# AlgorithmWatch Response: Guidelines for Providers of Very Large Online Platforms and Very Large Online Search Engines on the Mitigation of Systemic Risks for Electoral Processes

## Key Points & Proposals

We see these guidelines as playing a potentially vital role in ensuring the DSA, and other legislation, plays its long-proposed role in protecting many upcoming European elections, including the imminent European Parliament Elections; this is particularly important given many other aspects of the DSA, including Articles 34, 35, and 40 are not currently as operational as we would have hoped in the absence of a delegated act on data access or guidelines on systemic risks.

However the task proposed by these guidelines is a complicated one, and will not be easily addressed with one round of comments on this draft. Moreover the task of directing VLOP/SEs towards generalised “best practices” for this complex, context-sensitive, and evolving field is not well suited to a static document which is reviewed once a year.

Our overarching proposals therefore are:

- These guidelines should be limited to providing clarity for platforms and external actors on how to meet requirements under the DSA and other related legislation, particularly those which are currently awaiting clarity via further guidelines and delegated acts.
- The guidelines themselves can also take a more modest approach of pointing to broadly accepted basic practices, without which VLOP/SEs run clear risks of being negligent in the face of election risks, for example: not having access to linguistic or country-specific knowledge; having failed to set up reporting, escalation, and decisionmaking processes which can effectively address emerging risks at sufficient pace; or failing to protect against highly predictable risks in a given election.
- Beyond this, the guidelines should not try to direct VLOP/SEs towards specific or novel approaches in a generalised fashion - e.g. saying platforms “should use watermarking” or “should apply inoculation measures”. The question of “best practices" against election risks in digital environments is too complex, context-sensitive, and fast-moving for such an approach.

- The guidelines should instead require VLOP/SEs, as part of preparations for a particular election, to *find, use, and share best research and practices at the given time and for the given context*; the guidelines should also propose methods by which this can be accomplished. The DSCs, and Digital Services Board, can play a valuable convening role here even in the absence of full legal designation.
- The guidelines must ensure that processes of collaboration and information-sharing, while valuable and welcome, are coordinated to ensure external experts and civil society groups are not unnecessarily overstretched by duplicate requests from multiple VLOP/SEs.
- For generative AI, based on our past research, we propose:
    - For information about elections, we propose that generative AI chatbots always return answers completely equivalent to traditional search engines: i.e. lists of links to sources, ranked for quality, relevance, and reliability, rather than attempt to summarise and reproduce information in a probabilistic fashion.
    - There should be straightforward and stable ways to automate generation of repeated answers to prompts to allow for external assessment; and access to volume-over-time and location-specific data on how many people are querying chatbots about particular topics or keywords, similar to the public Google Trends tool (though with absolute not relative numbers).

We appreciate the time pressures of the task, given the upcoming EU Parliament elections, but we believe that the above changes are feasible; arguably more feasible than attempting to decide upon static "best practices" from the wealth of available research.

To further elaborate on our reasoning: a fundamental issue of the framing concerns what is meant by "best practices". One interpretation is that these guidelines should, in a convenient form, lay out the "baseline expectations" to help platforms meet their legal requirements under the DSA (and other related legislation) related to elections. This would be important to fill current gaps in clarity around systemic risks in the absence of comprehensive guidelines around Articles 34 and 35. As found in numerous papers and expert workshops (for example by the Global Network Initiative and in our own work) there is considerable confusion around the term "systemic risks", exacerbated by the lack of transparency around Article 34 risk assessments. The guidelines should also further encourage platforms to follow important provisions, and clarify procedures, around data access and collaboration with researchers while we await the delegated act on Data Access. As such, a document laying out specific baselines which VLOPs and VLOSEs should follow, and allowing external parties to comment on these, seems helpful in the circumstances.

However a second interpretation of "best practices" is that this document proposes state-of-the-art approaches and encourages platforms to follow these. This is highly complicated,

and indeed risky, in the context of online environments and elections. It is complicated for the following reasons: There is an enormous body of research on the topic, and no clear consensus on broad questions like most severe risks or most effective mitigations. This is to be expected given the breadth and complexity of the topic, but is not conducive to laying out overarching recommendations of specific practices that are believed to be "best" or "state-of-the-art". Research can obviously be useful for guiding decision-making in specific contexts - for instance trying to protect elections which are particularly at risk from interference from specific hostile states can obviously learn from research on methods these states have used in other countries. As another example, responding to informational risks in countries with high levels of trust in media and institutions will be different to those with low levels, and preparations may benefit more from examples from similar rather than dissimilar countries. However attempting to recommend specific best practices *in general* is complicated, due to the factors already outlined. It also comes with risks. For example, inoculation or "pre-bunking" may be effective in many circumstances, but could risk raising awareness of hostile narratives.

Our answers to further questions below are all framed in terms of (i) our distinction between interpreting "best practices" as "clarifying expectations under the DSA" vs. as "adopting specific state-of-the-art practices" and (ii) our proposal that this document should stick to the first interpretation, while supporting methods for sharing and adopting state-of-the-art research and practices as appropriate for a given context, and in a dynamic fashion which can adapt in a fast-changing field.

Below follows our more detailed responses to the specific consultation questions. We have reordered and lightly edited to assist readers.

**Author**: Dr. Oliver Marsh, Project Lead "Auditing Algorithms for Systemic Risks"

**Contact**: marsh@algorithmwatch.org

## Mitigation measures linked to Generative AI

*Do you agree with the recommended best practices in this section?*

Our response to this section draws on our research with AI Forensics into election-related misinformation produced by Microsoft Bing Chat (now Copilot), as referenced in footnote 23 of the guidelines. We are continuing this research with a view to mitigating risks in the upcoming elections. This has included submitting a data access request under Article 40(4), however due to the lack of a delegated act - without which DSCs and platforms are unlikely to respond to requests - we believe it is highly unlikely this approach will succeed, which is a disappointment given the potential of these data access provisions to support protection of elections.

Specific ideas from the guidelines we support on grounds of ensuring basic levels of risk mitigation (under our first interpretation of "best practices" in opening):

- Further information to be provided under Article 40(12). In particular we request straightforward and stable ways to automate generation of repeated answers to prompts to allow for external assessment, as in our research with AI Forensics; and, to allow for research into scale of potential risks, access to volume-over-time and location-specific data on how many people are querying about particular topics or keywords, similar to the public Google Trends tool (though unlike the public Google Trends, numbers should be absolute, not relative, and allow for complex Boolean keyword queries as this is essential for isolating specific election-related queries).
- Warnings for potential errors should be very clear, not small footnote text. This should be particularly the case for any settings which may set higher expectations for accurate information (e.g. Copilot's "more precise" setting). Our proposal (below) that generative AI chatbots mimic traditional search when used for election-related queries should be explained to users by reference to risks of inaccuracies.
- While the precise approach may take many forms, we strongly support references to companies clearly testing their models for inaccuracies, with context-specific risks of elections part of this testing. Data and metrics used for testing, and results of tests, should be available (at the very least to vetted researchers).

Beyond these we would refer to our first answer about avoiding trying to set down best practices in this document, particularly in such a new and fast-moving field. It is also not clear the extent to which generative AI produces genuinely new problems; strategies such as reproducing old images in new contexts to mislead people, or artificially increasing the supply of hostile or divisive rhetoric, are still widely used without requiring AI. While the attention to misinformation and risks from hostile actors that GenAI has sparked is welcome, the specifics of generative AI should not be allowed to dominate the discussion

to the exclusion of other TTPs or technologies.

*Which risks of Generative AI for electoral processes should additionally be considered in this section?*

Again, we would recommend that there be access to dynamic information on specific risks. However we propose some additional high-level risks that should be noted based on our research.

In relation to searching for information about elections, we propose that generative AI chatbots always return answers completely equivalent to traditional search engines: i.e. lists of links to sources, ranked for quality, relevance, and reliability, with exactly the same text snippets as the traditional search engine would produce, rather than attempt to summarise and reproduce information in a probabilistic fashion (which comes with risks of inaccuracy and non-transparency). This is due to the following issues. If users wish to find specific information (e.g. polling numbers, candidate policies, information on how to vote) then the probabilistic nature of generative AI poses risks of returning confidently-stated inaccuracies. If users wish to find more value-based or politically-themed information, then the lack of transparency makes it very challenging to check whether a user is being only presented a specific point of view dressed up as fact.

1. Models may be trained to base probabilities of outputs too heavily on past data, which poses risks given that elections may involve new candidates, policies, or issues which may not be sufficiently accounted for in training data. For example, in our research on misinformation produced by Bing (now Copilot) in advance of German and Swiss elections, the actual candidates for elections were sometimes mis-named in favour of more famous politicians, or (more worryingly) had incorrect scandals associated to them.
2. For text-based questions and responses, generative AI chatbots produce a much narrower range of information compared to the list of outputs from traditional searches. While chatbot responses may include links to outside sources, the nature of the responses may encourage users to believe their questions have been “answered” with no sight of alternative views.

*What additional evidence-based best practices on risk mitigation for electoral processes related to the dissemination of Generative AI content should be considered?*

See previous answers; dissemination is a particular example of where it is unclear the extent to which GenAI raises new questions over and above dissemination of other

potentially risk-related content.

*What are best practices for providers of VLOPs and VLOSEs to ensure that their risk mitigation measures keep up with technological developments and progress?*

As proposed, this question should inform all aspects of these guidelines not simply generative AI.

## Election Specific Risk Mitigation Measures

*Do you agree with the recommended best practices in this section?*

Paragraph 12 is important: Using "local context-specific risks and Member State specific information" is absolutely vital, and would fit into our proposal that these guidelines focus on "best practices" in the sense of "would be negligent not to do". It is good to see the references to use of relevant partners. However were VLOP/SEs to individually conduct analyses of local contexts, with each potentially engaging the same or similar partners, this could create substantial additional work for all parties. We would strongly recommend that the guidelines propose joined-up efforts between parties to avoid this duplication. The DSCs, supported by expert advisory bodies, should play a role in coordinating these for individual Member States, and the Board in relevant information-sharing mechanisms between Member States. As other responses to this consultation have argued, such engagements should be transparent and accountable; we have created the 5 E's Framework for Stakeholder Engagement to address questions of how to ensure legitimate forms of engagement https://algorithmwatch.org/en/stakeholder-legitimacy-framework/

Paragraph 18 is also key for us, as we see these guidelines as filling a major gap while we await the delegated act on data access and as such are unable to use 40(4), and while issues with 40(12) are being addressed. The language should reflect this reality, providing legal clarity that vetting researchers and providing data access (for the purposes of researching and mitigating risks to upcoming elections) can and must proceed in the absence of the delegated act. It could also allow for DSCs and proto-DSCs to vet researchers to allow transparent and secure collaborations with VLOP/SEs for the specific purposes of researching and mitigating risks to upcoming elections, with an understanding that decisions can be modified following later adoption of the delegated act. The ability to vet researchers also plays a valuable role in supporting trusted and transparent collaborations with VLOP/SEs. Allowing for vetting and data access for election-related work would also play an important role of allowing for the upcoming delegated act to learn from

real examples, rather than hypotheticals. The guidelines should also ensure there are effective and rapid appeals mechanisms against insufficient compliance with 40(12) with VLOP/SEs, either in terms of who is granted access or the data provided.

Paragraph 16(d) is an example of where clarity of the DSA is needed, as decisions of whether something "threatens the integrity of the electoral process itself" (which we presume is intended to factor into whether the observed phenomenon is a "systemic risk" to elections) is important for DSA compliance but currently unclear. Examples of phenomena which would and would not be considered threats, or systemic risks, could be provided. AlgorithmWatch are crowdsourcing and analysing real observations for this purpose https://eu.jotform.com/233514485703052 , however the Commission itself should be able to provide examples (even if hypothetical) that guide its thinking as the enforcer of Article 34.

Paragraph 23: We strongly support the proposal to make more of the assessments conducted by VLOP/SEs available to external scrutiny. As we have noted the lack of transparency around risk assessments, in conjunction with a currently delayed and limited data access regime, drastically limits the power of external expertise to support risk mitigation around elections. However the text of this paragraph about fundamental rights impact assessments does not seem to align with the actual requirements specified in the DSA (either recital 90 or elsewhere). Nonetheless we would strongly support alignment with the text of Recital 90 (which also appears in the delegated act on auditing) that VLOP/SEs "test their assumptions with the groups most impacted by the risks and the measures they take. To this end, they should, where appropriate, conduct their risk assessments and design their risk mitigation measures with the involvement of representatives of the recipients of the service, representatives of groups potentially impacted by their services, independent experts and civil society organisations. They should seek to embed such consultations into their methodologies for assessing the risks and designing mitigation measures".

Paragraphs 11, 13, 15-17, 24, fall into our category of recommendations which seem to direct platforms towards particular ends or examples. This is not to say we disagree with the proposed themes, but there is not a clear sense of why *these* themes and specific examples have been chosen as priorities; whether these priorities should change over time or for different elections; and whether and how VLOP/SEs can appropriately involve themselves while mitigating attendant risks. Some of the proposed tasks are very specific, others extremely broad and high-resource (with attendant risks of overloading partners). As proposed in our first answer we would prefer to see briefer and higher-level requirements in these guidelines, more clearly connected to the DSA and related legislation, with more detailed recommendations coming through more dynamic and contextually-sensitive mechanisms.

For 16(f), it may be helpful to briefly summarise key requirements of the upcoming regulation on political advertising.

Paragraph 20: “As transparent as possible” runs risks of helping hostile actors game systems and avoid mitigation measures; a balance should be struck, potentially allowing vetted researchers much greater transparency.

Paragraph 22 could also reference risks of silencing candidates.

*What additional factors should be taken into account by providers of VLOPs and VLOSEs when detecting systemic risks related to electoral processes?*

See our first answer: this is best answered by dynamic and contextually sensitive expertise and resources.

*Are there additional mitigation measures to be considered as best practices on the basis of their proven effectiveness mitigating risks to electoral processes?*

See our first answer regarding “proven effectiveness”.

*How should providers of VLOPs and VLOSEs measure effectiveness of their risk mitigation measures in a reliable and conceptually valid way for electoral processes?*

As above, this is a complex question which is best addressed in a dynamic and contextually sensitive manner.

## Cooperation with national authorities, independent experts and civil society organisations

*Do you agree with the recommended best practices in this section?*

The types of cooperation listed in this section will be vital to our proposal that dynamic advisory mechanisms, not guidelines in documentation, should inform best practices. We therefore welcome this section. We would point to a greater role for the DSCs than is currently envisaged to select and coordinate relevant expertise, supported (particularly in their early work as DSCs) by bodies like EDMO and the Working Group on Elections of the Code of Practice. As noted previously, this should be conducted in a manner which

facilitates collaborations involving all VLOP/SEs and minimises duplicated efforts for partners; and also in a manner which ensures transparency, accountability, and legitimacy when engaging external parties.

*What other mechanisms should be considered to foster more effective collaboration with relevant stakeholders, such as national authorities and civil society organisations?*

This section can draw on proposals for expert advisory mechanisms proposed for DSA enforcement such as in https://dsa-enforcement.vergnolle.org/ (Dr Suzanne Vergnolle) and Here is what a strong Digital Services Coordinator should look like (Dr. Julian Jaursch) .

Paragraph 37: For countries with short electoral periods, putting in preparations one month before the electoral period seems narrow. There should also be consideration of how to address snap elections. We welcome that the measures continue for a month after elections, given potentially heightened risks to e.g. civic discourse in the event of contentious results.

Paragraph 41: The "follow the sun" emphasis on time zones seems out of place for Europe (which only has three time zones); the pertinent concern would rather be that processes are able to, where necessary, access relevant expertise and escalate decisionmaking irrespective of time.

Paragraph 42: Article 84 refers to professional secrecy, not crisis protocols.

*How can rapid response mechanisms be improved for handling election-related incidents on VLOPs or VLOSEs?*

More clarity as to how this intersects with the crisis response mechanism in Article 36, in the event that election-related activity poses a threat to public safety; and clearer understanding of accountability and potential ramifications for parties responsible for consequential and avoidable delays.

*What other mechanisms should be considered to foster more effective collaboration with national authorities and civil society organizations?*

As previously answered, our main requests would be (i) more transparency, if necessary via vetting procedures and (ii) coordination to ensure partners are not overwhelmed by simultaneous (and potentially duplicate) requests from many VLOP/SEs.

*Are there any additional resources that help providers of VLOPS and VLOSEs identify relevant organisations/experts at the national level?*

## After an electoral period

*Do you agree with the recommended best practices in this section?*

We strongly welcome the proposal to publish a public version of post-election review documents.

*What elements should be included in voluntary post-election review by providers of VLOPs or VLOSEs to assess the effectiveness of their risk mitigation strategies?*

Evaluations of how effectively knowledge- and expertise-sharing worked with partners and between VLOP/SEs; any further evidence generated on risks and effectiveness of mitigation measures in particular contexts, including where unclear or inconclusive.

## Specific guidance for the elections to the European Parliament

*What are your views on the best practices proposed in this section?*

As noted in previous answers, it should be acknowledged that various important aspects of the DSA will not be fully in place (Article 40), and/or fully clarified (Articles 34 and 35), in advance of the European Parliament elections in 2024 and therefore (as discussed in previous answers) the guidelines should provide measures by which these gaps can be addressed in the context of the European Parliament Elections; also to provide tests from which to learn for the delegated act on data access, guidelines on risk assessments, and other future clarificatory documents.

## Conclusion

*What additional feedback or suggestions do you have regarding these guidelines?*

As in our first answer, it is important that structures for further feedback, and dynamic adaptation of these guidelines, is possible beyond the proposed yearly reviews of the document. Despite our concerns about the framing of the guidelines, we are positive about the opportunities they present for us and for others and grateful for the work underlying them.