Syriza party, was one of the exceptions. In most cases, these were profiles of politicians, not parties, which indicates considerable celebritisation and personification of opinions. The exception is Sweden, where the Sweden Democrats were the most numerous party profile in the monitored period. However, the research also featured countries where politicians and political parties did not have much relevance on social networks, namely Bulgaria and Portugal: the most successful political contribution in Bulgaria came from a successful rapper Itso Hazarta, announcing his entry into politics, while in Portugal, out of the 676 most popular posts examined, only 35 were published by politicians or political parties.

On the other hand, politicians and influencers who are publicly exposed exhibit a consistent pattern of influencing their followers and other users by expressing their opinions in a decisive manner. This behaviour encourages users to transition from passive observers to active participants by engaging on social media platforms. A similar pattern can be observed in the analysis of posts made by non-publicly exposed personalities among Facebook and Twitter users. The common thread linking such social media content is its emotional charge, as well as the expression of specific opinions and stances on various issues. This mode of communication grabs the attention of other users and contributes to the virality of the

content. Several studies support this point, examining the trend of emotional contagion on Facebook and Twitter (Kramer et al., 2014; Larsen et al., 2015; Li et al., 2016; Sailunaz et al., 2019).

Throughout Europe, the prevailing topics of discussion centred around the COVID-19 pandemic, the associated restrictions, the health crisis, and the economic consequences of anti-pandemic measures. Criticism towards the European Union and its institutions was commonplace, although some countries expressed a predominantly positive attitude towards the institution. Notably, in Turkey, two-thirds of the Facebook posts were contributed by politicians from the Republican People's Party (CHP), the largest opposition party in the Turkish parliament. The CHP occupies a position within the political middle-left spectrum in Turkey, and its program is largely oriented towards a pro-European stance.

Another intriguing finding is that news media content on social media platforms receives less engagement compared to content from other users. Media outlets tend to focus more on health and climate-related publications, while political actors are more inclined to discuss the economy. Despite the politicization of topics related to the COVID-19 pandemic, political actors were not dominant in discussions about health; instead, they appeared to concentrate their efforts on political content related to economic and European matters. This focus was not solely on economic and European issues per se but rather reflected a greater emphasis on the internal political struggles that could arise from those issues. It sheds light on how political actors utilize social media platforms to serve localised political conflicts (Fuchs, 2021; Highfield, 2017). In a sense, this highlights the way in which political actors appropriate and adapt these social media platforms (Bucher & Helmond, 2018; Silverstone, 2005).

To sum up, this deliverable reveals significant differences among Facebook, Twitter, and YouTube platforms in terms of content and user profiles. Facebook appears as the most popular platform, while Twitter is dominated more by media and political actors. Facebook also appears, in many countries, as the most politicised platform among the three. Platforms seem often used for criticizing those in power and expressing resentment towards the establishment. Most posts focus on organizations other than political actors and media, in particular public authorities. European institutions and law are rather frequently mentioned, but the social and cultural European social aspects receive limited attention. European issues are often leveraged for national or local debates. Most notably, populist politicians seem to effectively use the platforms, especially Facebook, which indicates potentially toxic encounters and debates. The slice of social media posts examined thus appears as a platform for emotional and polarizing content, which may not be detectable by the linguistically grounded moderation efforts, but still posing challenges for pluralism in Europe.

2.3 Anti-EU fake news and the quest for reliable information: Fact-checking policies

EUMEPLAT project's deliverable [D2.5 of WP2](#) entitled *Anti-European Fake News and What to Do* addresses the increasing concern over the spread of misinformation and polarization on social media platforms, specifically in relation to European issues; on the case study of Brexit. More precisely, using a quantitative approach, the study examined the production of news content and user engagement related to Brexit, with the aim of exploring the presence of "anti-EU" news sources in the debate and identifying the ones more likely to spread misinformation. As highlighted before, debated topics tend to foster the clustering of users into ideologically homogeneous groups which get informed from similar sources. Based on this assumption, the study utilised retweet data and employed a latent ideology estimation technique to infer the ideological stance of users and news sources in the Brexit debate.

The results indicate the clustering of news outlets based on their opinions on Brexit, with left-leaning outlets primarily retweeted by pro-EU politicians and right-leaning outlets shared by anti-EU proponents. News sources with synthetic opinion values above zero are identified as anti-EU news sources. However, variations exist among countries, with different distributions of synthetic opinions and variations in the presence of right-leaning news outlets.

Polarisation around a topic renders an issue prone to misinformation and toxic content. As discussed in the study, being able to timely identify such polarisation dynamics unfolding around debated issues is helpful in determining in advance the targets of hoaxes and fake news. Furthermore, as polarization is one of the main drivers behind toxicity on social media, a better understanding of polarization processes could also lead to an effective strategy for promoting reasonable deliberation in the online public sphere. Last but not least, the primary strategy today to combat fake news and misinformation concerns the implementation of fact-checking policies and the safeguarding of fact-checking agencies' independence.

This is one of the main issues addressed in the deliverable [D2.3 of WP2](#), entitled *Positive and Negative Externalities of News Platformization*, which aimed at investigating the fact-checking policies and the role of independent fact-checking agencies in different countries. The power and political biases of fact-checkers, as well as their role in regulating information flows, are also discussed in this framework.

As in many European countries, the Swedish national government initiated a mission to combat misinformation, propaganda, and online hate through outward-facing initiatives.

Regarding the autonomy of fact-checkers, Sweden's Source Criticism Bureau focuses on falsification and operates as an independent journalistic office. In Germany, Correctiv, the first donation-funded research centre for investigative journalism, sets an example for other media organizations. The Press publishers' ancillary copyright in Germany also contributes to independent fact-checking efforts. Greece faces challenges in combating fake news due to perceptions of media bias, low trust in the press, and a fragmented online news market. Fact-checking platforms such as Hellenic Hoaxes and Check4facts aim to address the issue through crowdsourcing, machine-learning algorithms, and traditional research procedures.

Spain's strategy against disinformation is the EUvsDisinfo project, which aims to identify and expose the cases of disinformation. Portugal seems to lack an official strategy for dealing with fake news, but independent fact-checking agencies like Polígrafo and Observador Factchecks operate in the country. There is also the digital media observatory IBERIFIER, jointly operated in Spain and Portugal, which is promoted by the European Commission and linked to the European Digital Media Observatory (see Badillo-Matos, et al., 2023). In Italy, the early warning approach proposed by data scientists analyses web environments and discussion topics to detect potential polarization and misinformation cascades. Fact-checkers in Italy face challenges with autonomy, as some agencies have links with political or institutional bodies.

In the Czech Republic, experts and national institutions collaborate to develop strategies to fight disinformation. Fact-checking platforms like Demagog.cz, Manipulátoři.cz, and Ověřovna.cz play important roles in verifying claims and debunking misinformation. Turkey's best strategy is exemplified by Teyit, an internationally recognised fact-checking organization. In Belgium, initiatives such as Textgain and FactRank Pro employ technology to combat disinformation, while independent fact-checking platforms like Factcheck.Vlaanderen play a significant role. Factcheck.bg is the only platform in Bulgaria dedicated solely to fact-checking, an initiative of the Association of European Journalists-Bulgaria (AEJ).

Overall, maintaining the autonomy of fact-checkers is crucial, since they provide a vital service in debunking misinformation and keeping in check the potentially toxic content. However, there are concerns regarding public-private partnerships and the collaborations between platforms and fact-checking companies. Successful institutional cooperation against fake news and the application of EU frameworks, such as the European Digital Media Observatory, seems a goal for the future, which has to be specified and adjusted to cultural differences.

2.4 Civil Society on European platforms

The country reports within EUMEPLAT's Work Package 2 analysing the posts about Europe, economy, climate, and health do not provide many threads on the relevance of civil society and pluralistic values on European platforms. The Turkish country report just states that European new social movements do not play any roles (Peschke & Gümüş, 2022, p. 13). The impression that social media are primarily the place for non-professional voices to speak up and debate is also not confirmed, at least for Germany. In this sense, the reality is more conservative than the constant hyping of revolutions and disruptions would make us believe. In contrast, "the impact of the radical right on the digital social sphere seems to be disproportionately larger than in the parliamentary arena" (Grassmuck & Thomass, 2022, p. 1). This finding is confirmed in Work Package 4 research even without looking for political leaning, as gender and migration proved to be two topical arenas strong predictors of AfD and like-minded actors in the debate on Facebook and Twitter. The data for the Czech Republic indicate a high level of polarisation and the dominance of populist and extreme-right voices, but cannot be generalised to all periods, since the research period was dominated by the Czech parliamentary elections (Doudaki et al., 2022, p. 13).

The fil rouge that binds the posts published by Facebook and Twitter users in Spain who are part of a broader term of civil society is the emotional charge, it is also having a precise opinion and taking a position in favour or against a given issue (Latronica et al., 2022). This way of communicating seems to attract the attention of other users and tends to make the content viral. It is evident in the country report of Belgium that civil society groups bring up specific issues that are pertinent to daily life – e.g., the consequences of the floods in July 2021 in Wallonia (De Sutter et al., 2022, p. 7). The authors of the country report from Greece conclude that "there is strong evidence of social media's function as agenda setters, fostering the introduction of news topics that are different or neglected by the legacy media. Both on private users' pages as well as on Facebook group pages the topics of most popular posts were completely different than those found on the respective media accounts" (Papathanassopoulos et al., 2022, p. 11).

This overview of the democracy-promoting civic voices suggests that the picture in our research countries is diverse and slightly bleak. While Europe is brought to the social media agenda by right-wing and populist politicians, it does not appear much within the concerns and discourse of civic actors. For all country reports in Work Package 2, it can be said that popular posts in all platforms may mention Europe or the EU in general, but mentions to specific aspects or dimensions of the EU as indicated are scarce.

In sum, this oversight from platformisation of political discourse in Europe indicates the competing values, ideologies and interests in the online public sphere. Looking at the

question who speaks out in the platformed debates, it indicates, within a particular slice, the range of topics which is brought in as well the dangers, in the sense that pluralistic values can get increasingly under the pressure of toxic debates. As algorithms do shape the debate according to their own logics, further analysis with the help of results within the topic of destructive technologies is needed.

2.5 Summary

The second part of the report has delved into various aspects of the EUMEPLAT Project's research findings from the perspective of toxic debate and pluralistic values. It is important to repeat that the previous work packages, which were re-examined through a different lens, were not originally designed to address toxicity. Nevertheless, the selected findings uncovered significant challenges that could jeopardize the future of civil public communication on online platforms in Europe.

We stated by discussing that the media concentration and lack of pluralism in Europe is in fact partially linked to the plurality of regulation, as the EU's policy has been to provide independence to member states in assuring pluralism a responsibility of individual countries. A linked and salient issue lies in the fact that the media market is on the whole profit-oriented, producing tensions between regulation and competition. Steps in achieving more coherent and compact legislation towards media pluralism have now, however, been taken by the European Commission.

Another way to express the fundamental value of pluralism with specific regard to democratic debate in the platformed era is inclusiveness, i.e., the openness to the voices participating in the public debate. Social media platforms seem to open some room for common people to raise their voices and participate in the public debate on issues at stake. Particularly, Facebook group pages seem to offer an opportunity for citizens to exchange their views. As the popularity of anti-vaccination Facebook groups reveal, these groups may also be conducive to the formation of initiatives and strategies against political measures. The fact that health issues like vaccination emerged in the public sphere as polarised, provides evidence that under specific circumstances, any issue could serve as a trigger of uncivil or toxic comments. The data also provides evidence that politicians and the legacy media have adapted to the new rules of discussion facilitated by the settings and algorithms, resulting in the popularity of their posts. Further, the populist far-right political actors seem to have found an ideal venue to engage the public. With their emotionally charged posts gaining more attention, they effectively create a more polarised and hostile online environment. In this regard, future work could explore the relationship between the ideology of the political actors

and the tonality of their posts.

The creation of so-called echo chambers in social media platforms seems to have paved the way for an escalation of social and political polarization. Since the chambers facilitate the spread of information – sometimes false and unreliable – aligned with commonly shared beliefs, they make the rejection of opposite views easier. In turn, increasing polarization and segregation of users in echo chambers amplify the diffusion of false and unreliable content to the public sphere. As a result, a rapid increase in polarization around a particular topic might serve as a “warning sign” that the topic has generated a fragmented information environment in which a debate unfolds. In this context, the intentional circulation of false information, misinformation and fake news are considered to contribute to the toxicity of public debates. In this regard, the case of Brexit provides a view of the perils of anti-EU fake news. The reviewed study confirms that the primary path employed in tackling sorts of misinformation is fact-checking, with the varied organizations addressing the problem operating independently. While their autonomy is crucial, new perspectives on cooperation among the companies, public initiatives and public organisations are needed to take the prised efforts to the next level.

III. Future scenarios

The analysis presented in this section investigates the Future Scenario Essays produced within the framework of EUMEPLAT Project's Work Package 5, which focused on the externalities of social media platforms. In the design of the scenario-building workshops, Delphi+ method, a method for future scenario-building and forecasting with a long history (Gordon, 2009), was used. Developed in the early stages of the Cold War in order to predict the impact of technology on warfare (San-Jose & Retolaza, 2016), Delphi method's initial aim was to predict the probability and intensity of enemy attacks. Currently, it is used as a technique that offers a "systematic means of synthesizing the judgments of experts" (Gordon, 2009, p. 11) and is used across various academic disciplines. Landeta (2006) defines the Delphi method as "a method of structuring communication between a group of people who can provide valuable contributions to resolve a complex problem" (p. 468). As Gordon (2009, p. 4) summarizes it, the Delphi method is grounded in a "controlled debate" which allows for the establishment of consensus among experts, through a series of iterations. This implies that expert participants can discuss the responses of others and the work of the group as a whole, but also that they can alter their own positions during the process.

Despite its limitations (Winkler & Moser, 2016, p. 63), the Delphi method is often used in future studies, while it is met also in other fields (Poli, 2018). The field of future studies has moved "from predicting the future to mapping alternative futures to shaping desired futures" (Inayatullah, 2012, p. 37). Its three components refer to three different approaches – with different ontological assumptions – namely, forecasting (to predict the most likely future), scenario-building (to explore alternative futures), and backcasting (to assess the feasibility of a desired future).

Drawing on Inayatullah's (2012) description of the field of future studies, we set our goal for the present study as "the systematic study of possible, probable and preferable futures including the worldviews and myths that underlie each future" (p. 37). In what follows, we use the "myths" as a summarising dimension that underlies the scenarios and summarises their analysis.

3.1 Method and Material

We adjusted the Delphi method into 3-and-a-half-hour face-to-face scenario-building workshops, which focussed on five pre-given themes: surveillance and resistance,

algorithms and choice, toxic debate and pluralistic values, destructive technologies and war, and gender in society. Four workshops were organised in three different European cities, with a total of 29 participants (see Table 1 for an overview). The workshop participants were selected from a variety of societal fields, on the basis of their affinity with, and knowledge about, digital media platforms in Europe. As the workshops were organised, for practical reasons, at the moments and locations of the EUMEPLAT consortium meetings, participants were also selected on the basis of their proximity to the meeting locations. The four workshops were organised in two stages. Stage one consisted of small group discussions, with one moderator for each of the subgroups, with the aim of producing three future scenarios for each theme. In stage two, which was a plenary stage, the participants introduced a selection of scenarios to the entire group. While the first stage aimed at developing a series of future scenarios, the second part at achieving consensus about the nature and importance of the different scenarios. As a method, these time-compressed workshops approximate what Pan et al. (1996) called a mini-Delphi, although we prefer to label these four workshops 'Delphi+' workshops.

Table 1. The EUMEPLAT Delphi+ workshops

Number	Date	Location	Participants
1	5 July 2022	Malmö, Sweden	Science fiction writers and foresight researchers, experts on science communication or philosophy of science, and specialists in digital marketing and applied predictive models (6 participants)
2	4 October 2022	Sofia, Bulgaria	A theatre artist, a Roma activist, a journalist, and a former representative of the Bulgarian government in the field of culture (6 participants)
3	13 April 2023	Rome, Italy	Expertise ranging from cultural relations, bioethics and AI to political science and the futures of electronic music (7 participants)
4	23 June 2023	Sofia, Bulgaria	A filmmaker and producer, a TikTok influencer, journalists, media studies professors, and chatbot and new media experts (10 participants)

In the Delphi+ workshops, each subgroup of participants was asked to produce 3 future scenarios, which resulted in a total of 33 scenarios produced this way (see Table 2). In addition, to complement the Delphi+ workshop scenarios, the authors of this text also wrote

8 scenario essays, making the total number of scenarios analysed N=41. The analysis presented in this section used different types of data, namely (1) the scenario cards that the Delphi+ workshop participants filled out during their discussions (summarizing each scenario in keywords), (2) the transcriptions of the Delphi+ workshop participant discussions and (3) the essays generated in written form.

For their analysis, we used the procedures of qualitative content analysis, informed by narratology and the narrative paradigm (Fisher, 1987; van den Hoven, 2017). The unit of analysis was each scenario, and our coding grid included the following coded fields: Title of the scenario, Question raised (if any), Scene (that the depicted future takes as the background), Main actor (that brings the significant change), Main Event (that took place in relation to Toxic Debate and Pluralism), Value (that grounds the aspired or unwelcome future), Prescription (if any), The role of the EU (if any), and Pessimism/Optimism. Except for the last one (Pessimism/Optimism), all fields were coded with an inductive or bottom-up approach. That is, rather than imposing top-down categories, we first coded particular values, actors, events, etc. Once the initial coding was finished, we grouped these particular figures into simple categories (e.g., human actors vs. non-human actors), and when necessary, into further and more diversified categories.

Table 2. The EUMEPLAT Delphi+ workshops

Delphi+ workshop location	Number of scenarios
Sofia	7
Malmö	9
Rome	10
Sofia 2	7
Essays	8
TOTAL	41

In the process of categorizing and comparing codes, we continuously tried to connect the common patterns and identify the “underlying myths” (Inayatullah, 2012) that weave the coded content together. In other words, we used the concept of myth as a summarizing notion, in that it provides an analytical summary of the otherwise very complex relations that connect the present to the futures envisaged in the scenarios. In the following subsection,

we report first the categorised codes that make the most sense of the scenarios and then the myths underlying them.

3.2 An overview of the future scenarios: “Educate people, not machines!”

Actors. We start the overview of the future scenarios from one of the most relevant codes in understanding the agency involved in constructing the future of toxic debates and pluralism: actors. This code aimed to register the actor/actant that brings the significant change in each scenario. It was recategorised into three actor categories (see Table 3), besides the null category. The latter, namely the *No actors identified*, included the cases where a passive voice dominated the conversation, e.g., “Everybody will be anonymised. Like the memes you lose track of everything.” (16, Malmö)¹.

Table 3. Categories for the code Main Actor

Actor	Number of scenarios
Digital/technical	19
Political/institutional	9
Media	5
No actors identified	8
TOTAL	41

The outstanding finding in this coding concerns the predominance of non-human actors (58% when the null category is omitted), which are specifically digital or technical actants, such as “Chatbots”, “Artificial Intelligence”, “Algorithms”, “Interface”, “Platforms”, “(Journalistic) machines”, “WeChat” and “Technology” at large. This predominance may be seen to indicate the hype built around the rise of generative AI and algorithms at large. It simultaneously shows the preoccupation of our participants with the enormous social impacts of the recent developments in the computational sciences.

¹ Number 16 refers to the count of the scenario, Malmö refers to the workshop location.

In the second place, we can consider the *Media* actors together with the broader category of *Political/institutional* actors (42%). To chunk these into political and media actors, following Table 3, we can distinguish the actors such as “Right-wing and populist parties”, “Alternative and marginalised voices”, “Colonizers”, “Acceleration”, “The public”, “Europe”, “Media literacy programs”, and “Some authority” as examples of the main human agency, which can be recognised as political actors. Among these, notably, colonizers were used for denoting the human actors behind the algorithms regulating public opinion and human consciousness.

The media actors, considered specifically, occupy a rather limited place among the actors overall (15%). The five occurrences of this category are “Media”, “Niche media”, “Fake news” (two times) and “Public Service Media Organisations”. Note that fake news is a category that partially belongs to the political domain, for them being often used for illegitimate political interests. Without these two occurrences that essentially pertain to pessimistic scenarios, niche media and public service media appear as the sole actors that are devised to bring some change from the media domain into the world of toxic communication.

Values. In second place we report the values that pertain to communication and that the scenarios resort to explicitly. The values typically ground the evaluations of the imagined futures, more precisely, the actions/impacts brought about by the key actors, and they can be grouped into four categories (see Table 4).

Table 4. Categories for the code Values

Values	Number of scenarios
Intellectual values	13
Ethical values	14
Sociopolitical values	8
Technological values	3
No values identified	3
TOTAL	41

In some contrast with the code Actors, values related to technology occupy a very small place in the values associated with communication. Instead, Intellectual (34%) and ethical

(37%) values take precedence in the ideas that ground the conversation on toxic communication and pluralism. To better understand these, we can exemplify intellectual values as follows: “Critical thinking”, “Media critical thinking”, “Media literacy”, “Solid starting points”, “Tolerance comes from knowledge”, and “Substance of debate”. Notably, there were no negative values among those that relate to the intellect, suggesting our participants’ interest and esteem in the powers of intellect in tackling toxicity and probably a view of pluralism as a sort of intellectual virtue.

Ethical values, which occupy a significant place in the scenarios, can be exemplified by “Peaceful communication”, “Respect”, “Tolerance”, “Pluralism”, “Identity politics”, and “Sensationalism”. The latter two of these are negative values in the sense that they are related inversely to pluralistic values and regarded as contributing to toxicity.

On the third place are the values we designated as sociopolitical (21%), a minority of which were negative values. While “Public good”, “Transparency”, “Universal citizen rights”, and “Legitimate authority” count as positive values, “Corporate interest” and “Authority” exemplify the negative values.

Finally, the three occurrences of technological values can be captured as “Autonomy of technology”, “Lack of face communication” (in online communication), and “Mobility” (between bubbles as a capacity that is achieved technologically). Notice that the first two of these are negative values – with autonomy of technology referring to the loss of human control over technological change. This suggests that when technology is linked to values grounding decisions or actions, it does so rather negatively.

Prescriptions. This code aimed to register the proposals and suggestions that the scenarios may involve. It is typical of the pessimistic scenarios, in the sense that most of these scenarios devise an issue or problem – e.g. fake news, deep bubbles, the demise of the notion of truth – and then offer certain ways out of the predicament. N=25 out of the total of 41 scenarios involved such ways towards positive change or prescriptions. We initially coded these into two categories, namely two fundamental aspects of social change, structure and agency (Best, 2014). The output of the coding process was rather extraordinary, with all but one of the prescriptions being categorised as Structural change (N=24). Since this figure did not distinguish much, and building on the previously reported codes, we re-coded the Structural change to distinguish the prescriptions that involved technology centrally. This way we achieved three categories for the code prescriptions (Table 5).

Even after the attempt to chunk the code Structural change into two codes, there still is an overwhelming weight of the prescriptions of structural change (N=21, 84%). This reflects the locus of change as pointed out by our participants. Rather than individual or ethical action

prescriptions – except for one case – all scenarios involving such action-guiding proposals expected the change to come from or originate in the structure, namely institutions and regulations. Let us see some examples:

“...Yes, encouraging pluralism. So, first to distinguish what are the hidden forms of dialogue that we can encourage and then to provide the tools for the people to be able to participate with them, because, the first one is how they can break this you and me contradiction model.” (41, Sofia 2)

“An obligatory continuous media education had been implemented in schools of all types (...) The compulsory information and media education is a part of educational systems among Europe in all stages of education.” (5, Essay)

In the first of these two excerpts, the participant aligns herself with a top-down agenda of sorts that provides tools for the public, encouraging novel formats of dialogue. The second excerpt also exemplifies the scenarios in which encouragement is envisaged in a more structured educational reform. Such a position echoed in most of the scenarios, where education at large, and “encouraging and innovating in activism and participation” (39, Sofia 2) or “democratisation of culture and knowledge worldwide, and algorithm knowledge” (24, Rome), in particular, were offered as the locus of the solution(s) to online communicative predicaments.

Table 5. Categories for the code Prescriptions

Prescriptions	Number of scenarios
Structural change	21
Technological change	3
Agential/Personal change	1
No prescriptions identified	16
TOTAL	41

The goals of the educational intervention were elaborated into four aspects in one case (5, Essay): first, development of critical thinking for evaluating (online) content, second, encouraging empathy and respectful online interactions, third, encouraging responsible digital citizenship, and fourth, addressing online hostility. Notably, these prescriptions seem

to align well with the three categories of values identified above, namely intellectual, ethical, and sociopolitical values. Even the fourth and final measure identified by Participant 5, “addressing online hostility”, may be aligned with the fourth and least salient value emerging from this study, technological values.

To further emphasize the weight of digital literacy and education in prescriptive statements, more examples of the following kind can be offered: “...very close to this Critical thinking. Progress through education, consensus through education and through developing critical thinking” (27, Sofia). Such calls for “progress through education” should not be regarded as unreflexive prescriptions of simple modernization. Some participants recognised the fact that education policies are sort of old-fashioned or “boring”. For instance, in the conversation around the scenario “Breaking the bubble” (35, Sofia 2), platforms were depicted as “dividers and opinion shapers” and that “digitalization” would bring more bubbles and segregation to societies. The participant who uttered the prescriptive element at this point in the scenario-building workshop stated: “And the only bridge for me, I’ll bring the boring word, education”. Designating education as “boring” in this way suggests that the participants were well aware of the limits and failures of education as a policy to deal with social problems. Yet, they seem to be unable to come up with other proposals, probably due to the recognition of the necessity to approach such communicative problems in a bottom-up fashion.

To a lesser extent than the prescriptions on what may be called *critical thinking and digital literacy*, there were also prescriptions for more and extensive *regulations*. These were typically top-down measures to control and restrain the corporate power reigning in social media platforms and digital communication at large. Examples are “Regulation of commercial platforms” and:

“...interventions in business models, aligning with democratic principles (...) platforms cannot be operated with the same profit margins as before (...) Political support must come both from the nation-states and from the European level. ” (2, Essay).

In a realistic manner, the need for regulation is recognised as an integral task for nation-states. Rather than imagining some new and innovative agency, for instance at the global level – except for good algorithms – the prescriptions recorded ascribed responsibility to current authorities and governments. This seems to suggest that for our participants toxicity is a problem to be dealt with and a phenomenon that can be regulated today, rather than in an imagined future.

After examining the prescriptions for structural interventions, let us also briefly look at the outlier: the only scenario that included aspects of agential/personal change as a response

to toxicity. This prescriptive statement also comprised algorithmic knowledge and digital literacy:

“I’m done having choices made for me. You will have to extricate yourself from a lot of systems (...) Media literacy influences the debate, not just being offline or away from the keyboard, but learning more about how things work, like how algorithms for how media works and so forth...” (17, Malmö).

Notice that while the source of salvation is the same with the majority of the prescriptive points marked just above, in this excerpt the predicate is to “learn” – rather than “encourage” – and it signals the powers or agency of the users in a bottom-up fashion. While it plays the agential tune in reverse, in regard to the content, the outlier also falls in the broad domain of education, with an emphasis on acquisition and self-instruction of how algorithms work.

In sum, two major messages can be drawn from the prescriptive statements involved in the scenarios analysed: educate and regulate. In this regard, perhaps the most salient direction that can be drawn from the experts involved in the scenarios is summarised in a slogan that popped up in one of the sessions: “Educate people, not machines!” (20, Sofia).

The role of Europe in building the scenarios and addressing the predicament was explored at varying degrees in the scenarios. Europe was mentioned only in N=11 of the total of 41 scenarios. In these, it appeared typically with the role to “safeguard democracy”, “defend the institutions” (1, Essay), and as “among the institutions most likely to foster, and cultures most prepared to sustain, such an open public debate on issues of common concern” (6, Essay). The EU made an important part of the public education and digital literacy efforts mentioned above: “Under the coordination of European institutions, specific modules designed to combat toxicity could be established in schools across Europe” (8, Essay). Besides these, there were also few mentions of a “stronger European identity”, and more precisely, the appeal “The EU should empower its tech and media industry to take the lead, even to import know-how from abroad, since most European AI companies are still at an early stage.” (7, Essay). Generally speaking, the EU was not a defining actor in the scenarios, but there were some calls for it to become one if toxicity and fragmentation of society were to be tackled.

3.3. Discussion of the future scenarios

We used the notion of “underlying myths” (Inayatullah, 2012) as a summarizing dimension that weaves the coded content together, consisting of the connections among the common patterns and storylines. This section discusses three underlying myths that characterize the

future scenarios. We take these myths as scripts or common patterns of reasoning that connect the present day to the futures envisaged in the scenarios. As such, they provide visions of the future rather than the futures themselves, and thereby they can be used to extract certain lessons or implications for today.

The first myth can be called *technological disruption*, or more specifically, AI and algorithmic disruption. It underlies the descriptions of a brave new world where the integration of digital technologies into all aspects of human communication brings numerous challenges that the public cannot even fully comprehend. This myth is grounded in the findings that almost no agency is ascribed to the individual and the public in the prescriptions examined, and also in the overwhelming predominance of the AI/digital actants among the actors that bring the change in the 20-year time horizon. In other words, a central preoccupation of our participants is that digital and generative technologies will bring a sweeping change that will disrupt manifold aspects of human life and communication. Rapidly evolving digital technologies are thus envisaged as the villain and the main cause of future predicaments.

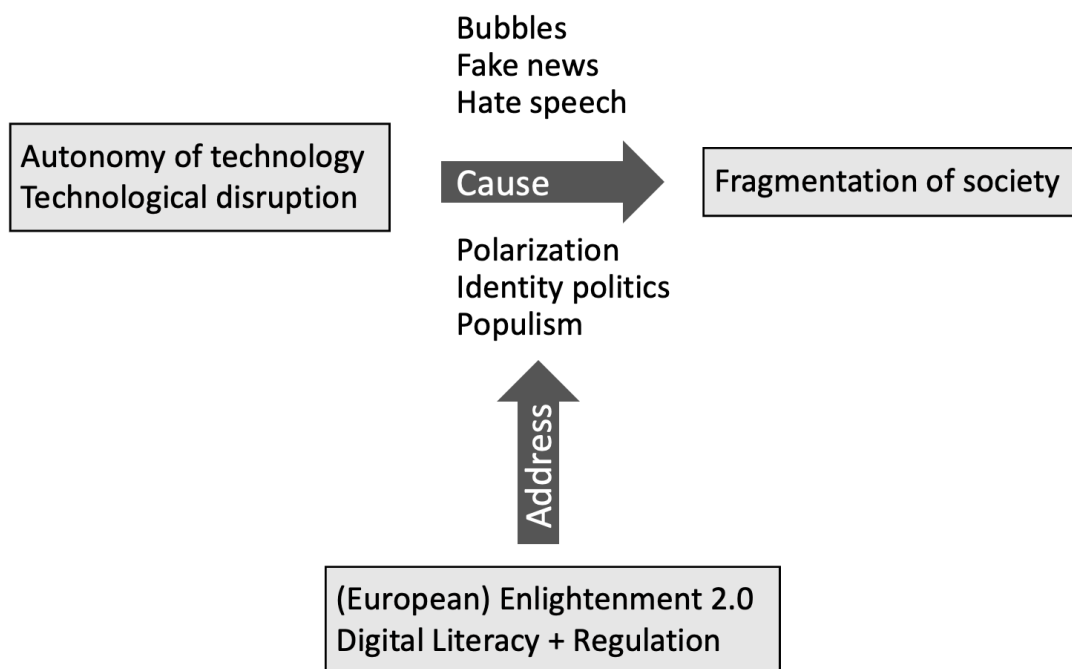
The second myth can be called the *fragmentation of society*, or with its sociological metaphor, anomie. As the corporate deployment of algorithms, AI and other technologies amplify existing cleavages, nothing short of the breakdown of common grounds and communicative frameworks is regarded as the peril ahead. It is the immediate impact of the disruptive technologies, mentioned as the first myth, that brings about this second and more central myth underlying most of the concerns over toxicity. Several cascading factors and issues can be aligned in this causal link: lack of multidimensional communication and facework, bubbles, fake news, polarization, blurring boundaries of the real and virtual, and the complete loss of the sight of truth. In short, in the platform media designed and maintained by non-human – not to say inhuman – values and interests, the central worry is to lose the foundational elements of human interaction and communication, being locked in conflicts and contradictions that attract the most attention.

The third myth can be called *Enlightenment 2.0* and it is connected to Europe more directly than the others. This myth mostly underlies the prescriptive statements examined in Section 3.2. The enlightenment involved in this debate is a decidedly digital enlightenment – hence the 2.0 designation – in which the authorities are invited to encourage digital literacy, public knowledge on algorithms, critical thinking to evaluate online information, and support the epistemic quality or substance that grounds the public debate. Notably, the Enlightenment is not just about enhanced critical thinking on the part of the users but also about regulating the platform business and the corporate interests behind algorithmic distortion. The regulations mentioned concern also upholding and innovating in public service media, opening alternative paths to media institutionalization, and innovation in the design of online debate and interactions. In this sense, Enlightenment 2.0 incorporates both bottom-up and

top-down measures to address the yet not-well-known impacts of algorithms and platforms on human communication and society at large.

To address the question concerning the relations among the three underlying myths, two causal relationships may be discussed (See Figure 1). The first causal link lies between the first two myths, depicting that rapid advances in digital technologies and AI constitute a challenge for the already complex communicative predicaments that are captured by the notion of toxic debates. In other words, issues such as filter bubbles and polarisation are expected to exacerbate with the further involvement of social media platform technologies. The impacted end of the causal link is the human society at large, and an associated worry is that the public is not ready to handle, nor comprehend, the challenges human nature and institutions are faced with.

Figure 1. A basic model of relationships among the three underlying myths



It is important to note that, contrary to what Figure 1 may suggest, technology is not the only cause that brings about the second myth. It may even be seen as merely exacerbating the already existent social and political problems. In this sense, the locus of the relationships among the three myths are the six problems that connect all three imaginaries: bubbles, fake news, hate speech, polarisation, identity politics and populism.

If the link between the first and the second myths was causal, the link between the second and the third myths can be designated as negative causation. That is, the third myth impacts

the envisaged impact of digital technologies on society in a way to avert these, namely to protect the communicative and social relations from the impacts of technological disruption. To put it in other words, to avert the ongoing fragmentation of society due to the polarizing designs of online platforms, fake news and online hate, the recommendation is to launch a public campaign of sorts, to enhance digital literacy and regulate social media platforms in such ways to enhance democratic and accountability. In this regard, the calls for regulation, associated with the institutional level, may be seen to indicate a certain concern or fear of the AI-powered algorithmic distortion as a “symptom” of deregulation and neoliberalism.

Several limitations of this study must be recognised. First, our corpus consisted of two types of scenarios: those produced in groups orally and those produced by individual persons in writing. The latter had more structure, a clear beginning and an end, with prescriptions attached. The former, with much less structure and many times not much relevant codable content, may be said to limit the potential of the analysis. In relation to this, second, the codes Scene, Event, and The role of the EU did not work well: in many cases, there was either not enough substance to code, or many varied but not very relevant ideas and information that we could not bring together within those codes. Finally, our interpretations may not be well-grounded as our access to the context of the Delphi+ Group Discussions was limited, as none of the authors were actually present in the Delphi+ workshops. That said, the simple coding grid and coding strategy may be said to have worked well, perhaps at the expense of a more detailed and richer description of the future scenarios.

Conclusion

Based on research on externalities of media platformisation in Europe, EUMEPLAT research has identified problem areas and issues that can be considered a threat to public communication. One of these problem areas is the toxic debate, relevant in the context of pluralistic values in European societies. In order to place the results of our empirical research into this context, this document offered a dense description of the different definitions, manifestations, and dimensions of toxicity, as well as the essential norms that apply to public communication on social media platforms.

Above all, our study of the platformisation of news indicates the success of far-right populist politicians. On the other hand, politicians and influencers who are publicly exposed exhibit a consistent pattern of influencing their followers and other users by expressing their opinions in a decisive manner. With respect to European issues, we found that they are often leveraged for national or local debates. Overall, populist politicians seem to effectively use the platforms, especially Facebook, and the posts that are examined thus appear as a platform for emotional and polarizing content, posing challenges for pluralism in Europe. Looking into our findings about the democracy-promoting civic voices the picture is worrying as well: While Europe is brought to the social media agenda by right-wing and populist politicians, it does not appear much within the concerns and discourse of civic actors.

Our results do not allow us to clearly identify the amount of toxicity in the data. Even though we could not hard-code toxic speech or the degree of toxicity in our research, we saw trends that can pay into toxicity. We relate this insight to the literature indicating that the expression of emotions to a particular content on social media impacts the experience of such content and emotions, providing experimental evidence of large-scale contagion through social networks (e.g., Han & Brazeal, 2015; Kim et al, 2021). Accordingly, to take a role in the creation of toxic debates, a post does not have to involve hate or toxic speech. Merely breaching any of the norms identified at the micro, meso, and macro levels may suffice. This is not just due to the modelling effects and emotional contagion mentioned just above but also relates to the enormous differences in the topical content knowledge (Goodwin, 2019) with which users engage in a debate – or rather, in different façades of a debate, without engaging in one another’s perspective at all.

In this regard, toxic debates are those encounters which the parties fail to encounter each other. This failure may be topic-driven (Salminen et al., 2020a), in the sense that certain issues, such as abortion, bring about a crystallised background over which parties merely score their points associated with their identity positions (Fogelin, 1985). The “failure”, alternatively, may be due to an organised move using toxicity strategically to bring about

political effects (Paliewicz & McHendry, 2017; Zarefsky, 2016), as captured in some instances of what is known as “cancel culture”. In both cases, there is a clear repudiation of pluralistic values and a disregard for the other, many times exacerbated by the technological design and means used in the medium.

This is in line with the experts’ concerns expressed in the future scenarios. Participants in the future scenarios examined in Part III saw the immediate impact of the disruptive technologies, that bring about several cascading factors and issues which have the potential to damage the democratic debate: lack of multidimensional communication and facework, bubbles, fake news, polarization, blurring boundaries of the real and virtual, and the complete loss of the sight of truth. The central worry that emerged from the scenarios was to lose the foundational elements of human interaction and communication, being locked in conflicts and contradictions that attract the most attention.

What should have become apparent so far is that “language on social media is not inconsequential” (Hiaseshutter-Rice & Hawkins, 2022, p. 3). As the authors explain, if users are accustomed to negative and toxic political discussions regarding current political affairs, this is likely to trigger further tensions about what is happening around the world (p. 14). After all, as Taboada (2021) critically remarks “what we do with language and how we weaponize it may be the measure of our society”. If we aspire to change the status quo on social media platforms deliberation, we need to resort to the principles of reasoning and argumentation, as well as highlighting values of high significance in globalised and multicultural settings like, inclusion, respect, and acceptance of the Other, in other words, we need to embrace all the premises of a truly participatory and democratic society.

It is also important to once again highlight the architecture and the interaction design of online platforms, with the significant effects they have on the shape of social media discussion. By comparing four discussion arenas on three different platforms, Jakob and colleagues (2023) indicate that online toxicity is more pronounced in settings where discussions revolve around diverse issues, blurring the boundaries of the institutional, public, and private. In contrast, in platforms that keep public and private conversations separate, toxic outrage tends to be lower. The authors suggest that user-generated discussions thrive in environments that prioritize pertinent issues, encourage participants to seek compromise, and create a space for public discourse somewhat removed from purely social interactions.

As practice has shown, content moderation has also an important role to play in safeguarding pluralism and a healthy public sphere, as well as providing users with a safe experience online. As the experience with Reddit’s 2015 anti-harassment policy has shown, taking down offensive webpages and posts can effectively limit the volume of toxic speech online, albeit with some expenses to the platform. After Reddit started to remove problematic

subreddits, more accounts than expected discontinued using the site, and those that stayed drastically decreased their hate speech (Carlson, 2021, p. 141). Yet, the differences in size, reach, design, and business model of social media platforms are significantly involved in how content moderation works (Gillespie et al., 2020). With content moderation left completely to the platforms themselves, and such moderation is by and large automated, it proves difficult to even discuss some common grounds. Perhaps more importantly in this context, neither the public nor the national state authorities have the capacity to address the speed, scale and global reach of toxicity on social media platforms. With the new challenges such as organised networks, orchestrated attacks, bot armies, new generation avatars, and a variety of AI tools for content generation, an uneven battle for silencing any opposition may be in sight without any public scrutiny.

This point can be linked to the prescriptions involved in the future scenarios, which point to two significant demands from public and private institutions: education and regulation. While our research did not deal with issues like media literacy or critical thinking, we elaborated on how EU legislation and directions worked on the topic of safeguarding pluralism as one of the fundamental norms for public communication. The outcomes of the latest development with respect to common standards to tackle issues such as disinformation and online hate messages as it is regulated in the Digital Service Act still must be awaited.

According to Habermas (2004) there is an *ostensible paradox*, a definitive point of tolerance that establishes a code of conduct that everybody must acknowledge, defining a boundary for what is not permissible. Essentially, it is this defined intolerance that enables the existence of a tolerant society: when intolerance is permitted, it is the tolerant society that ultimately suffers its consequences. On this principle, one could argue that the European legal framework to address toxicity is on the right track. By taking value from the various platforms' best practices to tackle disinformation and violence online, the EU should consider imposing standard guidelines for content moderation and users' rights common to all platforms, according to their size and operational capacity. In this way, new social media designs could also have the necessary support in order to implement better architectures that otherwise would need much empirical research on each company's part. A first step in this direction could be to request from platforms transparency in their actions and data regulations. An independent observatory is also of crucial importance to assist in coordination and document how the legal framework is implemented by the platforms.

References

- Ackerman, B. (1989). Why dialogue? *The Journal of Philosophy*, 86(1), 5-22. <https://doi.org/10.2307/2027173>
- Adams, C.J., Sorensen, J., Elliott, J., Dixon, L., McDonald, M., Nithum, Cukierski, W. (2017). Toxic Comment Classification Challenge. Kaggle. <https://kaggle.com/competitions/jigsaw-toxic-comment-classification-challenge>
- Anderson, A. A., Brossard, D., Scheufele, D. A., Xenos, M. A., & Ladwig, P. (2014). The “nasty effect:” Online incivility and risk perceptions of emerging technologies. *Journal of Computer-Mediated Communication*, 19(3), 373-387. <https://doi.org/10.1111/jcc4.12009>
- Anderson, A. A., Yeo, S. K., Brossard, D., Scheufele, D. A., & Xenos, M. A. (2018). Toxic talk: How online incivility can undermine perceptions of media. *International Journal of Public Opinion Research*, 30(1), 156-168.
- Arenhövel, M. (2000). Zivilgesellschaft, Bürgergesellschaft. *Wochenschau II*, no 2 March/April 2000.
- Badillo-Matos, A., Baldi, V., Arteaga, F., Paisana, M., Crespo, M., Cardoso, G., Rementería, M.J., Philippe, O., Calvo, B., Buslón, N., Hernández-Escayola, P., Gómez-Romero, J., Molina-Solana, M. (2023). Analysis of the Impact of Disinformation on Political, Economic, Social and Security Issues, Governance Models and Good Practices: The cases of Spain and Portugal. Pamplona: IBERIFIER. <https://doi.org/10.15581/026.002>
- Best, S. (2014). Agency and structure in Zygmunt Bauman's modernity and the holocaust. *Irish Journal of Sociology*, 22(1), 67-87.
- Bobba, G., Alberto, C. C., & Cremonesi, C. (2017). *The Age Of Populism?* ECPR General Conference, Oslo.
- Boswell, J. (2015). Toxic narratives in the deliberative system: how the ghost of Nanny stalks the obesity debate. *Policy Studies*, 36 (3), 314-328. <https://doi.org/10.1080/01442872.2015.1065966>.
- Bradshaw, S., & Henle, A. (2021). The gender dimensions of foreign influence operations. *International Journal of Communication*, 15, 23-41.
- Brooks, D. J., & Geer, J. G. (2007). Beyond negativity: The effects of incivility on the electorate. *American Journal of Political Science*, 51(1), 1-16.
- Bucher T and Helmond A (2018) The Affordances of Social Media Platforms. In: Burgess J, Poell T, and Marwick A (eds), *The SAGE Handbook of Social Media*, London: Sage, pp. 233-253.
- Burnap, P. & Williams, M. L. (2015). Cyber hate speech on twitter: An application of machine classification and statistical modeling for policy and decision making. *Policy*

& *Internet* 7(2),223–242.

- Cammaerts, B. (2009). Radical pluralism and free speech in online public spaces: The case of North Belgian extreme right discourses. *International Journal of Cultural Studies*, 12(6), 555-575. <https://doi.org/10.1177/1367877909342479>
- Cardoso, G., Álvares, C., Moreno, J., Sepúlveda, R., Crespo, M., Foà, C. (2021). A Framework and Methodological Protocol for analyzing the platformization of news https://www.eumeplat.eu/wp-content/uploads/2022/02/D2.1_Framework-and-methodological-protocol.pdf
- Carlson, C. R. (2021). *Hate Speech*. MIT Press.
- Carpentier, N. & Cammaerts, B. (2006). Hegemony, Democracy, Agonism and Journalism: An Interview with Chantal Mouffe. *Journalism Studies* 7(6): 964-75.
- Chen, G.M. (2017). *Online incivility and public debate: Nasty talk*. Springer.
- Chen, G. M., Muddiman, A., Wilner, T., Pariser, E., & Stroud, N. J. (2019). We should not get rid of incivility online. *Social Media +Society*, 5(3).<https://doi.org/10.1177/2056305119862641>.
- Cinelli, M., Pelicon, A., Mozetič, I., Quattrociocchi, W., Novak, P. K., & Zollo, F. (2021). Dynamics of online hate and misinformation. *Scientific reports*, 11(1), 22083.
- Coe, K., Kenski, K., & Rains, S.A. (2014). Online and uncivil? Patterns and determinants of incivility in newspaper website comments. *Journal of Communication*, 64(4), 658-679.
- Crespo, M., Álvares, C., Cardoso, G., Moreno, J., Sepúlveda, R., Foà, C. (2022). Catalogue of Best Practices in Citizen Journalism. From citizen journalism to the platformization of news EUMEPLAT WP2 Deliverable https://www.eumeplat.eu/wp-content/uploads/2022/11/D2.4_Catalogue-of-Best-Practises-in-Citizen-Journalism.pdf
- Davidson, T., Warmesley, D., Macy, M. W., & Weber, I. (2017). Automated Hate Speech Detection and the Problem of Offensive Language. *Proceedings of the 11th ICWSM*, 512-515.
- Doudaki, V., Hroch, M., Carpentier, N. (2022). Citizen Journalism in Ten Countries. EUMEPLAT WP2 Country report: Czech Republic.
- Dubiel, H. (1994). *Ungewißheit und Politik*. Frankfurt: Suhrkamp.
- Duggan, M. (2017). Online Harassment 2017. Pew Research Center. Accessed at: <https://www.pewresearch.org/internet/2017/07/11/online-harassment-2017/>
- Dunn, S., Vaillancourt, T., & Brittain, H. (2023). Supporting Safer Digital Spaces. Suzie Dunn, Tracy Vaillancourt & Heather Brittain, "Supporting Safer Digital Spaces", Centre for International Governance Innovation.
- Edwards, M. (2004). *Civil Society*. Cambridge, England: Polity Press.

- van Eemeren, F. H. & Grootendorst, R. (2004). *A Systematic Theory of Argumentation: The Pragma-Dialectical Approach*. New York: Cambridge University Press.
- Fan, W., & Gordon, M. D. (2014). The power of social media analytics. *Communications of the ACM*, 57(6), 74-81.
- Ferrara, E., & Yang, Z. (2015). Quantifying the effect of sentiment on information diffusion in social media. *PeerJ Computer Science*, 1(26).
- Fisher, W. R. (1987). *Human communication as narration: Toward a philosophy of reason, value, and action*. University of South Carolina Press.
- Fogelin, R. (1985/2005). The logic of deep disagreements. *Informal Logic*, 25, 1, 1-7.
- Fortuna, P., & Nunes, S. (2018). A Survey on Automatic Detection of Hate Speech in Text. *ACM Comput. Surv.*, 51(4), 85, 1-30. <https://doi.org/10.1145/3232676>.
- Freelon, D. (2015). Discourse architecture, ideology, and democratic norms in online political discussion. *New Media & Society*, 17(5), 772–791. <https://doi.org/10.1177/1461444813513259>
- Friess, D., & Eilders, C. (2015). A systematic review of online deliberation research. *Policy & Internet*, 7(3), 319–339. <https://doi.org/10.1002/poi3.95>
- Fuchs, C. (2021). *Social media: A critical introduction*. London: Sage.
- Funiok, Rüdiger (2002). Medienethik: Trotz Stolpersteinen ist der Wertediskurs über Medien unverzichtbar [Despite stumbling blocks, the discourse on media values is indispensable]. In: Karmasin, Matthias (Ed.) (2002). *Medien und Ethik [Media ethics]* . Stuttgart: Reclam. S. 37-58.
- Gautam, V. & Taboada, M. (2019). Constructiveness and Toxicity in Online News Comments. Discourse Processing Lab Simon Fraser University. Accessed at: https://www.sfu.ca/~mtaboada/docs/research/Constructive_News_Comments_Report.pdf
- Gervais, B.T. (2015). Incivility online: Affective and behavioral reactions to uncivil political posts in web-based experiment. *Journal of Information Technology & Politics*, 12(2), 167-185.
- Giddens, A. (1984). *The Constitution of Society*. Cambridge, England: Polity Press.
- Gillespie, T. (2020). Content moderation, AI, and the question of scale. *Big Data & Society*, 7(2). <https://doi.org/10.1177/2053951720943234>.
- Godden, D., & Brenner, W.H. (2010). Wittgenstein and the logic of deep disagreement. *Cogency: Journal of Reasoning and Argumentation*, 2(2), 41.
- Goodwin, J. (2019). Sophistical refutations in the climate change debates. *Journal of Argumentation in Context*, 8(1), 40-64.
- Goodwin, J. (2020). Should climate scientists fly? A case study of arguments at the system level. *Informal Logic*, 40(2), 157-203.

- Gordon, T. (2009). "Delphi", in Jerome C. Glenn and Theodore Gordon (eds.), *Futures Research Methodology Version 3.0. Washington: The Millennium Project*, pp. 1-29.
- Gosewinkel, D., Ruch, D., van den Daele, W. & Kocka, J. (2004). *Zivilgesellschaft - national und transnational*. Berlin: Edition sigma.
- Gosewinkel, D. (2012). Civil Society, European History Online, Mainz: Institute of European History, <http://ieg-ego.eu/en/threads/transnational-movements-and-organisations/civil-society/dieter-gosewinkel-civil-society> (Retrieved 23.3.2016)
- Grasmuck, V. & Thomass, B. (2022). Media and radical political voices dominate the social media sphere. Eumeplat WP2 Country report Germany draft 24.10.2022
- Guberman, J., Schmitz, C., & Hemphill, L. (2016). Quantifying Toxicity and Verbal Violence on Twitter. *Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion (CSCW '16 Companion)*, 277-280. New York: Association for Computer Machinery.
- Guess, A., Kim, J., Nyhan, B. & Reifler, J. (2021). The Distorting Prism of Social Media: How Self-Selection and Exposure to Incivility Fuel Online Comment Toxicity. *Journal of Communication* 71, 922-946. Oxford University Press.
- Habermas, J. (2004). Religious tolerance - the pacemaker for cultural rights. *Philosophy*, 79(1), 5-18.
- Hameleers, M., Bos, L., & de Vreese, C. H. (2017). "They did it": The effects of emotionalized blame attribution in populist communication. *Communication Research*, 44(6), 870-900.
- Han, S. H., Brazeal L. M. (2015). Playing nice: Modeling civility in online political discussions. *Communication Research Reports*, 32, 20–28.
- Han, S. H., Brazeal, L. M., & Pennington, N. (2018). Is Civility Contagious? Examining the Impact of Modeling in Online Political Discussions. *Social Media + Society*, 4(3). <https://doi.org/10.1177/2056305118793404>
- Hede, A., Agarwal, O., Lu, L., Mutz, D.C, & Nenkova, A. (2021). From Toxicity in Online Comments to Incivility in American News: Proceed with Caution. *arXiv preprint arXiv:2102.03671*.
- Hiaeshutter-Rice D., & Hawkins, I. (2022). The Language of Extremism on Social Media: An Examination of Posts, Comments, and Themes on Reddit. *Front. Polit. Sci.* 4:805008. doi: 10.3389/fpos.2022.805008.
- Highfield, T. (2017). *Social media and everyday politics*. John Wiley & Sons.
- Hmielowski J. D., Hutchens M. J., Cicchirillo V (2014). Living in an age of online incivility: examining the conditional indirect effects of online discussion on political flaming. *Information, Communication & Society*, 17:10, 1196-1211
- Hopp, T., Vargo, C. J., Dixon, L., & Thain, N. (2018). Correlating Self-Report and Trace Data Measures of Incivility: A Proof of Concept. *Social Science Computer Review*,

38(5), 584–599. <https://doi.org/10.1177/0894439318814241>

- Hummel, D. (2016). Shrinking Cities and the Possibilities of E-Participation. *National Civic Review*, 105, 2, pp. 8-12.
- Inayatullah, S. (2013). Futures studies: theories and methods. *There's a future: Visions for a better world*, 36-66.
- Jakob, J., Dobbrick, T., Freudenthaler, R., Haffner, P., & Wessler, H. (2023). Is constructive engagement online a lost cause? toxic outrage in online user comments across democratic political systems and discussion arenas. *Communication Research*, 50(4), 508-531.
- Jamieson, K. H., Volinsky, A., Weitz, I., & Kenski, K. (2017). The political uses and abuses of civility and incivility. In K. Kenski & K. H. Jamieson (Eds.), *The Oxford handbook of political communication* (pp. 205–218). Oxford University Press.
- Judson, E., Atay, A., Krasodomski-Jones, A., Lasko-Skinner, R., & Smith, J. (2020). Engendering hate: The contours of state-aligned gendered disinformation online. Accessed at: <https://demos.co.uk/wp-content/uploads/2023/02/Engendering-Hate-Oct.pdf>
- Kies, R. (2010). Promises and limits of web-deliberation. Palgrave Macmillan.
- Kim, J. W., Guess, A., Nyhan, B., & Reifler, J. (2021). The Distorting Prism of Social Media: How Self-Selection and Exposure to Incivility Fuel Online Comment Toxicity. Accessed at: <https://cpb-us-e1.wpmucdn.com/sites.dartmouth.edu/dist/5/2293/files/2021/03/comment-toxicity.pdf> (Accessed, June 25, 2023).
- Kim, J. W., Guess, A., Nyhan, B., & Reifler, J. (2021). The distorting prism of social media: How self-selection and exposure to incivility fuel online comment toxicity. *Journal of Communication*, 71(6), 922-946.
- Kocka, J. (2002). Das Bürgertum als Träger von Zivilgesellschaft – Traditionslinien, Entwicklungen, Perspektiven. In: Enquete-Kommission „Zukunft des Bürgerschaftlichen Engagements“ Deutscher Bundestag (ed.): Bürgerschaftliches Engagement und Zivilgesellschaft. Opladen: Leske + Budrich, pp. 15-22.
- Konikoff, D. (2021). Gatekeepers of toxicity: reconceptualizing Twitter's abuse and hate speech policies. *Policy & Internet*, 13(4), 502–521.
- Krabbe, E. C. W., and J. A. van Laar (2011). The Ways of Criticism. *Argumentation*, 25, 2, 199–227. doi:10.1007/s10503-011-9209-8.
- Kramer, A. D., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, 111(24), 8788-8790.
- Kumar, R., Reganti, A.N., Bhatia, A., & Maheshwari, T. (2018). Aggression-annotated

- Corpus of Hindi-English Code-mixed Data. *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, 1425-1431. Miyazaki, Japan: European Language Resources Association (ELRA).
- Landeta, J. (2006). Current validity of the Delphi method in social sciences. *Technological Forecasting and Social Change*, 73(5), 467–482.
- Larsen, M. E., Boonstra, T. W., Batterham, P. J., O’Dea, B., Paris, C., & Christensen, H. (2015). We feel: mapping emotion on Twitter. *IEEE journal of biomedical and health informatics*, 19(4), 1246-1252.
- Latronico V., Jim Ingebretsen Carlson, Francisco Lupiañez-Villanueva (2022). WP2 National Report: Spain.
- Lewiński, M., & Aakhus, M. (2014). Argumentative polylogues in a dialectical framework: A methodological inquiry. *Argumentation*, 28, 161-185.
- Li, Q., Zhou, B., & Liu, Q. (2016). Can twitter posts predict stock behavior?: A study of stock market with twitter social emotion. In 2016 IEEE International Conference on Cloud Computing and Big Data Analysis (ICCCBDA) (pp. 359-364).
- Liboiron, M., Tironi, M., & Calvillo, N. (2018). Toxic politics: Acting in a permanently polluted world. *Social Studies of Science*, 48(3), 331–349.
- Mohammed, D. (2019). Standing Standpoints and Argumentative Associates: What is at Stake in a Public Political Argument? *Argumentation*, 33(3). 10.1007/s10503-018-9473-y
- Molina R. G., Jennings F. G. (2017). The role of civility and metacommunication in Facebook discussions. *Communication Studies*, 69, 42–66.
- Mondal, M., Silva, L.A., & Benevenuto, F. (2017). A measurement study of hate speech in social media. *Proceedings of the 28th ACM Conference on Hypertext and Social Media*, 85-94. <https://doi.org/10.1145/3078714.3078723>.
- Mouffe, C. (2005). *On the Political*. London: Routledge.
- Muddiman, A., & Stroud, N.J.(2017). News values, cognitive biases, and partisan incivility in comment sections. *Journal of Communication*, 67(4), 586-609.
- Munn, L. (2020). Angry by design: toxic communication and technical architectures *Humanities and Social Sciences Communications*, 7(1), 1–11. <https://doi.org/10.1057/s41599-020-00550-7>.
- Mutz, D. (2016). *In-your-face politics: the consequences of uncivil media*. Princeton University Press.
- Nobata, C., Tetreault, J., Thomas, A., Mehdad, Y., & Chang, Y. (2016). Abusive Language Detection in Online User Content. *Proceedings of the 25th International Conference on World Wide Web (WWW ‘16)*, 145–153. <https://doi.org/10.1145/2872427.2883062>.
- Oswald, S., Lewinski, M., (2014). Pragmatics, cognitive heuristics and the straw man

- fallacy. In: Herman, T. & Oswald, S. (Eds.), *Rhetorique et cognition: perspectives theoriques et strategies persuasives*. Peter Lang, Bern, Switzerland, pp. 313e343.
- Paliewicz, N. S. & McHendry, G. F. (2017). When good arguments do not work: Post-dialectics, argument assemblages, and the networks of climate skepticism. *Argumentation and Advocacy*, 53(4), 287-309.
- Paliewicz, N. S. & McHendry, G. F. (2020). Post-dialectics and fascistic argumentation in the global climate change debate, *Argumentation and Advocacy*, 137-154.
- Paliewicz, N. S. (2022). Arguments of green colonialism: a post-dialectical reading of extractivism in the Americas. *Argumentation and Advocacy*, 1-17.
- Pan, S. Q., Vega, M., Vella, A. J., Archer, B. H., Parlett, G. R. (1996). A mini-Delphi approach: An improvement on single round techniques, *Progress in Tourism and Hospitality Research*, 2(1), 1-109.
- Pankoke, E. (2004). Voluntary Associations and Civic Engagement: European Traditions, Discourses and Perspectives for Voluntary and Intermediary Networks. In: Zimmer, A. & Priller, E. (eds.). *Future of Civil Society*. Wiesbaden: VS Verlag, 57-76.
- Papacharissi, Z. (2004). Democracy online: Civility, politeness, and the democratic potential of online political discussion groups. *New Media & Society*, 6(2), 259-283.
- Papathanassopoulos S., Ioanna Archontaki, Iliana Giannouli, Achilleas Karadimitriou (2022). WP2 Country Report: Greece WP2_Greece Report_UPD.docx (1) (1).pdf
- Pascual-Ferrá, P., Alperstein, N., Barnett, D. J., & Rimal, R. N. (2021). Toxicity and verbal aggression on social media: Polarized discourse on wearing face masks during the COVID-19 pandemic. *Big Data & Society*, 8(1).
<https://doi.org/10.1177/20539517211023533>
- Patel A., Cook Ch. L., Wohn D. Y. (2021). User Opinions on Effective Strategies Against Social Media Toxicity. *Proceedings of the 54th Hawaii International Conference on System Sciences*. <http://hdl.handle.net/10125/70980>
- Petljutschenko, N., Petranova, D., Stashko, H. & Panasenko, N. (2021). Toxicity phenomenon in German and Slovak media: contrastive perspective. *Lege artis. Language yesterday, today, tomorrow*, 2, 105-164.
- Peschke L. & Gümüş Y. A. (2022). EUMEPLAT WP2 Country Report: Turkey.
- Pieper, A. (1991). *Einführung in die Ethik [Introduction to Ethics]*. Tübingen: Francke.
- Poli, R. (2018). A note on the classification of future-related methods. *European Journal of Futures Research*, 6(1), 15.
- Posetti, J., Harrison, J., & Waisbord, S. (2020). Online attacks on women journalists leading to 'real world' violence, new research shows. *International Center for Journalists*.
- Pratchett, (1993). *Theatre of Cruelty*. <https://www.lspace.org/books/toc/index.html>

- Putnam, R. D., Leonardi, R. & Nanetti, R. Y. (1994). *Making Democracy Work: Civic Traditions in Modern Italy*. Princeton University Press.
- Rajadesingan, A., Resnick, P., & Budak, C. (2020). Quick, Community-Specific Learning: How Distinctive Toxicity Norms Are Maintained in Political Subreddits. *Proceedings of the International AAAI Conference on Web and Social Media*, 14(1), 557-568. <https://doi.org/10.1609/icwsm.v14i1.7323>.
- Rösner, L., Winter, S., & Krämer, N. C. (2016). Dangerous minds? Effects of uncivil online comments on aggressive cognitions, emotions, and behavior. *Computers in Human Behavior*, 58, 461-470.
- Rossini, P. (2019). Toxic for Whom? Examining the Targets of Uncivil and Intolerant Discourse in Online Political Talk. In P. Moy & D. Mathlson (Eds), *Voices: Exploring the Shifting Contours of Communication* (pp. 221-242). New York: Peter Lang.
- Sailunaz, K., & Alhajj, R. (2019). Emotion and sentiment analysis from Twitter text. *Journal of Computational Science*, 36, 101003.
- Salminen, J., Sengün, S., Corporan, J., Jung, S-g, & Jansen, B. J. (2020a). Topic-driven toxicity: Exploring the relationship between online toxicity and news topics. *PLoS ONE*, 15(2): e0228723. <https://doi.org/10.1371/journal.pone.0228723>.
- Salminen, J., Hopf, M., Chowdhury, S.A., Jung, S-g., Almerexhi, H., & Jansen, B.J. (2020b). Developing an online hate classifier for multiple social media platforms. *Human-centric Computing and Information Sciences*, 10 (1). <https://doi.org/10.1186/s13673-019-0205-6>.
- San-Jose, L. & Retolaza, J. L. (2016). Is the Delphi method valid for business ethics? A survey analysis. *European Journal of Futures Research*, 4(1), 19.
- Sheth, A., Shalin, V. L., & Kursuncu, U. (2022). Defining and detecting toxicity on social media: context and knowledge are key. *Neurocomputing*, 490, 312-318. <https://doi.org/10.1016/j.neucom.2021.11.095>.
- Shmargad, Y, Coe, K, Kenski, K., & Rains, S.A. (2021). Social norms and the dynamics of online incivility. *Social Science Computer Review* (p. 0894439320985527).
- Silverstone, R. (2005). 12 Domesticating domestication: Reflections on the life of a concept. *Domestication of media and technology* (Eds. Berker et al.) Open University Press.
- Sklan, A. D. (2013). @Socialmedia: Speech with a click of a button? #socialsharingbuttons. *Cardozo Arts & Entertainment Law Journal*, 32(1), 377–410.
- Sobieraj, S., & Berry, J. M. (2011). From incivility to outrage : Political discourse in blogs, talk radio, and cable news. *Political Communication*, 28(1), 19-41.
- Stieglitz, S., & Dang-Xuan, L. (2013). Emotions and information diffusion in social media—sentiment of microblogs and sharing behavior. *Journal of Management Information Systems*, 29(4), 217-248.

- Suler, J. (2004). The online disinhibition effect. *Cyberpsychology & Behavior*, 7(3), 321–326.
- de Sutter, F., Biltreyst, D. van Bauwel, S. (2022). EUMEPLAT WP2 National Report: Belgium.
- Sydnor, E. (2019). *Disrespectful democracy: The psychology of political incivility*. Columbia University Press.
- Taboada, M. (2021). The banality of online toxicity. *Policy, Option, Politiques*. <https://policyoptions.irpp.org/fr/magazines/october-2021/the-banality-of-online-toxicity/>
- Thomaß, B. (1998). *Journalistische Ethik. Ein Vergleich der Diskurse in Frankreich, Großbritannien und Deutschland [Journalistic Ethics. A Comparison of Discourses in France, Great Britain and Germany]*. Opladen: Westdeutscher Verlag.
- Thomaß, B. (2013): Normativität in der internationalen und interkulturellen Kommunikation [Normativity in international and intercultural communication]. In: Karmasin, M., Rath, M. & Thomaß, B. (Eds.). *Normativität in der Kommunikationswissenschaft [Normativity in Communications Studies]*. Springer VS.
- Tirrell, L. (2017). Toxic Speech: Toward an Epidemiology of Discursive Harm. *Philosophical Topics* 45(2), 139-161.
- Titley, G., Keen, E. & Földi, L. (2014). *Starting Points for Combating Hate Speech Online*. Strasbourg, Council of Europe, Youth Department.
- Üzelgün, M. A., Lewiński, M., & Castro, P. (2016). Favorite battlegrounds of climate action: Arguing about scientific consensus, representing science-society relations. *Science Communication*, 38(6), 699-723.
- Üzelgün, M. A. (2023). Expanding deep disagreements: incompatible narratives as interpretative repertoires in political disagreements. *Proceedings of the 2022 European Conference on Argumentation*.
- Vaccari, C., & Valeriani, A. (2021). *Outside the bubble: Social media and political participation in western democracies*. Oxford University Press.
- van den Hoven, P. (2017). Narratives and pragmatic arguments: Ivens' The 400 million. In *Narration as Argument* (ed. P. Olmos) pp. 103-121, Springer Cham.
- van Eemeren, F. H., & R. Grootendorst, (2004). *A Systematic Theory of Argumentation: The Pragma-Dialectical Approach*. New York: Cambridge University Press.
- Walton, D. (2006). Poisoning the well. *Argumentation*, 20, 273-307.
- Waseem, Z., & Hovy, D. (2016). Hateful Symbols or Hateful People? Predictive Features for Hate Speech Detection on Twitter. *Proceedings of the NAACL Student Research Workshop*, 88–93. San Diego, California: Association for Computational Linguistics.
- Waseem, Z., Davidson, T., Warmusley, D., & Weber, I. (2017). Understanding abuse: a

- typology of abusive language detection subtasks. *Proceedings of the First Workshop on Abusive Language Online*, 78-84. Vancouver, BC, Canada: Association for Computational Linguistics.
- Webster, F. (2014). *Theories of the information society*. Routledge.
- Williams, J. (2018). *Stand Out of Our Light: Freedom and Resistance in the Attention Economy*. Cambridge University Press.
- Winkler, J. & Moser, R. (2016). Biases in future-oriented Delphi studies: A cognitive perspective, *Technological Forecasting and Social Change*, 105, 63–76.
- Zarefsky, D. (2016). On deep disagreement. In von Burg A. (Ed.). *Dialogues in Argumentation*. Accessed at:
www.windsor.scholarsportal.info/omp/index.php/wsia/catalog/book/12
- Zenker, F., van Laar, J. A., Cepollaro, B., Gâță, A., Hinton, M., King, C. G., ... & Wagemans, J. H. (2023). Norms of public argumentation and the ideals of correctness and participation. *Argumentation*, 1-34. <https://doi.org/10.1007/s10503-023-09598-6>
- Ziegele, M., Weber, M., Quiring, O., & Breiner, T. (2018). The dynamics of online news discussions: Effects of news articles and reader comments on users' involvement, willingness to participate, and the civility of their contributions. *Information, Communication & Society*, 21(10), 1419-1435.

Get in touch

 info@eumeplat.eu

 www.eumeplat.eu



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101004488

