European Partnership for Democracy

Is election integrity integral to the Artificial Intelligence Act?

The Artificial Intelligence Act (AI Act), the new EU Regulation that introduces rules for AI systems according to their risk level, was published in the Official Journal. The rules are going to be implemented over the next years with a phased application period.

As a horizontal framework, the Artificial Intelligence Act has the potential to **impact a wide range of issues linked to fundamental rights, including the right to vote** as in Article 39 of the Charter of Fundamental Rights of the European Union[1]. Furthermore, the Regulation itself mentions in the Recitals (Recital 1, 2 and 8 for example) as well as in Article 1 that it has as an **objective, among others, the protection of democracy and the Rule of Law** as well as the protection of fundamental rights included in the Charter of Fundamental Rights of the European Union. This includes **Article 39 of the Charter on the Right to vote and to stand as a candidate at elections to the European Parliament**. Recital 48 also sets criteria to define high-risk AI systems, including harms to fundamental rights among the risks, including the right to vote.

Based on the above considerations, while the Artificial Intelligence Act has no specific purpose to protect election integrity against the use of AI systems, it is still **susceptible to being used as a tool to ensure free and fair elections**, by protecting them from the potential negative impact of certain AI systems. How exactly, that is quite a different matter.

In the AI Act, there are several sections that refer either explicitly or implicitly to **AI systems with the potential to impact election integrity** and related solutions to limit such impact **according to their risk level**.

[1] Article 39 - Right to vote and to stand as a candidate at elections to the European Parliament
1. Every citizen of the Union has the right to vote and to stand as a candidate at elections to the European Parliament in the Member State in which he or she resides, under the same conditions as nationals of that State.
2. Members of the European Parliament shall be elected by direct universal suffrage in a free and secret ballot.

In particular:

- **prohibited AI systems**, which cannot be deployed on the EU market;
- **high-risk AI systems**, which will need to comply with specific obligations such as conducting risk assessments and putting forward mitigation measures;
- **limited-risk AI systems**, which will have to comply with specific transparency requirements;

Many AI systems often mentioned as a threat to elections including deepfakes, General Purpose AI (GPAI) and chatbots are mostly considered in the limited risk category - which also provides the lowest level of protection.

The question we will try to answer in this paper therefore is: **what kind of AI systems linked to elections would be included in the other two (more protective) categories**?

Which AI systems are (not) prohibited

When it comes to the **prohibited systems**, there are a few articles that might refer to AI systems used in the context of elections, namely **Article 5.1a on subliminal techniques**[2]; **Article 5.1b on exploiting vulnerabilities to distort the behaviour of a person**[3]; and **Article 5.1g on categorisation of natural persons based on political opinions**[4]. Article 5.1a and 5.1b are particularly controversial.

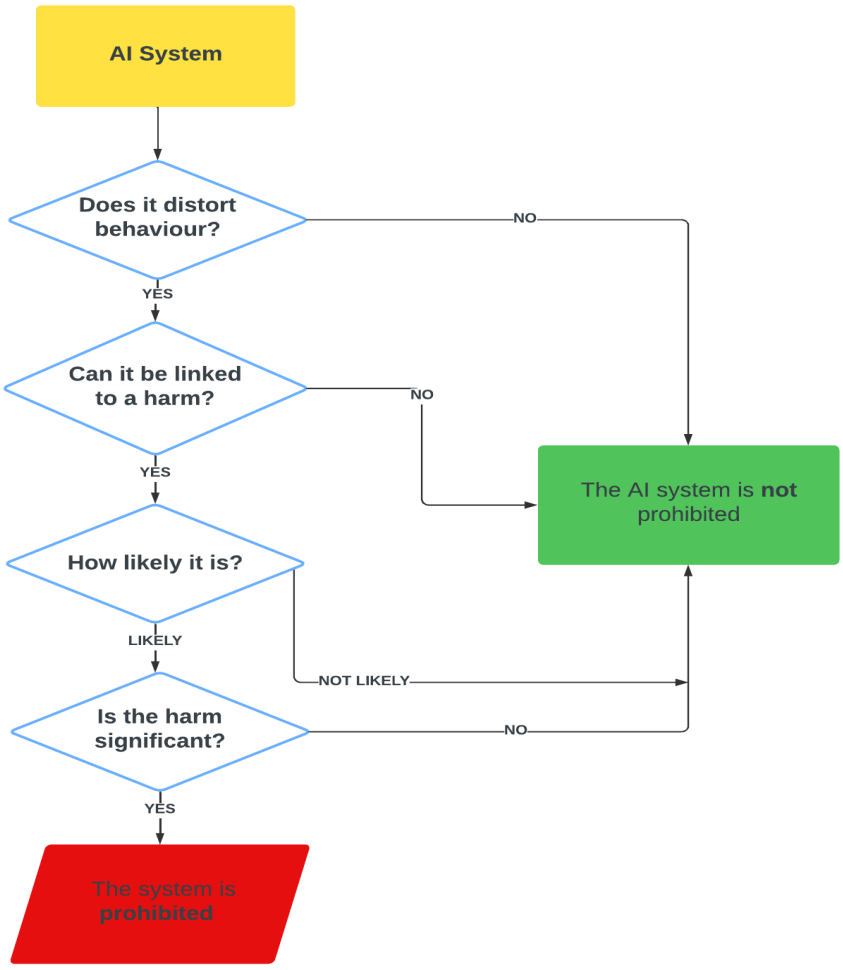
Some of the systems that could be considered under these categories are AI systems for voter data analysis and predictive analytics to perform microtargeting (e.g. Cambridge Analytica), ad delivery systems, recommender systems and chatbots manipulating an individual or even more abstract dystopian systems that authoritarian governments might put in place in the future to manipulate voters.

---

[2] **Article 5.1a**: *"subliminal techniques beyond a person's consciousness or purposefully manipulative or deceptive techniques, with the objective to or the effect of materially distorting a person's or a group of persons' behaviour by appreciably impairing the person's ability to make an informed decision, thereby causing the person to take a decision that that person would not have otherwise taken | **in a manner that causes or is likely to cause that person, another person or group of persons significant harm**;"*

[3] **Article 5.1b**: *"the placing on the market, putting into service or use of an AI system that exploits any of the vulnerabilities of a person or a specific group of persons due to their age, disability or a specific social or economic situation, with the objective to or the effect of materially distorting the behaviour of that person or a person pertaining to that group | **in a manner that causes or is reasonably likely to cause that person or another person significant harm**;"*

[4] **Article 5.1g**: *"the placing on the market or putting into service for this specific purpose, or use of biometric categorisation systems that c**ategorise individually natural persons based on their biometric data to deduce or infer their** race, **political opinions**, trade union membership, religious or philosophical beliefs, sex life or sexual orientation. This prohibition does not cover any labelling or filtering of lawfully acquired biometric datasets, such as images, based on biometric data or categorising of biometric data in the area of law enforcement;"*



While it is not impossible to conceive that some of these applications might fit the definition in the first part of each article, the **main problem is posed** by the requirement that these systems pose a **'significant harm'** to **'the person', 'another person'** or **'a group of persons'**.

It is **legally very difficult to prove the existence of significant harm in the context of elections, to measure it as 'significant' and to demonstrate how likely it is that a certain AI system causes a certain harm** (see flowchart above). The harm for the individual could be the changing of voting behaviours (including not voting); and the harm for 'a group of persons' could be a societal harm such as election unrest (e.g. Capitol Hill); consequences on election results (e.g. Cambridge Analytica scandal); or more broadly the instalment of an authoritarian regime. It is however

questionable whether these impacts could be considered as a 'harm' and doubts remain on how to measure them, on how to demonstrate the link between the AI system and the harm, and on how likely the harm is to occur.

The table below summarises how the reasoning just outlined could be applied to different AI systems to identify whether they fall into the category or not - and in most cases they seem to hardly fit the category.

Possible applications	Fits the first part of the definition	Can it be linked to a harm?	How likely is it?	Is the harm 'significant'?
Dystopian government system to manipulate voters	Yes	Yes, societal harm, potentially also financial.	Not very because these systems have not been documented as of now.	Societal could be significant per se. Financial can be measured.
AI systems for voter data analysis and predictive analytics to perform microtargeting (e.g. Cambridge Analytica)	Yes	Yes, societal harm, potentially also financial.	Likely because there are precedents (even though GDPR should already avoid this happening).	Societal could be significant per se. Financial should be measured.
Chatbots & virtual assistants	Yes	Yes, financial harm if the bot extorts money for campaigns or political parties.	Not really likely at this stage, no real precedents in the political context.	Financial harm should be measured.
Recommender systems	Yes	Ideally yes, but the link is unclear / indirect.	Likely (it exists already).	Difficult to measure, more evidence needed.
Ad delivery systems	Yes	Ideally yes, but the link is unclear / indirect.	Likely (it exists already).	Difficult to measure, more evidence needed.

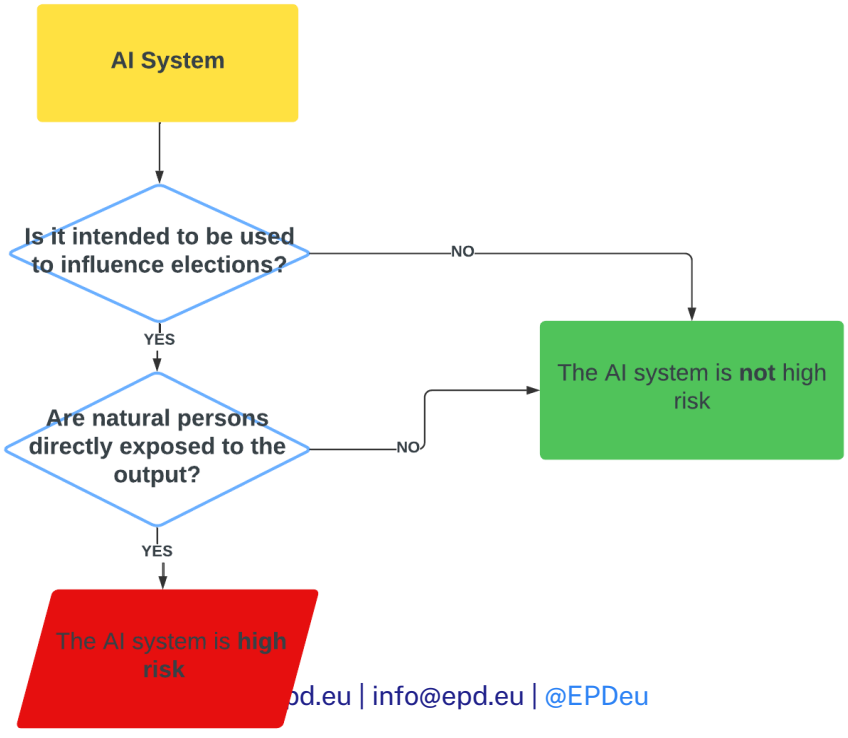
Which AI systems are (not) high-risk

While the prohibited systems category seems to have little use in protecting elections from specific AI systems, the high-risk category seems to provide a much better hook, because Annex III 8b mentions explicitly AI systems linked to elections, more specifically:

> *"AI systems **intended** to be used for influencing the outcome of an election or referendum or the voting behaviour of natural persons in the exercise of their vote in elections or referenda. This does not include 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 logistic point of view."*

Once again, however, the **threshold to prove is very high and in many cases the combination of the two requirements would exclude AI systems**, even ones that are very closely related to elections.

Some AI systems that might be considered are for example: AI systems used to deliver political advertising, profile voters; including with microtargeting and amplification techniques; AI systems used to process or count voting ballots or maintain voting lists; AI systems used to identify cybersecurity attacks against IT systems allowing elections to take place; Chatbot-based AI systems to provide voter assistance; AI systems to perform voter data analysis and predictive analytics; AI systems used to counter biased content and for electoral content moderation.



The **most difficult element to prove in the wording of Annex III 8b is the *intentionality*** ('AI systems *intended* to be used for influencing the outcome of an election [...]'), which represents quite an important loophole, as many AI systems are not originally intended to be used in this way. The 'organisational' criteria provides yet **another loophole to exclude specific kinds of applications**.

The table below summarises how the reasoning just outlined could be applied to different AI systems to identify whether they fall into the category or not - and once again in most cases they seem to hardly fit the category.

Possible applications	Is it 'intended' to be used to influence elections?	Are natural persons directly exposed to the output?
Ad delivery systems	Not necessarily, as the technology per se is neutral and could be used for other kinds of ads too, not just political ones.	Yes (the ad).
Microtargeting of political ads	Yes	Not necessarily, as targeted mechanisms do not always have a direct access to the distribution mechanism.
AI systems to perform voter data analysis and predictive analytics ('Cambridge Analytica style systems')	Typically yes - but in some cases like research of journalistic information, not necessarily.	Yes when that leads to a political campaign.
AI systems used to process or count voting ballots or maintain voting lists	No - it's just intended to record the results of the elections, but not to influence them. It might influence them (people can change vote behaviour based on the results), but that is not the 'intention'.	Yes, voting results impact natural persons directly.
AI systems used to identify cybersecurity attacks against IT systems allowing elections to take place	No - it is intended to protect IT systems used for elections. Most likely the technology is neutral.	Yes in a sense because they allow elections to take place - but still an indirect link.

Chatbot-based AI systems to provide voter assistance	No - it is intended to provide information.	Yes, because they interact with the chatbot.
AI systems used to counter biased content and for electoral content moderation	Not necessarily, as once again the technology per se could be a neutral one, also used to moderate other kinds of content.	Yes, the social media feed that is displayed to them.

Conclusions and recommendations

Based on the above analysis, the **AI Act does not seem protective enough when it comes to election integrity, as it is very difficult to include any sort of AI application related to elections in either the prohibited or the high-risk category**. Even some AI systems that would naturally be included (e.g. microtargeting of political ads) hardly fit in the very narrow definitions provided by the AI Act.

To enhance the effectiveness of the Regulation in this area, we would recommend a broader interpretation of different concepts, such as the one on 'significant harm' for the prohibited systems and 'intentionality' for the high-risk systems. We also believe that it will be important to evaluate the different AI systems *ex-post* based on concrete cases and incidents and based on the methodology questions outlined above.

The link between these tools and existing ones is also unclear and **possible infringements of these provisions would conceptually and more easily already be caught under the General Data Protection Regulation, Digital Services Act,** or the **Regulation on Transparency and Targeting of Political Advertising** (once it is in place).

Some of these recommendations could be taken into account throughout the implementation, in particular with the **Guidelines on high-risk and non-high-risk use cases** on AI systems under Article 6.5 and the **Guidelines on prohibited practices** referred to in Article 5, according to Article 96.1(b).

Key recommendations

- ➔ Provide examples of unclear concepts such as 'significant harm' under Articles 5.1a and 5.1b and 'intentionality' under Annex III 8b and possibly use a broad understanding of the concept based on due diligence.
- ➔ Evaluate AI systems ex-post based on concrete cases and incidents based on methodology questions outlined above. Civil society should also contribute here by keeping an eye on the most recent developments and flagging them to the Commission during the implementation to inform the drafting of the guidelines.
- ➔ Clarify the link between the AI Act provisions with DSA and GDPR and the potential added value of the AI Act.