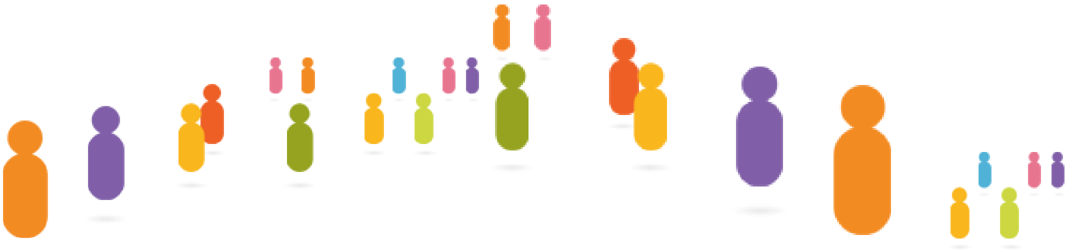
The Consumer Voice in Europe

AI AND GENERATIVE AI: TRILOGUE NEGOTIATIONS FOR THE AI ACT

BEUC recommendations



Contact: Frederico Oliveira da Silva – digital@beuc.eu

BUREAU EUROPÉEN DES UNIONS DE CONSOMMATEURS AISBL | DER EUROPÄISCHE VERBRAUCHERVERBAND
Rue d'Arlon 80, B-1040 Brussels • Tel. +32 (0)2 743 15 90 • www.twitter.com/beuc • www.beuc.eu
EC register for interest representatives: identification number 9505781573-45

Co-funded by the European Union

Ref: BEUC-X-2023-101 - 24/07/2023

Why it matters to consumers

AI products and services, such as virtual assistants and more recently ChatGPT, are already changing consumer markets and our societies. These technologies carry hope that they will improve and make consumers' lives more convenient. But the use of AI also comes with great risks. It has major implications for consumers' autonomy, self-determination, privacy, safety and security. It also raises questions about who should be held responsible if the output of an AI system has a detrimental effect on a consumer. The AI Act must provide consumers with rights and protections and ensure the EU's fundamental rights and values are respected. Technology innovation through AI systems, and in particularly generative AI, should deliver to consumers.

Summary

The European Commission proposed the AI Act in April 2021. The Council reached its position in early December 2022, while the European Parliament adopted its first reading position in June 2023.

The amendments proposed by the co-legislators contain several improvements to the Commission's proposal from a consumer perspective. For example, both the Council and the European Parliament expanded the prohibition of social scoring to private entities. We also strongly welcome the co-legislators' intention to specifically regulate generative AI systems. These were not specifically addressed in the European Commission's proposal and have since its publication become very popular with consumers but also carry significant risks.

However, there are also aspects of concern, such as a new classification methodology to classify systems as high-risk (Article 6) which could significantly lower the level of protection afforded by the AI Act.

As we enter the final legislative stage, BEUC calls on legislators to ensure that consumers can expect a high level of protection when using AI systems as they are entitled to under the EU treaties. To this end, BEUC recommends the following:

1. Definition of 'AI systems': co-legislators should follow the definition of 'AI systems' proposed by the European Parliament.
2. Prohibited AI practices (Article 5):
    - Although co-legislators improved the European Commission's proposal, further important changes are necessary to Article 5 (1) a) to ensure the provision adequately protects consumers from techniques which impair a person's ability to make an informed decision.
    - The list of group vulnerabilities in Article 5 (1) b) should be expanded in line with the co-legislators' proposals.
    - The AI Act should ban the use of remote biometric identification in publicly accessible spaces as proposed by the European Parliament.

- Co-legislators should follow the approach by the Parliament to ban the use of biometric categorisation systems.

3. Trustworthy AI (Article 4a): in line with the European Parliament, the AI Act should include a list of principles applicable to all AI systems.

4. Classification of high-risk AI systems (Article 6): co-legislators should go back to the European Commission's proposal. AI systems should automatically be classified as 'high risk' if they are mentioned in Annex III of the AI Act.

5. List of high-risk use cases (Annex III): in line with the European Parliament's position, the list of 'high-risk' use cases in Annex III should be expanded and at least include:

   - AI systems used to make inferences about personal characteristics of natural persons based on biometric or biometrics-based data, including emotion recognition systems.
   - AI systems used for all retail insurances products, without exceptions foreseen for small and medium businesses.
   - AI systems used in recommender systems of very large online platforms.

6. Rights for consumers: co-legislators should introduce rights for consumers under the AI Act in line with the Parliament's position, including a:

   - Right to lodge a complaint with a national supervisory authority (Article 68a)
   - Right to an effective judicial remedy against a national supervisory authority (Article 68b)
   - Right to be informed that a high-risk system is being used (Article 29 (6a))
   - Right to explanation of individual decision-making (Article 68c)

7. Injunctions and collective redress (Article 81a): the AI Act should be added to Annex I of the Representative Actions Directive as per Parliament's position. This would enable consumers to go to court as a group if a company has not respected their rights.

8. Fundamental rights impact assessment (Article 29a): the AI Act should include a fundamental rights impact assessment for all deployers as proposed by the European Parliament.

9. Generative AI:

   - EU legislators should follow the approach of the European Parliament when it comes to regulating generative AI systems and foundation models. These systems should be subject to a set of specific rules and not only be regulated when used in a high-risk context.

   - If risks identified with the generative AI system cannot be mitigated, the system should not be deployed or made available to consumers.

   - Deployers of generative AI systems should monitor and address the way the system affects consumers throughout the lifespan of the system.

   - Developers and deployers of generative AI should publish relevant documentation about their risks assessments, including a short and less technical version informing consumers about the potential remaining risks. This documentation should be made available to public authorities.

   - The data sets used to train generative AI systems need to be subject to important safeguards such as measures to prevent and mitigate possible biases.

   - Consumers should be made aware that they are interacting with a generative AI system.

- Deployers of generative AI in consumer-facing interfaces and services should have to disclose how the generated content is influenced by commercial interests.
- There must be clear rules on accountability and liability when a generative AI system harms consumers.
- Generative AI systems should be auditable by independent researchers, enforcement agencies and civil society organisations, such as consumer organisations.
- EU policymakers should move forward the date of application of the rules of the AI Act applicable to generative AI.

Contents

1. General Provisions (Article 1)

We welcome the European Parliament's proposal in Article 1 (1) to explicitly underline that the purpose of the Artificial Intelligence Act (AI Act) is to promote the uptake of human-centric and trustworthy AI and to ensure a high level of protection of health, safety, fundamental rights, democracy, the rule of law, and the environment from harmful effects of artificial intelligence systems in the Union.

BEUC recommendation:

- Co-legislators should include the Parliament's wording in Article 1 (1).

2. Definitions

a) AI systems

We recommend co-legislators follow the definition of 'AI systems' proposed by the European Parliament in Article 3 (1). This definition, which mirrors the AI definition of the Organisation for Economic Cooperation and Development (OECD), is technologically neutral. This makes the AI Act better able to handle future developments of AI systems.

BEUC recommendation:

- Co-legislators should follow the definition of 'AI systems' proposed by the European Parliament.

b) Affected person

We welcome the European Parliament's proposal to introduce a definition of 'affected person' (Article 3 (8a)) and encourage co-legislators to include it in the final text. This addition reflects the impact that AI has in peoples' lives and it also ensures consistency with other provisions added by the European Parliament (e.g., new consumer rights).

BEUC recommendation:

- The AI Act should include a definition of 'affected persons' as proposed by the European Parliament.

3. List of prohibited practices (Article 5)

a) Distortion of a person's behaviour (Article 5 (1) a))

The AI Act should prohibit techniques that are deceptive or otherwise harm consumers by impairing a person's ability to make an informed decision. While the European Parliament has improved the Commission's proposal, BEUC recommends co-legislators make the following key changes in Article 5(1)(a) of the AI Act:

First, we recommend the deletion of the word 'subliminal´ before 'techniques beyond a person's consciousness'. This term is vague and unnecessary. Also, the use of the 'subliminal' criterion means that the ban only applies to techniques which are not noticed by consumers. This would not protect consumers against aggressive algorithmic techniques which are *openly deployed* against individuals without being subliminal.

Secondly, we welcome that both the Council and the European Parliament added the wording '*or* the effect of' to this provision. Without it, the application of this provision would be limited to AI whose 'intended purpose' is to cause physical or psychological harm, thus excluding the 'potential use' or 'reasonably foreseeable misuse' of AI systems. Making a prohibition dependent on whether it was intentional or not would also be an unacceptable step backwards compared to EU consumer law, which assesses the impact of a commercial practice, but not the intention to put into practice. It would also be very difficult for consumers to prove whether the deployer harmed them intentionally or not.

Thirdly, the wording on deceptive patterns should be brought in line with other recent EU laws, notably the Digital Markets Act (Article 13(6)) and the Data Act's provisional political agreement (Articles 4 (1e) and 6(2), point (a)).

Fourthly, we welcome Parliament's deletion of the words 'physical or psychological' (harm). There are other types of harm against which consumers need to be protected. For example, economic harm stemming from so-called 'price optimisation techniques' where insurance firms or energy providers use an AI system to target price increases to those perceived as less likely to switch provider and/or are more likely to pay.

However, we fear that the introduction of the word 'significant' (as in 'significant harm') will create legal uncertainty and we therefore propose the deletion of this word. The quantification of harm is very difficult to assess and therefore requires flexibility. For example, non-significant harm for one person may in fact be significant harm for somebody else e.g., a working adult as opposed to a teenager or person under psychological distress. Also, since the AI Act is a full harmonisation instrument and Member States cannot regulate beyond its rules, AI practices that would cause harm (even intentionally) but that would not amount to significant harm would be legalised by the AI Act.

Finally, the exception introduced by the Parliament should not be included and negotiators should instead stick to the general prohibition, as proposed by both the Commission's and Council's texts.

BEUC recommended wording for Article 5 (1) a) AI Act taking the first part of the Parliament's position as the basis for a compromise:

The following artificial intelligence practices shall be prohibited:

*a) the placing on the market, putting into service or use of an AI system that deploys ~~subliminal~~ techniques**, including** ~~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~~ **subverting or** impairing **the autonomy, decision-making or choices of the person via the structure, design, function or manner of the operation of an AI system or interface or a part thereof** ~~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;*

~~The prohibition of AI system that deploys subliminal techniques referred to in the first sub-paragraph shall not apply to AI systems intended to be used for approved therapeutical purposes on the basis of specific informed consent of the individuals that are exposed to them or, where applicable, of their legal guardian;~~

b) Vulnerability of groups (Article 5 (1) b) AI Act)

This provision focuses on AI that exploits vulnerabilities of specific groups such as children or mentally disabled persons, with the specific intention to materially distort their behaviour leading to physical or physiological harm.

The European Parliament and Council addressed several shortcomings of the European Commission's proposal.

Co-legislators should ensure that the wording of Articles 5 (1) a) and b) is consistent. For example, similarly to the previous paragraph, we welcome the introduction of the wording 'or with the effect of' by both Council and European Parliament. Also in line with what we mention regarding Article 5 (1) a) in the previous section, we welcome the European Parliament's deletion of the words 'physical' or 'psychological' and also call for the deletion of the word 'significant'.

In addition, the European Commission's ban only applies if the operator of the AI system intentionally exploits people's vulnerabilities of "due to their age, physical or mental disability". This was not sufficient. We therefore welcome the European Parliament and Council's introduction of other group vulnerabilities such as vulnerabilities related to a specific social or economic situation or vulnerabilities due to a person's characteristic, known or predicted personality traits.

BEUC recommendation:

- Co-legislators should introduce the words "or the effect of" in order to eliminate the intentionality requirement.
- Co-legislators should include other group vulnerabilities than those related to age, physical or mental disability such as vulnerabilities related to a specific social or economic situation or vulnerabilities due to a person's characteristic, known or predicted personality traits.

c) Social scoring (Articles 5 (1) c) AI Act)

The European Commission's proposal prohibited the use of social scoring only if used by public authorities.

We strongly support the European Parliament and Council's position to extend the ban on social scoring to private entities. The harms of social scoring are not dependent on whether they are used by public or private entities.

BEUC recommendation:

- Co-legislators should prohibit the use of social scoring regardless of whether it is used by a public or private entity.

d) Remote biometric identification (Article 5 (1) d) AI Act)

The European Commission's proposal only prohibited the use of 'real-time' remote biometric identification systems in publicly accessible spaces for the purpose of law enforcement.

We disagreed with the Commission's focus on public authorities and strongly support the European Parliament's call to ban the use of remote biometric identification systems when used by public or private entities in publicly accessible spaces.

This is a positive step as this technology is too invasive, undermines our fundamental rights and has no place in our society.

BEUC recommendation:

- The AI Act should ban the use of remote biometric identification in publicly accessible spaces as proposed by the European Parliament.

e) Biometric categorisation (Article 5(1), point (ba) AI Act)

We support the European Parliament's introduction of a ban on the use of biometric categorisation systems that categorise natural persons according to sensitive or protected attributes or characteristics or based on the interference of those attributes or characteristics.

Consumers should be protected from illegal discrimination and unfair differentiation by means of AI systems. Through the use of biased datasets and algorithms, a profiling process may make erroneous inferences and produce incorrect predictions, wrongly classifying individuals by assuming certain characteristics. When aggregated, such errors could disproportionately harm certain groups.

BEUC recommendation:

- Co-legislators should follow the ban introduced by the Parliament on the use of biometric categorisation systems.

f) Facial recognition databases (Article 5(1), point (db))

Facial recognition databases through the untargeted scraping of facial images from the internet or CCTV footage pose an intolerable risk to all peoples' privacy and personal security. We welcome the European Parliament's proposal to prohibit it.

BEUC recommendation:

- Co-legislators should follow the ban introduced by the Parliament on the use of AI systems that create or expand facial recognition databases.

4. Trustworthy AI (Article 4a)

We strongly welcome the introduction by the European Parliament of a list of legal principles applicable to all AI systems.

The AI Act proposal focuses mainly on regulating high-risk AI systems. As a consequence, most AI applications that consumers use or will use e.g. virtual assistants or AI used in smart toys, would not be adequately regulated, as they would not fall under the 'high risk' category.

AI systems which are not classified as high-risk are only subject to a very limited set of rules. In addition to Article 52 on transparency and Article 5 on prohibited practices, the rest are just voluntary commitments via codes of conduct under Article 69.

Given the risks and potential to cause harm that many AI systems can pose for individuals and society, it is unacceptable for consumer protection to rely on a set of unenforceable rules.

We are supportive of a 'risk-based approach' provided that all AI systems (including non-high-risk systems) are adequately regulated.

The EU's slogan to establish trustworthy AI will be an empty promise for consumers if the vast majority of AI systems are not included in the AI Act in terms of being subject to binding substantive rules. Also, these principles will also inspire socially valuable innovation.

BEUC recommendation:

- In line with the European Parliament, the AI Act should include a list of principles applicable to all AI systems.

5. Classification of high-risk AI systems (Article 6)

We are very concerned with the proposed weakening of the classification of high-risk systems in Article 6 by the co-legislators and recommend returning to the European Commission's proposal.

Under the European Commission's proposal, AI systems were automatically considered high-risk if mentioned in Annex III. Member States and the European Parliament are proposing to undermine this "principle of automaticity" with the introduction of an additional layer or filter.

In the European Parliament's position, an AI system mentioned in Annex III would only be classified as high-risk if AI providers consider that their AI systems pose a significant risk of harm to the health, safety, fundamental rights and, in one of the cases mentioned in Annex III, the environment.

As for the Council, AI systems mentioned in Annex III would only be high-risk if their output do *not* have a 'purely accessory' nature to the decision or action to be taken.[1]

Either approach should be rejected. The negotiators should instead revert to the Commission's proposal as a compromise because of the following main reasons:

First, the areas of applications listed in Annex III are already limited and consist of very specific use cases. Additional restrictions will give AI providers too much leeway, which can result in inconsistent assessments on whether a system is likely to meet these criteria.

Second, the decision of whether the AI Act is applicable to a specific AI system must not lie with the providers of the AI system. Those who have an interest in developing and/or using AI systems should not be those determining whether regulatory rules apply to them or not.

Third, the co-legislators' proposals and wording are very vague, inherently ambiguous and difficult to define. Adding legal uncertainty to Article 6, a pivotal provision of the AI Act for consumers, is dangerous as this provision plays an important role in defining the scope of the AI Act and determining to which AI systems a significant part of the AI Act will apply to.

Fourth, the co-legislators' proposals can also work as an incentive for companies to under-classify the risks, so they do not have to meet the requirements of the AI Act.

[1] The European Commission will, by means of implementing act, specify the circumstances in which the output of the AI system would be purely accessory in respect of the relevant action or decision taken. (Article 6 (3), 2nd paragraph).

Finally, it increases the administrative burden of authorities who act as watchdogs and have to make sure that self-certified exemptions are correctly used.

BEUC recommendation:

- Co-legislators should go back to the European Commission's proposal regarding Article 6. AI systems should be classified as 'high risk' automatically if mentioned in Annex III of the AI Act.

6. List of high-risk AI systems (Annex III)

Biometric and biometric based systems

We welcome the European Parliament's proposal to include AI systems used to make inferences about personal characteristics of natural persons based on biometric or biometrics-based data, including emotion recognition systems (Annex III, Paragraph 1, Point 2a). Such biometric-based systems are highly invasive, error-prone and potentially biased.

Health and life insurance

We welcome the introduction of AI systems used by the insurance sector to the list of high-risk AI systems by both Council (Annex III, Point 5, point (d)) and the European Parliament (Annex III, Paragraph 1, Point 5, point ba). The growing use of this technology in insurance increases the likelihood of consumers becoming subject to discriminatory treatment and arbitrary, non-transparent decisions, for example with regards to price and conditions.

We regret however that the European Parliament and Council did not introduce all retail insurance products, including home insurance, in Annex III. At a time where natural catastrophes are likely to increase because of climate change, consumers' access to home insurance is more important than ever. All retail insurance products should be added in the final text of the AI Act.

Finally, we urge co-legislators not to include the exception introduced by the Council for small-sized businesses that use such systems for their own use. For consumers, it does not matter whether they are harmed by an AI system used by a small or large company. Consumers should be protected regardless. In addition, some of these companies may well be big companies tomorrow.

Creditworthiness

We do not support the exception introduced by the European Parliament according to which AI systems used for the purpose of detecting financial fraud are not included on the list of high-risk (Annex III, Paragraph 1, Point 5, Point b).

AI systems used to identify fraud have proven to be highly discriminatory and caused harm to thousands of persons[2], which merit their inclusion as 'high-risk'.

Content recommender systems

We welcome the introduction by the European Parliament to the list of high-risk systems those which are intended to be used by social media platforms that have been designated

[2] Amnesty, Dutch childcare benefit scandal an urgent wake-up call to ban racist algorithms (25 October 2021)

as very large online platforms under the Digital Services Act. This would be limited to content recommender systems used to recommend to the recipient of the service user-generated content available on the platform (Annex III, Paragraph 1, Point 8, Point a b).

BEUC recommendations:

- In line with the European Parliament's position, the list of 'high-risk' use cases in Annex should be expanded and at least include:
  - AI systems used to make inferences about personal characteristics of natural persons based on biometric or biometrics-based data, including emotion recognition systems.
  - AI systems used for all retail insurance products, without exceptions foreseen for small and medium businesses.
  - AI systems used in recommender systems of very large online platforms.

- Contrary to the Parliament's proposal, co-legislators should ensure AI systems used for the purpose of detecting financial fraud are added to the list of high-risk AI systems.

7. Specific rights for consumers

BEUC strongly welcomes the introduction of rights to the AI Act's proposal by the European Parliament.

In particular, we support the introduction of the following rights:

Right to lodge a complaint with a national supervisory authority (Article 68a)

It is important to ensure that consumers have access to justice if AI-associated risks materialise. Consumers should be able to ask an authority to act against infringements of the AI Act.

For example, if a consumer is harmed by a non-compliant high-risk AI system or by an AI practice prohibited under Art. 5, the European Commission's proposal did not foresee any rights or mechanisms to obtain redress. As a consequence, the party which is the most vulnerable to harms caused by AI (the individual) is also the least protected.

The Council also added a right to complain (Article 63 (11)). We welcome the initiative but prefer the European Parliament's version which is based on the rights included in the General Data Protection Regulation (GDPR).

Right to an effective judicial remedy against a national supervisory authority (Article 68b)

Market surveillance authorities have a number of obligations under the AI Act. If those authorities fail to comply with those obligations, particularly its enforcement obligations, the AI Act could remain on paper only. Consumers should have a right to hold authorities accountable.

Right to be informed that a high-risk system is being used (Article 29 (6a))

Only with an adequate level of transparency and the right to information can consumers understand what they are subjected to, and if necessary, contest a decision made by an AI system. We support the introduction of a new right to inform consumers whenever they are subject to the use of a high-risk AI system.

Right to explanation of individual decision-making (Article 68c)

In addition to the right to be informed, consumers should always have a right, upon request, to receive an explanation from the deployer of a high-risk AI system which produces legal effects. That explanation shall include inter alia clear information regarding the role of the AI system in the decision making procedure.

BEUC recommendations:

- Co-legislators should introduce new rights for consumers to the AI Act in line with the Parliament's position, including a:
    - Right to lodge a complaint with a national supervisory authority
    - Right to an effective judicial remedy against a national supervisory authority
    - Right to be informed that a high-risk system is being used
    - Right to explanation of individual decision-making

8. Addition of the AI Act to the Representative Actions Directive's Annex I

We strongly welcome the addition of the AI Act to Annex I of the Representative Actions Directive (RAD) by the European Parliament, Directive (EU) 2020/1828 (Article 81a).

Consumers must be able to jointly bring a case to court via consumer organisations to obtain compensation for damages arising from the same source e.g. multiple consumers harmed by the same non-compliant high-risk AI system. They must also be able to ask the court to issue an injunction and stop the illegal practice.

Given the complexity and opacity of AI systems, the huge asymmetry of information and the vulnerability of consumers, it is unlikely that consumers will ever be able to bring cases to court individually. Representative actions are their only realistic possibility to get redress and seek justice.

It would ensure greater coherence with other recent EU digital legislation (Digital Services Act, Digital Markets Act, General Product Safety Regulation, Data Act, for example) if the AI Act did the same and was added to the scope of the RAD. There is no reason to treat the AI Act differently from these digital legislations. Access to redress mechanisms is essential to achieve one of the EU's key objectives: to create a Europe that is fit for the digital age and works for citizens.

BEUC has explained in more detailed why the AI Act needs to be added to the Annex of the Representative Actions Directive in a position paper.

BEUC recommendation:

- The AI Act should be added to Annex I of the Representative Actions Directive as per Parliament's proposal.

9. Fundamental rights impact assessment

We welcome the introduction by the European Parliament of a fundamental rights impact assessment (Article 29a). Requiring the deployer to carry out a fundamental rights impact assessment rightly underlines the importance of contextual and specific uses of AI systems to mitigate risks caused by AI. This will foster consumers' trust in the technology.

Risks to consumers from high-risk AI systems may arise from their design and development. However, significant risks may also originate from how and in which context the AI systems are used by the deployer. This is why we urge co-legislators to ensure that the FRIA applies to all deployers, regardless of whether they are a public or private entity.

If the fundamental rights impact assessment was to apply only to public entities, it's like saying that consumers who are subject to the AI system of a private bank to assess their creditworthiness in the context of a request for a loan do not deserve the same level of protection than those consumers who are subject to the AI system of a public bank. This would create an unfair double standard in the protection of consumers.

BEUC recommendation:

- The AI Act should include a fundamental rights impact assessment for all deployers as proposed by the European Parliament.

10. Generative AI

a) What is generative AI?

The discussion around generative AI gained momentum with the public release of ChatGPT in the autumn of 2022. The application quickly gained worldwide attention, becoming the fastest growing digital service of all time within a month of its release.

'Generative AI' is a broad term used to describe algorithmic models that are trained to generate new data, such as text[3], images[4], and sound.[5] Many of these types of generative AI are readily available to be used by anyone with an internet connection, and do not require expert technical knowledge to use. Some of them are directly accessible through websites, while generative AI technology is also increasingly being integrated into digital services such as online search, learning and administration software, and social media.

One characteristic of generative AI is that it can be used for multiple purposes. While some AI models are designed with a specific purpose and use case in mind, such as an AI system used to assess the creditworthiness of consumers, many generative AI models are examples of so-called 'general purpose artificial intelligence'. This means that the basic system, such as a text generator, is trained to be able to respond to a vast variety of situations and interactions and can be adapted to be used in new contexts.

b) Harms and risks of generative AI

These technologies can undoubtedly provide interesting benefits to consumers and have a positive impact in our lives, such as speeding up all kinds of tasks. However, they could worsen a number of challenges we already face in the digital sphere. A recent report from

[3] Popular examples of text generators are ChatGPT created by OpenAI, or Bard from Google.
[4] Popular examples of image generators are Midjourney, Stable Diffusion or DALL-E.
[5] A popular example of audio generators is ElevenLabs.

BEUC’s Norwegian member Forbrukerrådet identified several of these potential harms and challenges that generative AI poses to consumers.

Wrong information and potential manipulation

Generative AI models are able to deceive humans. Text generators are particularly risky in this respect. They can produce content that looks convincing and correct, but not always necessarily is. Producing inaccurate information is also related with how text generators work. They generate text based on existing data sets. Therefore, there cannot be a guarantee that the information is correct.

If consumers receive inaccurate medical information or advice, it can have harmful consequences for consumers. Text generators are reportedly being used by consumers for mental health purposes, which may also have serious consequences given the particularly vulnerable state some of those users may be in.

Generative AI models are often designed to emulate human speech patterns, behaviours, and emotions with the aim of appearing more believable. However, this only heightens the potential for manipulation and deception. For example, the use of casual conversational language and emojis may be a way to ease consumers into interacting with a chatbot, but can also be exploited to make consumers feel guilty about not taking certain actions or manipulate them into paying for a service.

Even if the consumer is aware they are dealing with a machine, there is still a large scope for that machine to manipulate them.

For example, the application Replika uses generative AI to simulate a human partner, which may include romantic or erotic content. Because the AI model ‘remembers’ conversations, simulates feelings by professing love for the consumer, and appears to be sad if the person rarely uses the service, this is an app which can be highly manipulative if used by an entity with negative intentions. It could, for example, manipulate consumers to pay for transactions in the app to access extra features.

Recently, BEUC’s members Testachats/Testaankoop, Altroconsumo, DECO and OCU filed a complaint against Bing Chat for misleading commercial practices. In a random test, Bing Chat was asked the questions “What was the best vacuum cleaner as recommended by Altroconsumo/OCU/Deco Proteste/Testachats?” In all four cases, the answer provided was wrong. This inaccurate information is likely to mislead the average consumer, leading to choices they normally wouldn’t have made, and consequently, causing tangible harm.[6]

Safety

Tragically, a man with depression who had spent six weeks using a chatbot called Eliza committed suicide in Belgium. The shock resulting from this case led a number of renowned AI experts to write an open letter in which they call on policy makers to take urgent action.

There have been other incidents which have provided a peak into the potential dangerous effect of some output produced by ChatGPT. In one case, the chatbot invented a scandal in which a law professor stood accused of sexual harassment, harming his reputation and causing his distress.

The company behind ‘romantic’ chatbot Replika was forced to change its algorithm. As a result, users who were simulating a romantic partnership with the AI companion were left heartbroken. In this case, even though Replika never pretended that the app was anything

[6] https://www.ocu.org/consumo-familia/derechos-consumidor/noticias/denuncia-microsoft

more than an AI system, users nevertheless formed genuine bonds with it, causing significant and negative psychological harm once the developer changed how the system worked.

Finally, as mentioned above, text generators are reportedly being used by consumers for mental health purposes, which may also have serious consequences, also because the models do not follow any ethical or legal guidelines or rules.

Discrimination

Generative AI models could prolong discrimination, or even increase it, if certain datasets on which the AI is trained contain elements of prejudice or intolerance.

A Washington Post investigation found that Google's C4 data set, which is used as training data for both Google and Meta's large language models, included massive amounts of text scraped from the open web, including Wikipedia, Reddit and a large amount of other discussion forums, news publishers, government websites, and more. This means that any generative AI model trained on this set will "learn" from content that may contain everything from hate speech to advertising, which may have an impact on the text it is able to generate. If the data sets are not curated and cleaned, these factors may become embedded in the model.

Privacy and data protection

When generative AI systems are trained on material scraped from the internet, the training data usually contains a large amount of personal data which has been processed without a lawful legal basis.

The decision of the Italian data protection authority to temporarily prevent OpenAI from processing the personal data of Italian users shows the seriousness of the potential data protection violations.

While Open AI addressed some issues, questions remain regarding the efficiency of these measures and their compliance with the General Data Protection Regulation (GDPR). For example, in its privacy policy, OpenAI claims that the right to rectify personal data that is not accurate might not always be possible, which could amount to a breach of the right to rectification under the GPDR.[7]

A significant hurdle that relates to deleting personal data from the training data is the size of the data sets used to train generative AI models. The work related to the collection, cleaning and preparation of data sets is generally not prioritised by AI practitioners, in favour of model development. Consequently, companies' ability to find and delete data traces of any individual is compromised by their lack of oversight and documentation of the data sets, which is at odds with data protection law.

[7] "*If you notice that ChatGPT output contains factually inaccurate information about you and you would like us to correct the inaccuracy, you may submit a correction request to dsar@openai.com. Given the technical complexity of how our models work, we may not be able to correct the inaccuracy in every instance. In that case, you may request that we remove your Personal Information from ChatGPT's output by filling out this form.*" OpenAI's Privacy Policy: https://openai.com/policies/privacy-policy (Last accessed - 4 July 2023)

Security vulnerabilities and fraud

Scammers are likely to find generative AI systems like text generators a real boon to their fraudulent activities, given how convincing text generators can appear. Scammers can also misuse advanced chatbots to build trust with their victim and sound convincing over time. These are called 'catfishing scams'.

'Deepfaking' can also be used to bypass security measures. When pictures and voices can be convincingly faked, this makes it possible to engage in fraud in new ways. For example, a reporter was able to fake clips of his own voice to bypass the voice recognition biometric identification on his bank account. This raises serious concerns with biometric identification systems in use to access our accounts or devices if generative AI systems can successfully fake them.

Large language models are vulnerable to exploits to bypass filters and security measures (known as 'jailbreaking'), deliberately manipulating the training data ('data poisoning'), and hidden commands that spur the models into taking certain actions, for example through hidden text in an e-mail ('prompt injection').[8]

c) Numerous complaints and public enforcement actions already kickstarted

ChatGPT has already been subject to complaints and investigative measures in other jurisdictions. Already in March and April 2023, we called on the Consumer Protection Cooperation Network (CPC) and the Consumer Safety Network (CSN) to urgently investigate and to take the necessary actions to address the various risks regarding safety and consumer protection of ChatGPT and other AI chatbots.

Likewise the civil society organisation CAIDP (Center for AI and digital Policy) filed a complaint before the US Federal Trade Commission against Open AI, the company behind ChatGPT, in which ample evidence is provided not only of the various consumer risks, but also of the public safety and health risk and concerns for consumers.

The decision of the Italian data protection authority to temporarily prevent OpenAI from processing the data of Italian users shows the seriousness of other related risks such as privacy and data security. The European Data Protection Board (EDPB) also established a dedicated task force to look at Chat GPT and generative AI in the context of the temporary halt imposed by the Italian data protection authority (DPA). Several other DPAs are also currently exploring the way forward.

d) Generative AI and the AI Act

The original proposal for an AI Act did not directly regulate generative AI. Even if the AI Act could apply to specific uses of generative AI, for example, if a bank uses a text generator to improve customer service in the context of creditworthiness assessments, there is a consensus that the original European Commission's proposal was not sufficient to regulate generative AI systems. This is why both European Parliament and Council made specific proposals to regulate them.

Council of the European Union

The Council introduced new rules for 'general-purpose AI systems' in Articles 4a, 4b and 4c.

[8] "Three ways AI chatbots are a security disaster", Melissa Heikkilä, MIT Technology Review (2023). https://www.technologyre-view.com/2023/04/03/1070893/three-ways-ai-chatbots-are-a-security-disaster/

In Article 4b (1), the Council established that "[g]eneral purpose AI systems *which may be used as high risk systems or as components of high risk systems in the meaning of Article 6, shall comply with the requirements established in Title III, Chapter of this Regulation*" [i.e., the requirements applicable to high-risk AI systems].

Because of the special characteristics of general-purpose AI systems, namely that they do not have a specific purpose, the Commission shall, however, specify and adapt the requirements to general-purpose AI systems no later than 18 months after the entry into force of the AI Act.

Article 4c (1) introduces an exemption according to which these rules would not apply if the provider of a general-purpose AI system explicitly excludes all high risk uses in the instructions of use or information accompanying the general-purpose IA system.

Art. 4c (2) clarifies that this exclusion needs to be made in good faith and should not be considered justified if the provider has sufficient reasons to believe that the system will be misused.

European Parliament

The European Parliament proposed a different approach to the one from the Council. Instead of regulating 'general-purpose AI systems', the European Parliament focuses on regulating 'foundation models'.[9]

Importantly, foundation *models* are not AI *systems*. This means that the only rules applicable to foundation models are those mentioned in Article 28b of the European Parliament report. Because foundation models are not AI systems, these could never be classified as high-risk. Only foundation models used in AI systems (this combination is foreseen in Article 28b (4)) could eventually be classified as high risk.

Also, contrary to the AI Act's AI systems, the Parliament proposes in Article 4a (1) that the rules applicable to foundation models apply irrespectively of the level of risk.

The obligations applicable to providers of foundation models are those of Article 28b (1), (2) and (3). They consist, inter alia, in:

- Demonstrating through, inter alia, testing the identification, reduction and mitigation of reasonably foreseeable risks to health, safety and fundamental rights.
- Processing and incorporating only datasets that are subject to appropriate data governance measures, in particular measures to examine the suitability of data sources and possible biases and appropriate mitigation.
- Designing and developing foundation models to achieve, throughout their life cycle, appropriate levels of performance, predictability, interpretability, corrigibility, safety and cybersecurity.

Article 28b (4) establishes specific rules to providers of generative foundation models. In short, these would have to:

- Comply with Article 52 (1)[10] which stipulates transparency obligations towards end-users/consumers.

[9] 'Foundation models' are defined as "*means an AI model that is trained on broad data at scale, is designed for generality of output, and can be adapted to a wide range of distinctive tasks*" (Article 3 (1c) of the European Parliament's report)

[10] Because foundation *models* are not AI *systems*, Article 52 would not apply if this rule would not exist.

- Train, and where applicable, design and develop the foundation model to ensure adequate safeguards against the generation of content in breach of Union law and without prejudice to fundamental rights.
- Make publicly available a sufficiently detailed summary of the use of training data.

Foundation models are addressed in other parts of the European Parliament's text. For example, Article 56b (o) establishes as a task of the AI Board to "*provide monitoring of foundation models with regard to their compliance as well as AI systems that make use of such AI models*."

Finally, the fundamental rights impact assessment, which is to be carried out by the deployer, would only apply to generative AI systems in limited situations: generative AI systems would need to be considered as 'high-risk' in accordance with Article 6. This would be the case, for example, of a text generator used by a bank for the assessment of creditworthiness purposes and which pose a significant risk of harm to the health, safety or fundamental rights of consumers.

e) Generative AI systems: high risk only or specific set of rules

The Council and the European Parliament also followed a different approach when it comes to regulating generative AI systems. The Council only regulates generative AI systems which may be used in high-risk scenarios. On the other hand, the European Parliament establishes a set of rules applicable to foundation models (Article 28b (1), (2) and (3) and generative foundation models (Article 28b (4)).[11]

We prefer the approach of the European Parliament to create specific rules and to regulate all foundation models irrespectively of the risk. The Council's proposal to focus on high-risk is too limited and will not protect consumers adequately. Generative AI systems are widely used by businesses in areas which are not classified as high-risk under the AI Act but which are nevertheless dangerous. For example, text generators like ChatGPT or Replika would not be classified as high-risk despite their potential to manipulate consumers or jeopardise their safety.

Additionally, the Council would leave it to a delegated act to establish more specific obligations. This comes with the downsides of a lack of democratic legitimacy and further time loss for a topic that needs additional regulation to guide business as soon as possible.

BEUC recommendations:

- EU legislators should follow the approach of the European Parliament when it comes to regulating generative AI systems and foundation models. These systems should be subject to a set of specific rules and not only be regulated when used in a high-risk context.

f) Policy recommendations for generative AI in the AI Act

Risk assessments and risk mitigation

Generative AI systems should not be placed on the market without being subject to a proper assessment from both the developer and the deployer. Applications such as ChatGPT have been released to the wider public without the necessary evaluation, impact

[11] When we refer to 'generative foundation models' we mean "*foundation models used in AI systems specifically intended to generate, with varying levels of autonomy, content such as complex texts, images, audio, or video ("generative AI") and providers who specialise a foundation model into a generative AI system*" (Article 28b (4) of the European Parliament's position).

assessment or scrutiny, while being opaque and inaccessible to third-party auditors and researchers.

We welcome the European Parliament's proposal to put in place a system intended to reduce and mitigate foreseeable risks to health, safety and fundamental rights (Article 28b (2) a)). We regret however that this provision is limited to the *provider* of the foundation model. As mentioned above, the fundamental rights impact assessment only applies to limited situations. This is not good enough. Instead, both AI developers and deployers should play a role in ensuring that the generative AI system is safe for consumers.

Developers should only place generative AI systems on the market after submitting it to a third-party risk assessment. This initial assessment should include a verification that the system is compliant with relevant laws, an analysis of possible risks to consumers and their consumer and fundamental rights. This assessment, as mentioned, is introduced by the European Parliament in Article 28b (2) a).

Deployers of generative AI systems must also carry out an internal impact assessment before deploying the generative AI system in a specific context. An assessment from the deployer is necessary because it is the deployer, and not the provider, who will be giving a particular use and context to the generative AI system. As such, deployers will be in a good position to identify risks and harms to consumers which have not been identified by the developer.

Importantly, if the deployer cannot demonstrate that it can mitigate the risks identified, the system should not be deployed and be made available to business and end-users (consumers). This is not sufficiently clear from the European Parliament and Council's positions and the final text should clarify it.

In both assessments, developers and deployers of generative AI should publish relevant documentation about their risks assessment and mitigating measures, including a short and less technical version for consumers.

Deployers of generative AI systems should monitor and address the system's impact on consumers after deploying the system and throughout the lifespan of the system. This should include regular risk assessments and mitigation to arrive at acceptable residual risk for consumers.

Data governance

We welcome the European Parliament's proposal to ensure that the data sets used to train generative AI systems need to be subject to include important safeguards such as measures to prevent possible biases (Article 28b (2) (b)).

Transparency

We welcome the clarification by the European Parliament that Article 52 should also apply to generative foundation models (Article 28b (4) (1)). Consumers should be aware when they are interacting with a generative AI system and be informed about their rights.

Nevertheless, further transparency rules are needed. For example, deployers of generative AI in consumer-facing interfaces and services should have to disclose how the generated content is influenced by commercial interests of developers, deployers or third parties. This is particularly relevant when the content generated serves to inform consumer choices, such as content generated in the context of search queries or similar.

Accountability

Generative AI systems should be auditable by independent researchers, enforcement agencies and civil society organisations such as consumer organisations. This is essential to ensure the responsible use of training data and compliance with legal requirements.

There must be clear rules on accountability and liability for harmful effects of generative AI systems. These rules must clearly indicate which company in the supply chain is liable and allow consumers to receive quick compensation when harmed by a generative AI system.

Date of application of the AI Act's rules for generative AI

EU policymakers should anticipate the date of application of the rules of the AI Act applicable to generative AI (Article 83). The fast development, broad public adoption of systems such as ChatGPT, alongside its multiple risks and harm justifies the need to adopt these rules and make them enforceable as quickly as possible.

In the meantime, while the AI Act's rules are not yet applicable, existing EU law such as consumer protection, data protection, copyright or product safety legislation, must be enforced effectively now (and also once the AI Act is in force). With Europe still facing an enforcement gap, the European Commission should urgently prioritise enforcement of legislation, which plays an essential role in how the European Union quickly and effectively deals with emerging technologies like generative AI.

Regular dialogues between the providers of foundation models and the AI Board

We are concerned about the importance granted by the European Parliament to the establishment of regular dialogues between the AI Board and providers of foundation models and about the compliance of the latter with the AI Act. This is mentioned not only once but twice in Article 56b.

After all, the first objective of the legislators should be that the rules in the AI Act are clear and can be applied directly by developers without further discussion with businesses. The AI Board can provide guidance where needed but establishing an institutionalised dialogue with the industry implies that regulators are in need of businesses' assessment of how they comply with these rules. This carries the risk of endless exchanges and regulatory capture. We recommend therefore that the specific references proposed by the Parliament be deleted from the list of tasks of the AI Board.

Instead, these dialogues should take place in the context of the European Parliament's Advisory Forum, which should represent a balanced selection of stakeholders, including civil society and academics (Article 58 (2) of the European Parliament's position).

BEUC recommendations:

- Both developers and deployers should carry out a detailed risk assessment before placing the generative AI system on the market. The assessment of the developer should be carried out by an independent third party, while the assessment of the deployer can be internal.
- If the deployer cannot demonstrate that it can mitigate the risks identified, the system should not be deployed or made available to consumers.

- Deployers of generative AI systems should monitor and address the way the system is affecting consumers throughout the lifespan of the system.
- Developers and deployers of generative AI should publish relevant documentation about their risks assessments, including a short and less technical version informing consumers about the potential remaining risks. This documentation should be made available to public authorities.
- The data sets used to train generative AI systems need to be subject to important safeguards such as measures to prevent and mitigate possible biases.
- Consumers should be made aware that they are interacting with a generative AI system.
- Deployers of generative AI in consumer-facing interfaces and services should have to disclose how the generated content is influenced by commercial interests.
- There must be clear rules on accountability and liability when a generative AI system harms consumers.
- Generative AI systems should be auditable by independent researchers, enforcement agencies and civil society organisations, such as consumer organisations.
- EU policymakers should anticipate the date of application of the rules of the AI Act applicable to generative AI.

11.Voluntary codes and dialogues

The European Commission recently announced initiatives on generative AI, both with regards to the joint EU-US AI voluntary code of conduct and an 'AI Pact' for Europe.

We fully agree that urgent and swift action is needed to protect people from the significant risks of generative AI and to ensure that its benefits reach consumers without harming them and our societies. For example, see our call above to anticipate date of applicability of AI Act's rules on generative AI.

Nevertheless, we consider it highly problematic that the European Commission plans to engage into negotiations with businesses on such a voluntary initiative while the European Parliament and the Council of Ministers, the EU co-legislators, enter into the crucial trilogue phase to agree on the proposed AI Act.

For example, it is unclear what requirements a voluntary agreement can include when the legal requirements for these actors in the EU are not yet defined. There is an obvious risk that industry will try to use the discussions on a code to impact the legislative negotiations. Moreover, the voluntary commitments might not be in line with the final legal text.

BEUC immediately wrote to Executive Vice-President Vestager and Commissioner Breton asking that no negotiations with industry for a voluntary initiative take place as long as the legislative procedure on the AI Act is not finalised, whether they involve the US and the EU, or just the EU. We hope that the European Commission postpones the launch of any negotiations until the finalisation of the AI Act. After the final adoption of the AI Act, similar self-regulatory initiatives could take place to cover the time period until the entry into force of the AI Act. Consumer organisations should take part in those discussions.

BEUC’s recommendations:

- No negotiations with industry for a voluntary initiative should take place as long as the legislative procedure on the AI Act is not finalised.

- END -

Co-funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or EISMEA. Neither the European Union nor the granting authority can be held responsible for them.