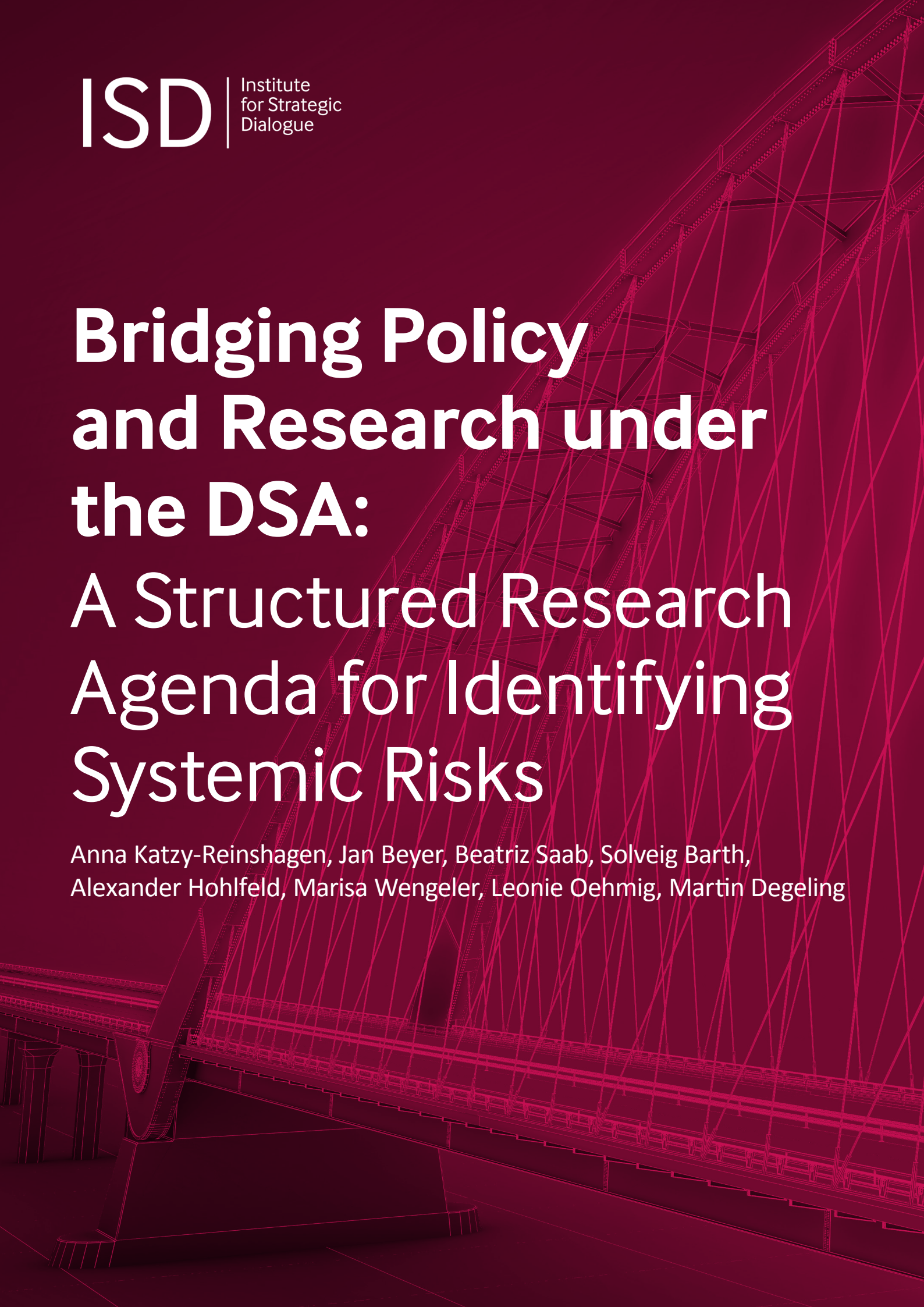
ISD | Institute for Strategic Dialogue

Bridging Policy and Research under the DSA:

A Structured Research Agenda for Identifying Systemic Risks

Anna Katzy-Reinshagen, Jan Beyer, Beatriz Saab, Solveig Barth, Alexander Hohlfeld, Marisa Wengeler, Leonie Oehmig, Martin Degeling



About the project AHEAD.tech

ISD's project AHEAD.tech aims to develop refined methods for analysing systemic risks to support both researchers and regulators in navigating the new regulatory framework introduced by the Digital Services Act (DSA). It is funded by the Mercator Foundation.

The project includes two studies on the asymmetric amplification of political content on TikTok during the Brandenburg state election in 2024 and the German federal election in 2025. Both studies were complemented by a policy brief and workshop series focussing on understanding and mitigating systemic risks as outlined in the DSA. Participation included a diverse stakeholder group including regulators, policymakers and researchers from both civil society organisations and academia. A separate workshop with researchers concentrating on researching risks on TikTok resulted in the publication of a dataset of political accounts that summarises different datasets from the participants.

Drawing on insights from both the empirical studies and the multi-stakeholder workshops, this report builds a bridge between hands-on research and broader policy implications.

www.isdglobal.org

Executive summary

The Digital Services Act (DSA) obliges Very Large Online Platforms (VLOPs) and Very Large Online Search Engines (VLOSEs) to identify, assess and mitigate so-called systemic risks arising from the design, functioning and use of their services. However, the category of systemic risks is deliberately left to be defined through the engagement of stakeholders in an iterative process.

The DSA's risk governance architecture gives researchers a direct and legally recognised function, including both an agenda-setting and control function. This allows them to broaden the scope of what could fall under the definition of systemic risk, supporting providers and regulators.

However, the ambiguity of the systemic risk definition also creates uncertainties. Researchers may find it difficult to understand which research gaps they could prioritise and how to effectively communicate their findings to regulators. Regulators are confronted with broad, diverse empirical and conceptual findings that may be difficult to implement.

This report argues that these challenges can be addressed through a structured research agenda centred on assessing the *severity* and *likelihood* of systemic risks. By providing concrete examples of severity indicators, the research agenda demonstrates how regulators can organise and interpret diverse empirical findings, and how researchers can identify potential research gaps. We argue that no single method can adequately capture the indicators needed to assess the severity and likelihood of systemic risks. Instead, a combination of observational and experimental approaches is required.

We demonstrate how severity can be operationalised across three dimensions (scale, scope and remediability). By mapping concrete indicators and associated methods, we show how the research agenda structures evidence on when risks linked to asymmetric amplification may become systemic.

To illustrate this approach, we use a case study of asymmetric amplification of political content (defined as the unequal distribution of political content, caused by algorithmic ranking systems that surface some content more than others when compared to a neutral or alternative baseline). Although asymmetric amplification is associated with a variety of risks to civic discourse and electoral processes, there is an ongoing discussion as to whether these risks would also fall under the scope of the DSA. It is therefore a useful example to showcase what potential evidence would need to look like.

The conclusion outlines how the research agenda could be applied to other potential systemic risks and recommends a variety of potential measurements. These include long-term funding of research projects, the adoption of a harmonised and broad definition of public data, stronger collaboration to identify research gaps and more transparency by platforms and regulators regarding investigations and risk assessments.

Contents

Glossary

Asymmetric amplification: The unequal distribution of political content, caused by algorithmic ranking systems that surface some content more than others when compared to a neutral or alternative baseline.

Bot networks: A group of automated accounts or devices controlled to perform coordinated actions online, often used to spread disinformation, amplify content or manipulate public discourse.[1]

Content-related risks: Risks arising from the creation, dissemination or amplification of illegal content.[2]

Disinformation: False or misleading content spread with an intention to deceive or secure economic or political gain, and which may cause public harm.[3]

Foreign information manipulation and interference (FIMI): behaviour that is generally not illegal but threatens or has the potential to negatively affect values, procedures and political processes. Such activity is manipulative in nature. It is carried out deliberately and in a coordinated manner by state or non-state actors (including proxies) globally.[4]

Filter bubble: An online phenomenon where people are mainly exposed to opinions and information that reinforce their existing beliefs. Filter bubbles are driven by an ecosystem of information shaped by algorithmic personalisation, in which platforms selectively display content based on user data. This limits exposure to differing viewpoints and can lead to users' isolation.[5]

For You feed: The primary, personalised feed on TikTok that curates videos based on users' interests, behaviour and engagement history. Displayed by default when the app is opened, the For You feed attempts to provide content aligned with the users' viewing preferences.[6]

Misinformation: False or misleading content shared without harmful intent, e.g. when people share false information with friends and family in good faith. However, the effects can still be harmful.

Recommender system: A fully or partially automated system used by an online platform to suggest specific information to recipients of the service or to prioritise that information in its online interface. This includes information displayed as a result of a search initiated by users or otherwise the determining of the relative order or prominence of information shown.[7]

Remediability: A severity dimension that captures how feasible it is to correct or reverse the negative effects of a risk.

Scale: A severity dimension that captures the intensity of the impact that a risk may produce, independent of how many people are affected.

Scope: A severity dimension that captures the breadth of impact, i.e. how widely a risk manifests across communities, geographies, languages or platforms.

Systemicness: Term used in this report to refer to the systemic nature of a risk, applied in the context of the DSA.

Virtual Computing Environment (VCE): A digital infrastructure provided to researchers that encapsulates access control and data processing resources. Data, access and resources are maintained and completely under the VCE provider's control. They are made available through an online interface.

Very Large Online Platform (VLOP)/Very Large Online Search Engine (VLOSE): Online platform or search engine with more than an average of 45 million monthly active recipients in the EU designated by the European Commission under the DSA.[8]

Relevant legislative documents

Short name	Full title
Digital Services Act (DSA)	Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act) (Text with EEA relevance)
Guidelines on the mitigation of systemic risks for electoral processes	Commission Guidelines for providers of Very Large Online Platforms and Very Large Online Search Engines on the mitigation of systemic risks for electoral processes pursuant to Article 35(3) of Regulation (EU) 2022/2065
Delegated Regulation on Data Access	Commission Delegated Regulation (EU) 2025/2050 of 1 July 2025 supplementing Regulation (EU) 2022/2065 of the European Parliament and of the Council by laying down the technical conditions and procedures under which providers of very large online platforms and of very large online search engines are to share data with vetted researchers

Figures and tables

1. Introduction

The Digital Services Act (DSA) represents a significant evolution in the EU's regulation of the online environment. It incorporates the existing liability rules of the e-Commerce Directive but also introduces a risk-based layer of due diligence obligations. This requires VLOPs and VLOSEs to identify, assess and mitigate so-called systemic risks arising from the design, functioning and use of their services. However, the category of systemic risks is deliberately ambiguous. Art. 34 DSA only sets out broad risk categories, such as illegal content and negative effects on the exercise of fundamental rights, as well as on civic discourse, electoral processes, and public security.

The concept of systemic risks provides flexibility for emerging technological and societal challenges. However, determining when a risk reaches a level of societal significance to warrant regulatory intervention remains a major challenge for DSA enforcement. It is difficult to prove a causal link between platform dynamics and societal effects across the Member States of the EU and the sprawling digital ecosystem. Furthermore, the functioning of platform services remains highly opaque, which is also why they are referred to as "black boxes".[9]

Against this background, the DSA has assigned civil society and academic researchers a direct and legally recognised role in the risk governance architecture. These stakeholders are empowered to generate independent evidence capable of reducing the opacity of VLOPs and VLOSEs, identifying persistent or emerging risks potentially overlooked by platforms and regulators. The use of broad risk categories allows researchers significant freedom in their choice of investigations.

At the same time, the vague concept of systemic risks also creates uncertainties for regulators and researchers. As systemic risks remain conceptually fluid and contested, researchers often report uncertainty about which research gaps to prioritise and how to effectively communicate their findings to government stakeholders.[10] Regulators may also struggle: it can be difficult for them to incorporate diverse methods and findings from researchers into an enforcement framework. This challenge is heightened by the pluralism of research approaches, methodological diversity and ongoing scientific debate that characterises research practices.

The report proposes the development of a structured research agenda which would form a coherent and useable basis for both regulators and researchers. It builds off work by Renn et al. (2022) who note that systemic risks cannot be captured through a single measurement. They instead suggest it should be understood through multiple indicators and dynamic dimensions which provide a richer picture of emerging harms when taken together.[11] The research agenda follows this logic and shows how to integrate this multiplicity of data points, measurements and methods.

The report first clarifies the terminology of systemic risks and the role assigned to researchers under the DSA. It then introduces asymmetric amplification of political content as a case study that illustrates the challenges researchers face. In this report, asymmetric amplification is not understood as a risk in itself, but rather as a phenomenon that is associated with various risks in research. We then present the research agenda and demonstrate how diverse strands of existing research can be organised into a coherent structure. This structure is intended to both support regulators in evaluating and responding to emerging risks and to guide researchers in identifying relevant research questions.

The report shows that a new, single methodology for assessing systemic risks is neither feasible nor necessary. Instead, a plurality of methods is required to capture the multiple dimensions that shape what we describe as the systemicness of a risk. This plurality also compensates for constraints that are inherent to certain methods and limitations on data access. The overarching aim is to make the process of defining and evidencing systemic risks more transparent, coherent and collectively informed, allowing for proportionate, effective and research-driven enforcement of the DSA.

2. Systemic risks and the role of researchers under the DSA

The following sections aim to clarify the notion of systemic risks within the DSA framework and to analyse the implications this has for the role of researchers. The ambiguity of the systemic risk definition affects how platforms approach their risk assessments; it also shapes whether researchers can effectively fulfil the role assigned to them in the DSA.

2.1 The concept of systemic risk

As a central organising principle of the DSA's risk-based approach, systemic risk lacks the definitional precision typically required in a regulatory context. While risks can be broadly understood as "adverse events that may occur in the future,"[12] it is much harder to settle on a definition of "systemic" in this context.

In general, systemic risks can be understood as "potential threats that endanger the functionality of systems of critical importance for society",[13] with impacts that may extend across systems and over time and geographic location. Such risks are further characterised by complexity, uncertainty and ambiguity.[14] In the financial sector, systemic risks are clearly described in reference to a situation in which shock at one financial institution threatens to destabilise the entire industry.[15]

Such definitions cannot be directly translated into the context of the DSA. The "shock" model at the core of systemic risks in finance does not capture the dynamics of online risks: they are shaped by the interplay between recommendation systems, content moderation architectures, interface design and the intentions of malicious actors. As harms on digital platforms often emerge gradually over time, they require a broader and more flexible understanding of systemic risk.[16]

This absence of a precise definition of systemic risk in the DSA might not be an oversight but a deliberate regulatory choice.[17] The financial sector can rely on a long-established and relatively stable understanding of how systemic events unfold, which allows regulators to codify specific triggers, contagion channels and prudential responses. By contrast, the DSA operates in an environment where harms are sociotechnical, rapidly evolving and highly context dependent.

The meaning of systemic risk under the DSA therefore emerges iteratively through regulatory practice, oversight cycles and the engagement of multiple stakeholders who interpret harms to fundamental rights in real time, civic discourse, public health or the protection of minors. **In short, its coherence is meant to be built through use rather than prescribed from the outset.** The presence of multiple actors in the process is thereby also linked to the presence of diverging interests that in turn shape the definition of what can be considered a systemic risk.

VLOPs and VLOSEs tend to adopt content-based, compliance-oriented definitions of systemic risks. The first published reports with the results from the risk assessment have shown that such an interpretation can lead to a focus on symptoms, such as the prevalence of illegal content, rather than underlying causes or broader societal impacts. Platform elements, such as the design as a driver for risks, for example, have been widely neglected so far.[18]

Researchers argue for a broad, multi-dimensional definition of systemic risks that supports a diverse research ecosystem. Their argument is that narrow definitions that require causal proof or quantitative thresholds risk excluding valuable methodologies and perspectives.[19] Yet, other researchers caution that broad and vague definitions could grant platforms as well as the European Commission (as the primary regulator for VLOPs and VLOSEs) considerable discretionary power. This could undermine the legality principle (that restrictions on speech must be provided by the law) and incentivise platforms to impose overly strict mitigation measures stifling freedom of expression.[20]

In addition, unclarity exists whether risks posed by VLOPs and VLOSEs are inherently systemic, given that they are described in Rec. 76 DSA as "different in scope and impact from those caused by smaller platforms", or whether additional, more specific thresholds must be met to qualify as having major social, political or ethical effects.

Systemic risks remain conceptually fluid and open to competing interpretations, creating uncertainty for both assessment and enforcement. The next section focuses on researchers, who play a crucial role in navigating this ambiguity within the DSA's enforcement framework.

2.2 The role of researchers under the DSA

As previously established, the definition of a systemic risk in the DSA will be the result of an iterative process involving a variety of stakeholders, including researchers.[21]

The DSA fundamentally expands the role of researchers within EU platform regulation.[22] Whereas earlier regulatory frameworks drew researchers only in indirect and peripheral ways, the DSA introduces an entirely new model of involvement. In established areas of EU risk regulation, researchers typically contribute as external experts whose knowledge is incorporated only later in the process: for instance, through expert testimony, scientific publications consulted by regulator, or participation in advisory committees. This meant that their role was limited to offering input when invited, without any official structural position within the regulation itself. The DSA instead gives researchers a direct and legally recognised place within the risk governance architecture.

This role has two parts: a control function and an "agenda-setting" function. First, the DSA empowers researchers to take on a control function. Online platforms, for a long time, have been described as "black boxes" [23] to which only they themselves had access. The DSA not only acknowledges the importance of "bridging information asymmetries" (Rec. 96 DSA) through research, reflecting the opacity of platforms, but also obliges VLOPs and VLOSEs to provide researchers with data necessary to conduct research on the "detection, identification and understanding of systemic risks" (Art. 40 DSA). The results of such research should be shared in the risk management process through stakeholder participation. More concretely, Rec. 90 DSA emphasises that platforms' risk assessments and implemented mitigation measures should be conducted "with the involvement of [...] independent experts and civil society organisations." The DSA therefore explicitly includes researchers as control element in the risk management process.

Additionally, the role of researchers under the DSA entails a second, less explicit function, which we describe as an "agenda-setting function." While Art. 40 DSA makes a strong case for researchers to contribute to risk assessment, it is up to researchers to decide which risks they focus on in their research. When researchers uncover evidence of systemic risks that are not addressed in platforms' risk assessments, they can give the European Commission "an initial suspicion to further investigate the matter." This creates "an incentivising effect" [24] for platforms to take research findings into account, even if they are not legally obliged to.

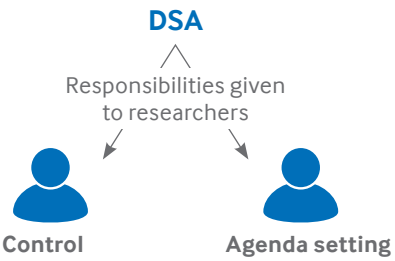


Figure 1: The role of researchers under the DSA.

In practice, however, translating research findings into the regulatory process is difficult. Researchers are expected to provide evidence on systemic risks, yet the concept they are asked to investigate remains underdefined in the DSA. As a result, they must work without clear criteria for what counts as a systemic effect, which complicates decisions about what to measure, how to interpret empirical findings and how to present conclusions. Also, different methodological approaches can generate complementary but also contradictory results, which is challenging to reflect in policy measures. Finally, academic research often uses different terminology from platform risk assessments or policy publications, making it challenging to bridge research results and policy.

Uncertainty around systemic risk is heightened by the ambiguity around certain risk categories (Art. 34(1) DSA). Risks such as the dissemination of illegal content are easily defined legally or operationally (based on the legal frameworks of Member States) – risks such as "fundamental rights," "public security" or "civic discourse" are, however, not.[25] The next section illustrates how this conceptual uncertainty appears in practice and introduces a research agenda that can help structure diverse findings despite the unclear definition of systemic risks.

3. Asymmetric amplification as a case study

While the broad concept of systemic risk already poses challenges for translating research into regulatory practice, the complexity is further amplified by the fact that risk categories (Art. 34(1) DSA) themselves are not always clearly delineated. For example, risks such as the dissemination of illegal content are more easily defined, as they are anchored in the legal frameworks of EU Member States. In contrast, risks related to "fundamental rights," "public security," or "civic discourse" are often value-laden and difficult to capture in legal or operational terms.[26] Yet, this definitional ambiguity does not make such risks any less significant. On the contrary, they may pose serious threats to democratic processes and public trust, underscoring the need to integrate diverse research perspectives into regulatory decision-making.

The following sections explore the concept of asymmetric amplification of political content during elections as a complex and ambiguous phenomenon associated with a variety of risks. These have the potential to significantly influence civic discourse and the integrity of electoral processes, but it is ambiguous as to whether they qualify as systemic risks under the DSA. This very ambiguity makes it a valuable case for exploring what kind of empirical evidence regulators would need to make that determination.

3.1 The definition of asymmetric amplification

The visibility of online content on digital platforms is closely linked to recommender systems, which are the central technical systems that underpin the business model of most VLOPs. Recommender systems function as information filtering systems designed to predict and suggest items or content that a user might be interested in. As such, they shape what content is surfaced, promoted and (conversely) deprioritised.

There are different types of recommender systems. On social media platforms, engagement-based recommender systems are commonly used. These are at least partly-automated systems which prioritise content that is believed to maximise user engagement and time spent on platform, based on past user behaviour, expressed preferences or patterns "derived from similar users".[27] This takes into account both explicit[28] and implicit feedback.[29] Explicit feedback describes preferences users communicate in form of direct input, for example by answering questions about their interest when opening an application for the first time. Implicit feedback describes user interactions with the system that are not direct statements of preferences but still tell something about the user's interests. These include swiping, saving or liking of content.

However, there are also other approaches that have been discussed as alternatives to the engagement-based algorithms. For example, researchers have tested the effect of so-called reverse-chronological feeds on user behaviour, in which content is presented in the order it is being posted.[30] Reverse chronological feeds were the default prior to engagement based feeds. Another type of algorithms that has recently gained more attention are bridging-based algorithms.[31] These recommend content that is liked by two users that usually do not share all preferences. Hence it is seen to be "bridging" differing political opinions instead of pushing users into filter bubbles.

No matter which type of recommender system is implemented on platforms, service providers have an important influence on the visibility of online content, including during politically sensitive periods, such as elections. It is in this context that the phenomenon of asymmetric amplification has become central in the public debate around algorithmic systems. The debate has led to both regulatory action and research, for example on the question whether platform algorithms systematically amplify certain political viewpoints over others,[32] but also regulatory action. In 2024, for example, the European Commission opened proceedings against Meta for demoting political content on the feeds of Instagram and Facebook.[33]

In the ongoing debate about what is meant by the term "asymmetric amplification" the question of an appropriate baseline is central. This raises the issue of comparison, specifically what amplified content is relative to. Thus, most definitions of amplification draw on the idea of relative amplification. They define it as the extent to which the algorithm gives certain content greater reach than it would have received with some other neutral baseline algorithms, such as a reverse-chronological feed[34] or the change in the distribution of content with engagement-based algorithms compared to alternatives, such as bridging-based algorithms while holding user behaviour constant.[35]

While asymmetric amplification is not constrained to a specific type of content, this report focuses on asymmetries in the amplification of political content during

election periods. In the context of elections, asymmetric amplification refers to an unequal coverage or promotion of political content from electoral contenders or political accounts relevant to the election on online platforms during an election period.[36] The reasons for varying levels of coverage are diverse, and unequal coverage does not inherently constitute a systemic risk.

For the purpose of this report, asymmetric amplification in the context of elections is defined as **the unequal distribution of political content caused by algorithmic ranking systems that surface some content more than others when compared to a neutral or alternative baseline.**

Asymmetric amplification can appear on personalised user feeds, suggested accounts to follow, search engine result rankings, autocomplete suggestions or any other element of a platform where suggestions to users are generated algorithmically. It can play an important role during elections, as even small changes in visibility may influence how political information reaches voters. The DSA identifies several election-relevant risks as potentially systemic. These include disruptions to electoral processes, civic discourse and citizens' ability to access reliable information when making political choices. Asymmetric amplification can interact with these risks in multiple ways, yet it remains uncertain when such dynamics rise to the threshold of systemic concern under Art. 34 and 35 DSA.

3.2 Risks linked to asymmetric amplification: systemic relevance?

Digital platforms are an important source for accessing information[37] and can influence individual opinion formation.[38] This entails the risk that algorithmic distortions can have systemic implications on electoral fairness and democratic legitimacy.[39] If platforms disproportionately amplify sensationalist, extreme or divisive content, this could deepen societal polarisation.[40] Furthermore, if algorithms asymmetrically amplify some political actors while de-amplifying others, it could significantly influence voters' knowledge and attitudes towards candidates.

So far, research has found evidence for asymmetries in content distribution on several major platforms. Research on X (formerly Twitter) found that in "six out of seven countries studied, the mainstream political right enjoys higher algorithmic amplification than the mainstream political left".[41] Other studies found asymmetries in content distribution and skews towards Republican content on TikTok during the 2024 presential election.[42] And studies on YouTube found skews towards both left-leaning content in the US between December 2022 and January 2023 on the platform,[43] as well as towards far-right content over a two-week period in January/February 2019.[44] Another study found that YouTube would direct right-leaning users in particular to ideologically biased and increasingly radical content.[45] This dynamic may not only privilege certain voices but also effectively silence others.

The risks extend to broader societal concerns, including the targeted dissemination of misleading information about voting procedures. In electoral contexts, this represents a further pathway through which asymmetric amplification can translate into concrete harm. Such practices can generate confusion, depress participation or even suppress turnout among specific communities. Research on what Cirone and Hobbs (2023) call "asymmetric flooding" illustrates this dynamic: targeted communities are inundated with large volumes of apolitical or distracting content, exploiting how recommender systems prioritise recency, volume and engagement. By doing so, coordinated influence campaigns can crowd out mobilising political information and suppress participation.[46] These dynamics highlight the wider risk that algorithms can be manipulated by users or malicious actors, potentially through botnets or automation.[47]

Asymmetric amplification has further been linked to the elevated presence of polarising rhetoric,[48] reduced exposure to diverse viewpoints,[49] and the circulation of misinformation[50] or hateful narratives (particularly during periods of heightened political tension). For example, Bail (2021) argues that social media could act as a "prism" refracting identities, creating distorted perceptions of society by amplifying extremists and silencing moderates.[51] This can contribute to perceived polarisation by portraying politicians and parties as more extreme than they actually are.[52] A reason could lie in the very design of recommender systems as "[e]ngagement-based ranking rewards conflict creators over bridge builders".[53] A study on X compared a reverse-chronological baseline to the platform's engagement-based timeline.[54] It found that emotionally charged content hostile towards political out-groups was relatively amplified; this was despite users reporting that such posts made them feel worse

about their political output and expressed a preference against them.

Some of these concerns are directly reflected in the DSA which notes that recommender systems "can have a significant impact on the ability of recipients to retrieve and interact with information online [...]". It also states that "[t]hey also play an important role in the amplification of certain messages, the viral dissemination of information and the stimulation of online behaviour" (Rec.70 DSA). The DSA explicitly addresses the issue of algorithmic amplification in its Recitals, stating that "[w]here the algorithmic amplification of information contributes to the systemic risks, those providers should duly reflect this in their risk assessments" (Rec. 84 DSA).

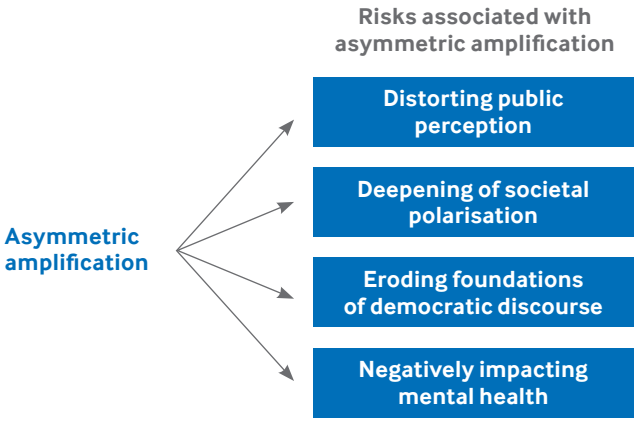


Figure 2: Overview of risks associated with asymmetric amplification.

Recommender systems are defined by the DSA as "fully or partially automated system[s] used by an online platform to suggest in its online interface specific information to recipients of the service or prioritise that information, including as a result of a search initiated by the recipient of the service or otherwise determining the relative order or prominence of information displayed" (Art. 3(s) DSA). The DSA considers recommender systems as an influencing factor for systemic risks (Art. 34(2)(a) DSA), as well as a key mitigation measure (Art. 35(1)(d) DSA). Risk assessments shall consider whether and how risks are "influenced by intentional manipulation of their service, including by inauthentic use or automated exploitation of the service" (Art. 34(2) DSA). This is not limited to illegal content: platforms also need to take into account "how their services are used to disseminate or amplify misleading or deceptive content, including disinformation" (Rec. 84 DSA).

As stated in Art. 35 DSA, providers are required to mitigate such risks sufficiently. This explicitly includes the testing and adapting of recommender systems, to "correct the criteria used in their recommendations" (Rec.88 DSA) and to "prevent or minimise biases that lead to [...] discrimination of persons in vulnerable situations" (Rec. 94 DSA). In its guidelines on the mitigation of systemic risks for electoral processes, the European Commission has provided further guidance on potential measures. These include, among others, ensuring that recommender systems are designed and adjusted to give users meaningful choices and control over their feeds, with due regard for media diversity and pluralism. Other measures include the limited amplification of deceptive content and red-teaming exercises to identify potential risks, such as those stemming from biases. Although not legally binding, the guidelines set out best practices and potential mitigation measures to help providers of VLOPs and VLOSEs comply with Art. 35 of the DSA. They serve as a key reference point for assessing the adequacy of mitigation measures under Art. 35 DSA.

This highlights that risks falling under the umbrella of asymmetric amplification can also be included in the scope of the DSA as potential systemic risks to identify and assess (Art. 34 DSA), as well as to mitigate (Art. 35 DSA).

While the DSA clearly recognises the potential adverse role of algorithms, the empirical evidence presented above still falls short of establishing systemicness. The main challenge is that recommender systems do not operate in isolation. Amplification outcomes are shaped by both supply (in this case posts from political candidates) and user demand. Hence, proving that the reason for asymmetric amplification lies in the algorithmic design is hardly possible in the current data access environment.

A further limitation is that each study represents a single data point. Their findings may be contradictory, shaped by distinct methodological choices or simply not comparable, which produces an overall inconclusive picture. If there is a reasonable suspicion that asymmetric amplification could be linked to systemic risks, the question arises of how researchers can generate the evidence needed for regulators to make this judgement.

4. Introduction of the research agenda and its implementation in practice

4.1 Introduction of the agenda

The research agenda is a structured framework that organises the many different types of evidence researchers produce into a clear and comparable format. It translates the broad ideas of the DSA into concrete analytical dimensions, using two concepts rooted in the act: severity and likelihood. This approach provides a common structure for researchers to locate their findings within a shared analytical map. It also allows regulators to understand how individual studies relate to one another.

It is not the intention of this agenda to determine whether a given risk is systemic - a judgement that lies with regulators and, where contested, the courts. Instead, it clarifies the empirical evidence, methodological approaches and analytical reasoning needed to support such regulatory decisions. This helps ensure platform risk assessments are transparent and evidence-based, allowing regulators to situate studies within a broader analytical framework and enabling researchers to contribute their findings to a shared body of evidence.

Structuring the agenda, severity and likelihood constitute the central concepts in the presented approach. Art. 34 DSA requires platforms to identify potential negative effects arising from their systems, and Rec. 85 DSA explicitly anchors likelihood assessments in historical or evidence-based analysis. These concepts also mirror international risk assessment standards used by regulators, industry, academia and civil society, including Ofcom's Online Safety guidance[55], the World Economic Forum's Global Coalition for Digital Safety[56], ISO 31000[57] and established human rights impact assessment methodologies.[58]

To make these concepts more usable for empirical research, the agenda translates severity into three analytical dimensions. These dimensions are scale, scope and remediability. They are widely used in rights-based due diligence and align with the DSA's requirements.[59]

The agenda also translates likelihood into three complementary dimensions. This reflects the DSA's emphasis on understanding the probability of risks materialising, including through evidence-based assessments grounded in historical data. These dimensions are platform characteristics, historical patterns and contextual factors.

The six dimensions operationalise the DSA's core risk assessment logic as set out in Art. 34 DSA and Recitals 79 and 85. They support a consistent interpretation of observable signals across datasets, platforms and risk categories.

Severity	Likelihood
Scale identifies how widespread or intensive a potential effect is,	**Platform characteristics** identify how the design and functioning of a service influence the probability that a risk emerges,
scope identifies whose rights, interests or communities may be affected,	**historical patterns** capture whether similar harms have occurred before and whether past mitigation efforts prevented recurrence,
remediability assesses how preventable or correctable the effects are, given platform design, architecture and policy choices.	**contextual factors** reflect how external conditions, such as elections or crises, affect the likelihood that certain risks arise.

Table 1: Overview of the methodology-driven risk assessment framework.

These six dimensions provide a structured way to gather, compare and interpret evidence relevant to a systemic risk analysis. While they cannot alone determine if a risk is systemic, they help reveal where patterns of systemicness may emerge. As such, they can highlight whether evidence is sufficient to support regulatory conclusions.

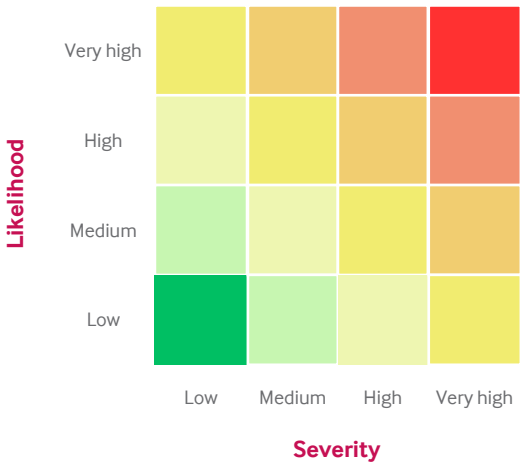


Figure 3: Risk matrix with levels of likelihood and severity.[60]

Assessing systemic risks requires engaging with both severity and likelihood dimensions. For this reason, the agenda emphasises the need for a coordinated interplay of multiple methodologies, including quantitative, qualitative, experimental, architectural and comparative approaches. Organising these approaches within a shared structure allows researchers and regulators to:

- map where evidence is strong or incomplete,
- identify knowledge gaps and information asymmetries,
- understand how methodological limitations shape conclusions and
- collectively build a cumulative evidence base that can support assessments of systemicness.

4.2 Application of indicators in our case study of asymmetric amplification

Building on the agenda framework developed in section 4.1, this section focuses on the severity dimension. It also exemplifies the implicit assumption of the agenda, as identifying an adverse phenomenon is only the starting point. A phenomenon becomes relevant from a systemic risk perspective not simply because it occurs, but only when the risks it produces or enables reach a certain severity. Since this section focuses on operationalising severity, the analysis centres on the three elements through which severity can be evidenced: how large the phenomenon is (scale), how far its effects extend across users, groups and contexts (scope), and how difficult it is to prevent, limit or reverse the associated harms, given existing platform architectures and mitigation measures (remediability).

Without clarifying the severity dimension in this way, research on asymmetric amplification risks remains fragmented. This can make it difficult for regulators to judge when observed patterns may become sufficiently extensive, persistent or consequential to warrant regulatory attention. The agenda therefore insists on a plurality of complementary measurements rather than a single decisive metric and offers a shared structure for relating these measurements to potential regulatory decisions.

The sub-sections that follow therefore do not introduce new concepts of risk. They do not attempt to decide whether asymmetric amplification already meets the DSA's threshold of systemicness. Instead, they set out the types of questions and empirical aspects that studies would need to address under each dimension of severity to understand how risks emerging from asymmetric amplification might become systemic.

In doing so, they illustrate the complexity of the research agenda itself: even when focusing solely on impact, a meaningful assessment already presupposes a dense set of indicators and lines of inquiry. At the same time, by specifying which kinds of indicators would be needed under each dimension, the agenda makes it possible to locate individual studies within a broader evidentiary grid and to see where important questions have not yet been addressed or cannot be answered with existing data. After this is done, section 4.3 will then link this agenda to the methodological toolbox required to answer these questions in practice and to assemble diverse forms of evidence into a coherent basis for regulatory decision-making.

Indicators for risks associated with asymmetric amplification

Scale

Scale captures the severity or intensity of the impact that a risk may produce, independent of how many people are affected. Because asymmetric amplification is not a risk in itself but a phenomenon linked to multiple risks, such as polarisation, distortions of public discourse or certain forms of voter manipulation, the task is to determine how asymmetric amplification may intensify these risks under different electoral contexts.

Scale-related indicators help determine how strongly a risk linked to asymmetric amplification manifests. For instance, if the relevant risk is polarisation, researchers might analyse the skewness of public debate by measuring visibility differentials between actors or narratives.[61] Metrics such as impressions, engagement velocity or content share can quantify the intensity of such imbalances.

Broader signals of severity often materialise through visibility imbalances. For example, disproportionate exposure may be a proxy for exclusion from debate or amplified influence by certain actors. This could potentially affect opinion formation or democratic fairness.[62] This is directly relevant to the examples in section 3.1. When read through the research agenda, such signals can be treated as scale-oriented in the matrix that link concrete measurements of visibility to broader questions of democratic fairness and rights.

Crucially, visibility imbalance is only meaningful when assessed alongside other dimensions, such as scope and remediability, and when considering contextual evidence such as inauthentic behaviour or coordinated cross platform activity. Their primary function is to ensure that assessments of asymmetric amplification are grounded in a plurality of concrete measurements of intensity. These can then be combined with other dimensions rather than treated as a single decisive threshold.

The following table presents a deliberately broad selection of scale-related indicators that researchers can use to assess the intensity of asymmetric amplification impact across different electoral scenarios. It demonstrates how even a limited subset of indicators already spans multiple analytical traditions and data requirements. This underscores the agenda's central premise that no single approach can capture the full severity of amplification dynamics. The sources provided in the columns on associated risks indicators are example studies that aim to understand such risks.

Dimension	Indicator (signal)	Metrics (measures)	Associated risks to which indicators contribute
Scale	**Content volume imbalance:** disproportionate amount of political/ election content from one actor/narrative	• Number of posts, comments and hashtags, • content volume by actor/party, • percentage share of election-related posts.	This indicator maps which actors or narratives dominate the flow of election-related information and can help to identify overrepresentation. If certain content keeps appearing across time or platforms this could indicate a structural visibility advantage that shapes the broader information environment. In the context of asymmetric amplification, this indicator contributes to the following potential measuring aspects: • Skewed public discourse where competing viewpoints are less visible, • perception of broad support where little exists (manufactured consensus), • heightened polarisation if one-sided narratives saturate feeds, • voter manipulation if one political actor overwhelms the information space during sensitive periods.
	Reach asymmetry: one actor/narrative consistently achieves higher reach than others	• Impressions, • watch time, • unique viewers, • engagement ratio between viewpoints (likes/shares/ comments)	Reach asymmetry compares content diffusion across competing political actors or ideological camps. When disparities persist, despite similar posting frequency or audience size, this may suggest algorithmic bias, network-based boosting or coordinated inauthentic behaviour rather than organic user preference. In the context of asymmetric amplification, this indicator contributes to the following potential measuring aspects: • Disproportionate influence on political opinion formation, • distorted marketplaces of ideas where alternative narratives have limited reach, • unequal visibility in electoral debates (risk of democratic imbalance), • intensified polarisation if extreme narratives gain amplified reach.
	Amplification velocity: sudden acceleration of visibility for certain narratives	• Time-to-rapid spread, • posting frequency, • engagement velocity metrics (likes/shares, per minute/hour).	Amplification velocity measures how quickly election narratives gain traction, marked by sudden visibility spikes. Rapid surges indicate possible algorithmic promotion or coordinated boosting, highlighting brief but intense bursts that can shape public discourse during critical election periods. In the context of asymmetric amplification, this indicator contributes to the following potential measuring aspects: • Rapid spread of misleading or harmful content before moderation reacts, • accelerated diffusion of voter suppression messages, • emotional reactions and waves of outrage contributing to polarisation, • agenda hijacking during election cycles or crises.

Table 2: Overview of indicators and metrics, contributing for measuring scale of asymmetric amplification in the context of elections.

Continuation next page

Continuation **Table 2:**

Dimension	Indicator (signal)	Metrics (measures)	Associated risks to which indicators contribute
Scale	**Volume of boosted content:** algorithmic or coordinated signals amplify certain narratives disproportionately	• Percentage of top-N trending posts/topics dominated by one actor, • share of algorithmically recommended vs. organically reached content, • boost intensity (ratio of reach via recommender vs. follow-based reach).	This indicator assesses whether amplification is dominated by top-performing posts. Disproportionate visibility of certain actors/narratives may indicate algorithmic bias/manipulation. This is measured by comparing recommended versus organic reach. In the context of asymmetric amplification, this indicator contributes to the following potential measuring aspects: • Systemic distortion of public debate, • misleading impression of narrative dominance ("amplified realities"), • higher exposure to harmful-but-legal political content, • systemic risk where harmful narratives bypass normal friction or scrutiny.
	Growth velocity of actors: rapid expansion of political accounts benefiting from amplification	• Percentage of top-N trending posts/topics dominated by one actor, • share of algorithmically recommended vs. organically reached content, • boost intensity (ratio of reach via recommender vs. follow-based reach).	Growth velocity of actors assesses the broader, cumulative impact of amplification by tracking how exposure translates into audience expansion. In the context of asymmetric amplification, this indicator contributes to the following potential measuring aspects: • Biased perceptions of credibility or popularity, • increased dissemination power for actors exploiting amplification, • accelerated mobilisation or radicalisation pathways via artificially-inflated influencers, • unequal competitive conditions in d emocratic processes.

Table 2: Overview of indicators and metrics, contributing for measuring scale of asymmetric amplification in the context of elections.

Scope

This section shows how scope can be operationalised when assessing risks associated with asymmetric amplification. As previously introduced, scope captures the breadth of impact, i.e. how widely a risk manifests across communities, geographies, languages or platforms. In the context of asymmetric amplification, scope helps determine whether amplification effects spread across populations and information environments in ways that broaden the potential for harm.

In electoral settings, scope-related indicators allow researchers to assess which groups or regions are most exposed to amplified narratives and whether certain communities experience disproportionate influence. Studies examining political content recommendations across user profiles on TikTok and other platforms show how exposure can vary by political orientation, age or region.[63] By analysing audience breakdowns, recommendation patterns or demographic exposure ratios, researchers can begin to understand which groups are reached by amplification reaches and how broadly its effects extend.

Such measurements help clarify whether the amplification patterns, such as the preferential boosting of certain political actors, affect broader demographic groups. They also help identify whether misleading information about voting procedures circulates mainly within targeted communities or whether it spills into wider population segments. This in turn indicates the potential extent of voter confusion or suppression.

Scope indicators help determine how far a risk linked to asymmetric amplification extends and how many communities may be affected, forming another essential part of the broader evidentiary puzzle. If polarising content or hostile out-group narratives spread far beyond their initial audiences, scope indicators help determine whether these dynamics are shaping perceptions across entire regions or linguistic communities. Similarly, if coordinated tactics such as asymmetric flooding are detected across multiple demographic groups rather than isolated segments, scope indicators can indicate potential influence on public discourse or electoral engagement at a societal level.

As with scale, indicators of scope do not establish systemic risk in isolation. A narrative that spreads widely may simply reflect organic user interest, platform demographics or uneven content production across communities. Scope becomes meaningful only when interpreted alongside scale and remediability. It should also be considered in light of contextual evidence such as coordinated posting, foreign-linked activity or synchronised cross-platform dissemination.

The following table presents a deliberately varied selection of scope-related indicators that researchers can use to assess how widely asymmetric amplification extends across electoral contexts. It illustrates how even a limited set of indicators draws on multiple analytical perspectives: these include demographic exposure analysis, geographic distribution patterns, linguistic reach and cross-platform dissemination.

Dimension	Indicator (signal)	Metrics (measures)	Associated risks to which indicators contribute
	Cross-platform spread: narratives gain traction across multiple platforms	• Cross-platform presence index, • cross-platform repost ratio, • referral traffic intensity, • topic/hashtag migration velocity, • cross-platform narrative coherence.	Cross-platform spread signals the ecosystem-level reach of amplified narratives. When identical or near-identical content appears across multiple platforms, it may suggest that amplification effects exceed a single service's recommender logic, revealing coordinated dissemination or platform spillover effects. In the context of asymmetric amplification, this indicator contributes to the following potential measuring aspects: • Wider polarisation as narratives reach new audiences, • increased visibility of coordinated inauthentic behaviour across platforms, • faster and broader diffusion of voter suppression narratives, • agenda manipulation, as fringe or misleading narratives gain cross-platform legitimacy.
Scope			
	Geographic asymmetry: uneven amplification across regions or foreign infiltration	• Domestic vs. foreign engagement share, • exposure ratio by region, • geographic clustering score, • foreign-linked account footprint, • regional sensitivity index.	Geographic asymmetry evaluates whether amplification is uneven across regions or whether there are signs of foreign interference in domestic information spaces. This helps determine whether amplification may distort equitable access to political information. In the context of asymmetric amplification, this indicator contributes to the following potential measuring aspects: • Regionally targeted voter suppression campaigns, • localised polarisation driven by repeated exposure to conflict-oriented content, • greater vulnerability to foreign interference in specific regions, • socio-political tension where certain areas experience disproportionate manipulation.

Table 3: Overview of indicators and metrics, contributing for measuring scope of asymmetric amplification in the context of elections.

Continuation next page

Continuation **Table 3:**

Dimension	Indicator (signal)	Metrics (measures)	Associated risks to which indicators contribute
Scope	**Demographic skew:** certain groups exposed disproportionately	· Exposure ratio per demographic group, · reach concentration index, · engagement asymmetry, · group-level sentiment shift after amplification spikes.	Demographic skew highlights when certain population groups (e.g. youth, linguistic minorities or politically marginalised communities) suffer disproportionate exposure to amplified narratives. Audience metrics might reveal unequal information access or targeted influence operations. In the context of asymmetric amplification, this indicator contributes to the following potential measuring aspects: · Targeted voter suppression (e.g. discouraging messages disproportionately shown to youth or minorities), · group-specific polarisation as certain communities receive more intense exposure, · discriminatory harms if harmful narratives about protected groups gain disproportionate reach, · entrenched misinformation within particular demographic segments.
	Language imbalance: stronger amplification or weaker moderation in some languages/ dialects	· Volume of harmful/ violating content per language, · moderation lag times across languages, · labelling accuracy by language, · share of unlabelled/ unremoved content by language, · amplification differences by language, · safety resource ratio (staff and automated detection coverage).	Language imbalance captures how differences in content moderation capacity or algorithmic behaviour across languages and dialects can expose speakers of under-resourced languages to greater amounts of unlabelled or harmful content, resulting in structural asymmetric amplification. In the context of asymmetric amplification, this indicator contributes to the following potential measuring aspects: · Higher visibility of harmful or misleading narratives in under-moderated languages, · increased spread of political manipulation or disinformation targeting linguistic minorities, · unequal protection from hate speech across language communities, · amplified polarisation where harmful narratives persist in specific linguistic groups.

Table 3: Overview of indicators and metrics, contributing for measuring scope of asymmetric amplification in the context of elections.

Remediability

Remediability captures how feasible it is to correct or reverse the negative effects of a risk. This includes how quickly platforms can intervene, how effective these interventions are and how long they last. In the context of asymmetric amplification, the core question is whether platforms can detect, limit and undo amplification dynamics once they occur, particularly in the time-sensitive environment of elections. Low remediability indicates that harmful amplification persists despite interventions, increasing the severity of the risk and its systemic potential.

In electoral contexts, remediability-related indicators help determine whether platforms can meaningfully mitigate the risks associated with asymmetric amplification. This involves assessing the timeliness of detection, the effectiveness of moderation or design adjustments and the likelihood that harmful narratives resurface even after action is taken. Experimental studies that test the impact of algorithmic versus chronological feeds illustrate one way researchers can assess the remediability of amplification dynamics. For example, switching users from personalised to chronological feeds allows researchers to observe whether amplification effects can be reduced through design interventions and how such changes affect user attitudes and behaviour, as shown in a study on Facebook and Instagram during the 2020 US election.[64]

These types of measurements help clarify, for instance, whether platforms can meaningfully dampen the amplification of hostile out-group rhetoric or sensationalist political content described in section 3.2, or whether such narratives rapidly reappear despite corrective design changes. They also help assess whether misleading electoral information resurfaces after removal, which is critical for understanding the extent to which voter suppression or confusion risks can genuinely be mitigated in practice.

As with scale and scope, remediability indicators must be interpreted in context. Delays or limitations in mitigation may stem from operational constraints, adversarial adaptation, or technical limitations, rather than intentional design. The following table therefore presents a focused selection of remediability-related indicators that researchers can use to evaluate how reversible or persistent amplification effects are in electoral scenarios.

Dimension	Indicator (signal)	Metrics (measures)	Associated risks to which indicators contribute
Remediability	**Moderation success/ failure:** extent to which amplification of content that breaks platform's policies is corrected	· Percentage of flagged content labelled/ removed, · ratio of labelled content that continues spreading, · time harmful content remains visible.	Moderation success or failure measures whether harmful amplification is effectively reduced after detection. Persistent spread of labelled or corrected content can signal weak enforcement or limited algorithmic reach reduction. In the context of asymmetric amplification, this indicator contributes to the following potential measuring aspects: · Prolonged exposure to harmful narratives, increasing downstream impact, · persistent visibility advantages for actors benefiting from amplification, · accelerated spread during sensitive moments (e.g., elections), · structural bias where some content types remain amplified longer.
	Persistence of harmful content: whether removed/ demoted content remains impactful	· Reappearance of removed content (mirrors, screenshots) · visibility after "shadow removals", · residual engagement after downranking.	Persistence of harmful content reflects how long audiences remain exposed. Even after removal or demotion, narratives can persist via mirrors, screenshots, or reposts, highlighting the temporal limits of platform mitigation. In the context of asymmetric amplification, this indicator contributes to the following potential measuring aspects: · Sustained influence of harmful narratives despite interventions, · difficulty containing virality of manipulated political content, · further polarisation if content reappears in new forms, · undermining of trust in platform enforcement.
	Recurrence of narratives: harmful content/ narratives resurface repeatedly	· Frequency of reposts/ duplicates, · reactivation of banned accounts, · cyclical patterns in narrative reappearance.	Narratives recur when previously mitigated content reappears, often altered or reposted. It signals that moderation may be insufficient and revealing potential structural weaknesses in detection or enforcement. In the context of asymmetric amplification, this indicator contributes to the following potential measuring aspects: · Cyclical resurfacing of harmful political narratives · reactivation of coordinated networks or banned actors, · long-term entrenchment of misinformation, · challenges in preventing repeated election interference.

Table 4: Overview of indicators and metrics, contributing for measuring remediability of asymmetric amplification in the context of elections.

Continuation next page

Continuation **Table 4:**

Dimension	Indicator (signal)	Metrics (measures)	Associated risks to which indicators contribute
Remediability	**Policy efficacy:** whether interventions reduce amplification asymmetries	· Impact of fact-checking labels on shares, · algorithmic downranking effects, · reduction in engagement after interventions.	Policy efficacy evaluates whether platform interventions such as fact-checking labels effectively reduce amplification asymmetries in measurable ways. In the context of asymmetric amplification, this indicator contributes to the following potential measuring aspects: · Continued high reach of harmful/false content, · unequal enforcement across actors or narratives, · limited impact of corrective measures during crises or elections, · algorithmic systems reinforcing amplification instead of mitigating it.
	Effectiveness of user reporting and removal tools: whether users can meaningfully flag harm	· Average time-to-takedown after report, · user satisfaction with reporting outcomes, · share of reports acted on.	Effectiveness of user reporting and removal tools evaluates the agency of users in addressing amplified harms to assess whether reporting mechanisms function as meaningful remediation channels or remain symbolic. In the context of asymmetric amplification, this indicator contributes to the following potential measuring aspects: · Slower response times to harmful narratives, · inconsistent removal of damaging political or hateful content, · frustration or loss of trust among users, · disproportionate harm to vulnerable groups who rely more on reporting.
	Policy responsiveness: speed and consistency of enforcement	· Platform response time during elections, · appeal outcomes, · enforcement consistency across regions/languages, · responsiveness before vs. after elections.	Policy responsiveness measures the speed and consistency of enforcement, particularly during sensitive periods such as elections. Slow, inconsistent or election-specific enforcement can allow harmful narratives to spread unchecked. In the context of asymmetric amplification, this indicator contributes to the following potential measuring aspects: · Higher reach of harmful content during critical periods, · geographic, linguistic or political enforcement disparities, · strategic exploitation by malicious actors aware of platform delays, · reduced user trust in the fairness of enforcement.

Table 4: Overview of indicators and metrics, contributing for measuring remediability of asymmetric amplification in the context of elections.

Taken together, the indicators across scale, scope and remediability show that assessing the systemic potential of asymmetric amplification does not require inventing new methodologies. It involves assembling a diverse set of existing methods, datasets, and analytical tools, each illuminating a different facet of the risk.

No single indicator can determine systemicness; it emerges only when signals across these dimensions are interpreted in context, alongside evidence of adversarial behaviour, platform design effects and electoral timing. This multiplicity of methods is not a weakness but a necessary feature of systemic-risk assessment under the DSA.

4.3 Technical implementation of indicators

The introduced research agenda provides a conceptual framework to structure empirical research findings. Its implementation, however, is currently limited by the actual research environment under the DSA. In this section, we argue that the research agenda cannot rely on a single methodological pathway. Instead, evidence has to rely on a variety of methods to deal with the constraints in the research environment and the complexity of systemicness.

4.4 Observational vs. experimental approaches

Because the indicators discussed further above capture different kinds of evidence, analysing them requires multiple methodological approaches. Indicators linked to scale often rely on descriptive information that reflects how amplification unfolds on platforms. This type of information can be obtained through observational approaches: these use systematic, non-interventionist methods such as descriptive statistical analysis, temporal activity mapping or network analysis to document how content appears, circulates and gains visibility. Such approaches reveal visibility imbalances or circulation patterns and are well suited for indicators such as content volume imbalance, reach asymmetry, amplification velocity or growth velocity.

Observational approaches commonly include descriptive comparisons such as average views or retweets across different political actors, as well as network analyses that identify which communities were exposed to which posts during a given political event (e.g. a protest).[65] Such approaches do not introduce any intervention by the researcher (apart from the decisions, researchers need to take during the data collection process) and therefore reflect amplification as captured through available data. Thus, these techniques build a granular picture of the distribution, spread or intensity of amplification but do not explain why these patterns arise.

Observational approaches

Descriptive statistics offer a systematic way to summarise and interpret patterns within social media data, helping to capture how different types of political content are distributed across user experiences. They allow for an overview of the general presence of political material within users' feeds and enable comparisons between user groups in their exposure to different parties or viewpoints. By focusing on averages, proportions and distributions, descriptive analyses can reveal how frequently users encounter political videos, how exposure differs between supporters and non-supporters of specific parties and how overall visibility of party-related content varies across the platform.

Other indicators address questions that observational approaches cannot answer. For example, those indicators that aim to capture who is affected by amplification, or whether certain patterns could be reversed, require evidence on mechanisms that observational data do not reveal. These questions require controlled variation, which is central to experimental approaches. Experimental methods deliberately manipulate and simulate user characteristics or platform conditions to observe how changes in inputs produce changes in exposure outcomes.

Experimental approaches

Experiments are a method that test causal relationships by systematically manipulating one or more variables and observing their effects on another variable. In a typical experiment, participants are randomly assigned to two different groups: a control group that experiences standard or baseline conditions, and one or more treatment groups that are exposed to a specific intervention or altered condition. By comparing outcomes between these groups, researchers can determine whether the treatment caused any observed differences. Experiments are often combined with surveys that allow researchers to get information on the user experience before and after the treatment.

Different types of experiments include A/B-testing, Simulation-based experiments[66], offline experiments and crowdsourced experiments.[67]

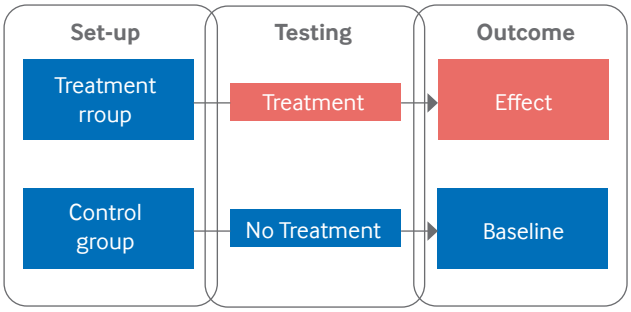


Figure 4: Overview experimental set-up.

Another research approach refers to the use of sock-puppet accounts, that is artificial user profiles with predefined attributes or behavioural scripts. They enable researchers to test how recommender systems respond to systematically varied user signals by e.g. comparing how users with varying political preferences are recommended or shown content.[68]

Sock-puppet accounts

Sock-puppet accounts (or simply sock-puppets) are artificial online accounts that imitate real user behaviour and are controlled by individuals or organisations. While they can be used deceptively, they can also serve legitimate research purposes. In algorithmic auditing, researchers use sock-puppet accounts to test what content platforms recommend to different users. Accounts with varying characteristics (e.g., gender, location, political orientation) can be used to compare personalised feeds and analyse how these traits shape recommendation outcomes.

Studies typically create control accounts (neutral profiles) and treatment accounts (profiles with specific attributes, which are trained to behave according to predefined interests). The training can be conducted either manually (e.g. users scrolling according to rules) or automatically (e.g. bots watching selected political content). Control accounts receive no such training. In the last stage, researchers either manually or automatically scroll through a feed with the conditioned accounts. The content shown to the sock-puppet accounts constitutes the outcome data which can then be further analysed.

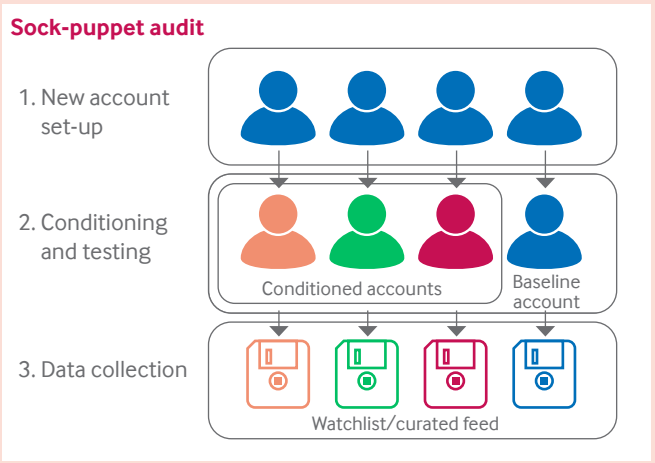


Figure 5: Overview set-up for sock-puppets.

While the data collection with sock-puppet accounts allows researchers to simulate user experiences, it is important to note the limitations that come with this approach. First, human user behaviour might deviate from the simulated user behaviour, which heavily depends on the training decisions the researchers take. And second, researchers cannot test for the effects of exposure to certain content types (e.g. political content) on political attitudes. For this reason, experiments with actual users are needed.

One example is the previously referenced study on Facebook and Instagram during the 2020 US presidential election campaign. Although participants who were assigned chronological feeds spent less time on the platforms, the experiment did not detect significant changes in political attitudes or knowledge.[69] The experiment therefore allowed researchers to understand behavioural responsiveness and behavioural stability in ways that neither simulated accounts nor purely observational designs can capture.

Experimental approaches in general are essential for indicators connected to remediability. They allow researchers to test whether changes in platform design, moderation settings or ranking systems materially alter user experiences or downstream outcomes. They also support scope-related indicators by showing how different groups of real users respond to the same interventions, revealing whether platform adjustments have uneven effects across communities.

4.5 Why data collection requires multiple tools: the reality of access constraints

The need for multiple methodological approaches often begins before the analysis stage as the data environment itself is fragmented. Although the DSA provides a formal framework for researcher access under Art. 40(12) and 40(4) DSA, in practice, access to data remains constrained. Procedures for non-public data have only recently begun, and public APIs often do not expose the datapoints that the indicators require. Even well-functioning APIs provide only partial visibility into the data necessary to support indicators across scale, scope and remediability.

Research APIs

Research APIs are a type of application programming interface (API) that provides structured access to data for the purpose of academic or civil society research. APIs are often accessible through HTTP requests, the basic protocol of the internet. They rely on a small set of standards operations.

API providers, in this case platforms that make their data accessible through an API, define a resource endpoint (e.g. http://research.api/posts?hashtag=btw2025) where information about specific data objects, such as posts, can be retrieved. In this example, the request would return a list of metadata (meaning descriptive information about each post such as timestamps and number of likes) that contains all posts containing this hashtag in a structured format.

Queries can also include multiple constraints, meaning researchers can filter the data they request by adding specific conditions. These conditions might limit results to posts published within a certain time period. Combining such constraints allows researchers to narrow down large datasets to only the most relevant material. To manage large data volumes, APIs typically return results in batches (or "pages"), rather than delivering the full dataset in a single response.

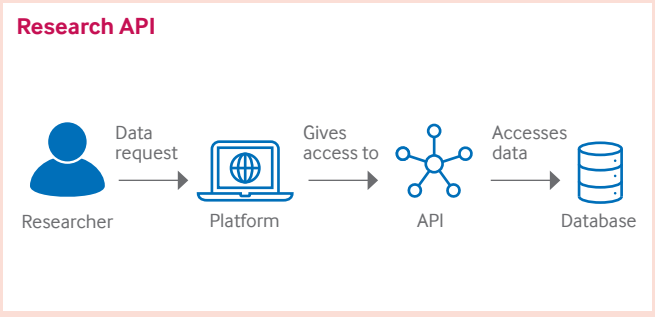


Figure 6: Overview of the process for obtaining data access via the research API.

For scale, for example, APIs often deliver only basic engagement metrics; they do not provide impressions, unique viewers or complete content volumes. For scope, essential datapoints such as demographic breakdowns, geographic reach or cross-platform diffusion are not systematically available. For remediability, information on enforcement accuracy, response times, internal escalation or the presence of downranking and shadow removals is not accessible externally. Hence, these gaps affect every indicator class and create a structural mismatch between what the APIs provide and what systemic risk assessment according to the above illustrated agenda requires. These limitations mean that indicators cannot be supported by API data alone.

Dimension	Indicators	Typically available via APIs	Commonly missing/restricted
Scale	Content volume imbalance	Number of posts, comments, hashtags retrieved, API query counts	Hard caps on full-volume access, deleted/ edited posts not systematically exposed
	Reach asymmetry	Engagement counts (likes, retweets/ shares, comments)	Impressions, watch-time, click-through data, unique viewers
	Amplification velocity	Timestamps of posts/interactions → possible to infer posting speed	Latency in API feeds, real-time virality signals
	Volume of boosted content	Shares/retweets counts (sometimes top-shared content endpoints)	Hidden boosts (algorithmic prioritisation, bot amplification not visible)
	Growth velocity of actors	Follower counts (sometimes available)	Detailed follower growth over time, shadow-followers, un-follow dynamics
Scope	Cross-platform spread	Outbound links, hashtags, mentions of other platforms	Cross-platform reposting metrics, systematic visibility into coordinated campaigns
	Geographic asymmetry	Sometimes geotags on posts (if user opts in)	Systematic geo-data, inferred location of users, geographic reach of impressions
	Demographic skew	Limited: inferred from self-disclosed profile info (age, gender rarely exposed)	Verified demographic breakdowns, ad-target-ing demographics, audience composition
	Language imbalance	Text language field (often auto-detected by platform)	Multilingual amplification visibility (esp. low-resource languages, dialects)
Remediability	Moderation success/ failure	Labels on content (fact-check, misinformation, sensitive content)	Enforcement accuracy (false positives/ negatives), moderation decisions' rationale
	Persistence of harmful content	Content status (available/deleted) sometimes retrievable	Shadow removals, de-boosting, downranking not exposed
	Recurrence of narratives	Ability to track re-posted or duplicate content via IDs	Cross-post recurrences, recycling of memes/ images across accounts
	Policy efficacy	Some transparency reports (aggregate)	Granular per-content/per-user policy enforcement data
	Effectiveness of user reporting/removal tools	Rarely exposed (only if users share reports)	Internal responsiveness, time-to-takedown, complaint outcomes
	Policy responsiveness	Occasional transparency reporting	Internal escalation data, appeals handling, systemic responsiveness

Table 5: Overview of datapoints commonly available through APIs and those frequently absent, categorised by impact dimension and indicator.

These gaps are not merely technical inconveniences. They directly limit the ability of researchers and regulators to assess asymmetric amplification along the three required systemic-risk dimensions. Missing data on impressions and reach restrict assessments of scale; absent demographic, linguistic or cross-platform signals inhibit evaluation of scope; and missing information on moderation pipelines or algorithmic adjustments prevents meaningful assessment of remediability.

Scraping, a method used to automatically collect large datasets from social media platforms, can partially compensate for these gaps: the technique enables collecting information that is publicly visible on platforms but not exposed through APIs. For example, it can help infer signals such as geographic clustering or approximate audience characteristics. However, scraping also faces significant limitations. It depends on the stability of platform interfaces, which can change without notice. It does not provide access to internal processes or hidden ranking mechanisms. It cannot capture metrics that platforms do not display at all, including time-to-takedown or algorithmic downranking. Furthermore, scraping often operates under technical and legal constraints; its coverage may remain incomplete when platforms restrict automated access.

Data donation is a data collection method based on voluntary contributions of personal platform data by users for research purposes. It can play a complementary role by providing information that only users can supply. Data donations offer access to feeds, recommendations and reporting interactions that cannot be reconstructed from external observation. This makes data donations particularly relevant for indicators related to remediability and aspects of scope that depend on user-level exposure. These methods, however, also have inherent limitations. They rely on users being able to export their data easily or on tools such as browser plugins that often function only on desktop devices. Participation is voluntary, which can introduce selection effects. Donations also cannot reveal information that platforms do not include in user-accessible data exports, such as internal moderation decisions or hidden ranking adjustments.

Together, these methods respond to the structural limits of official access channels. Scraping extends what can be observed externally while data donations reveal what users encounter directly. However, platform compliance remains necessary for indicators that depend on internal processes. No single approach can cover all indicators; only a combination of methods can approximate the evidentiary requirements the above-mentioned agenda demands.

5. Conclusion

The regulatory framework of the DSA requires VLOPs and VLOSEs to identify systemic risks and implement appropriate mitigation measures. As this report highlights, the term "systemic risks" is intentionally broad, but it leaves important questions unanswered, particularly regarding when a risk qualifies as *systemic*. While this openness allows various stakeholders to contribute to the discussion, it also presents challenges for researchers and regulators in applying the DSA effectively.

This report derives indicators for the severity of asymmetric amplification of political content, based on three assessment dimensions: scale, scope and remediability. It argues that these three dimensions could serve as the guiding elements for a research agenda that supports researchers in developing relevant research questions and help them understand which methods they can use. It also helps regulators in organising the large amount of empirical evidence. Furthermore, given the complexity of systemicness, a mosaic of evidence drawing on diverse methods is required. It is important to note that the report does not intend to present any specific benchmarks or thresholds that would define when a risk becomes systemic. It also does not make any claims on whether risks associated with asymmetric amplification would qualify as systemic. Decisions on concrete benchmarks will ultimately be up to regulators and, where contested, courts.

Although the research agenda focuses on asymmetric amplification of political content as a phenomenon associated with a variety of risks to civic discourse, it is not limited to this category of risks. The suggestion to operationalise severity and likelihood of risks through indicators to structure evidence on when a risk becomes systemic is readily applicable to risks from other risk categories. The tables in section 4.2 may serve as an exemplary template for the dimensions linked to the severity of risks. Also, we argue that a multiplicity of methods is needed to cover indicators from the three dimensions applies to all types of risks. To prove the systemicness of any kind of risk, no single study or metric can sufficiently provide the diverse forms of evidence needed.

Limitations

Even though methodological diversity and comprehensive evidence are critical for advancing risk analysis, practical implementation remains difficult due to the restrictive conditions shaping today's research landscape. Researchers are often constrained by short project durations or a lack of funding for more extensive methods such as experiments. Additionally, as discussed in section 4.5, missing or limited data access due to a contested understanding and differing interpretations of what qualifies as publicly available data significantly hinders researchers in fulfilling their role as foreseen in the DSA. In the case of ISD research conducted in the context of this project, data access provided through the Virtual Compute Environment was so limited that it ultimately could not be used for the intended research purposes.[70]

Finally, at a more conceptual level, this report assumes that complex risk categories, such as civic discourse, can be operationalised, and will be further defined in an exchange with different DSA stakeholders including researchers and CSOs. This presupposes a shared understanding of what kind of civic discourse societies would like to ultimately reach. However, such questions are inherently normative and cannot be answered only by evaluating empirical evidence.[71]

Recommendations

In order to address such limitations for future work and to translate the suggested research agenda into practice this report recommends taking the following measures.

1. **EU institutions and member state governments are recommended to ensure long-term funding for researchers**

High-quality research takes time and resources. Experiments, for example, produce valuable insights into user behaviour on VLOPs and VLOSEs but are resource-intensive due to the complex set-up required. Also, going beyond anecdotal evidence would require measuring indicators such as those set out in this report over a longer timeframe to assess and evaluate data meaningfully. It is therefore recommended to expand funding for long-term research projects at both the European Union level through initiatives from EU institutions, and the national level via member state governments, to provide the stability needed to produce valuable research outputs.

2. EU regulators should adopt a harmonised, broad definition of public data

High-quality research can only produce reliable results with reliable data. As seen in section 4.5, VLOPs and VLOSEs currently interpret Art. 40 (12) of the DSA in different ways, resulting in the provision of inconsistent data points. Regulators should adopt a harmonised, broad definition of public data, covering all content and metadata that an ordinary user can view, including when logged in, and mandate stable machine-readable access (e.g. APIs) to these data.

3. EU regulators should leverage investigatory powers to foster meaningful transparency

VLOPs and VLOSEs themselves hold the most detailed evidence on asymmetric amplification and other hard-to-research potential risks, as documents published from whistleblowers like Frances Haugen illustrate. While researchers have their role to play, this report recommends that regulators leverage their investigatory powers as much as possible to access such information and share findings externally in adequate ways, if possible. Moreover, we recommend platforms and regulators to disclose more information on both the investigations as well as the risk assessments (e.g. metrics and benchmarks used in the risk assessments, which are currently confidential), to allow researchers to fulfill their functions in the context of the DSA.

4. Structured collaboration between the DSCs Board, the European Commission and researchers

Stakeholder engagement, as foreseen in the DSA, requires ongoing and structured exchange. The first report of the European Board for Digital Services already recognises that a stronger engagement with external stakeholders, including researchers is planned for 2026. To ensure that the work of the different actors is well aligned, this report recommends strengthening collaboration between the DSCs Board, the European Commission and the research community. One practical step could be a joint workshop where stakeholders review the current state of literature, identify critical knowledge gaps and prioritise areas for further investigation. The research agenda proposed in this report could serve as a starting point for such a discussion.

References

1 Naja Bentzen, *Computational Propaganda Techniques* (European Parliamentary Research Service, 2018), https://www.europarl.europa.eu/RegData/etudes/ATAG/2018/628284/EPRS_ATA(2018)628284_EN.pdf.

2 Institute for Strategic Dialogue, *Identification, Assessment and Mitigation of Systemic Risks in the Context of the Digital Services Act* (2025), https://www.dsc.bund.de/DSC/DE/Aktuelles/studien/KeyFindings_ISD.pdf?__blob=publicationFile&v=2.

3 Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions on the European Democracy Action Plan (2020), https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020DC0790.

4 European Union External Action, *1st EEAS Report on Foreign Information Manipulation and Interference Threats* (Strategic Communications, Task Forces and Information Analysis, 2023), https://www.eeas.europa.eu/sites/default/files/documents/2023/EEAS-DataTeam-ThreatReport-2023.pdf.

5 Giacomo Figà Talamanca and Selene Arfini, 'Through the Newsfeed Glass: Rethinking Filter Bubbles and Echo Chambers', *Philosophy & Technology* 35, no. 1 (March 2022): 20, https://doi.org/10.1007/s13347-021-00494-z.

6 TikTok, 'For You', accessed 24 November 2025, https://support.tiktok.com/en/getting-started/for-you#1.

7 Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and Amending Directive 2000/31/EC (Digital Services Act) (2022), Art.3(s), https://eur-lex.europa.eu/eli/reg/2022/2065/oj/eng.

8 European Commission, 'Supervision of the Designated Very Large Online Platforms and Search Engines under DSA', 2025, https://digital-strategy.ec.europa.eu/en/policies/list-designated-vlops-and-vloses.

9 Michèle Finck, 'Digital Regulation: Designing a Supranational Legal Framework for the Platform Economy', *SSRN Electronic Journal*, ahead of print, 2017, https://doi.org/10.2139/ssrn.2990043.

10 Oliver Marsh, *Researching Systemic Risks under the Digital Services Act* (AlgorithmWatch, 2024), https://algorithmwatch.org/en/researching-systemic-risks-under-the-digital-services-act/.

11 Ortwin Renn et al., 'Systemic Risks from Different Perspectives', Risk Analysis 42, no. 9 (2022): 1902–20, https://doi.org/10.1111/risa.13657.

12 Robert Baldwin and Julia Black, 'Driving Priorities in Risk-Based Regulation: What's the Problem?', *Journal of Law and Society* 43, no. 4 (2016): 566, https://doi.org/10.1111/jols.12003.

13 Renn et al., 'Systemic Risks from Different Perspectives', 1903.

14 Renn et al., 'Systemic Risks from Different Perspectives'.

15 Sally Broughton Micova and Andrea Calef, *Elements for Effective Systemic Risk Assessment under the DSA* (Cerre, 2023), https://cerre.eu/publications/elements-for-effective-systemic-risk-assessment-under-the-dsa/; GNI and DTSP, *European Rights & Risks: Stakeholder Engagement Forum Summary* (2024), https://globalnetworkinitiative.org/eu-stakeholder-engagement-forum-summary/.

16 Broughton Micova and Calef, *Elements for Effective Systemic Risk Assessment under the DSA*.

17 Mateus Correia de Carvalho, 'It Will Be What We Want It to Be: Sociotechnical and Contested Systemic Risk at the Core of the EU's Regulation of Platforms' AI Systems', *JIPITEC – Journal of Intellectual Property, Information Technology and E-Commerce Law* 16, no. 1 (March 2025): 70–71, https://www.jipitec.eu/jipitec/article/view/420.

18 DSA Civil Society Coordination Group, *Initial Analysis on the First Round of Risk Assessment Reports* (2025), https://cdt.org/wp-content/uploads/2025/03/RA-Report-Assessment-Report.pdf.

19 Marsh, *Researching Systemic Risks under the Digital Services Act*, 5.

20 Agustina Del Campo, Nicolas Zara, and Ramiro Álvarez-Ugarte, 'Are Risks the New Rights? The Perils of Risk-Based Approaches to Speech Regulation', SSRN Scholarly Paper no. 5161173 (Rochester, NY: Social Science Research Network, 28 February 2025), 247, https://doi.org/10.2139/ssrn.5161173.

21 Marsh, *Researching Systemic Risks under the Digital Services Act*; Rachel Griffin, 'The Politics of Risk in the Digital Services Act: A Stakeholder Mapping and Research Agenda', *Weizenbaum Journal of the Digital Society* 5, no. 2 (June 2025), https://doi.org/10.34669/wi.wjds/5.2.6.

22 Anna Liesenfeld, 'The Legal Significance of Independent Research Based on Article 40 DSA for the Management of Systemic Risks in the Digital Services Act', *European Journal of Risk Regulation* 16, no. 1 (March 2025): 184–96, https://doi.org/10.1017/err.2024.61.

23 Ibid.

24 Ibid.

25 Griffin, 'The Politics of Risk in the Digital Services Act'.

26 Ibid.

27 Shaina Raza et al., 'A Comprehensive Review of Recommender Systems: Transitioning from Theory to Practice', arXiv:2407.13699, preprint, arXiv, 16 October 2025, https://doi.org/10.48550/arXiv.2407.13699.

28 Dietmar Jannach, Lukas Lerche, and Markus Zanker, 'Recommending Based on Implicit Feedback', in *Social Information Access*, ed. Peter Brusilovsky and Daqing He, Lecture Notes in Computer Science (Cham: Springer International Publishing, 2018), 10100:510–69, https://doi.org/10.1007/978-3-319-90092-6_14.

29 Wenqi Li et al., 'Beyond Explicit and Implicit: How Users Provide Feedback to Shape Personalized Recommendation Content', *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, ACM, 26 April 2025, 1–17, https://doi.org/10.1145/3706598.3713241.

30 Smitha Milli et al., 'Engagement, User Satisfaction, and the Amplification of Divisive Content on Social Media', arXiv:2305.16941, preprint, arXiv, 7 December 2024, https://doi.org/10.48550/arXiv.2305.16941.

31 Aviv Ovadya, *Bridging-Based Ranking How Platform Recommendation Systems Might Reduce Division and Strengthen Democracy* (Harvard Kennedy School. Belfer Center for Science and International Affairs, 2022), https://www.belfercenter.org/sites/default/files/pantheon_files/files/publication/TAPP-Aviv_BridgingBasedRanking_FINAL_220518_0.pdf.

32 Ferenc Huszár et al., 'Algorithmic Amplification of Politics on Twitter', *Proceedings of the National Academy of Sciences* 119, no. 1 (December 2021): e2025334119, https://doi.org/10.1073/pnas.2025334119.

33 European Commission, 'Commission Opens Formal Proceedings under DSA', Text, 30 April 2024, https://ec.europa.eu/commission/presscorner/detail/en/ip_24_2373.

34 Kristian Lum and Tomo Lazovich, 'The Myth of The Algorithm: A System-Level View of Algorithmic Amplification', *Knight First Amendment Institute*, 13 September 2023, http://knightcolumbia.org/content/the-myth-of-the-algorithm-a-system-level-view-of-algorithmic-amplification.

35 Jonathan Stray, Luke Thorburn, and Priyanjana Bengani, 'Making Amplification Measurable', *Tech Policy Press*, 28 April 2023, https://techpolicy.press/making-amplification-measurable.

36 Sofia Calabrese and Orsolya Reich, *Identifying, Analysing, Assessing and Mitigating Potential Negative Effects on Civic Discourse and Electoral Processes* (Civil Liberties Union for Europe, European Partnership for Democracy, 2024), https://www.liberties.eu/f/mpdgy5.

37 Nic Newman et al., *Reuters Institute Digital News Report 2025* (Reuters Institute for the Study of Journalism, 2025), https://doi.org/10.60625/RISJ-8QQF-JT36.

38 Anna Sophie Kümpel, *Individuelle Meinungsbildung und -äußerung auf sozialen Medien* (2024), https://www.bpb.de/themen/medien-journalismus/soziale-medien/545487/individuelle-meinungsbildung-und-aeusserung-auf-sozialen-medien/.

39 The Electoral Commission, 'Response to the Science, Innovation and Technology Committee Call for Evidence on Social Media, Misinformation and Harmful Algorithms', 28 January 2025, https://www.electoralcommission.org.uk/news-and-views/our-responses-consultations/response-science-innovation-and-technology-committee-call-evidence-social-media-misinformation-and.

40 Antoine Banks et al., '#PolarizedFeeds: Three Experiments on Polarization, Framing, and Social Media', *The International Journal of Press/Politics* 26, no. 3 (July 2021): 609–34, https://doi.org/10.1177/1940161220940964.

41 Huszár et al., 'Algorithmic Amplification of Politics on Twitter', 1.

42 Hazem Ibrahim et al., 'TikTok's Recommendations Skewed towards Republican Content during the 2024 U.S. Presidential Race', arXiv:2501.17831, preprint, arXiv, 7 May 2025, https://doi.org/10.48550/arXiv.2501.17831.

43 Hazem Ibrahim et al., 'YouTube's Recommendation Algorithm Is Left-Leaning in the United States', *PNAS Nexus* 2, no. 8 (August 2023): pgad264, https://doi.org/10.1093/pnasnexus/pgad264.

44 Joe Whittaker et al., 'Recommender Systems and the Amplification of Extremist Content', *Internet Policy Review* 10, no. 2 (June 2021), https://doi.org/10.14763/2021.2.1565.

45 Muhammad Haroon et al., 'Auditing YouTube's Recommendation System for Ideologically Congenial, Extreme, and Problematic Recommendations', *Proceedings of the National Academy of Sciences of the United States of America*, no. 120 (50) (2023), https://doi.org/10.1073/pnas.2213020120.

46 Alexandra Cirone and William Hobbs, 'Asymmetric Flooding as a Tool for Foreign Influence on Social Media', *Political Science Research and Methods* 11, no. 1 (January 2023): 160–71, https://doi.org/10.1017/psrm.2022.9.

47 Stacie Hoffmann, Emily Taylor, and Samantha Bradshaw, *The Market of Disinformation* (2019), https://oxtec.oii.ox.ac.uk/wp-content/uploads/sites/10/2019/10/OxTEC-The-Market-of-Disinformation.pdf.

48 Tiziano Piccardi et al., 'Reranking Partisan Animosity in Algorithmic Social Media Feeds Alters Affective Polarization', *Science* 390, no. 6776 (November 2025): eadu5584, https://doi.org/10.1126/science.adu5584.

49 Qazi Mohammad Areeb et al., 'Filter Bubbles in Recommender Systems: Fact or Fallacy—A Systematic Review', *WIREs Data Mining and Knowledge Discovery* 13, no. 6 (2023): e1512, https://doi.org/10.1002/widm.1512.

50 Eslam Hussein, Prerna Juneja, and Tanushree Mitra, 'Measuring Misinformation in Video Search Platforms: An Audit Study on YouTube', *Proc. ACM Hum.-Comput. Interact.* 4, no. CSCW1 (May 2020): 48:1-48:27, https://doi.org/10.1145/3392854.

51 Christopher Bail, *Breaking the Social Media Prism: How to Make Our Platforms Less Polarizing* (Princeton (N.J.): Princeton University press, 2021).

52 Banks et al., '#PolarizedFeeds'.

53 Ovadya, *Bridging-Based Ranking How Platform Recommendation Systems Might Reduce Division and Strengthen Democracy*, 8.

54 Milli et al., 'Engagement, User Satisfaction, and the Amplification of Divisive Content on Social Media'.

55 Ofcom, *Protecting People from Illegal Harms* Online (2023), https://www.ofcom.org.uk/siteassets/resources/documents/consultations/category-1-10-weeks/270826-consultation-protecting-people-from-illegal-content-online/associated-documents/annex-5-draft-service-risk-assessment-guidance?v=330403.

56 World Economic Forum, *Toolkit for Digital Safety Design Interventions and Innovations: Typology of Online Harms* (2023), https://www3.weforum.org/docs/WEF_Typology_of_Online_Harms_2023.pdf.

57 International Organization for Standardization, *Risk Management - Guidelines* (2018), https://www.iso.org/standard/65694.html.

58 Institute for Strategic Dialogue, *Identifikation, Bewertung Und Bekämpfung von Systemischen Risiken Im Kontext Des Digital Services Act* (Berlin, 2025), https://www.dsc.bund.de/DSC/DE/Aktuelles/studien/IdentifikationBewertungBekaempfungVonSystemischenRisiken.pdf?__blob=publicationFile&v=1.

59 Institute for Strategic Dialogue, *Identification, Assessment and Mitigation of Systemic Risks in the Context of the Digital Services Act*.

60 Ibid.

61 Ibrahim et al., 'TikTok's Recommendations Skewed towards Republican Content during the 2024 U.S. Presidential Race'; Paul Bouchaud, 'Algorithmic Amplification of Politics and Engagement Maximization on Social Media', 28 August 2023, https://hal.science/hal-04189507.

62 Fabian Baumann et al., 'Dynamics of Algorithmic Content Amplification on TikTok', arXiv:2503.20231, version 1, preprint, arXiv, 26 March 2025, https://doi.org/10.48550/arXiv.2503.20231; Timothy Graham and Mark Andrejevic, *A Computational Analysis of Potential Algorithmic Bias on Platform X during the 2024 US Election / Algorithmic Bias X* (2024), https://github.com/timothyjgraham/AlgorithmicBiasX/blob/3f4ca90ac156d88bdaf398caa518ac86f69c5885/A%20computational%20analysis%20of%20potential%20algorithmic%20bias%20on%20platform%20X%20during%20the%202024%20US%20election-4.pdf.

63 Johannes Wolfgram et al., 'TikTok Regional Election Monitoring 2024: Methods Report', kp25s_v1, preprint, OSF Preprints, 16 October 2024, https://doi.org/10.31219/osf.io/kp25s_v1; Simon Koenigsdorff and Jan Georg Plavec, 'Politik und Social Media: Filterblasen auf Tiktok – das sagen die Zahlen', *stuttgarter-zeitung*.

de, 11 July 2025, https://www.stuttgarter-zeitung.de/inhalt.politik-und-social-media-filterblasen-auf-tik-tok-das-sagen-die-zahlen.45465809-617f-4d8a-bf44-d2e40323f5a0.html; Jinyi Ye, Luca Luceri, and Emilio Ferrara, 'Auditing Political Exposure Bias: Algorithmic Amplification on Twitter/X During the 2024 U.S. Presidential Election', *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency* (New York, NY, USA), FAccT '25, Association for Computing Machinery, 23 June 2025, 2349–62, https://doi.org/10.1145/3715275.3732159.23 June 2025, 2349\\uc0\\u8211{}62, https://doi.org/10.1145/3715275.3732159.

64 Andrew M. Guess et al., 'How Do Social Media Feed Algorithms Affect Attitudes and Behavior in an Election Campaign?', *Science* 381, no. 6656 (July 2023): 398–404, https://doi.org/10.1126/science.abp9364; Stephanie Wang et al., 'Lower Quantity, Higher Quality: Auditing News Content and User Perceptions on Twitter/X Algorithmic versus Chronological Timelines', arXiv:2406.17097, version 1, preprint, arXiv, 24 June 2024, https://doi.org/10.48550/arXiv.2406.17097.

65 Mark Scott and Oliver Marsh, 'The Musk Effect: Assessing X's Impact on Germany's Election Discourse', *DFRLab*, 20 February 2025, https://dfrlab.org/2025/02/20/the-musk-effect-xs-impact-on-germanys-election/.

66 Petter Törnberg et al., 'Simulating Social Media Using Large Language Models to Evaluate Alternative News Feed Algorithms', arXiv:2310.05984, preprint, arXiv, 5 October 2023, https://doi.org/10.48550/arXiv.2310.05984.

67 Wang et al., 'Lower Quantity, Higher Quality'.

68 Ye, Luceri, and Ferrara, 'Auditing Political Exposure Bias'.

69 Guess et al., 'How Do Social Media Feed Algorithms Affect Attitudes and Behavior in an Election Campaign?'

70 Martin Degeling and Daniela Alvarado Rincón, 'Unpacking TikTok's Data Access Illusion', *Tech Policy Press*, 12 June 2025, https://techpolicy.press/unpacking-tiktoks-data-access-illusion.

71 Rachel Griffin and Riccardo Fornasari, 'Risky Business? Corporate Risk Management Obligations in Sustainability Due Diligence and Digital Platform Regulation', preprint, SSRN, 2025, https://doi.org/10.2139/ssrn.5140807.

ISD | Institute for Strategic Dialogue

Amman | Berlin | London | Paris | Washington DC

www.isdglobal.org