

## / Policy Brief

# Addition to the AI Act: Expansion of Article 5 on „Prohibited AI Practices“

April 2026

## Summary

We welcome the planned addition of a ban on non-consensual sexualized deepfakes to Article 5 of the AI Act. The addition would be an important step towards protecting victims of digital sexualized violence. The evidence is clear: such content overwhelmingly affects women, constitutes a form of gender-based violence, and has a “silencing effect” that limits democratic participation. However, such a ban must be carefully designed to ensure that it does not apply to legitimate uses and that liability is clearly assigned.

An overview of our demands:

- **Precisely define consent:** Free, informed, context-specific, and explicit; no inclusion of non-existent persons.
- **Clearly specify liability:** Only include AI systems that lack adequate safeguards.
- **Avoid unintended side effects:** Do not jeopardize open-source AI development through overregulation.
- **Strengthen security measures:** Reasonable safeguards with continuous monitoring and reporting mechanisms.
- **Consent request:** Mandatory for the depiction of real persons.

## Context, Evidence, and Demands

The EU's Artificial Intelligence Act (AI Act) is currently being revised (the so-called Digital Omnibus on AI). The proposals by the [European Parliament](#) (EP) and the [Council of the European Union](#) (Council) call for an addition to Article 5 of the AI Act, which defines prohibited practices. The creation of non-consensual sexualized deepfakes is to be included in this section as an additional application.

We strongly support this addition. Existing regulation, such as the Digital Services Act (DSA), primarily address how such content is disseminated on platforms. National initiatives, such as the law against digital violence currently being developed in Germany, aim to update criminal law by introducing new offenses. The landscape of national regulation specific to the context of digital sexualized violence still varies widely across EU Member States. An addition to the AI Act would close an important regulatory gap by specifically holding providers and deployers of AI systems accountable, thereby serving as another crucial building block in protecting victims of digital sexual violence.

### Problem

With the increasing prevalence of AI, particularly general-purpose AI (GPAI), it is becoming significantly easier to create or manipulate images, videos, and audio recordings. Of particular concern are AI models and tools based on them that make it easier to create non-consensual sexualized deepfakes. The spectrum ranges from applications that explicitly advertise their ability to sexualize images of real people (usually without verifying consent) to general-purpose software, such as "face swapping" apps, which can be misused for such purposes.

Most large-scale GPAI systems (such as ChatGPT, Claude, etc.) already have safeguards in place to prevent such content from being generated. Open-source AI can also include similar safeguards. However, it is virtually impossible to guarantee with absolute certainty that security measures cannot be circumvented. Users discuss new ways to bypass security measures, as can be read in relevant forums. Holding a provider or deployer liable for every conceivable use of its product therefore poses significant problems.

### Demands

Any additional ban in the AI Act must be drafted in such a way that it does not apply to legitimate methods of creating or editing depictions of nudity, i.e. when no real persons are involved or when the depictions were created consensually.

Nevertheless, there are technologies that clearly make it easier to create non-consensual sexualized deepfakes – and it is precisely their providers and deployers who must be held accountable. Since such accountability also entails corresponding penalties, a distinction must be made here: The law should not ban all AI systems used for image generation, but rather those that allow such content to be created without *reasonable* safeguards. Security measures should be as stringent as possible and must be regularly reviewed and tested by providers or deployers.

While temporarily bypassing security measures can indeed cause significant harm to those affected, liability should be limited to providers and deployers of AI systems who have failed to implement adequate safeguards. Systems that can only be bypassed with considerable effort should not be banned across the board, since it is impossible to completely rule out the possibility of disabling usage restrictions (jailbreaking).

If a third party uses an AI model as a basis for developing its own applications and, in doing so, removes or circumvents existing security measures, liability should not rest with the provider of the underlying model, but rather with those who removed the protective measures and misused the model. In such cases, it should be the newly created system that is impacted by Article 5.

In the following, we highlight the aspects that we consider particularly relevant – taking into account the wording already partially included in the proposals from the European Parliament and the Council:

- **Definition of consent:** A precise definition of consent is essential in the context of the creation of sexualized depictions. Consent must be given freely, in context, informed, unambiguously, and explicitly, so that consensual content can be clearly distinguished from non-consensual content. It should be made clear that a ban should generally not cover the creation of depictions of non-real persons.
- **Clarification of liability:** It must be clearly defined under what circumstances AI systems for image, video, and audio editing fall under the ban in the AI Act. The regulation should only apply to AI systems that facilitate the creation of non-consensual sexualized deepfakes without implementing adequate safeguards, and that neither implement adequate safeguards nor take corrective action in the event of misuse. The determining factors in this regard should be the system's functionalities, potential applications, objectives, and training data.
- **Counteracting unintended side effects:** A ban under Article 5 of the AI Act could go beyond the regulatory objective, particularly in the area of open-source AI: Providers and deployers who implement appropriate security and corrective measures and whose systems do not explicitly facilitate the creation of non-consensual sexualized deepfakes should not be held liable for misuse by third parties or in the case of jailbreak. Therefore, as outlined in the previous point, a clear allocation of liability is essential. In addition, consideration should be given to what additional measures would be appropriate to protect (open-source) providers and deployers who take reasonable precautions from the consequences of misuse of their systems.
- **Implementing safety measures:** AI systems and GPAI models that can be used to generate images, videos, and audio recordings must ensure, both at the level of their underlying models and through input options (prompting), that the generation of non-consensual sexualized content is prevented. Security measures must be continuously reviewed and, where necessary, expanded to make them harder to circumvent. Users of such systems must also be given the opportunity to report potential security vulnerabilities to the providers or deployers.
- **Mandatory consent requirement for depictions of real persons:** Apps and general-purpose AI (GPAI) that enable the editing of images, videos, and audio recordings must obtain the consent of the individuals depicted when creating content that shows real persons, except in areas covered by artistic freedom and satire. Furthermore, it should be transparent which AI models underlie the respective applications.

## Our organization

**AlgorithmWatch** is a human rights organisation based in Berlin and Zurich. Evaluating the social impact of algorithmic decision-making (ADM) and AI-based systems, we're dedicated to ensuring that these systems are used to strengthen human rights, democracy, and the rule of law. All too often, they restrict people's rights and are not deployed in their interest. To change this, we run campaigns, publish journalistic investigations, and scientifically analyse algorithmic systems and how they are overseen and controlled.