

Warsaw, 22.07.2026

Stakeholder Consultations Comment

Draft Guidelines on Classification of High-Risk AI Systems under Article 6 AI Act (Point 8(b) Elections)

Submitted by CEE Digital Democracy Watch

This submission responds to the Commission's draft guidelines on the classification of high-risk AI systems under Article 6 of Regulation (EU) 2024/1689 (AI Act), focusing in particular on Annex III point 8(b), which concerns AI systems used to influence the outcome of elections and voting behaviour.

As part of our work on digital democracy and AI-mediated electoral risks, we welcome the initiative to provide practical interpretative guidance and offer a set of clarifications and amendments that, in our view, are necessary to ensure effective protection of voters' rights and electoral integrity.

Drawing on recent research on algorithmic microtargeting, synthetic media, AI-driven disinformation and backend interference with electoral infrastructure, we argue that the current draft remains too narrow in three key respects: it focuses almost exclusively on systems explicitly "intended" for electoral influence; it relies on a direct-exposure condition that does not reflect how modern AI systems shape voter behaviour; and it lacks terminological precision regarding different forms of "influence on voting", including personalization, violations of information integrity and manipulative data practices.

We therefore propose targeted amendments to recitals 435–438 and 443–445, as well as a practical set of questions and implementation proposals to support regulators and providers in identifying high-risk AI systems in practice. Our overarching aim is to close avenues for regulatory arbitrage, prevent hidden backend threats to electoral processes, and provide clearer, risk-based criteria for distinguishing legitimately informative systems from those that undermine voters' autonomy and the democratic process.

1. General remarks

Article 8(b) of the AI Act currently classifies AI systems as high-risk if they are intended to influence the outcome of elections or the voting behaviour of voters and if voters are directly exposed to the system's outputs.

In support of ongoing efforts to implement the AI Act and assist relevant administrative bodies in fulfilling their obligations, we draw attention to the following issues, whose inclusion in the final guidelines would more fully ensure citizens' rights:

- The current guidelines refer only to systems dedicated to participation in the electoral process.
- The condition of direct impact.
- Lack of precision regarding “influence on voting”.

1.1 Systems “dedicated” to participation in the electoral process

The current approach, which limits the scope of Article 8(b) to AI systems dedicated to participation in the electoral process, is insufficient for effective protection of elections. Contemporary general-purpose AI systems can be configured and adapted to influence voters, regardless of the original intended use. The threat of electoral manipulation stems not only from purpose-built systems but also from widely available tools repurposed by political actors.

We therefore propose expanding the definition in Article 8(b) by replacing and supplementing the criterion of provider intent with the criterion of system capability to actually influence voters' conscious decision-making, irrespective of the system's initial functional categorisation.

A system designed for e-commerce with personalization capabilities can be instantly repurposed for voter microtargeting, yet the provider can claim the system was not intended for elections. Since e-commerce algorithms and political microtargeting systems rely on the same fundamental architecture, commercial tools can be adapted for elections while allowing providers to bypass strict campaign rules under the guise of selling a non-political product.

The current wording “intended to be used for” focuses exclusively on the primary intention of the system designer and creates a significant gap:

- Technology as a multifunctional tool: AI systems for content recommendation, microtargeting or sentiment analysis are not originally designed for electoral purposes, yet can be seamlessly adapted within campaigns (e.g. a general recommendation algorithm used to disseminate electoral content).
- No accountability for secondary use: A system designer can argue that the system was not “intended” for electoral purposes, despite it being actually used that way. This constitutes regulatory arbitrage.

1.2 The condition of direct impact

In practice, direct impact on voter behaviour often occurs entirely in the backend, invisible to the voter. A voter may receive a targeted advertisement, personalized news feed or recommendation, but does not see the system that analysed their data, identified their vulnerabilities or decided what message they should receive. They do not see the psychological profile built about them or understand that the content they encounter has been algorithmically selected to exploit specific cognitive biases. Yet the system directly shapes their voting decision. Therefore, the direct-exposure condition creates a significant loophole by enabling highly influential AI systems to be classified as operating outside the Act’s scope, potentially restricting the regime’s reach to a narrow subset of purpose-built systems.

Moreover, the current formulation restricts Article 8(b) to systems that directly influence individual voter behaviour, leaving a regulatory gap concerning threats to the technical and procedural integrity of electoral processes. Documented cases of electoral interference worldwide include:

- AI systems designed for cyberattacks against voting infrastructure or voter registration systems.
- Synthetic media technologies generating fraudulent representations of electoral officials and candidates.
- Algorithmic systems sabotaging voter registration databases.

We propose supplementing Article 8(b) with provisions addressing systems that interfere with electoral process integrity through data processing operations targeting voting systems, registration infrastructure or the authenticity of electoral communications, irrespective of whether interference occurs at the level of individual persuasion or systemic procedural disruption.

Originally, “natural persons must be directly exposed to the outputs of the AI system” meant a system was high-risk only if voters actually saw its outputs. A system hacking a voter registration database would not be high-risk because voters never see it. Yet the most dangerous threats to democracy are precisely those that voters do not see.

We welcome that newer drafting begins to extend high-risk classification to backend systems and voter-data manipulation, while seeking to avoid an overly broad classification that would be impractical or unfair to entities operating legitimate tools. The guidelines should make this extension explicit.

1.3 Lack of precision regarding “influence on voting”

The third gap stems from a lack of terminological precision regarding “influence on voting”. The current definition does not distinguish between fundamentally different threat types, such as:

- Systems disseminating false information (disinformation systems).
- Systems personalizing political messaging based on voter profiles (personalization engines).
- Systems violating information integrity of voter data to create behavioural profiles (behavioural data manipulation).

Without this distinction, implementation becomes uncertain. For example, a system using factual information to provide personalized summaries of a party programme would face identical requirements to a system disseminating synthetic deepfakes of a candidate, despite dramatically different threat levels to the democratic decision-making process.

This lack of precision enables divergent interpretations by member states, weakening harmonization and allowing circumvention of Article 8(b) through different system classifications. We therefore call for a threat categorization with clear criteria distinguishing violation of informational integrity, personalization and data manipulation, each with corresponding compliance levels.

The definition of Article 8(b) should explicitly categorize distinct forms of AI interference with electoral decision-making. Three primary mechanisms require separate regulatory treatment:

- Personalization: individual voters receive differentiated political messaging tailored to their profiles via microtargeting.

- Violation of information integrity: systematic dissemination of false, misleading or distorted factual content designed to alter perceptions of verifiable facts.
- Cognitive manipulation: exploitation of documented cognitive biases and behavioural vulnerabilities to circumvent rational deliberation.

These remarks underpin our proposed textual amendments.

2. Remarks on specific recitals

2.1 Recital (435) – suggested amendments

Original text (435)

(435)Point (8)(b) of Annex III lists as high-risk AI systems intended to be used for (i) influencing the outcome of an election or referendum or (ii) the voting behaviour of natural persons in the exercise of their vote in elections or referenda. That provision clarifies that this use case does not include AI systems the output of which natural persons are not directly exposed to, such as tools used to organise, optimise or structure political campaigns from an administrative or logistical point of view. Recital 62 AI Act explains that the classification of AI systems falling within this use case as high-risk aims to address the risks of undue external interference with the right to vote enshrined in Article 39 of the Charter, and of adverse effects on democracy and the rule of law.

Suggested new version

*Point (8)(b) of Annex III lists as high-risk AI systems intended to be used for **or capable of being deployed to:** (i) influence the outcome of elections or referenda; (ii) influence the voting behaviour of natural persons **through personalized information, violation of information integrity or harmful manipulation techniques;** or (iii) interfere with the **integrity of electoral processes through data processing concerning voters or electoral operations.** That provision clarifies that this use case does not include AI systems the output of which natural persons are not directly exposed to, such as tools used to organise, optimise or structure political campaigns from an administrative or logistical point of view. Recital 62 AI Act explains that the classification of AI systems falling within this use case as high-risk aims to address the risks of undue external interference with the right to vote enshrined in Article 39 of the Charter, and of adverse effects on democracy and the rule of law.*

Justification

The amendment strengthens protection of fundamental electoral rights by replacing open-ended definitions with precise risk categories that better reflect actual threats to democracy and the rule of law.

2.2 Recital (436) – suggested amendments

Original text (436)

For an AI system to be classified as high-risk under the use case listed in point 8(b) of Annex II AI Act, two conditions must be fulfilled: first, the intended purpose of the AI system must be the influence the outcome of an election or referendum, or to influence the voting behaviour of natural persons in the exercise of their vote in elections or referenda; second, natural persons must be directly exposed to the outputs of the AI system. These conditions and the concepts necessary to interpret them are analysed in the subsections below.

Suggested new version

*For an AI system to be classified as high-risk under the use case listed in point 8(b) of Annex III AI Act, two conditions must be fulfilled: first, the AI system must be intended for or capable of being deployed to influence the outcome of elections or referendum; **second, the AI system must be capable of influencing the voting behaviour of natural persons, by any applications – including but not limited to personalized targeting, violation of information integrity or manipulation with the exception of AI systems whose output natural persons are not directly exposed to, such as tools used to organise, optimise and structure political campaigns from an administrative and logistical point of view explained by the Recital 62 AI Act.***

Justification

The amendment introduces four key changes:

- (1) Expands intention from “intended purpose” to “intended for or capable of being deployed to”.
- (2) Specifies categories of threats.
- (3) Changes exposure from a positive condition to an exemption.
- (4) Explicitly references Recital 62.

Together, these changes make the provision more protective, more precise and more practically implementable.

2.3 Recital (437) – suggested amendments

Original text (437)

Only AI systems intended to be used to influence the outcome of an election or referendum or the voting behaviour of natural persons, i.e. the AI system would have to be specifically intended to have an effect on the electorate's choice or turnout, fall within the scope of the use case listed in point 8(b) of Annex III. The focus of the use case is thus on underlying objectives: the system must be directed towards electoral influence, not merely be incidentally capable of it. Therefore, AI systems with non-intentional influence on the outcome of an election or referendum or the voting behaviour of natural persons, such as general content recommender systems that, amongst other content, also recommend political content and thus, de facto, have the potential to influence political opinions, should not be classified as high-risk. General-purpose AI systems, which offer sufficient safeguards against a use influencing electoral processes, e.g. a chatbot clearly replying that it cannot give any voting advice when asked by a user, or providing only neutral, factual, and informational content about elections (voting procedures (how, when, where), registration procedures; the institutional and legal framework; objective information on parties and candidates (e.g. references to their official websites); and general civic education on democratic processes) would also not fall under high-risk category as not being intended to influence the outcome of an election or referendum or the voting behaviour of natural persons.

Suggested new version

For an AI system to fall within the scope of the use case listed in point 8(b) of Annex III, it must be intended for or capable of being deployed to influence the outcome of an election or referendum or the voting behaviour of natural persons through specific mechanisms of electoral influence, including personalized targeting, violation of information integrity or manipulation of voter data. The focus of the use case is thus on the system's capacity to affect electoral outcomes or voter autonomy through these mechanisms, not merely on incidental exposure to political content.

Therefore, general-purpose AI systems that are not specifically designed for electoral purposes, nor capable of being readily adapted for such purposes, should not be classified as high-risk, e.g. chatbot that explicitly declines to provide voting advice when asked, or a general content recommender system that like voting procedures, registration procedures, institutional and legal frameworks, objective information on parties and candidates with references to their official sources, and general civic education on democratic processes. should not be classified as high-risk, provided they do not apply personalization, microtargeting, or behavioral manipulation to electoral content.

As a sufficient safeguards in a purpose to receive an objective information should be understood: (i) AI system that does not possess verified or definitive data regarding an electorally sensitive query, it shall explicitly state its lack of knowledge instead of speculating, and shall immediately redirect the user to an authoritative electoral source by stating, for example: "I am not sure when the elections in your region take place. Please go here to check: [link to the official election website]". (ii) Providers shall maintain a comprehensive electoral error log to record every instance where the AI system has provided inaccurate or misleading information, including minor chronological discrepancies, such as incorrectly stating that an election is scheduled for June 15th instead of June 16th. (iii) Where an AI system has disseminated inaccurate electoral information on a large scale, the provider shall actively notify all affected users without undue delay, providing a corrective statement reading, for example: "We previously provided you with incorrect information regarding the elections. Here is the correct information: [correct information]".

An AI system shall benefit from a safe-harbour exemption from liability for errors contained within official data, provided that the system proactively and consistently redirects voters to authoritative channels by stating, for example: "Click here to check the official election schedule".

Justification

This calibrated reformulation resolves ambiguities while remaining implementable. It:

- (1) Expands the trigger from “specifically intended” to “intended for or capable of being deployed to”.
- (2) Anchors high-risk classification to identifiable mechanisms of electoral influence.
- (3) Clarifies that general-purpose systems are not automatically high-risk, but only where they employ influence mechanisms or are readily deployable for such purposes.
- (4) Provides concrete examples to reduce interpretive uncertainty.

2.4 Recital (438) – suggested amendments

Original text (438)

That the AI system should be intended to be used to influence the outcome of an election or a referendum or to influence the voting behaviour of natural persons in the exercise of their vote in elections or referenda means that the scope of the use case is both collective influence (the outcome of the elections or referenda as a whole) and individual influence (the behaviour of individual voters). Despite their alternative formulation, those two use cases are partially overlapping. That is because the influence of voting behaviour has as its necessary consequence an influence on the outcome of an election or referendum, whether by shaping the electorate's choices or by affecting voter turnout, and thus altering the distribution and the outcomes of votes. The wording of point 8(b) of Annex III explicitly ties the influencing of voting behaviour to the 'exercise of [a] vote in elections or referenda'.

Suggested new version

*That the AI system should be intended to be used to influence the outcome of an election or a referendum or to influence the voting behaviour of natural persons in the exercise of their vote in elections or referenda means that the scope of the use case is both collective influence (the outcome of the elections or referenda as a whole) and individual influence (the behaviour of individual voters) **as personalized targeting of differentiated messages to individual voters based on their profiles; violation of information integrity or deliberately misleading information about candidates, parties, or electoral procedures; manipulation of voter data to identify vulnerabilities and deploy targeted influence.** Despite their alternative formulation, those two use cases are partially overlapping. That is because the influence of voting behaviour has as its necessary consequence an influence on the outcome of an election or referendum, whether by shaping the electorate's choices or by affecting voter turnout, and thus altering the distribution and the outcomes of votes. The wording of point 8(b) of Annex III explicitly ties the influencing of voting behaviour to the 'exercise of [a] vote in elections or referenda'.*

Justification

The amendment specifies mechanisms of influence, addressing ambiguity in the original text and giving entities clearer guidance on which practices trigger high-risk classification.

CEE Digital Democracy Watch

Mokotowska 43/104, 00-551 Warsaw, Poland

www.ceeddw.org

KRS 0001090110 | EU Transparency Register 114284692189-05

2.5 Recital (443) – suggested amendments

Original text (443)

Point 8(b) of Annex III refers to AI systems intended to be used for a particular purpose, therefore requiring all providers of such AI systems to classify them as high-risk. The provision does not specify who provides, puts into service, or uses the AI system. The decisive factor is the intended purpose of the AI system as specified by the provider, not who the deployer will be.

Suggested new version

*Point 8(b) of Annex III refers to AI systems intended to be used for a particular purpose, therefore requiring all providers of such AI systems to classify them as high-risk. The provision does not specify who provides, puts into service, or uses the AI system. The decisive factor is **whether the AI system is intended for or capable of being deployed to electoral influence as determined by its technical design and functionalities, regardless of the provider's declared intent.***

Justification

Relying on declared intent creates regulatory arbitrage: providers can mischaracterize intended use to evade high-risk classification (e.g. a personalization algorithm built for voter targeting but labelled “general recommendation”).

2.6 Recital (445) – suggested amendments

Original text (445)

Point 8(b) of Annex III clarifies that it does not include AI systems to the output of which natural persons are not directly exposed, such as tools used to organise, optimise or structure political campaigns from an administrative or logistical point of view. This generally excludes AI systems used by political parties or candidates to increase back-office efficiency.

Suggested new version

CEE Digital Democracy Watch

Mokotowska 43/104, 00-551 Warsaw, Poland

www.ceeddw.org

KRS 0001090110 | EU Transparency Register 114284692189-05

*Point 8(b) of Annex III clarifies that it does not include AI systems, **such as tools used to organise, optimise or structure political campaigns from an administrative or logistical point of view.** This generally excludes AI systems used by political parties or candidates to increase back-office efficiency.*

Justification

The amendment removes the direct-exposure reference from this recital, aligning it with the proposed shift where exposure functions as an exemption rather than a primary condition.

Recitals (439–442), (444) and (446–447) are left without amendments.

3. Questions for the practical part – identifying high-risk AI models

To support implementation, we suggest the guidelines explicitly address the following questions:

- Will AI-driven bots on social media that impersonate real individuals and endorse specific political candidates be classified as high-risk under point 8(b) of Annex III?
- Will bots that mass-disseminate messages encouraging people to vote for a candidate, or that utilize personal characteristics such as voice cloning (e.g. the India robocall example), be considered high-risk?
- Will artificially generated podcasts simulating conversations with a real politician be classified as high-risk?
- Will an AI system that automatically alters political content based on regional issues and then automatically delivers tailored electoral advertisements to voters be classified as high-risk?

4. Elements missing from the document, unregulated areas, and proposals

Beyond the concrete textual amendments proposed above, it is essential that the classification of high-risk AI systems related to elections be treated as an evolving practice rather than a one-off technical exercise. The nature of AI systems and the pace of progress make it indispensable to continuously evaluate implementation, revisit risk assumptions and remain in structured dialogue with citizens, civil society organisations, public administrations, business and academia, so that emerging use cases and vulnerabilities are not overlooked.

Based on our work in election-integrity monitoring and exchanges with international bodies, we see particular value in developing practical tools and supervisory practices around three clusters of safeguards: clear and portable labelling of AI-generated political content; meaningful transparency of recommendation and targeting algorithms to competent authorities; and enforceable rights for users to appeal automated moderation decisions and obtain human review where their political expression is restricted.

At the same time, the guidelines should explicitly recognise that existing electoral law prohibitions on deceptive practices must fully apply to AI-mediated campaigning. In practice, this means clarifying that it is unlawful to generate or deploy deepfakes targeting political opponents; that political actors must disclose when AI systems are used for behavioural or demographic targeting, and that the operation of automated bot networks for the dissemination of disinformation constitutes a prohibited form of electoral interference rather than a neutral technical choice.

We remain at the disposal of the Commission and other stakeholders to discuss these developments and to contribute further evidence from our monitoring and research work.