CIVIL LIBERTIES UNION FOR EUROPE

Civil Liberties Union for Europe: Response to the targeted consultation on the draft Guidelines for the classification of high-risk artificial intelligence systems

23 July 2026

Executive Summary

Liberties is pleased to contribute to the European Commission's targeted consultation on the draft Guidelines for the classification of high-risk artificial intelligence systems. The AI Act's landmark role in shaping the future of AI use, regulation, and fundamental-rights protection should not be understated. The high-risk system classifications and the guidelines that accompany them are clearly intended to safeguard both the usability of AI and the highest standards of rights protection. Our submission aims to help ensure that this framework addresses the significant power asymmetry between individuals on the one hand, and the developers, providers and deployers that control broad-reaching AI systems on the other. This imbalance is further intensified by the opaque nature of AI systems. Without meaningful transparency, accountability, and safeguards, oversight may exist only on paper and will not work effectively.

This submission focuses on ways to reimagine the draft Guidelines' operationalisation of "intended use" and the Article 6(3) filter, along with the risks of interpretations too narrow and provider self-selection. We also examine loopholes, which, if employed, seem to be counter to the very goals of the Guidelines and regulatory framework, leaving individuals with potentially little recourse. Ultimately, we suggest introducing intermediate safeguards to increase public and civil society awareness, transparency, and prevent ex-post regulatory enforcement for areas impacting fundamental rights.
Liberties' comments focus on two elements of the draft Guidelines: the operationalisation of "intended use" as the impetus to high-risk classification, and the design and application of the Article 6(3) filter mechanism. We highlight the risks that arise when "intended purpose" and "intended use" are interpreted too narrowly, and, combined with expanded reliance on provider self-selection, create loopholes through which systems deployed in Annex III domains can avoid high-risk status despite their practical impact on fundamental rights.

Where these loopholes can have immediate, material consequences for fundamental rights — among others, in relation to elections and democratic processes, access to healthcare, social benefits and other essential services, and the administration of justice —are of particular concern to Liberties. In many of these settings, over-reliance on provider self-selection and a lack of transparency can exacerbate existing power imbalances between individuals and the providers and deployers of AI systems, leaving those individuals unable to seek meaningful redress.

We propose that the Guidelines:

1. **Intended use:** Shift the evaluation of "intended use" away from a provider-controlled monopoly toward a multi-factor test based on technical capabilities, high-probability foreseeable uses, and real-world deployment, rather than relying on narrow marketing labels or terms of service.
2. **Filter mechanism:** Narrow and tighten the Article 6(3) filter conditions (a–d) so that provider self-selection cannot be used to mischaracterize systems or rely on historically

biased data as merely "narrow," "preparatory," or "supportive" low-risk tasks. Tighten the Article 6(3) filter so that loopholes cannot allow Annex III decisions to be treated as low-risk;

3. **Elections and referendums:** Ensure General-Purpose AI (GPAI) chatbots and search interfaces providing election advice or voting-match scores are explicitly classified as high-risk under Annex III, Point 8(b), include Digital Services Act (DSA) risk assessments.
4. **Judicial Integrity:** Mandate an EU-wide AI disclosure certificate for criminal evidence, extend transparency and explanation rights to private legal-tech software, define "judicial authority" to close evidentiary oversight gaps across both criminal and civil proceedings.

These changes would help the Guidelines fulfil their stated role of supporting high-risk AI classification that is both rights-affirming and practicable in ever-evolving real-world deployment.

1. Closing potential loopholes in intended use

As underlined in the proposed Guidelines, General principles for classification of high-risk AI systems (section 2), how a system is "intended to be used" — its "intended purpose" — shall be the key criterion for determining whether it shall be designated a high-risk system. Article 3 (12) of the AI Act defines "intended purpose" as "the use for which an AI system is intended by the provider, including the specific context and conditions of use, as specified in the information supplied by the provider in the instructions for use, promotional or sales materials and statements, as well as in the technical documentation".

"Intended purpose" and "intended use" should be interpreted broadly to avoid the creation of loopholes through which systems deployed in Annex III domains can avoid high-risk status despite their impact on fundamental rights. Our main concern is that providers can strategically draft instructions for use, terms of service, and promotional materials using narrow and technical language, by explicitly stating that an AI system is *not* intended for high-risk uses (e.g., declaring a worker-monitoring system as "workflow optimisation" rather than "employment evaluation"). Currently, the framing offers an easy way for providers to bypass high-risk obligations such as conformity assessments and fundamental rights impact assessments. Furthermore, providers could claim legal immunity, arguing the deployer violated the system's "intended purpose".

Under Article 6(3), an AI system listed in Annex III is exempt from being classified as high-risk if it does not pose a significant risk of harm to health, safety, or fundamental rights. There is a potential loophole if developers carefully frame the "intended purpose" in documentation as a purely "preparatory", "supportive," or "narrow administrative" function, even in the field of sensitive domains like credit scoring or border control.

To prevent "intended purpose" from functioning as an ex-ante compliance bypass, we recommend that the guidance emphasise evaluating systems on actual *technical capabilities,* rather than provider marketing. *High-probability, foreseeable use*s, and *functionality* should be part of the legal scope regardless of the "intended purpose". The sector-specific sections included in the draft Guidelines already offer examples of such an approach (e.g., 3.3.2-3.5.5). In Sections 3.3 and 3.4 (education and employment), the Guidelines identify AI systems through the specific decision-making roles they play: "AI systems intended to be used to evaluate learning outcomes", "intended to be used for the purpose of assessing the appropriate level of education," "AI systems intended to be used for the recruitment

or selection of natural persons", and "intended to be used to manage work-related relationships". Here, it is not the market label, or a provider-driven focus; rather, the capabilities are explicitly defined through the intended outcomes and product users' desire to "manage", "evaluate", "recruit" and "assess" that matter.

As such, Liberties suggests a streamlined functional approach to "intended to be used". Additionally, the creation of shared provider-deployer responsibility when systems are distributed into sectors defined as high-risk under Annex III, through deployers also being required to disclose Annex-III deployments, would create additional buffers from potential loopholes.
As a general rule, Annex III systems are considered high-risk, and any deviation from this is exceptional and requires supporting evidence and documentation. However, this does not resolve the structural problem that the initial classification decision must be made by the regulated entity itself. The implementation architecture includes ex post corrective measures. Market supervisory authorities may reassess systems that the service provider has not classified as high-risk and, if necessary, treat them as high-risk. In addition, market supervision and fundamental rights authorities may also request documentation and, in justified cases, technical testing. All of this is important, but it is essentially an ex post protective mechanism. By the time reclassification takes effect, the system has often already had a real impact on people's rights or access.

Specific suggestions for closing the loopholes:

1. The most important dogmatic improvement would be to shift the concept of "intended use" away from a provider-controlled monopoly and toward a rebuttable, multi-factor legal test. We are of the opinion that the guidance should incorporate the principle that classification should consider not only the documented purpose but also reasonably foreseeable misuse. This approach would provide a more functional classification and an increase in the compliance workload.
2. It would also be important to strengthen the documentation requirements for Annex III systems classified as non-high-risk, including a detailed classification memo, a verifiable decision tree, a use-case inventory, a list of prohibited and permitted configurations, and a description of known real-world abuses.
3. It would be advisable to require mandatory independent preliminary testing, third-party testing, or randomised regulatory sampling audits, particularly in the areas of employment, essential services, the judiciary, and law enforcement.
4. Strengthened testing channels for fundamental rights authorities: a dedicated pool of technical experts, common investigation protocols, an expedited emergency reporting channel, and mandatory notification of the market supervisory authority in all cases where a fundamental rights authority disputes the system's actual decision-making power.
5. Strengthen the liability aspect. If a service provider intentionally or through gross negligence defines the intended use too narrowly, thereby causing the system to be excluded from the high-risk regime, this should result in specific legal consequences: a reversal of the burden of proof in certain legal disputes, more severe sanctions, and exclusion from the safe harbour. The right to file complaints, whistleblower protection, and market surveillance reclassification are existing foundations, but they are not sufficient on their own.

2. Filter Mechanism

While the intentions of the "filter mechanism" set out in Article 6 of the AI Act, and extrapolated upon in the Commission's Draft Guidelines, stress that Article 6(3)(a)-(d) be "narrowly" interpreted to

ensure the protection of fundamental rights, the mechanism's heavy reliance on providers' own self-selection, documentation, and system interpretation currently introduces too many gaps to assuage these protections and even regulatory efficiency. As noted previously, defining the product's "intended purpose" not by the system's capabilities but by the specified, intended use of the individual party at the time of drafting creates numerous problems. Since providers draft the intended-purpose statement and supply the primary evidence against which that statement is judged, the interpretive authority is almost entirely unilateral. Though the Guidelines' discussion of the natural person/legal person distinction (Section 2.2) provides some specificity through the lens of "intended use[s]", for instance, the variability of AI systems, their rapid iteration, and minimal documentation requirements present potential scenarios where the documentation requirements and initial definitions introduced are quickly outrun by the machinery itself.

Mechanically, this structural unilateralism extends to the four-filter conditions themselves.
Point (a) of Article 6(3) of the AI Act requires that systems allowed to be “filtered” out be those that perform “narrow procedural tasks”. The Guidelines that distinguish what counts as a narrow procedural task can raise concerns about arbitrariness. For instance, paragraph (92) specifies that AI systems that sort or alter the structure or presentation of data can be constituted as a narrow procedural task, whereas systems positing “value judgments for decision-making”, according to paragraph (93), cannot fall within such a filter. Changing the structure or presentation of data may very well extend to instances shared in paragraph (93), where evaluating whether data is “useful” or “less useful for human assessment” occurs.

Specific suggestion for point (a)

1. The Guidelines should clarify that the assessment under point (a) must hinge on the system’s actual function in modulating which information is presented to a user and how, rather than on the provider’s choice of labels. Systems that merely correct or standardise text, such as ordinary word-processors or spell-check tools, can appropriately be treated as narrow procedural tasks where they do not alter the substance of what is available to the human reviewer. By contrast, systems that prioritise, suppress, or filter information in a way that affects what the reviewer is likely to perceive or rely on should not qualify as performing only a narrow procedural task. The Guidelines should also indicate what mechanisms will verify that systems framed as data-restructuring tools are not substantively performing the aforementioned tasks under another label.

Point (b), covering AI systems “intended to improve the result of a previously completed human activity”, presents additional gaps. Specifically, if an AI system is “improving a previously completed human activity”, the sample use cases specify that qualifying systems are those that note contradictions or errors in already finalised human work, or that focus on mapping conclusions to evidentiary records, rather than systems that produce different results. The minutiae between these differentiations again depend on the providers’ own classifications rather than on the systematic tools themselves.

Specific suggestion for point (b)

2. The Guidelines should specify that point (b) applies only where the system is limited to identifying clerical errors or mismatches in already completed human work. If the same model can effectively rewrite the underlying account, however, the classification processes should focus on the system’s functional capabilities instead of the provider’s framing of the output. The Commission should clarify that systems

capable of materially altering the evidentiary or factual picture cannot be filtered out under point (b).

Point (c), which focuses on detecting decision-making patterns (systems detecting decision-making patterns or deviations from prior patterns, without replacing or influencing the prior human assessment), introduces potentially the greatest risks to fundamental rights protections. If prior decision-making use cases were inherently discriminatory, whether wittingly or unwittingly, such AI systems that address pattern deviations will treat any corrective pattern as anomalous.

Specific suggestion for point (c)

3. The Guidelines should clarify that systems falling under point (c) cannot benefit from the filter if there is a possibility that the baseline pattern against which deviations are measured reflects historical discrimination or other unlawful bias. Providers seeking to rely on point (c) for their filter should demonstrate that the historical decision pattern used as the comparator has been assessed for discriminatory effects and the system is absent from liability to identify corrective departures from discriminatory patterning as anomalous. The Guidelines should encourage accessible human review and internal challenge procedures for decisions influenced by such systems.

Point (d) rather nebulously points to the term “preparatory” as a defining, clear determinative. Yet, most AI functionalities can potentially be reclassified as preparatory to a “human act”. Moreover, the use cases specifying what does and does not classify as “preparatory” are, in some ways, arbitrary, as the classifications depend on the providers’ framing. The Guidelines should clarify that the term ‘preparatory’ cannot be satisfied merely because a human act formally occurs later in the workflow. The question asked, instead, should be whether the AI system can materially influence the human reviewers’ subsequent decisions, including score assignment. If a system meaningfully weighs particular inputs, it should not be treated as merely preparatory for the purposes of Article 6(3)(d).

Specific suggestion for point (d)

4. The Guidelines should clarify that an AI system cannot benefit from the Article 6(3)(d) filter simply because a formal decision by a human occurs later in the workflow. Providers should demonstrate that the system does not materially influence the follow-up human assessment. Where a system assigns scores, weights inputs, ranks or prioritises cases, recommends an outcome, or otherwise affects how the reviewer evaluates the case, it should not be regarded as performing a merely preparatory task.

The Guidelines should clarify that meaningful human involvement means they have sufficient information, competence, time, and authority to assess the system’s output independently and to ensure the possibility to depart from it without adverse consequences.

3. Elections and referendums

It is essential that the Commission takes further steps within the Guidelines to secure the integrity of elections, referendums, and other democratic processes. The draft Guidelines do not prioritise the functional and foreseeable impact on elections, but rather the objective described by the provider. This logic particularly favours search, recommendation, and chat systems that, while not marketed as “election AI” actually provide guidance on party or candidate selection during a campaign period.

Liberties' research paper[1] specifically recommends that such GPAI systems should be classified as either high-risk or, at the very least, subject to GPAI systemic risk and Digital Services Act (DSA) risk assessments. The final Guidelines must shift from the formal "intended purpose" toward the functional, foreseeable influence that arises in a campaign context. In the short term, this can begin by clarifying the Guidelines; in the medium term, by narrowing or amending the filter in Article 6(3). Liberties' research sheds light on the extent to which GPAI chatbots are unreliable sources of information during election campaigns, providing false or misleading information that has the potential to shape voters' opinions and behaviour (see also the section on intended use). The Guidelines should place greater emphasis on the election-specific use of GPAI and create a use-case example where election-related conversational AI models that provide election advice—including voting-match scores and suggestions intended to inform voting behaviour—are considered high-risk, even when such outputs are accompanied by a disclaimer that says the chatbot cannot provide voting advice. Such a disclaimer is clearly inconsistent with the rest of the text output and insufficient as a safeguard to inform users of the system's limitations and errors.

The mere fact that general-purpose AI systems clearly reply that they cannot give any voting advice when asked by a user is an insufficient safeguard, and the systems should fall under the scope of high-risk obligations in cases where this reply is augmented by additional text that offers voting advice. GPAI providers should be transparent, document their systems, and ensure they are evaluated, non-manipulative, secure, and continuously monitored for risks to fundamental rights, democracy, and the rule of law.

Specific suggestions for the draft Guidelines

AI systems intended to be used for influencing the outcome of elections or referendums:

1. The Guidelines should address GPAI models directly. They should largely keep the existing example of "AI-enabled voter advice applications" in the Guidelines on the classification of high-risk AI systems, ensuring that it is retained as a use-case example. Therefore, paragraph (437) of the Commission's Draft Guidelines should be updated to ensure that it includes a use case specifying that chatbots providing election advice—including voting-match scores and suggestions intended to inform voting behavior—must fall within the use case of point 8(b) of Annex III, even when such outputs are accompanied by a disclaimer stating that the system cannot provide voting advice.

2. The text stating, "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 [...] would also not fall under the high-risk category" must be clarified to exclude chatbot voting-advice use cases, unless the chatbot's only reply is that it cannot give voting advice.

3. In relation to elections and referendums, general-purpose AI systems should be transparent, well-documented, evaluated, non-manipulative, secure, and continuously monitored for risks to fundamental rights, democracy, and the rule of law. The draft Guidelines should reference

[1] Day, Simon, Should AI Tell You Who to Vote For?, July 2026, https://www.liberties.eu/f/uz6rz9

the Digital Services Act according to a system's use cases. Dual oversight under the DSA and the AI Act should help to close the regulatory gap in relation to GPAI voting advice.

In relation to the DSA, our specific suggestions are the following:

4. The involvement of the DSA enforcement team is important. They should interpret the DSA's "online search engines" functionally, rather than purely technically, in relation to GPAIs that retrieve, process, and present information from across the internet in response to user queries and perform a search-like function.

5. The Commission should clarify that AI-generated answers may constitute search results 'in any format' where they serve an 'access to information' function.

6. Where AI search interfaces or conversational GPAI tools perform search-engine functions, they should be designated as Very Large Online Search Engines (VLOSEs). The Commission should clarify that providers cannot avoid DSA oversight merely because information is presented as generated answers rather than hyperlinks.

7. AI-generated electoral guidance should fall within the DSA risk assessment framework for civic discourse and electoral processes, in relation to Article 34(1)(c) of the DSA. Risk assessments should cover misleading or biased party matching, unequal visibility of political actors, and inaccurate political recommendations. These risks should be assessed before and during election periods.

8. Regulators should combine the DSA's systemic risk oversight with the AI Act's GPAI risk management obligations.

4. Judicial Integrity

The AI Act's Annex III classifications present numerous law-enforcement use cases for AI systems. However, important gaps remain regarding evidentiary transparency and the way AI-assisted outputs feed into judicial decision-making. Law enforcement's actions, particularly in criminal investigations, are eventually reviewed by judges, and final decision-making is considered subject to human oversight; it is at this point that opacity around AI's role in producing and processing evidence most directly threatens fair-trial guarantees.

The Annex III Gap in Law Enforcement

The AI Act does contain some targeted safeguards. Absent the post-remote biometric identification (RBI) reporting and documentation requirements under Article 26(1) and Article 49, other high-risk law enforcement systems under points 6(a)–(e) of Annex III are subject only to the provider-side documentation regime, while a tailored evidentiary disclosure regime comparable to the post-RBI framework is missing from the safeguards.

The gap is even worse for AI-driven crime analytics tools. Broader law-enforcement analytics and predictive-policing use cases were discussed during the legislative process, but the final Annex III text

limits point 6 to five specific use cases and does not explicitly list generic crime-analytics systems among them. These tools, used to gather, organise and visualise investigative data that may go on to inform prosecution files, sit outside the high-risk tier and instead fall under the general regime for non-high-risk systems, where only voluntary codes of conduct are available under Article 95.

The AI Act does not provide judges or affected persons with a dedicated route to obtain meaningful information about the system's role, leaving any access largely to sectoral or national procedural law without proper EU-level oversight. This is a glaring omission, given the Act has acknowledged such systems can be prone to repurposing under the Act's own "reasonably foreseeable misuse" concept: a tool built for pattern-based crime analytics can shift toward individual-based profiling without triggering any change in its regulatory classification. Without reporting requirements, tailored or general, judges tasked with evaluating evidence may lack any understanding of its origin or provenance.

The trust of a free and fair trial is implicated where judges themselves are left unaware of the particularities of AI's role in processing evidence. ECtHR doctrine directly links the equality-of-arms and adversarial-hearing guarantees under Article 6 ECHR to a defendant's ability to contest the admissibility, reliability, and completeness of evidence—the very function that point 6(c) systems are meant to perform.

Civil and criminal asymmetry

The structure of Annex III creates a distinction between criminal and non-criminal proceedings. Point 6 concerns law-enforcement activities connected with the prevention, investigation, detection, or prosecution of criminal offences, the execution of criminal penalties, and certain public-security functions. Its safeguards therefore do not generally apply to civil proceedings.
Point 8(a) is broader in subject matter. It covers AI systems intended to assist judicial authorities in researching and interpreting facts and law and in applying the law to a concrete set of facts. It also extends to systems used in a similar manner in qualifying alternative dispute-resolution processes whose outcomes produce legal effects for the parties.

Point 8(a) applies only where the system is used by or on behalf of a judicial authority or relevant dispute-resolution body. AI tools used independently by litigants, law firms, insurers, litigation funders, or other private actors may therefore fall outside this category, even where they influence the evidence submitted to a court or the strategy adopted in proceedings.

Institutional Overlap

Additionally, the responsibilities of judicial authorities and law enforcement are not fully delineated in the AI Act. Article 3 defines a "law enforcement authority" as any public authority competent for the prevention, investigation, detection or prosecution of criminal offences, but offers no parallel definition of "judicial authorities," despite Annex III point 8(a)'s reliance on that term. In Member States where there is institutional overlap between prosecutorial and judicial functions, this creates a genuine classification problem.

Our specific suggestions are the following:

1. Establish an EU-wide AI disclosure requirement for criminal evidence. The draft Guidelines should require a certificate. Disclosure should contain the tool's intended purpose, known

error rates, and processing methodology to ensure compliance with Article 6 ECHR guarantees regarding equality of arms and adversarial hearings.

2. Update Annex III, Point 6 to explicitly classify generic crime-analytics and predictive-policing software as high-risk AI systems.

3. Clarify the legal definition of "judicial authority". Clear criteria must delineate when an authority acts in an investigative or prosecutorial capacity (Point 6) versus a judicial fact-finding/interpretive capacity (Point 8(a)) to prevent gaps.

4. Extending high-risk status beyond tools used strictly "by or on behalf of" courts ensures that private legal-tech software undergoes mandatory transparency, bias auditing, and risk management before its outputs influence judicial strategy or judicial fact-finding.

5. Add to the Guidelines the right to explanation for law enforcement outputs and clarify that the individual right to explanation under Article 86 applies directly to affected persons.

About the Civil Liberties Union for Europe

The Civil Liberties Union for Europe (Liberties) is a Berlin-based civil liberties group with 24 member organisations across the EU campaigning on human and digital rights issues, including the rule of law, media freedom, SLAPPs, privacy, targeted political advertising, AI, and mass surveillance.

Transparency register number: 544892227334-39

Address: c/o Publix
Hermannstraße 90
12051 Berlin, Germany

Authors: Eva Simon, Emma Siegel, Jonathan Day
Contact: eva.simon@liberties.eu